



TUM School of Education

Susanne Klatten-Stiftungslehrstuhl für Empirische Bildungsforschung

**An Evaluation of Simulated Conversations
as an Assessment of Pre-Service Teachers' Communication Competence
in Parent-Teacher Conversations**

Anne Birte Wiesbeck

Vollständiger Abdruck der von der Fakultät TUM School of Education der Technischen
Universität München zur Erlangung des akademischen Grades eines

Doktors der Philosophie (Dr. phil.)

genehmigten Dissertation.

Vorsitzende:	Univ.-Prof. Dr. Kristina Reiss
Prüfer der Dissertation:	1. Univ.-Prof. Dr. Manfred Prenzel
	2. Priv.-Doz. Dr. Johannes Bauer

Die Dissertation wurde am 1. Juli 2015 bei der Technischen Universität München eingereicht und durch die TUM School of Education am 30. Juli 2015 angenommen.

Content

Danksagung	VII
Abstract	VIII
Kurzfassung	IX
A Aims & Structure	1
B Theory	6
1 The Relevance of Parent-Teacher Cooperation for Students' School Success and Social Development	6
2 Parent-Teacher Conversations as a Central Component of Parent-Teacher Cooperation	10
2.1 Legal Regulations With Regard to Parent-Teacher Cooperation and Conversations in Germany.....	11
2.2 Parent-Teacher Conversation Praxis.....	13
2.3 Preparation for Parent-Teacher Conversations in Teacher Training Programs	15
2.4 Summary	19
3 Teacher Competencies for Successful Conversations With Parents	22
3.1 Definition of Communication Competence	22
3.2 Communication Competence in Parent-Teacher Conversations	30
3.3 Summary	38
4 Measuring Communication Competence of Teachers in Conversations With Parents	40
4.1 Goals and Challenges of Research on Competence Measurement in Teacher Education	40

4.2	Multimethod Measurements	42
4.3	Instruments for Diagnosing the Competence of Teachers to Communicate With Parents	44
4.4	Summary.....	53
5	State of Research on Simulated Conversations.....	56
5.1	Simulated Conversations in the Medical Domain	56
5.2	Summary.....	65
5.3	Simulated Conversations in the Educational Domain	67
5.4	Summary.....	73
6	Conclusions From the Theory for this Dissertation	76
C	Research Questions & Hypotheses.....	80
D	Method.....	86
7	Development of the Instruments	88
7.1	Development of the Case Vignettes	88
7.2	Recruitment and Training of the Simulated Parents.....	90
7.3	Construction of the Coding Manual	92
7.4	Construction of the Rating Scale for the Simulated Parents	93
7.5	Construction of the Self-Assessment Questionnaire for the Pre- Service Teachers.....	94
7.6	Recruitment and Training of the Raters	94
7.7	Adaptation of six Coding Manuals From the Medical Domain	98
8	Data Collection.....	108

8.1	Pilot Study.....	108
8.2	Main Study.....	111
9	Data Analysis Pilot and Main Study	115
E	Results - Pilot Study.....	119
10	Basic Analyses	120
10.1	Consistent Inter-Rater Agreement - Objectivity of Scoring	120
10.2	Fit of the Data to the Theoretical Construct - Internal Structure	122
10.3	Reliability of the Coding Manual - Composite Reliability / Internal Consistency	126
11	In-Depth Analyses on Objectivity, Reliability and Validity	129
11.1	Consistent Performance of the Actors in the Simulated Conversations - Objectivity of Application	129
11.2	Effect of the Case Vignettes - Generalizability	134
11.3	Perceived Authenticity of the Simulated Conversations - Response Processes	137
11.4	Multimethod Measurement - Relations of the Results of Simulated Conversations to Other Variables	137
12	Summary and Discussion of the Pilot Study Results	139
13	Implications for the Refinement of the Instrument	141
13.1	Case Vignettes	141
13.2	Actor Training.....	142
13.3	Coding Manual	142
13.4	Rater Training	144

F	Results - Main Study	146
14	Basic Analyses	147
14.1	Consistent Inter-Rater Agreement - Objectivity of Scoring.....	147
14.2	Fit of the Data to the Theoretical Construct - Internal Structure.....	148
14.3	Reliability of the Coding Manual - Internal Consistency / Composite Reliability	151
15	In-Depth Analyses on Objectivity, Reliability and Validity	154
15.1	Consistent Performance of the Actors in the Simulated Conversations - Objectivity of Application.....	154
15.2	Effect of the Case Vignettes - Generalizability	159
15.3	Perceived Authenticity of the Simulated Conversations - Response Processes.....	163
15.4	Multimethod Measurement - Relations of the Results of Simulated Conversations to Other Variables.....	163
16	Ancillary Psychometric Quality Criteria.....	171
17	Summary and Discussion of the Main Study Results	174
G	Discussion	179
18	Main Findings and Practice Implications.....	182
19	Limitations and Future Research Areas	190
20	Outlook.....	193
H	Appendix	197
21	List of Figures	197
22	List of Tables.....	198

23	List of Abbreviations.....	200
24	Materials.....	201
24.1	Case Vignette.....	202
24.2	Training of the Simulated Parents.....	208
24.3	Coding Manual Pilot Study.....	212
24.4	Coding Manual Main Study.....	229
24.5	Rating Scale for the Simulated Parents.....	247
24.6	Self-Assessment Questionnaire Pre-Service Teachers.....	248
24.7	Rater Training.....	249
I	Literature.....	251

Danksagung

Zum Gelingen dieser Arbeit und zu meiner fachlichen und persönlichen Weiterentwicklung haben sehr viele Menschen beigetragen. Ich freue mich sehr über die Gelegenheit mich an dieser Stelle ganz herzlich bei Euch allen bedanken zu können!

Zunächst gilt mein Dank meinem Erstbetreuer, Prof. Dr. Manfred Prenzel, für die Chance in einem äußerst inspirierenden Umfeld promovieren zu dürfen, seine Gabe mich zu motivieren und die sehr gut ausgewogene Mischung aus Handlungsfreiheit und Rückmeldung, wann immer ich sie gebraucht habe. Weiterhin möchte ich mich bei meinem Zweitbetreuer, PD Dr. Johannes Bauer, bedanken, der mir entscheidend dabei geholfen hat diese Arbeit zu strukturieren und der mit äußerst großem Engagement dafür gesorgt hat, dass ich mich fachlich weiterentwickle und über mich selbst hinauswachse. Ganz herzlich möchte ich mich auch bei meiner Mentorin, Dr. Katharina Müller, bedanken, die jederzeit sowohl fachlich als auch menschlich für mich da war.

Ein ganz großes Dankeschön gilt auch meinen Kollegen. Ich hatte das Glück in einem wunderbar inspirierenden und hilfsbereiten Umfeld arbeiten zu dürfen und möchte mich dafür bei dieser Gelegenheit bei unserem Lehrstuhlteam bedanken. Ihr seid großartig! Mein besonderer Dank gilt meinen Kolleginnen und Freundinnen Susanne, Franzi, Sara, Sandra, Jessi und Gloria – für Eure fachliche und emotionale Unterstützung, insbesondere aber dafür, dass ihr immer für mich da gewesen seid und die Promotionszeit zu einer unvergesslich schönen und lustigen gemacht habt!

Darüber hinaus möchte ich mich bei allen Projektpartnern für die gute Zusammenarbeit und insbesondere auch für Eure Unterstützung in Bezug auf meine Dissertation bedanken. Ebenfalls ganz herzlich bedanken möchte ich mich bei allen Studierenden und Hilfskräften, die mich während der letzten Jahre unterstützt und durch ihre Fragen und Ihr Interesse an meiner Forschung begeistert haben, insbesondere bei Regina Altmann.

Above this, I would like to thank the staff at Syracuse University, in particular, Ben Dotger and Steven Harris, for introducing me to their research and inspiring my work.

Zuletzt möchte ich mich an dieser Stelle bei meinen Freunden und meiner (Schwieger-) Familie bedanken, insbesondere bei meinen Eltern, dafür dass sie immer für mich da sind und mich darin unterstützen das zu tun was ich liebe.

Widmen möchte ich diese Arbeit dir Ferdinand, meinem Mann. Du machst mich jeden Tag zum glücklichsten Menschen dieser Welt!

Abstract

Parent-teacher conversations can have a positive impact on the academic success and development of pupils. However, German teachers feel insufficiently prepared for conversations with parents. Therefore, during the last few years, corresponding training programs for teacher education have been developed. Now, there is a need for instruments that can assess the effectiveness of teacher training with regard to parent-teacher conversations and permit to diagnose whether pre-service teachers are sufficiently prepared for parent-teacher conversations or where further training is required. Hence, the aim of this dissertation is to develop and validate an instrument for assessing the communication competence of pre-service teachers in conversations with parents.

Simulated conversations, an instrument from the medical domain, are adapted in this dissertation because they are promising for measuring pre-service teachers' communication competence in conversations with parents: they are performance-oriented, context-related and offer authentic measurement conditions. The newly-developed simulated conversations are evaluated in a pilot study with $N=49$ and a main study with $N=96$ pre-service teachers. The results of both studies are divided into analyses that target basic and in-depth aspects of psychometric quality criteria. Findings with regard to the basic analyses show that the conversations are rated with sufficient objectivity, that the structure of the data corresponds to the underlying theoretical construct and that the coding manual for analyzing the conversations is reliable. The in-depth analyses show that the portrayal of parents by different actors is consistent, the results of simulated conversations based on a certain case vignette are generalizable - though only to a limited extent - the pre-service teachers perceive the conversations as authentic and the relations between observer-ratings of the pre-service teachers' performance and other criteria gathered via a multimethod measurement correspond to theoretical expectations. In sum, the pilot and main study findings consistently indicate that simulated conversations are suitable to assess the competence of pre-service teachers to communicate with parents.

The dissertation demonstrates that simulated conversations can be used to ensure and refine the quality of teacher education with regard to preparing teachers for conversations with parents. Furthermore, it highlights possible areas for the incorporation of simulated conversations into teacher education and outlines factors which are decisive for their successful employment. In this way, the dissertation contributes to preparing teachers for conversations with parents and, in this way, to improving parent-teacher cooperation.

Kurzfassung

Internationale Forschungsbefunde zeigen, dass Gespräche zwischen Lehrpersonen und Eltern einen positiven Einfluss auf den Schulerfolg und die Entwicklung von SchülerInnen haben. Allerdings fühlen sich deutsche Lehrpersonen oft nicht ausreichend auf Gespräche mit Eltern vorbereitet. Infolgedessen wurden in den letzten Jahren entsprechende Trainingsprogramme entwickelt. Um zu überprüfen ob diese Programme effektiv sind und Lehrpersonen durch sie angemessen auf das Führen von Elterngesprächen vorbereitet werden, sind Verfahren zur Messung der Gesprächsführungskompetenz erforderlich. Ziel der vorliegenden Arbeit ist daher ein Instrument zur Messung der Gesprächsführungskompetenz von Lehramtsstudierenden in Elterngesprächen zu entwickeln und zu validieren.

Simulierte Gespräche, ein im Bereich der Medizin etabliertes Verfahren, werden in dieser Arbeit an den Lehramtskontext adaptiert, da sie es ermöglichen Gesprächsführungskompetenz handlungsorientiert, situationsspezifisch und unter realitätsnahen Bedingungen zu erfassen. In einer Vorstudie mit $N=49$ und einer Hauptstudie mit $N=96$ Lehramtsstudierenden werden die psychometrischen Gütekriterien der entwickelten Gespräche evaluiert. Beide Studien unterteilen sich in grundlegende und weiterführende Analysen. Die grundlegenden Analysen zeigen, dass Beobachter die Gespräche hinreichend objektiv auswerten, die Struktur der Daten das zugrundeliegende theoretische Modell widerspiegelt und das Kodiermanual zur Auswertung der Gespräche reliabel ist. Die weiterführenden Analysen liefern Evidenz dafür, dass verschiedene SchauspielerInnen die Eltern konsistent darstellen, die Ergebnisse verschiedener Gespräche generalisierbar sind – allerdings nur bis zu einem gewissen Grad – die Studierenden die Gespräche als authentisch wahrnehmen und die Ergebnisse der Gespräche entsprechend theoretischer Erwartungen mit anderen Kriterien, die im Rahmen einer multimethodalen Messung erhoben wurden, korrelieren. Die Befunde von Vor- und Hauptstudie sprechen dafür, dass simulierte Gespräche geeignet sind um die Gesprächsführungskompetenz von Lehramtsstudierenden in Elterngesprächen zu erfassen.

Die Arbeit zeigt, dass simulierte Gespräche zu einer evidenzbasierten Weiterentwicklung und Qualitätssicherung der Lehrerbildung beitragen können. Darüber hinaus werden mögliche Einsatzgebiete in der Lehreraus- und Weiterbildung ausgemacht sowie Faktoren, die für einen erfolgreichen Einsatz von simulierten Gesprächen entscheidend sind. Langfristig soll die Arbeit einen Beitrag dazu leisten Lehrpersonen auf Gespräche mit Eltern vorzubereiten und dadurch die Zusammenarbeit zwischen Lehrpersonen und Eltern verbessern.

A Aims & Structure¹

Evidence suggests that cooperation between schools and parents can have a positive impact on student development and school success (Chrispeels & Coleman, 1996; Jeynes, 2011; Kreider, Caspe, Kennedy, & Weiss, 2007). One of the most important tools for fostering parent-teacher cooperation is formal parent-teacher conversations with one or two parents. Consequently, in 2004 the *Kultusministerkonferenz*² declared institutionalized parent-teacher conversations an integral part of teacher tasks in Germany (Kultusministerkonferenz, 2004). Yet up to today, a lot of German parents report that they are not satisfied with parent-teacher conversations (Killus & Tillmann, 2012). This dissatisfaction might be due to the fact that many teachers find parent-teacher conversations challenging and feel insufficiently prepared by their education (Freyaldenhoven, 2005; Hertel et al., 2014). In recent years, several training programs pertaining to parent-teacher conversations have been developed and incorporated into teacher education in Germany (Aich, 2011; Gartmeier et al., 2015; Hertel, Bruder, & Schmitz, 2009). Hence, there is a need for instruments³ that can assess whether training measures are effective and whether pre-service teachers are adequately prepared for communicating successfully with parents. Furthermore, such assessments can yield valuable information about what future teacher training needs to target.

A promising approach for measuring the competence of teachers to conduct formal conversations with parents is simulated conversations⁴; an established method in the medical domain, since they offer authentic measurement conditions and are performance-oriented (Cleland, Abe, & Rethans, 2009; Shavelson, 2013; United States Medical Licensing Examination, 2015). In medical simulated conversations (pre-service) physicians interact with trained actors behaving like real patients. Research from the medical domain shows that

¹ As far as possible gender neutral language is applied. To facilitate readability and brevity sometimes the masculine form is used only. However, the female form is also always implied unless stated otherwise.

² Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany

³In this dissertation the term *instrument* is employed instead of the term *test* for the following reasons: According to the Standards published by the American Educational Research Association (AERA), the American Psychological Association (APA) and the National Council on Measurement in Education (NCME) in 2014 a test is “a device or procedure in which a sample of an examinee’s behavior in a specified domain is obtained and subsequently evaluated and scored using a standardized process” (p. 2). The AERA, APA & NCME Standards use the single term *test* to refer to all kinds of evaluative devices (American Educational Research Association, et. al., 2014). While this use of the term *test* also occurs in German, the term *test* is most frequently used to refer to written, standardized cognitive or motivational tests (Scheibe, Trittel, Klug & Schmitz, 2014). Thus, it might be misleading for German readers to use the term *test* for simulated conversations.

⁴ Simulated conversations are commonly referred to as conversations with simulated/standardized parents. For the use of terminology in this dissertation see chapter 5.1.2.

simulated conversations can objectively, reliably and validly measure competencies of (pre-service) physicians and are very well-accepted by them (Barman, 2005; Cleland et al., 2009; Newble, 2004). Thus, the aim of this dissertation is to transfer simulated conversations to the educational domain and to evaluate to what extent simulated conversations are suitable to diagnose the communication competence of pre-service teachers in formal parent-teacher conversations with one or both parents. The resulting overall research question is: To what degree are simulated conversations suited for diagnosing the communication competence of pre-service teachers in parent-teacher conversations? This question is broken down into its components with regard to the fulfillment of psychometric quality criteria. The result is seven sub-questions relating to main elements of objectivity, reliability and validity. These questions investigate, for example, to what degree independent observers agree regarding pre-service teachers' performance in simulated conversations; whether the different actors portraying parents or the different case vignettes on which the simulated conversations are based influence their difficulty; whether the pre-service teachers perceive the simulated conversations as authentic, or if the results of simulated conversations are related to other measurements, like a situational judgment test on parent-teacher conversations, and external variables.

To create a basis for this thesis and to single out areas in which (further) research is needed, chapters 1 - 6 provide the theoretical background. Chapter 1 contains a summary of findings about the effects of parent-teacher cooperation on students which highlights the relevance of parent-teacher cooperation and conversations for student school success and development.

Chapter 2 starts with a discussion of several modes of parent-teacher cooperation, which reveals the core relevance of formal parent-teacher conversations for parent-teacher cooperation. Parent-teacher cooperation and conversations can take a variety of forms and can be formal or informal. The focus of this dissertation is on formal, planned conversations between a teacher and one or two parents, since this type of conversation provides the key to synchronize educational processes at school and at home and, thus, has a lot of potential to support students' development and school achievement. If not stated otherwise, whenever the term *parent-teacher conversations* is used or the competence of teachers to communicate with parents is addressed, this refers to formal, individual parent-teacher conversations and the respective competencies for this type of conversation. In subchapter 2.1 the legal framework for parent-teacher conversations in Germany is sketched out in

order to make the institutional conditions under which parent-teacher conversations take place comprehensible. To clarify if and which points of parent-teacher conversations should be improved, the current state of German parent-teacher conversation practice is outlined and compared to legal postulates and parents' expectations (2.2). Chapter 2 ends by analyzing to what degree parent-teacher conversations are incorporated into teacher education programs (2.3). This analysis provides potential explanations for the fact that current parent-teacher conversation practice often fails to live up to theoretical and legal postulates, as well as parents' expectations. Moreover, it highlights the importance of a stronger integration of parent-teacher conversations into teacher education.

Developing and evaluating instruments to measure teachers' competence to communicate with parents that can provide evidence about whether corresponding teacher education is successful and about whether exiting pre-service teachers are equipped with all relevant competencies to communicate with parents, requires firstly an understanding of what constitutes a teacher's competence to successfully communicate with parents. As a basis for understanding what constitutes this specific competence, the terms *communication*, *competence* and *communication competence* are discussed with particular emphasis on communication competence in expert-layman conversations (3.1). Drawing upon existing models and definitions of parent-teacher and parent-school counselor communication (3.2), the Munich Model of Communication Competence in Parent-Teacher Conversations is introduced, which serves as a theoretical framework for this dissertation (3.3).

Conceptualizing and evaluating a new measurement instrument for diagnosing the competencies of (pre-service) teachers in conversations with parents, further necessitates an understanding of the goals and challenges of research on competence measurement in teacher education (4.1). Multimethod measurements can increase the validity of a measurement and provide important information with regard to the evaluation of newly developed instruments (4.2). Therefore, in the subsequent subchapter, advantages and disadvantages of five frequently used methods to assess teacher competencies and their suitability for assessing teachers' competence to communicate with parents are discussed (4.3). This process enables the identification of instruments that can complement simulated conversations with regard to measuring the communication competence of pre-service teachers in conversation with parents and, in this way, provide information for the validation process of simulated conversations in the context of a multimethod measurement in the empirical part.

For the purpose of employing simulated conversations with regard to parent-teacher conversations, a brief sketch of how this method has been applied in the medical context is useful to clarify terminology and potential uses as well as to get an overview of research findings on how the different components of simulated conversations should be designed (5.1). In the second part of chapter 5, current developments regarding simulated conversations in the educational domain are outlined and, in this way, future applications and research areas are identified (5.2). Moreover, the analysis of the current status of the use and research on simulated conversations in medical education and particularly in teacher education provides information regarding the design and evaluation of the simulated conversations for this dissertation and, consequently, forms the foundation for the deduction of research questions. In line with this, in the last part of the theoretical background conclusions for the empirical part are drawn from the theory (6).

The empirical part starts with an overview of the research questions and hypotheses (C). In the succeeding method part (D), firstly, an overview of all instruments employed is given. Next, the construction of the simulated conversations and the complementary instruments is described. The simulated conversations consist of four primary components: (a) the case vignettes the conversations are based on, (b) the actors portraying the parents and the training they receive, (c) the coding manuals, rating scales and self-assessment questionnaires for rating the conversations, (d) the independent raters and the training they received. The construction of the different components is explained in order of development and use. Following this, data collection is depicted (8). Data collection of the dissertation was embedded into the research project *ProfKom- Professionalisierung von zukünftigen ÄrztInnen und Lehrkräften im Bereich der Kommunikationskompetenz*⁵ funded by the German Federal Ministry of Education and Research. The aim of the project was to promote and assess the competence of pre-service physicians to communicate with patients and of pre-service teachers to communicate with parents. The simulated conversations developed in this dissertation served as an assessment for evaluating the developed training components for promoting the communication competence of the pre-service teachers in the project. In the last chapter of the methods part, the statistical procedures used for analyzing the data are outlined (9).

The results of the pilot study are described in section E and the results of the main study in section F. Both pre- and main study results are subdivided into basic analyses and in-depth

⁵ Professionalization of future physicians and teachers on the level of communication competence

analyses. The basic analyses address the fundamental aspects of building scores, such as clarifying inter-rater agreement, the factorial structure of the data and the reliability of the scales. The in-depth analyses are based on the previous analyses and inquire more deeply into aspects of the three main quality criteria objectivity, reliability, and validity. These analyses provide information about whether the actors perform consistently in the simulated conversations, to what degree the case vignettes the conversations are based on influence their generalizability, to what degree the pre-service teachers perceive the conversations as authentic and how observer ratings of the simulated conversations correlate with self-assessments of the pre-service teachers and external criteria. Based on the results of the pilot study, conclusions are drawn concerning the necessity and type of modifications of the instruments for the main study (13). In the main study, the pilot study analyses are replicated with a larger data set and additional in-depth analyses are conducted. These additional analyses further target the validity of the results of simulated conversations via investigating their relations to other measurements and variables through a multimethod measurement and a multitrait-multimethod matrix. As a last step in the results section, the fulfillment of ancillary psychometric quality criteria is outlined. The dissertation ends with a discussion of the results regarding the employment of simulated conversations in the educational domain. Moreover, the benefits and costs of simulated conversations are weighed and recommendations for the future employment of simulated conversations in research on teacher education and teacher training programs are developed (G).

B Theory

1 The Relevance of Parent-Teacher Cooperation for Students' School Success and Social Development

This chapter aims at clarifying the theoretical and practical relevance of parent-teacher conversations. Parent-teacher conversations are embedded into the broader framework of parent-teacher cooperation. Thus, the effects of this broader framework of parent-teacher cooperation on students are summarized as the point of departure for this dissertation.

Though factors, like peers or the media, tend to influence the development of students, family and school continue to be the most important influencing factors for the development and the school achievement of children and youth. This is the bottom line of a variety of international studies, meta-analyses and literature reviews that have been conducted in the past 50 years (Bloom, 1982; Busse & Helsper, 2007; Coleman et al., 1966; Epstein, 1990; Henderson & Mapp, 2002). This finding seems to be stable over time and applicable across countries. International comparative studies conducted from 2000 onwards, like the Programme for International Student Assessment (PISA), revealed that the amount to which the impact of socio-economic background, i.e. the family, on student achievement is compensated by schools, differs between countries. However, the most recent study confirms that up to today, there is a cross-national effect of socio-economic background on student achievement which continues to be particularly strong in Germany (Baumert & Schümer, 2001; Müller & Ehmke, 2013).

Since families are crucial for student development and school achievement, it is of utmost importance to involve them into the learning and school processes of their children. Epstein (1990) argues in her model of family school interconnections for promoting cooperation between families and school in order to merge these two entities. In line with Epstein, Krumm (1996) considers cooperation between parents and teachers as an effective measure to foster students' school success. This claim is supported by a consistent, positive and convincing body of evidence that shows that parent-teacher cooperation has an effect on social and emotional development as well as on school achievement for all students regardless of social background, age or nationality (Cox, 2005; Epstein, 1990; Henderson

& Mapp, 2002; Jeynes, 2007; Kohl, Lengua, & McMahon, 2000; Kreider et al., 2007; Manz, Fantuzzo, & Power, 2004).

A meta-analysis of 51 studies on the effect of parent-teacher cooperation by Henderson and Mapp (2002) revealed that when parents cooperated with schools and / or supported students regarding school issues at home, students, regardless of income and background, were more likely to achieve higher grades, be promoted and pass classes, attend school regularly, have a positive attitude towards school, have better social skills and graduate and go on to postsecondary education. The more families supported their children's learning processes, the better children tended to be in school (Henderson & Mapp, 2002). Those findings were replicated in a literature review by the Harvard Family Research Project in 2007 (Kreider et al., 2007). Moreover, the Harvard researchers found that family-school cooperation also has a positive effect on the self-esteem of students and a negative effect on substance abuse. In addition, a study by Fan & Williams (2010) showed that parental involvement facilitates intrinsic motivation towards school.

A huge amount of research shows that all children benefit from parent-teacher cooperation regardless of socio-economic, cultural and ethnical background (Desforges & Abouchar, 2003; Henderson & Mapp, 2002; Jeynes, 2007). However, a study by Hill and colleagues (2004) suggests that while all children benefit, the effect and purpose of cooperation between teachers and parents can vary depending on socio-economic, cultural and ethnical background. Some researchers even argue that parent-teacher cooperation is particularly helpful and beneficial for students with a low socio-economic or migration background and might be a means to provide more equal chances for all students (Hertel, 2009; Hill et al., 2004).

In their *Family Involvement Makes a Difference Series*, the Harvard Family Research Project highlights that parent-teacher cooperation matters not only for students from all societal subgroups but also for students of all age groups (Kreider et al., 2007). While early research on parent-teacher cooperation focused on the effects on younger children in kindergarten and primary school (Epstein, 1990), more recent studies gathered evidence for the continuing positive effect of parent-teacher cooperation on students in secondary school (Epstein & Sheldon, 2006; Henderson & Mapp, 2002; Hill et al., 2004; Jeynes, 2011). In their review of studies targeting this topic, the British Department for Children, Schools and Families (2008) comes to the conclusion that positive effects of parent-teacher cooperation persist into adolescence and even adulthood.

The majority of studies evidencing the effect of parent-teacher cooperation on student achievement and development have been conducted in the US (e.g., Cox, 2005; Epstein, 1990; Kreider et al., 2007; Henderson & Mapp, 2002; Jeynes, 2007). However, there are also several studies from Europe (Department for children, schools and families, 2008; Desforges & Abouchaar, 2003) and some studies from African (Erlendsdóttir, 2010) and Asian (Hui & Nirmala, 2000) countries that replicate the American findings. The diversity of countries in which the studies have been conducted with similar results indicates that parent-teacher cooperation is beneficial for students across cultures and nations. Moreover, the findings seem to be consistent over time since research for more than a quarter century shows that parent-teacher cooperation is beneficial for student achievement and social development (Chrispeels & Coleman, 1996; Cox, 2005; Eccles & Harold, 1993; Epstein, 1990; Henderson & Mapp, 2002; Jeynes, 2007; Kohl et al., 2000; Kreider et al., 2007; Manz et al., 2004).

Finally, family-school cooperation is not only beneficial for students, but also for teachers, schools and families. Epstein (1990) finds in her meta-analysis that teachers benefit in their understanding of families, report more positive feelings towards their job and are able to conduct more effective school programs. Parents report more knowledge about school and benefit in their role as educators (Epstein, 1990).

The first chapter has highlighted the central importance of families and schools for student development and school success. Consequently, cooperation between parents and teachers is of utmost importance for students. International studies from the past 30 years provide convincing evidence that parent-teacher cooperation is one key to promoting school achievement for all students regardless of social background, age or nationality (Chrispeels & Coleman, 1996; Department for children, schools and families, 2008; Henderson & Mapp, 2002; Kreider et al., 2007). Several studies show that those students with the greatest difficulties benefit most (Henderson & Mapp, 2002; Hill et al., 2004). In general, students benefit the more, the more their parents are engaged with the school (Henderson & Mapp, 2002). Institutionalized programs that connect schools and families can help form and sustain family-school cooperation (Henderson & Mapp, 2002). Most programs and family-school cooperation target younger students. However, research from the past years indicates that students from all age groups benefit from family-school cooperation (Department for children, schools and families, 2008).

Chapter 1 has set the frame for this dissertation by highlighting the relevance of parent-teacher cooperation and conversations for students. Chapter 2 continues by outlining the core relevance of parent-teacher conversations for parent-teacher cooperation and by analyzing institutional framework conditions for parent-teacher conversations, parent-teacher conversation praxis and teacher education regarding parent-teacher conversations.

2 Parent-Teacher Conversations as a Central Component of Parent-Teacher Cooperation

Chapter 2 starts by clarifying the relationship between parent-teacher cooperation and parent-teacher conversations in order to account for the central role of parent-teacher conversations in this dissertation. To make the institutional conditions under which parent-teacher conversations take place transparent, the legal framework for parent-teacher conversations in Germany is sketched subsequently (2.1). Next, the current state of German parent-teacher conversation practice is outlined and compared to legal postulates and parents' expectations to single out if and regarding which points parent-teacher conversations should be improved (2.2). In order to single out reasons for the suboptimal status quo, it is then analyzed how well teachers are prepared for conversations with parents by their education (2.3).

Parent-teacher cooperation and parental involvement can take a variety of forms: e.g. attending school events like open days or school festivals, volunteering at school, becoming a parent representative, participating in parent-teacher organizations, sitting in on students' classes or learning at home. For an overview of parental involvement forms in Germany see Textor (2009). The focus of this dissertation is on a very specific and effective aspect of parent-teacher cooperation, on formal, individual conversations between a teacher and one or two parents⁶. Those parent-teacher conversations are often considered the centerpiece of parent-teacher cooperation (Aich, 2011; Sacher, 2009; Textor, 2009) for the following reasons: firstly, parent-teacher conversations are a core practice into which teachers engage on a very frequent basis (Dotger, 2011a). On average, teachers have to conduct more than 3000 parent-teacher conversations during their career. Secondly, not all students talk to their parents about what happens at school and to their teachers about what happens at home. So, teachers as well as parents are often unaware of what happens in the other sphere (Textor, 2009). Consequently, parent-teacher conversations are the key to effective parent-teacher cooperation since they offer the chance to exchange relevant information about the student and to synchronize educational processes at school and at home and to inspire further forms of parent-teacher cooperation (Textor, 2009).

⁶ Whenever the term *parent-teacher conversations* is used or the competence of teachers to communicate with parents is addressed in the following this refers to formal, individual parent-teacher conversations and the respective competencies for this type of conversation if not stated otherwise.

In Germany, formal, individual parent-teacher conversations fall into two categories: conversations at parent-teacher conferences and during additional office hours. At parent-teacher conferences, teachers firstly introduce themselves to all parents, up-date them about the learning progress of the entire class and upcoming events and answer the parents' general questions. Subsequently, formal parent-teacher conversations take place so that parents receive the possibility to talk to teachers individually and in detail about their children. Parent-teacher conferences take place at the beginning of each school year and sometimes also during the second half of the school year. Normally, they take place sometime before the end of the school year in order to give parents the chance to react to teachers suggestions. Additionally, teachers offer office hours which either take place by individual appointment or regularly and are open for parents to come.

2.1 Legal Regulations With Regard to Parent-Teacher Cooperation and Conversations in Germany

Cooperation between schools and parents / legal guardians is regulated in article 6 §2 and article 7 §1 of the "Basic Law for the Federal Republic of Germany" (1949). According to article 6 §2, "the care and upbringing of children is the natural right of parents and a duty primarily incumbent upon them". In contrast, article 7 §1 states that "the entire school system shall be under the supervision of the state". In 1972, the Federal Constitutional Court of Germany ruled that article 6 and 7 are of equal rank (Bundesverfassungsgericht, 1972). Consequently, concerning the schooling of children, the spheres of responsibility of families and schools overlap and cooperation is imperative.

Germany consists of 16 federal states. The implementation and elaboration of cooperation between families and school resides with the federal states since they have *Bildungshoheit*, i.e. control over the federal educational systems. The *Kultusministerkonferenz (KMK)*⁷, a voluntary conference of the ministers of education and cultural affairs of the federal states, coordinates the education policy of the federal states. In 2004, the KMK declared parent-teacher cooperation an integral part of teacher tasks in its *standards for teacher education* (Kultusministerkonferenz, 2003). Equally, the KMK declared that teacher education needs to prepare (pre-service) teachers for cooperation with parents and, in particular, for parent-teacher conversations. However, the KMK only gives recommendations. It is the task of

⁷ Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany

the federal states to translate these recommendations into federal state laws. Yet, as a consequence of the KMK recommendations, today all federal states demand a close cooperation between parents and teachers in their education acts (Textor, 2009). In the following, federal state regulations will be exemplified with Bavaria, Germany's largest and second most populous federal state. Since federal state laws in Germany regarding parent-teacher cooperation and conversations differ with laws in Bavaria being comparatively restrictive, Textor (2009) is recommended for an overview of German federal state regulations with regard to parent-teacher cooperation in the other 15 federal states.

The Bavarian constitution, article 126 §1 and article 130 §1, rules analogously to the "Basic Law for the Federal Republic of Germany" that the education of children is the natural right of their parents while the Bavarian school system is under the supervision of the state (Bayerische Staatsregierung, 1946). The legal basis of the Bavarian school system are the *Bayerisches Gesetz über das Erziehungs- und Unterrichtswesen*⁸ (*BayEUG*), the *Dienstordnung für Lehrkräfte an staatlichen Schulen in Bayern (LDO)*⁹ and the *Schulordnung für die Gymnasien in Bayern (GSO)*¹⁰. The BayEUG article 74 §1 rules that education is a common task of school and legal guardians, which requires cooperation based on mutual trust (Bayerische Staatsregierung, 2000). This cooperation between families and schools is obligatory (BayEUG, 2000, art. 2 §3). According to §9a of the LDO, extra-curricular official duties of teachers include the constant development of cooperation with parents (Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst, 2008). Finally, the GSO §18 explicitly states that parent conference days and parent-teacher conversations are particularly important for cooperation with legal guardians (Bayerische Staatsregierung, 2007). Thus, according to GSO §18, parent conference days are obligatory in every school year and for every teacher. Moreover, full-time teachers have to offer weekly office hours. In 2013, the Bavarian State Ministry for Education, Cultural Affairs, Science and Arts declared the encouragement of parent-teacher cooperation their official goal (Bayerisches Staatsministerium für Unterricht und Kultus, 2014).

⁸ Bavarian law on education and teaching

⁹ Official regulations for teachers of Bavarian state schools

¹⁰ School regulations for secondary schools in Bavaria (there are analogous regulations for other types of schools)

2.2 Parent-Teacher Conversation Praxis

This subchapter starts with a short analysis of the frequency of and satisfaction with parent-teacher conversations in Germany in order to identify potential points for qualitative and quantitative improvements. Subsequently, a model that illustrates factors that influence the degree to which parent-teacher conversations take place is described to single out possibilities to increase the frequency of parent-teacher conversations.

2.2.1 *Frequency of and Contentment With Parent-Teacher Conversations*

Parent-teacher cooperation is an inherent part of school life in Germany. According to a Germany-wide, representative online survey¹¹ that questioned 1976 teachers and parents, 67.2% of parents reported to have had contact with the teachers of their children during the last year (Jäger-Flor & Jäger, 2009, p. 12). The most frequently used form of parent-teacher cooperation is parent-teacher conversations. Nearly every teacher offers parent-teacher conversations (Hertel, Bruder, Jude, & Steinert, 2013) on average 1.16 hours a week (Gartmeier, Gebhardt, & Dotger, in preparation) and 96% of German parents attend parent-teacher conversations either at parent-teacher conference days or individually during the school year, according to the 2nd representative, Germany-wide JAKO-O study in which Killus and colleagues (2012) interviewed 3000 parents and teachers guided by computer-based questionnaires¹².

Even though the large number of parent-teacher encounters is promising, parents complain that conversations are mostly initiated by them and that the vast majority of conversations are deficit-oriented and compelled because of school problems (Killus & Tillmann, 2012). Moreover, parents report that nearly half of the conversations do not contribute to changing the situation (Jäger-Flor & Jäger, 2009; Krumm, 1996). The atmosphere in parent-teacher conversations as well as the quality of parent-teacher relationships is generally considered acceptable to relatively good (Killus & Tillmann, 2012). Albeit, one third of parents feel that teachers are not sufficiently interested in their point of view, less than a half is completely satisfied with parent-teacher conversations and a significant number of conversations take a disagreeable course (Killus & Tillmann, 2012).

¹¹ Response rate 57%. Parents from all societal subgroups were randomly included in the survey. Females answered the questionnaire more frequently than males: 62.3%. Persons with higher school-leaving qualifications answered more frequently than persons with lower school-leaving qualifications.

¹² Response rate was not reported. Participation was voluntary. Parents were phoned and questioned until the targeted sample of 3000 participants was completed. Parents with children required to attend any type of school from all societal subgroups all over Germany were randomly included in the survey.

The dissatisfaction with parent-teacher cooperation applies especially to *Gymnasium*¹³ teachers. Several studies show that consulting services and parents' satisfaction with parent-teacher cooperation differs between school types with parents of *Gymnasium* students being the least content (Hertel et al., 2013; Killus & Tillmann, 2012).

2.2.2 Supply-Demand Model of Parent-Teacher Conversations

The last subchapter has shown that while a lot of parent-teacher conversations take place, they are mostly parent-initiated. Hertel and colleagues (2013) have developed and empirically validated a supply-demand model of parent-teacher conversations. This model might give suggestions on how to encourage teachers to be proactive and initiate more conversations with parents since it identifies factors which influence in how far teachers offer conversations and factors that influence to what degree parents embrace these offers.

The empirical results of Hertel and colleagues (2013) indicate that the relative importance attributed to parent-teacher cooperation at a specific school, is a predictor for the amount of conversations offered by teachers. Equally, the structural resources of schools, such as whether a conference room for conversations with parents is available, and the type of school seem to play a role (Hertel et al., 2013). In contrast, the composition of the student body seems to play a minor role (Hertel et al., 2013). In addition to the structural and process features on the school level, the individual features of teachers seem to be highly influential (Hertel et al., 2013). There is evidence suggesting that especially the perceived competence and the understanding of one's teacher role have an impact on the extent to which teachers offer parent-teacher conversations (Hertel et al., 2013). Moreover, it is likely that the demand for parent-teacher conversations also influences the supply of parent-teacher conversations. To what extent parents make use of parent-teacher conversations offers, depends on their socio-economic and migration background, as well as on the importance parents attribute to promoting their child (Hertel et al., 2013).

The research findings presented in subchapter 2.2 indicate that there is already an active culture of parent-teacher conversations in Germany in which the majority of German parents and teachers participate. However, there seem to be two points for improvement. Firstly, parents are not sufficiently satisfied with the quality of parent-teacher conversations and secondly, complain that most parent-teacher conversations are parent-initiated (Killus

¹³ Secondary school that leads to the *Abitur*, highest German school-leaving qualification. For more information about the German school system see: <http://www.kmk.org/bildung-schule/allgemeine-bildung/uebersicht-schulsystem.html>.

& Tillmann, 2012). The supply-demand model of parent-teacher conversations by Hertel and colleagues (2013) suggests that the extent to which teachers offer parent-teacher conversations is especially influenced by the perceived competence of teachers regarding the conducting of parent-teacher conversations and the understanding of one's teacher roles (Hertel et al., 2013). It is likely that the cornerstone for the understanding of one's teacher role and the perceived competence with regard to conducting parent-teacher conversations is laid in teacher education. Moreover, teacher education probably also influences the quality of parent-teacher conversations. Therefore, the next chapter analyses in how far teacher education prepares teachers for conversations with parents in order to elicit possible reasons for the suboptimal status quo of parent-teacher conversation practice and resulting possibilities for the optimization of teacher education regarding parent-teacher conversations.

2.3 Preparation for Parent-Teacher Conversations in Teacher Training Programs

A variety of German and international studies show that teachers are not adequately prepared for conversations with parents (Dotger & Smith, 2009; Flanigan, 2005; Hertel, 2009; Hertel et al., 2014; Jäger-Flor & Jäger, 2009; Kempen, 2008; Markow & Pieters, 2010; Shartrand, Weiss, Kreider, & Lopez, 1997). According to Freyaldenhoven (2005), the reason why German teachers are inadequately prepared is that while parent-teacher cooperation is required by law (see chapter 2.1) - though only on a very general level - preparation for parent-teacher conversations is not prescribed legally. Correspondingly, analyses of teacher education curricula and teacher surveys reveal that competencies needed for conversations with parents are not systematically integrated into the first and / or second phase of German teacher education yet (Behr, 2005; Friedrich, 2002; Hertel et al., 2013). A coherent educational concept seems to be missing (Aich, 2011). Mostly, conversation techniques are only broadly addressed through general communication trainings and classes that are not fitted to parent-teacher conversations (Hertel, 2009). Furthermore, US studies show that if parent-teacher conversations are part of the curriculum, they are usually related to early childhood certification (Shartrand et al., 1997; Weiss, Lopez, & Rosenberg, 2010). Up to now, there are no German studies investigating this question. However, some studies indicate that *Gymnasium* teachers conduct less conversations with parents than primary school teachers and parents seem to be least satisfied with the cooperation

with *Gymnasium* teachers and most satisfied with the cooperation with primary school teachers (Hertel et al., 2013; Killus & Tillmann, 2012). This might indicate that either those persons who choose to study primary education are more aware of the relevance of parent-teacher conversations or that parent-teacher conversations are also predominantly taught in early childhood education in Germany.

The insufficient incorporation of parent-teacher conversations into teacher education is reflected in teacher surveys. Most German teachers report that they feel inadequately prepared for parent-teacher conversations by their education (Hertel, 2009; Hertel et al., 2014; Kempen, 2008; Wild, 2003). This holds true for teachers who only recently entered the profession as well as for teachers with a lot of professional experience (Freyaldenhoven, 2005). In a teacher survey in 2009, 93% of teachers reported to be not adequately or rather not adequately prepared by their education for parent-teacher conversations (Hertel, 2009, p. 60). The following Figure 1 adapted from Hertel (2014), shows that the situation has not significantly improved during the last five years.

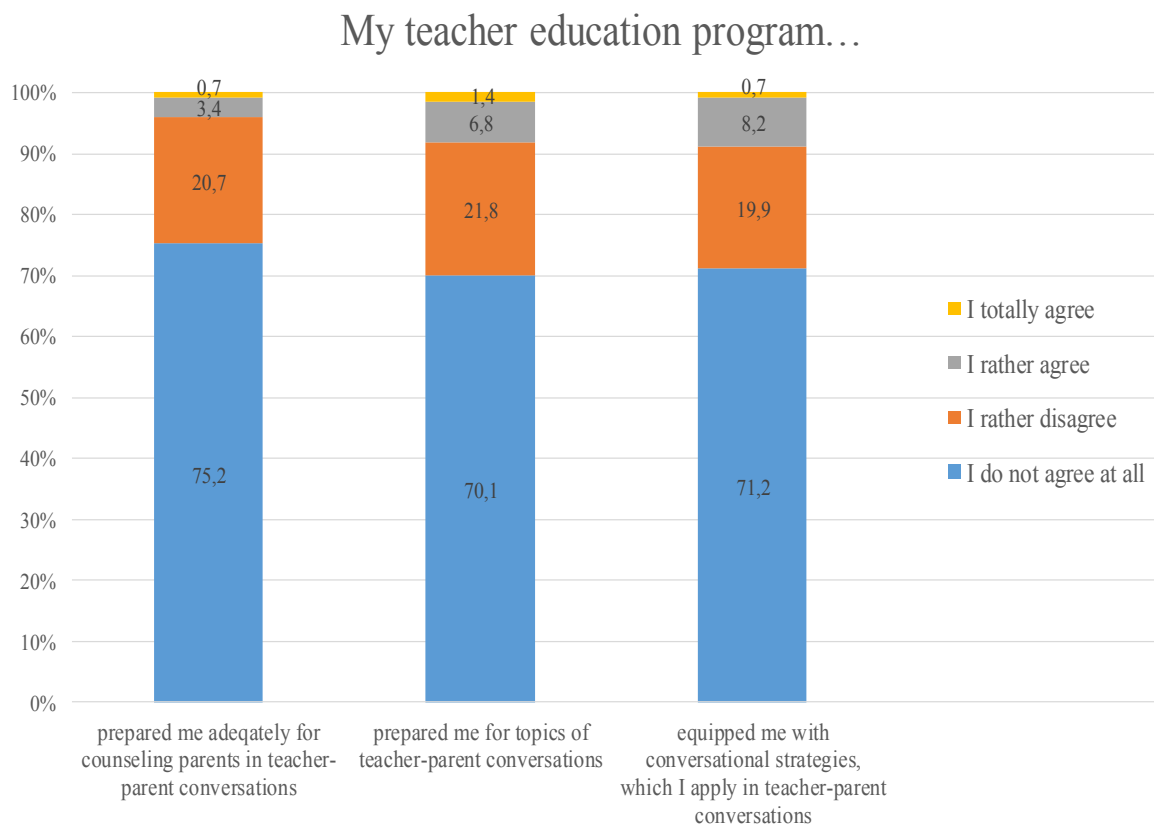


Figure 1. Preparation for parent-teacher conversations by teacher education programs (adapted from Hertel, 2014, p. 1)

Nevertheless, some progress has been made during the last years. Firstly, the competence of teachers to communicate with parents has been identified as important and was incorporated into theoretical concepts of teachers' professional knowledge, e.g. by Baumert & Kunter (2006). Subsequently, first models of communication competence needed for parent-teacher conversations (Bruder, 2011; Hertel, 2009), training programs for promoting this competence (Aich, 2011; Gartmeier et al., 2015; Hertel, 2009) and instruments for diagnosing it (Bruder, 2011; Bruder, Keller, Klug, & Schmitz, 2011; Hertel, 2009) have been developed. With regard to this development, a study by Djakovic and Hertel (2013) is promising since it showed that trainings and actions taken to professionalize teachers in the area of cooperation with parents can improve self-assessed competence, knowledge and professional belief of teachers. In line with the depicted developments, Bruder, Klug, Hertel and Schmitz (2010) found that teachers with less job experience (in years) showed better results in counseling parents than teachers with more job experience. Bruder (2011) suggests as one possible explanation for this seemingly paradoxical phenomenon that the newer teacher generation might be better educated in counseling than teachers who have been working for a longer time. Her conclusion is supported by survey results of the same group of teachers. Those teachers who were in the practical phase of teacher education at the time of the survey, reported significantly higher amounts of preparation for parent-teacher conversations than practicing teachers (Bruder, 2011).

The progress with regard to the preparation of teachers for conversations with parents is particularly important since teachers report that conversations with parents is one of the most challenging and threatening aspects of their job (Johns, 1992; Lawrence-Lightfoot, 2004; Markow & Pieters, 2010; Tacke, 1997). The strain and stress caused by parent-teacher conversations can lead to decreased job satisfaction and appears to be a major predictor of teacher's health (Unterbrink et al., 2008). An effective means to prevent negative effects of conversations with parents on teachers might be training for parent-teacher conversations. Hertel (2009) was able to show that a perceived high self-efficacy and level of competence regarding parent-teacher conversations are significantly negatively correlated to perceived strain, stress and burnout of teachers. Hertel's results are consistent with the findings of a study by Friedmann (2003). He found that teachers who perceive themselves as highly self-efficient and competent with regard to their profession, are less vulnerable to burnout (Friedman, 2003).

The teachers themselves are also convinced that preparation for parent-teacher conversations can play a crucial role as a protection against being over challenged and for improving parent-teacher cooperation (Hitziger, 1987). Consequently, most teachers are interested in education and further education with regard to parent-teacher conversations (Hertel, 2009; Wild, 2003). In the Germany-wide representative study by Jäger-Flor and Jäger, 76% of teachers and parents demanded to qualify teachers for parent-teacher conversations (Jäger-Flor & Jäger, 2009, p. 24).

With regard to the further incorporation of parent-teacher conversations into teacher education, it is important to note that teacher education is situated between the poles of scientific and professional orientation (Prenzel, 2009). On the one hand, scientific orientation should be the aim of all university studies since it prepares students for life-long learning (Wissenschaftsrat, 2006). An evidence-based, academic preparation at universities is particularly important with regard to the not precisely predictable challenges teachers will meet in their future work-life (Prenzel, Reiss, & Seidel, 2011). On the other hand, students often cite a gap between teacher education and the real lives of teachers (Putnam & Borko, 2000; Terhart, 2009). Thus, Prenzel and colleagues (2011) recommend that knowledge should be acquired contextualized, circumstanced and occupational whenever possible. Teacher students need the chance to apply occupational knowledge, gather experiences, receive feedback and reflect in order to build up scripts and routines (Prenzel et al., 2011). Due to the importance of both scientific and professional orientation, Prenzel (2009) and Ruthemann (2004) demand a systematic combination of scientific and professional orientation. In order to meet this demand, there is a need for new, action-oriented teaching and assessment methods, which allow a transfer of theoretical knowledge into praxis and, in this way, help bridge the gap between teacher education and practice (Dotger, 2011a).

The development of new methods for promoting and assessing parent-teacher conversations has to be constantly accompanied by research and evaluation (Caspé, Lopez, Chu, & Weiss, 2011). Evaluation provides information about how well prepared teachers are for parent-teacher conversations (Caspé et al., 2011). It allows singling out in which areas teachers are doing well and where teachers need further support (Spielberg, 2011). Moreover, evaluation makes it possible to identify promising training and assessment strategies and their impacts on teacher practices (Caspé et al., 2011). Empirically-based knowledge helps to refine curricula, training and assessment strategies and professional development

(Caspé et al., 2011; Weiss et al., 2010). To base teacher education consequently on research findings can ensure its quality (Prenzel, 2013).

2.4 Summary

The focus of this dissertation is on formal, individual conversations between a teacher and one or two parents since those kind of parent-teacher conversations are often considered the heart of parent-teacher cooperation for the following reasons: they offer the chance to exchange relevant information about the student, to synchronize educational processes at school and at home and to inspire further forms of parent-teacher cooperation (Textor, 2009). Parent-teacher cooperation and conversations are anchored in the national and federal constitutions. In line with legal postulates, there seems to be an active culture of parent-teacher conversations in Germany in which the majority of German parents and teachers already participate (Killus & Tillmann, 2012). However, parents complain that parent-teacher conversations are mostly parent-initiated and that they are often not satisfied with the conversations and their outcome (Killus & Tillmann, 2012).

The suboptimal status quo is particularly concerning since most researchers forecast that the need of parents for conversations with teachers will continue to rise due to societal changes and diversification of families (Bundesministerium für Familie, Senioren, Frauen und Jugend, 2010; Grewe, 2005; Jurczyk & Klinkhardt, 2014; Schnebel, 2007). These transformational processes will also make parent-teacher conversations more difficult (Wippermann, Wippermann, & Kirchner, 2013). Teachers already report that relationships with parents are progressively strained and conflicts rise (Wippermann et al., 2013). This tendency will probably continue since parents increasingly have the impression that education plays a crucial role for success in life and they feel under pressure to support their children to succeed in school evermore (Wippermann et al., 2013). As a result, the professional role of teachers is changing (Hertel, 2009; Wippermann et al., 2013). Future teachers will have to deal more often with parents under more difficult circumstances.

The supply-demand model of parent-teacher conversations by Hertel and colleagues (2013) suggests that the extent to which teachers offer parent-teacher conversations is especially influenced by the perceived competence of teachers regarding the conducting of parent-teacher conversations and the understanding of one's teacher roles (Hertel et al., 2013). It is likely that the cornerstone for the understanding of one's teacher role and the perceived competence with regard to conducting parent-teacher conversations is laid in teacher edu-

cation. Moreover, teacher education probably also influences the quality of parent-teacher conversations. Consequently, teachers have to learn about the importance of cooperation and need to be well-prepared for parent-teacher conversations by their education. Well-trained teachers can enhance motivation for and success of parent-teacher conversations (Anderson & Minke, 2007).

However, analyses of teacher education and teacher surveys indicate that German teachers are not yet sufficiently prepared for conversations with parents (e.g., Aich, 2011; Hertel, 2009; Hertel et al., 2013; Kempen, 2008). Still, in recent years, some progress has been made regarding the integration of parent-teacher conversations into teacher education. The competence to communicate with parents was incorporated into theoretical models of professional knowledge of teachers (Baumert & Kunter, 2006). Subsequently, first models of communication competence in parent-teacher conversations (e.g., Bruder, 2011), trainings for promoting this competence (Aich, 2011; Gartmeier et al., 2015; Hertel, 2009) and diagnostic tools for assessing it have been developed (e.g., Hertel, 2009). The progress made is particularly important because conversations with parents are a major source of stress and strain for teachers (Unterbrink et al., 2008) and preparation for conversations with parents can prevent negative effects on teachers (Hertel, 2009). Consequently, teachers perceive preparation for parent-teacher conversations as beneficial (Hitziger, 1987) and demand a stronger integration into teacher education and further education (Jäger-Flor & Jäger, 2009).

With regard to the future incorporation of parent-teacher conversations into teacher education, it is important to take into account that teacher education is situated between the poles of scientific and professional orientation (Prenzel, 2009). In order to meet the demands of both scientific and professional orientation, new teaching and assessment methods that can bridge the gap between theory and practice are needed (Dotger, 2011). The development and implementation of these methods need to be constantly accompanied by research and evaluation in order to validate if they are successful and to find out whether or in which areas (pre-service) teachers are already equipped with the necessary competencies for conversations with parents and where further training is needed (Caspé et al., 2011). Such an empirically-based, continuous development of teacher education can ensure its quality (Prenzel, 2013).

The findings presented in chapter 2 stress the need for a stronger integration of parent-teacher conversations into teacher education in order to improve the quality of parent-

teacher conversations and to encourage teachers to be proactive and initiate more conversations with parents. With regard to this stronger integration, there is need for assessment methods that can bridge the gap between theory and practice and provide information about the effectiveness of corresponding training programs developed in the past years, as well as about the readiness of teachers to conduct parent-teacher conversations and about starting points for further learning. Simulated conversations are an assessment method which possesses the potential to bridge the gap between theory and practice and provide the necessary information. Consequently, the aim of this dissertation is to develop simulated parent-teacher conversations. Since the development and implementation of new methods needs to be constantly accompanied by research and evaluation in order to ensure the future quality of teacher education, the newly-developed simulated parent-teacher conversations will be evaluated in the empirical part. In order to evaluate methods and measure competencies, a precise definition of the construct to measure is necessary (Koeppen, Hartig, Klieme, & Leutner, 2008; Maag Merki & Werner, 2011). Therefore, chapter 3 discusses what constitutes a teacher's communication competence in conversations with parents.

3 Teacher Competencies for Successful Conversations With Parents

Chapter 1 has shown that parent-teacher cooperation and conversations can positively impact student development and school success. Chapter 2 has highlighted the need for a stronger integration of parent-teacher conversations into teacher education and for methods that can assess the competence of teachers to communicate with parents in order to evaluate whether preparation for parent-teacher conversations through teacher education is successful. Developing and evaluating instruments to measure a teacher's competence to communicate with parents requires firstly an understanding of what constitutes a teacher's competence to successfully communicate with parents. In order to elaborate what constitutes this competence, at the start of this chapter, a general definition of communication competence is provided. In a second step, this definition is narrowed down to parent-teacher conversations since competencies have to be defined context-specifically in order to make them measurable. Finally, the Munich Model of Communication Competence in Parent-Teacher Conversation is discussed, which serves as a theoretical basis for this dissertation.

3.1 Definition of Communication Competence

This subchapter starts with a definition of the term *communication*. Subsequently, to embed parent-teacher conversations into theory and research on communication, an overview of influential communication theories from the educational field in which this dissertation is located follows. Since parent-teacher conversations can be regarded as expert-layperson conversations, the focus is on research on expert-layperson communication in particular. Two prominent approaches to expert-layperson conversations are presented: expertise research by Bromme, Jucks and Rambow (2004) and the *Client-Centered Therapy* by Rogers (1951). Additionally, the script-based learning theory by Schank & Abelson (1977) is introduced since it has also been applied in the context of expert-layman conversations, e.g. with regard to structuring physician-patient conversations (Bieber, 2007). From these three approaches, important features and aspects of communication competence can be deduced which have been integrated into the Munich Model of Communication Competence in Parent-Teacher Conversation, the theoretical foundation of this dissertation. Sub-

sequently, the term *competence* is discussed to then combine the two terms *communication* and *competence*. The combination of the two terms highlights that communication competence is not definable per se but needs to be defined context-specifically. Consequently, the second half of the chapter is dedicated to defining communication competence for conducting parent-teacher conversations.

3.1.1 *Communication*

The term *communication* stems etymologically from Latin *communicatio* which is the noun of action from the past participle stem of *communicare* which means *share, divide out, communicate, impart, inform, join, unite, participate in* (Harper, 2014, pp. communication). Today in the use of language in everyday life, communication stands for “the act or process of using words, sounds, signs, or behaviors to express or exchange information or to express your ideas, thoughts, feelings etc., to someone else” (Merriam-Webster, 2014, pp. communication). With regard to scientific language usage, there is no commonly agreed on definition of communication (Bentele, Brosius, & Jarren, 2006). The language and communication scientist, Merten, found 160 scientific definitions of communication with common and different features (Merten, 1977).

The variety of meanings of the term *communication* might be due to the fact that communication is an interdisciplinary research topic. The *International Encyclopedia of Communication* is currently the most comprehensive work about communication with contributions of more than 1100 international scholars (Donsbach, 2008). In its introduction, Donsbach (2008) states that communication is anything but a clearly defined field, it is interdisciplinary and quickly transforming. Communication is a major research and study topic in a vast number of disciplines ranging from politics over biology, technology and media studies to linguistics, psychology and education (Donsbach, 2008). For a complete overview of disciplines dealing with communication and a history on the research about communication see Donsbach (2008).

Since this dissertation is located in the educational context, the focus in the succeeding section is on selected communication theories that have influenced the discourse about communication in this field. Mathematical-technical communication models, like the one by Shannon and Weaver (1949), which explain the technical aspects of the transmittal process inspired the discourse on communication in the 1950s. From a pedagogical, social-scientific perspective these models were dissatisfying in the long run because they could

only explain the purely technical side of transmission processes. Consequently, subsequent researchers focused more on communication in the sense of spoken language (Schäfer, 2005). Prominent examples are the *Speech Act Theory* by Austin (1962) and Searle (Searle, 1962), which is concerned with the ways in which words can be used not only to present information but also to carry out actions, or the *Theory of Communicative Action* by Habermas (1984) that assumes that the linguistic structures of communication can help to establish a normative understanding of society. A very influential analysis was the *Interactional View* presented by Paul Watzlawick and his colleagues in the late 1960s (Watzlawick, Beavin, & Jackson, 2011). Drawing from the cybernetic tradition, Watzlawick and colleagues (2011) take into account that communication is a cyclic-bidirectional exchange process and highlight that both the content and the relationship level are important in conversations.

In his Client-Centered Therapy approach, Carl Rogers (1951) also highlights the importance of the relationship level but with regard to professional contexts. From the 1980s up until today, professional communication between experts and laypersons has increasingly become a focus of research within the scope of expertise research (Aich, 2011). Expertise research investigates the relationship between two conversational partners and the resulting requirements regarding conversations and is, thus, particularly interesting from a pedagogical perspective. Parent-teacher conversations can be regarded as expert-layman conversations and fall into the category of professional communication. Therefore, selected research findings about professional communication and expert-layman conversations are regarded in more detail in the following. See Schäfer (2005) for a more in-depth discussion of the history of research about communication in education and psychology in general and on current developments in particular (e.g. the influence of mass media on communication).

3.1.2 Communication Between Experts and Laypersons

Professionalizing experts, e.g. physicians, jurists, scientists, repairmen or teachers, for communication with laypersons, such as patients, clients, the public, customers or parents, is a central objective of expertise research (Stadtler, 2009). Bromme, Jucks and Rambow (2003) define experts as persons who can rely on a long education and job experience to solve complex problems. Laypersons are persons who do not have a systematic education regarding the focused knowledge domain (Bromme et al., 2003). There are three reasons

why expert-layman conversations are challenging. Firstly, they are characterized by a systematic knowledge asymmetry (Bromme et al., 2004). Thus, when communicating with laypersons, experts have to unzip their expert knowledge and rephrase it in everyday language (Bromme et al., 2004). Secondly, experts and laypersons have systematically different perspectives regarding this knowledge due to their different cognitive frames of reference (Jucks, 2001). In order to deal with this challenge, experts have to anticipate the layman perspective and adapt their communication to this anticipated perspective (Bromme et al., 2004). Anticipating what laypersons know is particularly challenging if the content of the conversation is located at the border of common and expert knowledge (Bromme et al., 2003) as it is often the case in parent-teacher conversations. What further complicates the anticipation process for experts is that lines between scientific and common knowledge are increasingly blurred due to the expansion of the internet. Thirdly, experts have to recognize that while they are experts with regard to the topic, the layperson is the expert for the problem, its specific context and the desired characteristics of the solution (Honal & Schlegel, 2002). This applies particularly to parent-teacher conversations since the teacher may be the expert for educational processes, but the parents are also experts in so far that they possess considerable knowledge about their child. What is further characteristic for parent-teacher conversations in comparison to other expert-layperson conversations is that the reasons for and the consequences of the conversation possibly affect both conversational partners and the expert more than in other expert-layman conversations.

Drawing on the psycho-linguistic communication theory of Herbert Clark (1996), (expert-layman) communication can be described as the trial to make two individual, cognitive frames of reference coincide sufficiently for establishing an adequate common ground that allows reaching the specific goal of the conversation (Bromme et al., 2004). Both conversational partners contribute information in order to enhance the common ground step by step (Bromme et al., 2004). Several grounding techniques enable the conversational partners to ensure if and in how far certain information is already part of the common ground (Bromme et al., 2004). Grounding techniques are, e.g., nonverbal signals or verbal agreement or disagreement (Bromme et al., 2004). In order to prepare experts for conversations with laypersons, it is important to teach them techniques for reaching common ground and to make them aware of the challenges of expert-layman conversations.

The second expert-layman approach presented here, the Client-Centered Therapy by Rogers (1951), aims at preparing professional counselors for conversations with their clients. The central ideas of the Client-Centered Therapy are that the climate of the conversation is essential for a positive outcome and that this climate is created by the therapist's attitude and behavior towards the patient (Rogers, 1951). Rogers (1951) specified three interdependent conditions which are essential for a positive conversation climate:

- 1) Congruence: counselors have to transparently and authentically relate to clients without hiding behind a professional or personal façade.
- 2) Unconditional positive regard: counselors should offer unconditional acceptance for their clients without conveying disapproving feelings or actions; they should listen attentively and without interrupting, judging or giving advice.
- 3) Empathy: counselors should try to understand and appreciate the perspective of their clients.

Although the Client-Centered Therapy has been criticized by psychoanalysts for providing a conditional relationship and by behaviorists for lacking structure (Prochaska & Norcross, 2010), it continues to be one of the predominant and empirically well validated approaches which contributes to a positive outcome of expert-layman conversations (Aich, 2011).

Script-based theory is a general approach to communication which has recently been applied for the specific purpose to train experts for communicating with laypersons. The term *script* was introduced by the American linguists Schank and Abelson (1977). A behavioral script is a sequence of expected behaviors for a given situation (Schank & Abelson, 1977). Up to 70% of adult communication is routinized and formulaic (Berger, 2008, p. 2483). With regard to this result, competent communicators may be those who, in a given situation, have rapid access to an effective and socially appropriate script that allows them to achieve their conversation goal (Berger, 2008). Therefore, in some disciplines like the medical domain, the idea emerged to make use of scripts with regard to expert-layperson conversations by teaching conversational scripts for frequent conversation types, such as shared decision-making with patients (Bieber, 2007).

From the three approaches to (expert-layperson) conversations presented above, three important features and aspects of communication competence can be deduced that are the basis of the Munich Model of Communication Competence in Parent-Teacher Conversations. Firstly, a conversation should contain a certain structure or sequence of expected behaviors which can be described in scripts. Secondly, for a positive outcome of a conver-

sation, it is decisive to establish a good relationship to the conversational partner by behaving congruently and showing unconditional regard and empathy. Thirdly, in order to solve a problem and reach a conversation goal, it is important to establish common ground, e.g. by using grounding techniques.

Every act of communication is open to evaluations of its quality and this quality usually matters for a real world outcome (Spitzberg, 2008). Therefore, a precise definition of communication competence is tantamount to every theory of communication (Spitzberg, 2008). Thus, in the following, what constitutes a competence in general and communication competence in parent-teacher conversations in particular, is discussed in more detail.

3.1.3 *Competence*

The term *competence* is frequently used in everyday language. Normally, it is used to indicate that a person is able to do something well (Merriam-Webster, 2014, pp. competence). In addition, it is used to express that a person or institution has the authority to do something or is responsible for something (Dudenverlag, 2014, pp. Kompetenz). Etymologically, *competence* comes from Latin *competentia* which means *meeting together, agreement or symmetry* (Harper, 2014, pp. Kompetenz). Weinert (2001) concludes that we know in general what the terms *competence*, *competent behavior* or *competent person* mean in everyday language, without being able to precisely define or clearly differentiate them.

In linguistics, pertaining to communication, the definition of competence by Noam Chomsky is widespread (Chomsky, 2006). Chomsky makes a distinction between *competence*, the system of linguistic knowledge possessed by speakers of a language, and *performance*, the way the language system is used in communication (Chomsky, 2006). In scientific language usage in the educational sciences, *competence* is a very popular term, especially since the inception of PISA (Prenzel et al., 2007). This is mirrored in the findings by Weinert (2001), who found 654 different key competencies in the literature on education and further education. There are different approaches to categorize these (key) competencies. Erpenbeck and von Rosenstiel (2003), for example, classify competencies into personal, action-oriented, subject-specific / methodical and social-communicative competencies. However, Erpenbeck and von Rosenstiel (2003) as well as a variety of other researchers come to the conclusion that the term *competence* is neither well-defined nor used consistently (Frey & Balzer, 2005; Hartig, 2008; Weinert, 2001).

In an overview article, Weinert (2001) depicts the wide range of meaning of the term *competence*. According to his analysis, the meaning of *competence* ranges from innate personality characteristics and motivational orientations over acquired domain-specific strategies, to comprehensive action ability for a profession, overarching key competencies and meta-competencies (Weinert, 2001). Based on his analysis, Weinert developed a frequently cited definition of *competencies*: “Competencies are the readily available or learnable cognitive abilities and skills which are needed for solving problems as well as the associated motivational, volitional and social capabilities and skills which are in turn necessary for successful and responsible problem solving in variable situations” (Weinert, 2001, p. 45). Furthermore, Weinert points out that *competencies*, *meta-competencies* and *key competencies* should be differentiated through a precise use of terminology. A competence should only be referred to as a *key competence* if it helps to solve a variety of different, equivalent requirements in one’s private as well as professional life (Weinert, 2001). According to Weinert’s definition (2001), communication competence can be considered a key competence since it helps to fulfill equivalent requirements in different areas of life. For a more detailed discussion of the term *competence* and a demarcation from other terms, such as *feature*, *characteristic* or *ability*, see Erpenbeck and von Rosenstiel (2003), Klieme and Hartig (2007) or Weinert (2001).

Weinert’s definition enables grasping of the concept *competence*. However, competencies are acquired and displayed in relevant, domain-specific situations (Koeppen et al., 2008). Therefore, Hartig (2008) states that a universal definition of *competence*, which fulfills scientific criteria, is unrealistic due to the variety of meanings of *competence*. In line with Hartig, most researchers agree that *competence(s)* should not be defined per se but regarding concrete contexts and requirements (Strasser & Gruber, 2008; Wollert, 1997). For answering empirical research questions and measuring competencies precisely, context-specific definitions of competencies are vital (Erpenbeck & von Rosenstiel, 2003; Hartig & Klieme, 2006). A precise, context-specific theoretical modeling helps to ensure that competencies are operationalized adequately, that different studies can be compared and that results of studies can be interpreted appropriately (Jahn, 2014).

Regarding the theoretical modeling of competencies, three types of competence models can be distinguished: models of competence structure, models of competence levels and models of competence development (Hartig & Klieme, 2006; Koeppen et al., 2008). Models of competence structures deal with the relations between performances in different

contexts and seek to identify common underlying dimensions (Koeppen et al., 2008). These models can explain performance in specific domains in terms of underlying basic abilities (Koeppen et al., 2008). Models of competence levels define the specific situational demands that can be mastered by individuals with certain levels of competencies (Koeppen et al., 2008). Models of competence development address how competence development takes place, e.g. whether it is continuous or not (Koeppen et al., 2008). While the three models relate to different aspects of competence constructs, they are not mutually exclusive but ideally complementary (Koeppen et al., 2008).

When developing competence definitions or models, it is important to contextualize them adequately and to differentiate the structure of a competence (Hartig, 2008). Moreover, if applicable, it can be helpful for the following measurement process to determine competence levels and / or to describe the competence development process. Models that identify the structure and the relations between different competence facets allow for a more valid diagnosis of competencies (Franke, 2005). With regard to the crucial role of the context, it is important to determine how narrowly the context should be defined (Jahn, 2014). Very broad as well as very narrow definitions of the context can result in a loss of value of a competence definition from a scientific perspective (Hartig, 2008). If the context is defined too broadly, the demarcation to other, more general constructs is blurred (Hartig, 2008). This hinders the development of instruments that can measure a certain competence (Hartig, 2008). A limitation of the context of the competence leads to an increase in precision and validity (Hartig, 2008). A specific context also provides hints regarding the operationalization of the construct (Klieme & Hartig, 2007). For example, it is possible to deduce possible test contents from relevant contexts (Klieme & Hartig, 2007). However, if the context is restrained too much, isolated skills or knowledge might be interpreted as competencies (Hartig, 2008). These would not comply with the high degree of complexity competencies are characterized by (Weinert, 2001).

Contextualizing competencies has the advantage that their praxis relevance is ensured. This complies with demands for a professional orientation of teacher education (cf. chapter 2.3). Yet, contextualization also implies that competencies are context / domain-specific and, consequently, definitions of competencies do not have universal validity (Hartig, 2008). Still, even though competencies are context- and / or domain-specific, they might be transferable to similar situations (Hartig & Klieme, 2006). However, their validity has to be reassessed for every new application context (Hartig, 2008).

3.1.4 *Communication Competence*

As other competencies, communication competence is not a generalized skill but specific to different kinds of communication goals and situations (Berger, 2008). Therefore, *communication competence* should be viewed as a theoretical term or domain of study rather than a single theoretical concept (Greene & Burlison, 2003). According to Berger (2008) and Spitzberg (2008), *communication competence* refers to the degree to which individuals are able to achieve their goals through communication (effectiveness) and the extent to which goal achievement is accomplished in an appropriate manner in a given context and culture (social appropriateness). Additionally, efficiency, e.g. how quickly a person reaches a certain goal, is often considered a component of communication competence (Berger, 2008). Berger's (2008) and Spitzberg's (2008) definition highlight the importance of context for communication competence. For the future and in order to measure certain types of communication competence, they demand the development of more specific definitions of communication competencies that take into account context and goals (Berger, 2008; Spitzberg, 2008). Considering this demand, in the next part four models that have recently been developed for describing the specific competencies teachers and school counselors need for successfully conducting different types of conversations with parents are presented.

3.2 **Communication Competence in Parent-Teacher Conversations**

Theory and research on the competence to communicate with parents are rare (Bruder, 2011). In particular, there is a lack of theoretically based, context-specific and empirically validated definitions and models of communication competence in parent-teacher conversations which are needed as a basis for the development of trainings and measurement instruments (Bruder, 2011; Hertel, 2009; Strasser & Gruber, 2008). This lack applies to the competencies of teachers as well as to the competencies of school counselors and psychologists (Strasser & Gruber, 2008). There are four exceptions, however, that model communication competence in parent-teacher conversations context-specific: the *Gmuender Gesprachsmodell* by Aich (2011), a definition of school counsellors' competence to communicate with parents by Strasser and Gruber (2008), a model of counseling competence in parent-teacher conversations regarding learning which was developed by Hertel

(2009) and refined by Bruder (2011) and the Munich Model of Communication Competence in Parent-Teacher Conversation by Gartmeier, Bauer, Fischer, Karsten and Prenzel (2011). There is some empirical evidence for the last three models. All four models are shortly presented and discussed in the following.

3.2.1 “*Gmuender Gespraechmodell*” for Parent-Teacher Conferences

One of the few models that are explicitly tailored to teachers’ competence to communicate with parents is the *Gmuender Gespraechmodell for Teachers to Improve their Communication Skills in Parent-Teacher Conferences*¹⁴ by Aich (2011). Based on Rogers’ Client-Centered Therapy (1951) and the Transaction Analysis by Berne (2002), he has developed a process model which singles out seven process stages teachers should try to conduct in parent-teacher conversations (Aich, 2011, p. 73). According to the model, the first step in a conversation should be starting the conversation by showing unconditional positive regard and by offering a symmetric relationship, for example (Aich, 2011). Further steps, such as offering subject-specific input or a joint rating of possible solutions, follow (Aich, 2011). The last step is to end the conversation with a concrete agreement (Aich, 2011). Aich’s model has the advantage that it is specifically contextualized for parent-teacher conversations. However, there is no empirical evidence for the model, yet (G. Aich, personal communication, September 12, 2014).

3.2.2 *School Counselors’ Competence to Communicate With Parents*

Drawing on expertise research, Strasser and Gruber (2008) single out three aspects of counseling competence: 1) information processing, e.g. how much relative importance is attributed to different sources of information, 2) formation of hypotheses and judgment, e.g. at which point in time hypotheses are formulated and 3) problem representation, e.g. how comprehensively the problem is captured (Strasser & Gruber, 2008). However, Strasser and Gruber (2008) do not focus on teachers but on school counselors and child guidance officers. They collected empirical evidence for their definition through a comparison of the competencies of experts (child guidance officers) and novices (school counselors) (Strasser & Gruber, 2008) according to the standard paradigm of expertise research (Ericsson, Charness, Feltovich, & Hoffman, 2006). Strasser and Gruber (2008) consider child guidance officers experts because it is their main task to council parents and, in con-

¹⁴ Translation by Aich (personal communication, September 12, 2014)

trast, only one of many tasks of school counselors. Additionally, child guidance counselors receive a more intense education. In their analysis, Strasser and Gruber (2008) find that expert counselors and novice counselors differ regarding the processing of information and the quality of the solution. Moreover, they report that experts proceed more systematically and goal-oriented than novices (Strasser & Gruber, 2008).

3.2.3 Counseling Competence in Parent-Teacher Conversations Regarding Learning

Hertel (2009) integrates the results of Strasser and Gruber (2008) as well as several theoretical approaches to counseling, such as the one from Schwarzer and Buchwald (2006), into a model of parent counseling competence for teachers. In order to ensure that the competence is conceptualized context- and domain-specific, she narrows the context down to counseling parents with respect to learning difficulties and strategies (Hertel, 2009). She differentiates the following five facets of counseling competence: 1) personal resources, 2) social collaboration, 3) counseling skills and pedagogical skills, 4) process skills and 5) coping (Hertel, 2009, p. 255). Based on this theoretical model, Hertel (2009) designed instruments for diagnosing the counseling competence of (pre-service) teachers in parent-teacher conversations regarding learning, such as a coding manual for written work-samples, a knowledge test, questionnaires for peer and observer ratings in role-plays and self-assessment questionnaires. In order to verify whether the internal structure of her data corresponded to the underlying theoretical model, she conducted confirmatory factor analyses (Hertel, 2009). However, those could not confirm the proposed five-dimensional structure of the model (Hertel, 2009). Nevertheless, Hertel (2009) found that a multidimensional model of counseling competence represents her data better than a one-dimensional model. She suggests that counseling competence consists of two or three competence facets and recommends further research regarding these facets and their interdependences (Hertel, 2009).

Based on the results of Hertel (2009) and a literature review regarding key aspects of counseling, Bruder and colleagues (2010) presented a modified model of teachers' counseling competence for conversations with parents. Just as Hertel does, Bruder and colleagues (2010) narrow the context down to conversations regarding learning difficulties and strategies. In a first step, Bruder (2011) investigates the modified, but still five-dimensional model with confirmatory factor analyses based on data collected with case scenarios and

self-assessment questionnaires. Similar to Hertel (2009), Bruder finds that a five-dimensional structure does not fit the data adequately (Bruder, 2011). Subsequently, Bruder (2011) suggests the following four-dimensional model: 1) counseling skills, 2) diagnostic / pedagogical skills, 3) collaboration / perspective taking, and 4) coping. Bruder (2011) calculated a structural equation model that shows that counseling competence concerning learning skills can be modelled by these four dimensions. However, the latent variable counseling competence had only moderate loadings on the subscales (Bruder, 2011). This applied especially to coping and counseling skills (Bruder, 2011). Still, Bruder (2011) concludes that her empirically validated construct can define teachers' counseling competence regarding conversations with parents about learning difficulties and strategies. As further evidence for the model, Bruder (2011) found that teachers in the practical phase of their education and practicing teachers had higher levels of counselling competence, as defined by the model, than teachers in the first phase of their education.

3.2.4 The Munich Model of Communication Competence in Parent-Teacher Conversation

The Munich Model of Communication Competence in Parent-Teacher Conversation is aimed at theoretically modeling the communication competence of teachers in formal conversations with one or both parents (Gartmeier et al., 2011). Gartmeier and colleagues (2011) conceptualize communication competence as a hierarchical and multidimensional construct that manifests differently in specific communication situations. Conversation types that occur frequently in the professional life of teachers are: (1) consulting / shared decision making, (2) conflict situations / handling complaints and (3) breaking bad news (Gartmeier et al., 2011).

The Munich Model of Communication Competence in Parent-Teacher Conversations is a topological model that aims at describing typical types of parent-teacher conversations in order to single out corresponding competencies needed for successfully managing these conversations. With regard to this, the model is subdivided into three competence facets that the authors consider as critical for the successful outcome of parent-teacher conversations: (1) establishing a positive interpersonal relationship with the conversational partner, (2) cooperative problem solving during the course of the conversation, and (3) transparent and adequate structuring of the conversation (Gartmeier et al., 2011). The model is dis-

played in the following Figure 2 with competence facets on the right side of the model and frequent types of parent-teacher conversations on the left:

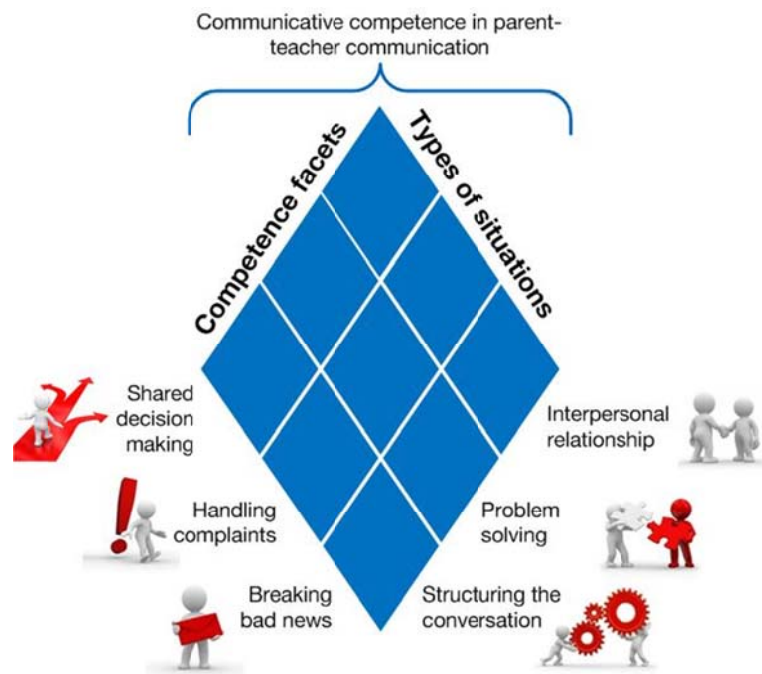


Figure 2. Munich Model of Communication Competence in Parent-Teacher Conversation (Gartmeier et al., 2011, p. 414)

The first competence facet aims at establishing a positive, long-term interpersonal relationship with the conversational partner that is based on mutual regard (Gartmeier et al., 2011). According to the communication theories by Schulz von Thun (1998) and Watzlawick and colleagues (2011), statements about the relation between the conversational partners are inherent in every message. Therefore, the competence to establish a positive interpersonal relationship with the conversational partner comprises behavior and strategies that show regard for the conversational partners and their perspective on the problem (Gartmeier et al., 2011). Drawing on Rogers' (1951) Client-Centered Therapy, Gartmeier and colleagues (2011) propose that teachers should act congruently and empathically and show unconditional positive regard, e.g. through active listening.

The second competence facet is aimed at cooperative problem solving during the course of the conversation (Gartmeier et al., 2011). In line with the research findings on expert-layperson conversations by Bromme and colleagues (Bromme et al., 2004; Bromme & Rambow, 2001), Gartmeier and colleagues (2011) consider the creation of common ground

as the first step for cooperative problem solving. Once a sufficient common ground is established, joint objectives should be agreed on and the problem solving process should be oriented towards these shared objectives (Bromme et al., 2004). Finally, it is important to give short intermediate summaries of the problem solving process (Allhoff & Allhoff, 2006) and to come to a concrete agreement (Jensen & Jensen, 2008).

The third competence facet is directed at structuring the conversation transparently and adequately (Gartmeier et al., 2011). Regarding the structuring process, it is important to ensure a comprehensible order of the different conversational phases as well as adequate transitions between two conversational phases (Gartmeier et al., 2011). Via metacommunication, the teacher should make the sense and the purpose of the different conversational phases and the transitions transparent to the parent (Gartmeier et al., 2011). Metacommunication are, e.g., advanced organizers or summaries. The structure of a conversation should also contain the most important problem solving steps outlined when explaining the problem solving competence facet (Gartmeier et al., 2011). While the focus of the competence facet problem solving is more on the quality of the different problem solving steps, the competence facet structuring the conversation focusses more on an adequate sequence of conversation phases and their transparency for the conversational partner (Gartmeier et al., 2011). Scripts for typical situations can be helpful for ensuring an adequate sequence of conversational phases as well as transitions between different phases (Gartmeier et al., 2011).

The Munich Model of Communication Competence in Parent-Teacher Conversation has four advantages. First of all, it is theoretically founded in research about expert-layperson conversations. The competence facet, interpersonal relationship, takes into account the importance of a good interpersonal relationship between experts and laypersons as stated by Rogers (1951). The competence facet, problem solving, highlights the importance of the creation of common ground as stated by representatives of expertise research, such as Bromme and colleagues (2004). The last competence facet, structuring the conversation, pays tribute to the fact that most conversations are routinized (Berger, 2008) and draws from the advantages of scripts (Gartmeier et al., 2011). It foregrounds the importance of a coherent and transparent conversational structure.

The model is not only theoretically founded but also takes into account context and praxis requirements, e.g. frequent types of parent-teacher conversations. Thus, its second ad-

vantage is that it complies with the demands of Prenzel (2009) for a scientific as well as a professional orientation of teacher education.

A third advantage of the Munich Model of Communication Competence in Parent-Teacher Conversation is that it is subdivided into competence facets as postulated by Franke (2005), to diagnose competencies in a more valid way. Moreover, it has three competence facets and, thus, ties in with the research findings by Hertel (2009), which suggest that the competence of teachers to communicate with parents is two- or three-dimensional. In a confirmatory factor analysis with observer ratings of the performance of 96 pre-service teachers and 72 pre-service physicians in simulated conversations, Gartmeier and colleagues (2015) were able to confirm the theoretically assumed structure of the model across two domains in which expert-layperson conversations are vital. They modelled general communicative competence as a second order factor and the three competence facets which are structuring, problem solving and establishing a relationship as first order factors (Gartmeier et al., 2015). This hierarchical model showed an acceptable fit ($\chi^2(24) = 48.88$, $p = .002$, $RMSEA = .079$, $CFI = .965$, $SRMR = .056$) (Gartmeier et al., 2015). The three first order factors had substantial factor loadings on general communication competence (structuring the communication $\lambda = .83$, problem solving $\lambda = .59$, interpersonal relationship $\lambda = .67$) (Gartmeier et al., 2015). The composite reliabilities were $\rho = .87$ for structuring the communication, $\rho = .83$ for problem solving, $\rho = .82$ for interpersonal relationship and $\rho = .71$ for general communication competence (Gartmeier et al., 2015). However, this finding needs to be replicated with a pre-service teacher sample only. A second finding by Gartmeier and colleagues (2015) was that pre-service teachers / physicians that had participated in a communication training on parent-teacher / physician-patient conversations had higher levels of communication competence as defined by the Munich Model of Communication Competence than a control group.

The fourth advantage of the model is that it is tailored specifically for parent-teacher conversations and thus, takes into account the importance of context. Still while being context-specific, the context is formulated more broadly than in Hertel's (2009) or Bruder's (2011) models. Nevertheless, it integrates all aspects of their models as displayed in Table 1.

Table 1

Allocation of the competence facets of the models by Hertel (2009) and Bruder (2011) to the competence facets of the Munich Model of Communication Competence in Parent-Teacher Conversations

Gartmeier and colleagues	Interpersonal relationship	Problem solving	Structuring the conversation
Bruder	<p>Collaboration / perspective taking</p> <p>(cooperative actions, perspective taking/ resources orientation)</p> <p>Coping</p> <p>(coping with criticism, dealing with difficult situations)</p> <p>Counselling skills</p> <p>(active listening)</p>	<p>Diagnostic / pedagogical skills</p> <p>(defining the problem, strategy, knowledge and goal orientation)</p>	<p>Counseling skills</p> <p>(structuring and paraphrasing)</p>
Hertel	<p>Social collaboration</p> <p>(cooperative behavior)</p> <p>Coping</p> <p>(coping with criticism, dealing with difficult situations)</p>	<p>Pedagogical skills</p> <p>Process skills</p> <p>(knowledge and adaption of strategies)</p>	<p>Counseling skills</p> <p>Personal resources</p> <p>(task-monitoring)</p> <p>Process skills</p> <p>(knowledge and adaption of strategies)</p>

The broader formulation of the Munich Model of Communication Competence in Parent-Teacher Conversation allows the application of the model to a variety of different parent-teacher conversations from shared decision making conversations regarding learning strategies, homework or the choice of elective subjects, to handling complaints in conversations

with regard to bad grades, difficulties with classmates or absenteeism. According to Gartmeier et al. (2011), it can also be transferred to other domains where expert-layperson conversations take place, e.g. to physician-patient conversations in the medical domain. The wide applicability is a great advantage and pays tribute to the variability of parent-teacher conversations and the importance of communication competence in different domains. Yet, it remains to be empirically tested whether the model is still sufficiently context-specific.

3.3 Summary

Communication can take a variety of forms and is an important, interdisciplinary research topic. Parent-teacher conversations fall into the domain of educational research and, more specifically, into the category of expert-layperson conversations. With regard to expert-layperson conversations, research has shown that it is important to establish common ground between experts and laypersons, e.g. via grounding techniques (Bromme et al., 2004) and to construct a positive conversational climate by communicating congruently, empathetically and with unconditional positive regard (Rogers, 1951). Additionally, it is helpful to structure (expert-layperson) conversations transparently, e.g. by following scripts (Schank & Abelson, 1977).

In order to measure competencies, they need to be modelled theoretically. Theoretical constructs of competencies should be adequately context-specific and their structure needs to be differentiated (Hartig, 2008). This applies also to communication competence (Berger, 2008), since communication competence per se is too broad a concept to work with (Greene & Burleson, 2003). Consequently, in order to construct measurement instruments and trainings for the communication competence in parent-teacher conversations, the competence of teachers regarding conversations with parents has to be theoretically modelled.

Up to now, there is only very few empirical evidence for models that target the competence of teachers to communicate with parents. The models by Hertel (2009) and Bruder (2011) pioneer in providing empirical evidence regarding the structure of teachers' communication competence in conversations with parents. However, Hertel's (2009) and Bruder's (2011) context is comparatively narrow. They conceptualize counseling competence of teachers only regarding conversations with parents about learning strategies and learning difficulties (Bruder, 2011; Hertel, 2009). Still some of their empirical findings might be transferable to conversations with parents in general. Among other things, the findings

suggest that the competence of teachers to communicate with parents is multidimensional (Bruder, 2011; Hertel, 2009). Hertel (2009) considers a two or three faceted model as most probable.

In line with the findings of Hertel (2009), the Munich Model of Communication Competence in Parent-Teacher Conversation conceptualizes communication competence as a three-dimensional hierarchical construct. The model is theoretically founded in research about expert-layperson conversations. In contrast to the models of Hertel (2009) and Bruder (2011), it has the advantage that it is context-specific but still applicable to a variety of parent-teacher and other expert-layman conversation situations.

Chapter 3 has shown that competencies need to be defined context-specific and with regard to their internal structure in order to be able to construct measurement instruments for diagnosing a certain competence, such as the competence of teachers to conduct conversations with parents targeted in this dissertation. The Munich Model of Communication Competence in Parent-Teacher Conversation has several advantages, such as a relatively wide range of applicability while still being context-specific and a subdivision into three competence facets that corresponds to prior research findings (Hertel, 2009). Therefore, the Munich Model of Communication Competence in Parent-Teacher Conversations will serve as the theoretical foundation for the construction of the instruments in the empirical part. Possible instruments for as well as goals and challenges of the measurement of the communication competence of (pre-service) teachers in conversations with parents are discussed in more detail in chapter 4 as a theoretical basis for the instrument development process in the method part.

4 Measuring Communication Competence of Teachers in Conversations With Parents

The aim of this chapter is to single out how to measure teacher competencies in general and in particular, the communication competence of teachers in conversations with parents. To inquire into these questions, the chapter starts with an evaluation of the central goals and challenges of competence measurement regarding research on teachers and teacher education. Subsequently, it is discussed how multimethod measurements can contribute to the validation process of new instruments, such as the simulated conversations developed in this dissertation. In order to single out instruments that could be part of a multimethod measurement that diagnoses the communication competence of teachers in conversations with parents and provides convergent and discriminant evidence for the validity of the newly-developed simulated conversations, an overview of four methods that are used to assess teachers' professional competencies is provided. The advantages and disadvantages of the four methods as well as their suitability for diagnosing the communication competence of teachers in conversations with parents are discussed. For the sake of completeness, simulated conversations – though not yet employed to measure teachers' competences - are also briefly discussed and compared to the other approaches with regard to their suitability for diagnosing the competencies of teachers to communicate with parents. Information on the state of research on simulated conversations needed for the development of the simulated conversations in this dissertation is provided in detail in chapter 5.

4.1 Goals and Challenges of Research on Competence Measurement in Teacher Education

Teacher education has to be consequently based on research in order to secure its quality and further development (Prenzel, 2009). While up to the year 2000 there was comparatively little research-based evidence on teachers and teacher education, this situation has changed in recent years (Prenzel, 2013). In the course of this development, several studies, which target the measurement of teachers' competencies, have been conducted (Kunter & Klusmann, 2010). Thereby, challenges regarding the measurement of teacher competencies have crystallized out. In particular the development of adequate instruments is crucial for

measuring (teacher) competencies (Koeppen et al., 2008). These instruments should be based on theoretical (cf. chapter 3) and psychometric models (cf. part D) and should take into account the intended usage of the measurement results (Klieme, Hartig, & Rauch, 2008; Koeppen et al., 2008).

When developing instruments one should take into account that competencies are complex and not directly testable. Instead, they have to be extrapolated from the performance in a corresponding test situation (Erpenbeck & von Rosenstiel, 2003) and can, thus, only be measured in a context-specific and performance-oriented manner (Koeppen et al., 2008; Maag Merki & Werner, 2011). Measurements should consequently either be conducted under real conditions or the selected context should be perceived as authentic by the test persons (Shavelson, 2012). Moreover, the resulting measurement score should allow the prediction of how the test person will solve real professional tasks (Hartig & Klieme, 2006; Klieme et al., 2008).

Principally, two main intended usages of measurement results can be differentiated: “assessment for learning” / “formative assessment” and “assessment of learning” / “summative assessment” (Schuwirth & van der Vleuten, 2011, p. 478). When assessing for learning, the assessment process is inextricably embedded within an educational process (Schuwirth & van der Vleuten, 2011). The assessment is maximally information-rich and serves to steer and foster the learning of each person to the maximum of his or her ability (Schuwirth & van der Vleuten, 2011). In contrast, assessment of learning takes place at or sometime after the end of instruction (Schuwirth & van der Vleuten, 2011). The purpose of assessment of learning is to determine whether a person has acquired sufficient knowledge or competencies for a certain purpose, e.g. when leaving university or entering a profession (Schuwirth & van der Vleuten, 2011). Furthermore, it is important to determine at which level information about the competence level is needed (Koeppen et al., 2008). In the pedagogical practice, e.g. in school or in teacher education, information is often needed at an individual level in order to give differentiated feedback regarding learning success to individual students / persons in order to foster their learning processes. Information at an individual level is also needed for selection processes, like choosing candidates for a study program, a vacant position or the issuance of a driving license. With regard to research, information is often needed at an aggregated level, e.g. in order to evaluate the effectiveness of seminars, study programs or interventions.

This subchapter has shown that instruments are the key to reliable and valid measurements. Multimethod measurements can provide information about the validity of individual measurement instruments and often allow the gaining of a more precise and complete picture of constructs. Thus, multimethod measurements are discussed in the next subchapter.

4.2 Multimethod Measurements

Multimethod measurements can play an important role in the validation process of measurement instruments since they can provide convergent and discriminant evidence for validity (Eid & Diener, 2006). Multimethod measurements go back to classical test theory, Brunswick's work on probabilistic functionalism, Jöreskog's approach to covariance structural equation modeling and Campbell's and Fiske's multitrait-multimethod-matrix (Schmitt, 2006). The core assumption behind multimethod measurements is that each measurement reflects on the one hand, the construct that is measured and on the other hand, the influence of the measurement method (Eid & Diener, 2006). Hence, in multimethod measurements different instruments are combined to measure one construct (Eid & Diener, 2006).

The multimethod measurement approach has two advantages. The first advantage is that results from different other measurements can be part of the validation process for a (newly-developed) instrument (Eid & Diener, 2006). A correlational pattern between the instrument and the other measurements which corresponds to theoretical expectations is an indicator of the validity of the measurement (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; Eid & Diener, 2006). With regard to this, it is important to differentiate that multimethod measurements can take place at two levels. The first level is the situation in which the measurement takes place. Several instruments can be applied in the same situation in order to provide different perspectives on the measured construct; e.g. it can be compared as how competent the parents with whom teachers have conducted a parent-teacher conversation perceive the teachers in comparison to independent observers that assess the same conversation. The second level is broader. It comprises all kinds of measurements that are closer or remotely related to the construct at stake (Eid & Diener, 2006). Continuing with the last example, measurements with a knowledge test on parent-teacher conversation or external criteria, such as previous knowledge of the teachers, could complement the assessments simulated parents and observers made of the teachers' competence in the

parent-teacher conversations. Both measurements undertaken in the narrower and broader context of a multimethod measurements can provide convergent and discriminant evidence for the validity of the measurement, depending on the form and content of the measurement. The crucial point in the validation process is to set up a hypothesis about the correlational structure and the expected strengths of the correlations between the different measurements and to then verify in how far the real correlational structure corresponds to theoretical expectations (American Educational Research Association et al., 2014).

The second advantage of multimethod measurements is that specific method effects of measurements can be singled out (Eid & Diener, 2006). In this way, the disadvantages of individual measurement methods can be compensated and the measurement of the construct can be improved with the resulting information (Eid & Diener, 2006). However, it is important to be aware of the fact that when human judgment is central in the combination of information from different measurements, the expertise of the person who is making judgments is decisive for the quality of the outcome (Schuwirth & van der Vleuten, 2011). Second opinions and careful note taking can enhance the quality of this judgment process (Schuwirth & van der Vleuten, 2011). Moreover, when compiling instruments for a multimethod measurement, it has to be ensured that the different instruments do not influence or interfere with each other and are adequate for measuring the construct. A last point that has to be taken into account is that multimethod measurements are comparatively time- and cost-intensive due to the inclusion of several measurement methods.

Multimethod measurements have become increasingly widespread in recent years since they provide information about the validity of individual measurements, allow the singling out of method effects and permit to get a more comprehensive and precise picture of competencies (Erpenbeck & von Rosenstiel, 2003; Hertel, 2009; Shernoff & Kratochwill, 2004). However, due to the cost-intensiveness and the expertise required to construct multimethod measurements, it should be carefully weighed in which situations multimethod measurements provide an additional benefit and should be used. An additional benefit through multimethod measurements is, e.g., to expect when information about the validity of a (new) instrument is needed or when multicomponent or multilevel phenomena are the target construct to be measured (Eid & Diener, 2006); e.g. interviews about emotions might provide other information than an analysis of facial expressions and / or a measurement of the endorphin level in the body. For a variety of other constructs, such as

knowledge about a certain topic, one instrument might provide an adequate amount of information in a more economical way.

Multimethod measurements in the broader and narrower sense have also been employed to assess the communication competence of teachers in conversations with parents. Hertel (2009) combined two kinds of self-assessments with work samples, participant observation in role-plays and expert ratings of the performance in role-plays. Bruder (2011) combined self-assessments with a knowledge test, case scenarios and a situational judgment test. Aich (2011) combined external ratings of the performance in role-plays with self-assessments and assessments of the conversational partner.

Multimethod measurements can contribute to the validation process of instruments. Thus, the newly-developed simulated conversations will be embedded in a multimethod measurement in the empirical part of this dissertation in order to gather convergent, discriminant and criterion-based evidence for their validity. In order to single out instruments that could be part of this multimethod measurement, an overview of five methods that are currently employed to measure (teacher) competencies is given. The advantages and disadvantages of the instruments and their suitability for assessing the competence of (pre-service) teachers to conduct conversations with parents are discussed.

4.3 Instruments for Diagnosing the Competence of Teachers to Communicate With Parents

There is consensus that there is a need for reliable and valid instruments that can assess the competence of (pre-service) teachers to communicate with and counsel parents (Bruder, 2011; Hertel, 2009; McLeod, 2003; Scofield & Yoxheimer, 1983). However, there are no universally suited instruments for the assessment of interpersonal communication competencies (Spitzberg, 2008) since their suitability and sometimes their advantages and disadvantages depend on several framework conditions and the purpose of the measurement. Scheibe, Trittel, Klug and Schmitz (2014) outline five criteria that should be considered on top of the research question when choosing a measurement method: 1) the examination object, 2) practical restrictions, 3) ethical considerations, 4) requirements regarding measurement reliability, and 5) the behavior modality of the dependent variable (e.g. reactions that are immediately observable vs. verbal estimations).

Concerning the behavior modality of the dependent variable, a classification by Shernoff and Kratochwill (2004) that orders techniques within behavioral assessment on a dimension of directness is helpful. Shernoff and Kratochwill (2004) define directness as the extent to which methods measure relevant behavior directly at the time and place of its natural occurrence. The more direct the observation, the lower the degree of inference necessary to extrapolate the target competence from the measurement results and the more direct the instrument is considered (Shernoff & Kratochwill, 2004). Shernoff and Kratochwill (2004) differentiate five behavioral assessment methods (presented in order of directness): direct observation, self-monitoring, analogue assessment, behavioral interviews and questionnaires. These methods can be used individually or information from different measurements, in the same situation / for the same construct, can be combined.

In the following the classification of Shernoff and Kratochwill (2004) is modified and amplified by an inclusion of frequently employed sub manifestations of certain methods, such as role-plays as one type of analogue assessments. The methods of the modified classification are presented in order of directness starting with the most indirect ones and modifications to the original classification are explained. The suitability of the methods for measuring the communication competence of teachers in conversations with parents is discussed as well as their strengths and weaknesses, taking into account the criteria for choosing a measurement method suggested by Scheibe and colleagues (2014).

4.3.1 *Questionnaires*

Questionnaires and interviews that involve self-reports are classified as the most indirect methods of assessment by Shernoff and Kratochwill (2004). In both of these methods, a particular behavior is reported so that the measurement result is a verbal representation of a relevant activity taking place at some other time or place (Shernoff & Kratochwill, 2004). Self-assessment questionnaires are the most frequently used method for measuring competencies (Kunter & Klusmann, 2010; Maag Merki & Werner, 2011) since they are economical and efficient and allow researchers to collect data of large samples (Frey & Balzer, 2005). In addition, they are comparatively objective and reliable and allow for the making of comparisons (Scheibe et al., 2014). However, self-assessments do not always correspond to directly observable behavior (Hertel, 2009). They are vulnerable to reporting bias, halo effects, reporting extremes and central tendencies (Eid & Diener, 2006; Shernoff & Kratochwill, 2004).

Aich (2011), Bruder (2011) and Hertel (2009) employed self-assessment questionnaires to assess the competence of teachers to communicate with parents. In his study, Aich (2011) reports a considerable difference between teachers' self-assessments of their competencies and their preparation for parent-teacher conversations through their education. Teachers do not feel adequately prepared for parent-teacher conversations by their education but believe that their competencies are good nevertheless (Aich, 2011). Aich (2011) attributes this result to a reporting bias due to the pressure and critique teachers feel regarding their competencies. He sees this as an indicator of the problems with self-assessments (Aich, 2011). Like Aich (2011), Hertel (2009) reports that self-assessments and observer ratings did not always concord in her study. Consequently, Aich (2011) as well as Bruder (2011) and Hertel (2009), complement self-assessments with other measurement methods.

4.3.2 *Interviews*

Similar to questionnaires, interviews can help assess an interviewee's beliefs and perceptions about a certain competence but do not allow direct competence assessment and are thus also considered as an indirect method (Shernoff & Kratochwill, 2004). Interviews differ at the level of structure and focus (Shernoff & Kratochwill, 2004). They can be conducted with single persons or groups, guideline-based or open (Maag Merki & Werner, 2011). Advantages of interviews vary with their degree of standardization (Scheibe et al., 2014). The more standardized interviews are, the more reliable and the less flexible they are. Interviews are not frequently employed in order to measure competencies since they are time-consuming (Maag Merki & Werner, 2011) and the interviewer may introduce bias and cause persons to answer in a socially desired way and may thus be a risk for objectivity and reliability (Bortz & Döring, 2006).

Subjective measurement methods, such as questionnaires and interviews, are suboptimal to measure competencies if they are not complemented by other methods (Hertel, 2009). Data are gathered retrospectively, require higher levels of inference and are no objective accounts of performance (Shernoff & Kratochwill, 2004). Consequently, it is controversially discussed whether self-assessments truly measure competencies or rather give us information about teachers' self-concepts or perception of competencies (Klieme & Hartig, 2007). Still, in combination with other methods they can provide valuable information, since the tested person is an expert for his or herself and self-assessments provide one perspective on the targeted construct.

Interviews, as well as questionnaires, can not only be used to have test persons self-assess their competencies but also to have the conversational partners rate their satisfaction with the conversation and the perceived competence of the (pre-service) teacher. Like self-assessments, judgments by the conversational partners are no objective accounts of observed behavior, but rather give information about whether the conversational partner is satisfied with the conversation and perceives the (pre-service) teacher as competent. Still, this second perspective provides additional information and is likely to be a little more direct and objective than self-assessments.

4.3.3 *Observation in Analogous Situations*

Analogue behavioral observation involves a situation designed, manipulated, or constrained by an investigator that elicits a measured behavior of interest (Shernoff & Kratochwill, 2004). A real-life situation is mimicked that triggers a target behavior and allows assessing this behavior under controlled conditions (Shernoff & Kratochwill, 2004). The real-life problem-solving or decision-making situations require test persons to develop solutions that involve the application and integration of knowledge, as well as multiple skills and strategies (Fuchs & Fuchs, 2000). In this way analogous situations provide authentic assessment conditions in which behavior and performance can be directly observed (Shernoff & Kratochwill, 2004). Thus, analogous assessments minimize the inferences needed to assess competencies (Shernoff & Kratochwill, 2004). However, analogous situations are like reality television. They provide an approximation of how real people would behave when placed in situations that are analogous to those in the real world and provide indicators for how an individual might react or behave in real life situations (Shernoff & Kratochwill, 2004). Thus, the generalizability of the behavior in analogous situations to that in natural settings is of concern (Shernoff & Kratochwill, 2004).

The major advantages of analogue behavioral assessments are that they measure performance in a comparatively authentic situation and allow meaningful real-world implications, e.g. with regard to the later profession (Shernoff & Kratochwill, 2004). Due to the authentic measurement conditions, the validity of the interpretation of their results is supposedly comparatively high. Moreover, analogous situations make it often possible to observe otherwise unobservable behavior (Shernoff & Kratochwill, 2004) and spare “real” persons from harm. However, instruments based on analogous situations are time- and

cost-intensive and understanding their psychometrics is critical (Shernoff & Kratochwill, 2004).

Analogous behavioral assessments exist in different variants. Three of these variants that are employed to assess or are promising for assessing the communication competence of teachers in conversations with parents are discussed in the following.

Paper and Pencil or Computer-Based Analogous Assessments

There are several written or computer-based assessments, such as case-based scenarios, key feature tests or situational judgment tests that are aimed at measuring competencies in a performance-based way in the context of prototypical situations (Scheibe et al., 2014). A situational judgment test, for example, consists either of short written case vignettes or video clips to which the test person has to take a stance based on prescribed answer alternatives (Scheibe et al., 2014). In contrast, a case scenario is constructed similarly but is usually less standardized and has open-ended questions (Bruder, 2011). There is a lot of research about situational judgment tests that indicate that the results in situational judgment tests are linked to real-life performance (Whetzel & McDaniel, 2009). This might be a reason for the acceptance / perceived authenticity by the test takers and explain the comparatively high validity of situational judgment tests (Whetzel & McDaniel, 2009).

Bruder (2011) developed a situational judgment test for assessing the counseling competence of teachers in conversations with parents. She was able to show that results in the situational judgment test were moderately to highly correlated to self-assessments of competence regarding real parent-teacher conversations (Bruder, 2011). For future research, she recommends investigating correlations with observer-rated performance in real conversations (Bruder, 2011).

Paper and pencil or computer-based performance-oriented assessments have the advantage that due to the realistic settings, interpretations of measurement scores are comparatively valid (Scheibe et al., 2014). However, their construction and scoring requires a lot of effort and objectivity of scoring often tends to be low (Scheibe et al., 2014) and this can possibly decrease reliability.

Role-plays

Like in paper and pencil or computer-based analogous assessments, in role-plays test participants are relocated into a situation that is analogous to a real professional situation. However, in role-plays test persons can neither select appropriate behavior from prescribed

answers nor do they have time to think about the correct answer. Instead, test persons have to show the appropriate competent behavior and have to react in real time.

Gallagher and Hargie (1989) aimed at determining the validity of role-plays to assess counselor skills. In their study with twelve counselors they found little difference between real and role-play situations and concluded that this is an indicator of the validity of the results in role-plays (Gallagher & Hargie, 1989). Hertel (2009) also employed role-plays to measure the counseling competence of teachers and pre-service teachers in conversations with parents. She found that for measuring the counseling competence of teachers, role-plays can provide valid results (Hertel, 2009). In contrast, role-plays did not provide valid measurement results for the pre-service teachers in her sample because the pre-service teachers did not perceive the role-play situations as authentic (Hertel, 2009). The perceived authenticity of role-plays seems to be a factor that can influence whether role-plays provide valid measurement results (Gallagher & Hargie, 1989; Hertel, 2009).

Aich (2011) also employed role-plays in order to measure teachers' competencies to communicate with parents. He compared the role-play results with expert-rated performance in audios of real parent-teacher conversations (Aich, 2011). Both the role-plays and the real audio-taped conversations depicted an increase in competence after communication training (Aich, 2011). However, the measurement results diverged significantly with competence levels in the real conversations being higher (Aich, 2011). As a possible explanation, Aich (2011) suggests that motivation is higher in real conversations. He recommends using videos instead of audios in order to additionally include non-verbal parameters in the future (Aich, 2011).

Simulated Situations / Conversations

Similar to role-plays, simulated situations serve to relocate persons into a situation which is analogous to a real situation. The difference is that in simulated situations, the degree of authenticity is usually higher since simulated situations are created in a way that is closer to reality. For example, role-play peers are substituted through trained actors. One type of simulated situations is simulated conversations. Simulated conversations with actors trained to portray a certain patient role are a very frequently used method to train and assess diagnostic and communication competence in the medical domain. A considerable amount of research in the medical domain has shown that simulated conversations are well-accepted, perceived as authentic and do fulfill psychometric quality criteria if constructed properly (Barman, 2005; Cleland et al., 2009; Newble, 2004). Simulated conversa-

tions seem to be a promising approach for measuring the communication competence of teachers in conversations with parents since they offer highly authentic measurement conditions and are performance-oriented. In order to scrutinize the potential of simulated conversations for assessing the competence of teachers to communicate with parents, the state of research on simulated conversations in the medical and the educational domain is discussed in detail in chapter 5.

Lane and Rollnick (2007) reviewed the literature on the use of simulated conversations and role-plays in medical training and compared the perceived authenticity and the acceptance of these two methods. There is only one study by Papadakis, Croughan-Minihane, Fromm, Wilkie and Ernster (1997) that directly compares role-play with simulated conversations (Lane & Rollnick, 2007). It focuses on simulated conversations and role-plays regarding the training of communication competence (Papadakis et al., 1997). Papadakis and colleagues (1997) found no significant difference in terms of increase in communication competence, but those students who consulted with a simulated patient rated the experience higher and more authentic than those who carried out role-plays with colleagues (Papadakis et al., 1997). This result is probably transferable to the measurement of communication competence with role-plays and simulated conversations. In sum, the literature review by Lane and Rollnick (2007) shows that simulated conversations are generally very well-accepted and considered authentic by pre- and in-service physicians whereas they are rather reluctant to take part in role-plays.

4.3.4 *Direct Observation of Behavior*

For direct observation of behavior, behavior is rated in a natural setting and at the time it occurs (Shernoff & Kratochwill, 2004). It is usually rated by trained impartial observers (Shernoff & Kratochwill, 2004). Recommendations for the assessment of communication competence favor direct observation of performance with predefined structured coding manuals according to a theoretical framework (Kauffeld, 2005; van der Vleuten et al., 2010). Coding manuals can require a low or high amount of inference depending on the type of items included in the coding manual (cf. chapter 5.1.3.4). Videotaping conversations allows an undisturbed course of action in the observed situation and provides the opportunity to repeat watching certain sequences in order to capture the complexity of an action (Maag Merki & Werner, 2011).

The direct nature of the observation of a behavior / performance has - independent from the nature of the coding manual - a clear advantage over procedures that require a higher level of inference regarding competencies (Shernoff & Kratochwill, 2004). Usually, the validity is higher (Shernoff & Kratochwill, 2004). However, there are also certain disadvantages; the context of real situations remains uncontrollable (Hertel, 2009), the observer can cause bias (Shernoff & Kratochwill, 2004), the training and the rating is demanding for the observers (Scheibe et al., 2014) and often - especially when relying on high inference coding manuals - the results are not as reliable as in more standardized assessment methods (Shernoff & Kratochwill, 2004). Moreover, direct observation has high demands on personnel and financial resources and is sometimes legally impeded.

Self-Monitoring

Self-monitoring is a special type of direct observation of behavior in which the test person observes and systematically protocols her own behavior normally immediately after it occurred (Hertel, 2009), e.g. on a checklist or via frequency ratings in a record booklet. Thus, self-monitoring is classified as a subtype of direct observation in the modified classification and not as a distinct measurement method as suggested by Shernoff and Kratochwill (2004). Self-monitoring has the advantage that it increases attention to one's own behavior and can function either as positive reinforcement or can cause behavior to change (Shernoff & Kratochwill, 2004). However, self-monitoring requires that an individual is able to identify and record occurrences of the targeted behavior (Shernoff & Kratochwill, 2004) and often does not fulfill psychometric quality criteria (Frey & Balzer, 2005). In addition, reactivity, i.e. unintended influences that result from self-recording occur (Shernoff & Kratochwill, 2004) and limit its application field. Self-monitoring is probably not ideally-suited for parent-teacher conversations since it would distract teachers to self-monitor their behavior during the conversation.

Figure 3 gives an overview of all presented methods including sub manifestations and the modified classification. In subchapter 4.3.6, advantages and disadvantages of the different instruments for measuring the professional competencies of teachers are summarized to lay the basis for the choice of instruments that can diagnose the competence of teachers to communicate with parents (in chapter 4.4) and that are part of the validation process of the simulated conversations in this dissertation.

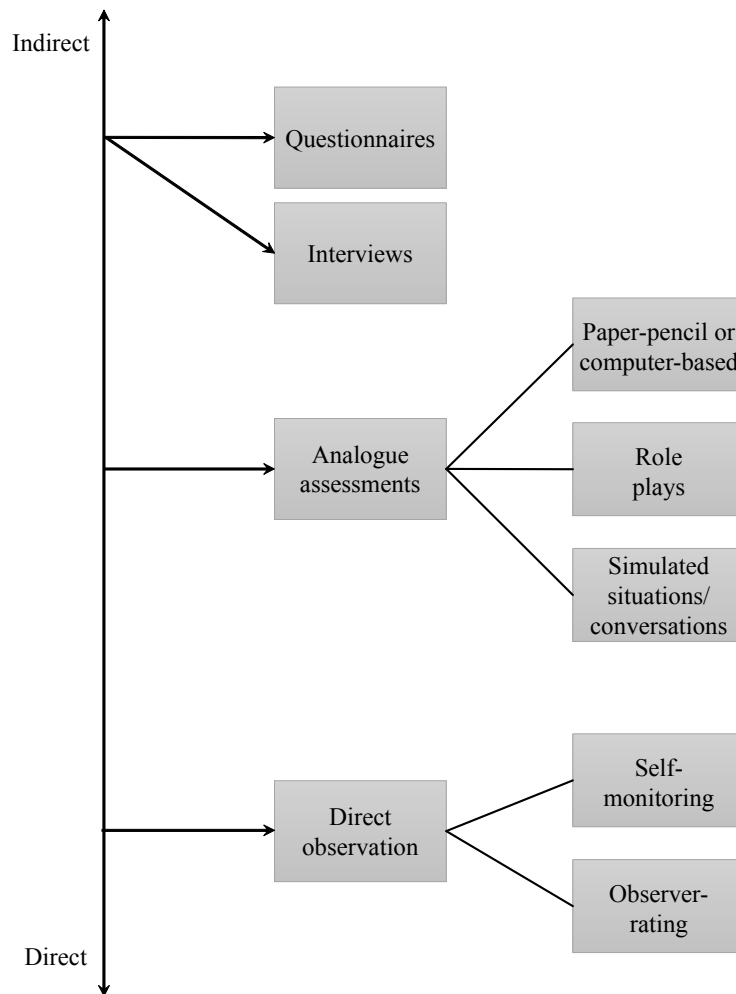


Figure 3. Classification of competence assessment methods ordered along a continuum of directness (adapted from Shernoff and Kratochwill, 2004, p. 371)

4.3.5 *Advantages and Disadvantages of Instruments for Diagnosing the Professional Competencies of Teachers*

The majority of studies on teacher competencies are based on indirect approaches, such as questionnaires or biographical data (Frey & Balzer, 2005). However, self-assessments and other rather indirect instruments cannot capture competencies in their entire complexity (Hertel, 2009). Since the interest and research on competence measurement of teachers has risen substantially, the definition of what constitutes a standardized instrument or measurement procedure has broadened significantly over the last few decades (American Educational Research Association et al., 2014). Various kinds of more direct performance assessments, such as simulations, have been developed to provide measures of constructs that might otherwise be difficult to assess (American Educational Research Association et

al., 2014). Yet, each step towards greater flexibility in the assessment procedure enlarges the scope of the variations allowed in replications over the testing procedure and therefore, tends to increase measurement error (American Educational Research Association et al., 2014). Nevertheless, some of these sacrifices in reliability may reduce construct irrelevance or construct underrepresentation and thereby improve the validity of the intended interpretations of scores (American Educational Research Association et al., 2014). For example, performance assessments that depend on ratings of extended responses tend to have a lower reliability than more structured assessments, such as questionnaires, but they can sometimes provide more direct measures of the attribute of interest (American Educational Research Association et al., 2014).

In sum, there is no such thing as a universally well-suited approach to measuring competencies (Kunter & Klusmann, 2010) but advantages and disadvantages of individual approaches have to be weighed with regard to the purposes and the requirements of the assessment. Ideally, an appropriate combination of instruments for a multimethod measurement should be compiled, which allows compensating the drawbacks of individual instruments and provides information about the validity of single instruments.

4.4 Summary

Instruments that can capture competencies in a contextualized and performance-oriented way are at the heart of research about teacher competencies (Koeppen et al., 2008; Maag Merki & Werner, 2011). Measurement instruments and methods, such as questionnaires, interviews, analogue assessments and direct observation, can be classified according to their directness, i.e. the degree of inference needed to extrapolate competencies from the measurement results (Shernoff & Kratochwill, 2004). Direct and indirect methods have distinct advantages and disadvantages. The more direct measurement methods are, the higher their validity tends to be. In contrast, the more indirect methods are, the more objective, reliable and time / money-efficient they tend to be.

Direct observation of behavior and self-monitoring in natural settings are classified as the most direct methods by Shernoff and Kratochwill (2004). Measurement results of direct observation by independent observers are usually highly valid. However, due to data protection laws, it is usually not possible to directly observe teachers in conversations with parents. Moreover, real parents would neither always represent the spectrum of cases which should be taught in teacher education nor necessarily be available as required or

willing to take part in teacher education. Self-monitoring in parent-teacher conversations is also not possible because teachers are busy conducting the conversation and are not able to self-monitor themselves or record behavior without producing reactivity. Consequently, observation of teachers' competence to communicate with parents has to take place in analogous situations, like simulated conversations, or based on more indirect methods.

Analogous situations, as well as real settings, are promising for measuring the competence of (pre-service) teachers to communicate with parents. They allow researchers to observe otherwise unobservable behavior, offer measurement conditions close to reality and are performance-oriented and thus, also tend to produce highly valid results. Simulated conversations are a particularly promising form of analogue assessment since they are not only performance-related and provide authentic and contextualized measurement conditions, but they are also very well / better accepted than role-plays and there is evidence from the medical domain indicating that simulated conversations are well-suited to diagnose communication competence (cf. chapter 5). Additionally, in comparison to real parents who differ from each other with regard to personality, socioeconomic background, attitudes towards school etc., simulated parents can be trained to consistently perform a certain parent role which increases the reliability of the measurement. For these reasons, simulated conversations are employed to measure the communication competence of teachers in conversations with parents in this dissertation. The conversations are videotaped since this provides the opportunity to repeat watching certain sequences in order to capture the complexity of an action (Maag Merki & Werner, 2011). Subsequently, they are rated by external observers since independent observer ratings have a higher validity than self-assessments or ratings by the conversational partners (cf. 4.3.1).

Multimethod measurements can play an important role in the validation process of measurement instruments, such as the simulated conversations developed in this dissertation, since they can provide evidence for an instrument's validity (Eid & Diener, 2006). In multimethod measurements, several instruments are combined to measure one construct (Eid & Diener, 2006). The degree to which the correlational pattern and strengths between an instrument and other measurements corresponds to theoretical expectations is indicative of its validity (American Educational Research Association et al., 2014). In order to compile evidence for the validity of the simulated conversations developed in this dissertation, they will be embedded in a multimethod measurement, i.e. complemented with other instruments discussed in this chapter. The designed multimethod measurement comprises two

levels. The first level is the situation in which the measurement takes place, the parent-teacher conversation. In addition to the independent observer ratings of the performance of the pre-service teachers in the simulated conversations, the pre-service teachers and the simulated parents also assess the performance of the pre-service teachers in questionnaires. While self-assessments and ratings by the conversational partner tend to be less valid than observer ratings (Aich, 2011; Hertel, 2009), they can provide two further perspectives and evidence for validity. Moreover, independent observers rate a subsample of the simulated conversations with six different coding manuals with different forms and content to provide further convergent and discriminant evidence for the ratings of the simulated conversations with the newly-developed coding manual (cf. chapter 7.7). The second level of the multimethod measurement goes beyond the actual parent-teacher conversation. Since Bruder (2011) was able to show that situational judgment tests can capture the communication competence of teachers in conversations with parents, a situational judgment test is chosen as a fifth component of the multimethod measurement in order to provide convergent evidence for the validity of the simulated conversations. Lastly, the results in the simulated conversations are compared to supposedly more or less influential external criteria, such as previous knowledge, in order to provide convergent and discriminant criterion-based evidence for validity. For an overview of all components of the multimethod measurement see Figure 7 in the method part. The hypothesized correlational structure is outlined in chapter C.

5 State of Research on Simulated Conversations

To date, simulated conversations have predominantly been used in the medical domain in order to teach and assess the clinical and communication skills of (pre-service) physicians. For the purpose of applying simulated conversations to the context of parent-teacher conversations and to understand future possibilities of simulated conversations regarding teacher education and research on teacher education, a brief sketch of how this method has been applied in the medical context is useful to clarify terminology, different potential uses and research findings on how the different components of simulated conversations should be designed (5.1).

Since costs and efforts to construct and employ simulated conversations are comparatively high, the use of simulated conversations has only recently spread to other domains. One of those is the educational domain. In the second part of chapter 5, current developments regarding simulated conversations in the educational domain are outlined and, in this way, future application and research areas are identified (5.2). Moreover, similarities and differences regarding simulated conversations in the educational and medical domain are discussed as a basis for the development and use of simulated conversations in the educational domain.

5.1 Simulated Conversations in the Medical Domain

This subchapter starts with a sketch of the history of simulated conversations in medical education in order to make current developments in the educational domain assessable. Since terminology regarding simulated conversations varies geographically and content-relatedly, next, frequent terminology is discussed and the use of terminology for this dissertation is defined. Simulated conversations consist of different components all of which contribute to and influence their adequacy and accuracy. Thus, possible influences as well as resulting requirements regarding the design of the components of simulated conversations, such as cases, actors, raters and coding manuals, are discussed in the following sections as a basis for the construction of simulated parent-teacher conversations in the method part.

5.1.1 *History of Simulated Conversations in the Medical Domain*

Medical training has traditionally depended on contact with patients (Cleland et al., 2009). However, deploying real patients for the training and assessment of (pre-service) physicians has a variety of disadvantages. Patients who stay at university and academic hospitals do not always represent the full spectrum of diseases which needs to be taught in medical education and are sometimes reluctant to participate in the formation of doctors (Cleland et al., 2009). Moreover, patients differ from each other with regard to severity of diseases as well as personality which wears on the reliability of assessments (Collins & Harden, 1998). Lastly, due to ethical reasons, the deployment of real patients into clinical teaching is sometimes inappropriate (e.g. when practicing to give a terminal diagnosis) and patients have to be protected from unnecessary harm (Gaba, 2004). Thus, in 1964, Barrows and Abrahamson introduced simulated patients to support clinical skills learning (Barrows & Abrahamson, 1964). According to Barrows (1987), “the Simulated / Standardized Patient (SP) is a person who has been carefully coached to simulate an actual patient so accurately that the simulation cannot be detected by a skilled clinician. In performing the simulation, the SP presents the gestalt of the patient being simulated; not just the history, but the body language, the physical findings, and the emotional and personality characteristics as well” (p. 17).

The simulated patients method has several advantages; simulated patients can portray a broad range of cases that students may not encounter in real patients, they are willing to undergo scenarios many times, their behavior is predictable, they can be trained to match their role to the student’s level of experience, they are available as and when required, are well-accepted by pre- and in-service physicians and can give feedback to medical students about their performance (Cleland et al., 2009). Thus, during the following years, the use of simulated patients increased and the method was developed further, especially with regard to assessment purposes (Cleland et al., 2009).

In 1975, dissatisfaction with the traditional assessment methods of clinical skills, such as essays or oral exams, on the part of teachers as well as students (Barman, 2005), led to the introduction of the Objective Structured Clinical Examination (OSCE) by Harden and colleagues (1975). OSCEs are very frequently used today. They are an examination format in which examinees rotate around a circuit of clinical task stations (Harden et al., 1975). OSCE tasks may or may not require interaction with simulated patients but mostly do

(Iramaneerat, Yudkowsky, Myford, & Downing, 2008). The dissemination of OSCEs further increased the employment of simulated patients.

During the 1990s, the number of US medical schools that employed simulated patients rose significantly to around 80% (May, Park, & Lee, 2009, p. 487). In consequence in 2001, the Association of Standardized Patient Educators (ASPE) was founded, aiming at promoting best practices in the application of simulated patient methodology for education, assessment and research and fostering the dissemination of research and scholarship in the field of simulated patient methodology (Association of Standardized Patient Educators, 2014). Three years later, in 2004, encounters with simulated patients became a compulsory part of the US Medical Licensing Examinations (United States Medical Licensing Examination, 2015). One part of the examination consists of the assessment of communication skills (United States Medical Licensing Examination, 2015). In consequence, communication skills are taught in nearly all medical schools in the US (Lurie, Mooney, Nofziger, Mel-drum, & Epstein, 2008) and simulated patients are used nation-wide today.

In Europe and particularly in Germany the employment of simulated patients started around 40 years later than in the US (Ortwein, Fröhmel, & Burger, 2006). Moreover, in contrast to the US, Canada and the UK (Adamo, 2003) assessments with simulated patients have not yet become a compulsory part of medical licensing in Germany. However, communication competence has been a compulsory part of German medical licensing since 2012 and 95% of German medical schools employed simulated patients for teaching and / or assessing communication competence in 2014 (Görlitz et al., 2014, p. 1).

In sum, today simulated patients are employed worldwide for teaching and assessing clinical and communication competence (May et al., 2009). Moreover, the use of simulated patients has expanded to other healthcare domains, such as pharmacy or dentistry (Association of Standardized Patient Educators, 2014) and recently also beyond the healthcare domain (cf. chapter 5.2).

5.1.2 *Use of Terminology*

A variety of descriptors, such as *programmed patients* or *prepared patients*, are used to refer to the simulated patient method and the simulated patients per se. However, *simulated* and *standardized patient* are the most commonly used terms (Adamo, 2003). While Asian and European educators tend to use the term *simulated patients*, in the US, the term *stand-*

ardized patient is predominantly used no matter to what degree the encounters are standardized (Cleland et al., 2009).

Collins and Harden (1998) differentiate three different types of simulated patients that differ in their degree of standardization: (1) Patients who only receive a rough outline of what is expected of them in a physician-patient encounter. (2) Patients who are given a scenario with which they must become familiar but beyond which they are free to respond as they wish; sometimes roles are adjusted to the patient's own background or personal experience and sometimes additional information can be made up. (3) Patients who are extensively trained and whose every response is thought through and rehearsed. Within this broad description, there is a continuum of training and preparation (Cleland et al., 2009). This continuum may have contributed to the range of terminology used including *simulated* and *standardized patient* (Cleland et al., 2009). The two terms are often used interchangeably but this can be misleading (Cleland et al., 2009). Thus, Cleland and colleagues (2009) argue that the term *simulated patient* should be used when the emphasis is on simulation (e.g. presenting the signs and symptoms of an actual patient but improvising background information). In contrast, the term *standardized patient* should be used when emphasis is on the standardization of the simulation process, i.e. when a patient is trained to give a consistent presentation which does not vary from student to student or patient to patient (Cleland et al., 2009). To sum up, according to this use of terminology, a standardized patient encounter is always a simulated patient encounter but a simulated patient encounter is not necessarily standardized (Adamo, 2003).

In the educational domain, simulated encounters are usually not entirely standardized (cf. chapter 5.2) and would rather fall into category one or two of the description of Collins and Harden (1998). Therefore, in line with the use of terminology proposed by Cleland and colleagues (2009), in this dissertation encounters are referred to as simulated unless they are fully standardized. In addition, since the conversational partners in this dissertation are no patients, the method is referred to as simulated conversations. The term *simulated conversations* has the advantage over the term *simulated parent* or *patient* in that it is more neutral and can be applied to a variety of domains and expert-layperson conversations independent of who the conversational partner is. Moreover, the term *simulated parents / patients* is ambiguous since it refers to the method per se as well as to the person playing a certain parent or patient role. The term *simulated conversation* helps to avoid this ambigui-

ty and raises awareness of the fact that simulated conversations consists of several, equally important components, such as cases, coding manuals or raters.

5.1.3 State of Research on Simulated Conversations in the Medical Domain

During the past 50 years, a considerable amount of research on medical simulated conversations has been conducted mainly focusing on the use of simulated conversations for assessments (May et al., 2009). The bottom line of meta-analyses and reviews that synthesize evidence on medical simulated conversations is that if constructed and conducted properly, they deliver results that are to a high degree objective, reliable and valid (Barman, 2005; Cleland et al., 2009; Newble, 2004). However, simulated conversations are vulnerable to many potential measurement errors that can be introduced by the raters (Chesser et al., 2009; Lurie et al., 2008), their way of scoring, i.e. the coding manuals the raters use and the training they get (Barman, 2005; Chesser et al., 2009; Iramaneerat et al., 2008), the content and number of the cases (Barman, 2005; Iramaneerat et al., 2008), the simulated patients and their training (Chesser et al., 2009; Iramaneerat et al., 2008) and the testing conditions under which the simulated conversations take place (Barman, 2005). The raters were often assumed to be the main problem undermining the reliability of simulated conversations either due to personal characteristics or to the lack of standardization of the tasks and scoring criteria (Newble, 2004). However, meanwhile a significant amount of evidence, collected e.g. via using Generalizability Theory, indicates that the problem of rater consistency is less important than the issue of case specificity (Barman, 2005; Guiton, Hodgson, Delandshere, & Wilkerson, 2004; Iramaneerat et al., 2008; Newble, 2004). In the following four subchapters, the most important sources of measurement error - cases, simulated patients, raters and coding manuals - will be elaborated.

5.1.3.1 Cases

Since cases are extremely important for the reliability and validity of simulated conversations, the case vignette is the heart of the simulated conversation (Dotger, Harris, & Hansel, 2008). A fundamental design principle that guides the formation of medical cases is that they are based on real-life cases (Barrows, 1987). If simulated cases duplicate patient problems that actually existed in every way, students are reassured of and motivated by their relevance (Barrows, 1987). Barrows (1987) recommends considering three tenets when crafting cases: 1) Prevalence: Cases should be chosen that students are likely to encounter frequently in their professional lives. 2) Clinical / social impact: Cases should be

chosen that might potentially have a strong impact on individuals or communities (especially if mishandled by the professional). 3) Instructional importance: Cases should be based on content that might not be otherwise addressed in their professional education.

Communication competence seems to be at least partly context-related and case-specific (Guiton et al., 2004) (cf. chapter 3). Consequently, sampling across problems is required if an adequate level of content validity and reliability is to be achieved (Newble, 2004; Reder, Anderson, & Simon, 1996). Therefore, OSCEs targeted at assessing or training communication competence, usually consist of multiple cases. Length, reliability and validity have to be weighed against practicability in order to identify an adequate number of cases. For example, the US Medical Licensing Examinations comprises thirteen and the Swiss licensing examination for human medicine, twelve simulated conversations (Gutormsen et al., 2013; United States Medical Licensing Examination, 2015). Lang, McCord, Harvill and Anderson (2004) demonstrated that the less cases are needed, the more routine cases are employed that require comparatively general communication competencies and provide multiple opportunities to demonstrate each core competence facet within one case.

5.1.3.2 Recruitment and Training of Patients

Besides the crafting of case vignettes, the recruitment and training of simulated patients is a cornerstone for the successful implementation of simulated conversations (Adamo, 2003). Cleland and colleagues (2009) outline four key factors that should be considered when recruiting simulated patients: ability, suitability, conscientiousness and credibility.

Ability: Simulated patients should be able to remember their roles, maintain focus or concentration over time and realize the importance of sticking to the script and guidance provided (Cleland et al., 2009). Especially in the case of standardized encounters, the number of facts and instructions to remember is cognitively challenging (Cleland et al., 2009). Additionally, if simulated patients are to rate the performance and give feedback, they have to have a profound understanding of the scoring criteria and must be able to give appropriate feedback (Cleland et al., 2009).

Suitability: The attitude of the simulated patients towards doctors and the reason why one wishes to be a simulated patient should be screened (Cleland et al., 2009). Patients that have a negative attitude towards doctors might potentially harm students that collect first experiences with patients (Cleland et al., 2009).

Conscientiousness: Simulated patients should be dependable since the organization of simulated conversations is logistically complex. Undependable patients would cause additional costs and could also decrease the reliability of assessments (Cleland et al., 2009).

Credibility: The simulated patient should be matched to case requirements in order to look and be as much as the actual patient to be simulated (Cleland et al., 2009). Characteristics that should be considered are age, language, gender, race / ethnicity, body habitus, findings on physical examinations and experience with the illness (Adamo, 2003). A lot of successful studies with children and adolescents show that even children as young as seven can successfully be deployed as simulated patients (for an overview see Adamo, 2003).

Though training of simulated patients is considered highly important there is no widely accepted standard for how much training is enough and there is little standardization between institutions and within or between countries (Adamo, 2003). One way of knowing how much training is enough is to test whether a simulated conversational partner can maintain a desired level of accuracy in performing (and reporting) in multiple, consecutive encounters (Adamo, 2003). Since the portrayal of roles changes over time, e.g., patients start to become expert-patients and cue students (Adamo, 2003), it is strongly recommended to constantly reassess the quality and if applicable retrain the simulated patients.

If simulated patients are trained well, they are usually not distinguishable from real patients (Beullens, Rethans, Goedhuys, & Buntinx, 1997; Cleland et al., 2009; Sanson-Fisher & Poole, 1980). In a meta-study Rethans, Gorter, Bokken and Morrison (2007) showed that 21 studies using incognito simulated patients have been carried out. In the majority of studies, simulated patients were retrospectively identified in less than 15% of visits (Rethans et al., 2007, p. 546). Non-detection is increased to up to 1% where there is a lengthy period between doctors' consent to participate in studies using simulated patients and the actual visit, use of authentic paperwork and careful preparation of simulated patients (Rethans et al., 2007, p. 546).

5.1.3.3 Raters

Performance in simulated conversations is usually rated during or immediately after the conversations (Lurie et al., 2008). Sometimes, conversations are video-taped and rated later on (Makoul, 2001b). The performance is either rated by external raters, e.g. faculty staff, examiners or experts, and / or by the simulated patients themselves (Huntley, Salm-

on, Fisher, Fletcher, & Young, 2012; Scheffer, 2009; Schirmer et al., 2005; tEACH Assessment subgroup, 2012).

Simulated patients are often used to give formative feedback on communication (or clinical) competence to support the students in reflecting on their competence development (Cleland et al., 2009). For summative assessment where pass / fail or grading judgments are required, consistent evaluation is more critical than for formative assessments (Cleland et al., 2009). When simulated patients are applied for summative assessments, their training has to be intensified (Cleland et al., 2009). In contrast to external raters, simulated patients have the twofold burden to perform a role and judge the performance of the conversational partner at the same time. Thus, in high-stakes exams external raters are usually used (Chesser et al., 2009). Still, there is evidence that both external raters and simulated patients can provide reliable results (Barman, 2005; Blake, Gusella, Greaven, & Wakefield, 2006). Schirmer and colleagues (2005) argue that simulated conversations should be rated by external raters and, additionally, by the conversational partners since this increases the validity of the judgment and adds an important dimension.

The reliability and validity of the scoring of both external raters and simulated patients depends on their training as well as on the coding manual they use (Barman, 2005; Schirmer et al., 2005). There is no gold standard for rater training since the best way and amount of training depends on the coding manual. Three types of coding manuals are discussed in the following section.

5.1.3.4 Coding Manuals

Seidel, Prenzel and Kobarg (2005) distinguish three types of methods to analyze (video-taped) behavior in educational settings: symbol systems (in the medical context usually referred to as checklists), category systems and rating scales. These three methods vary in their administration design, focus, psychometric properties, practicality and ease of use (Schirmer et al., 2005). A symbol system only codes whether or not a certain event occurs (Seidel et al., 2005). When using a category system, observers do not only have to note whether or not a certain event occurs, but also have to assign it to a certain category (Seidel et al., 2005). Finally, when using rating scales they also have to evaluate the quality of the event, taking into account interdependences between the observed behaviors (Seidel et al., 2005). With regard to the degree of inference needed for the coding, Seidel and colleagues (2005) refer to the directly observable symbol systems as “low inference”, to the category systems as “middle inference” and to the rating scales for which a lot of interpretation is

necessary as “high inference” (p.72, p.74). In recent years the question whether low inference or high inference methods are more appropriate to rate simulated encounters has been prominent (Newble, 2004).

With high inference ratings it is comparatively difficult to establish how ascertained instructional events and modes are to be identified and evaluated (Seidel et al., 2005). They demand qualitative decisions that often lead to a reduced reliability of the ratings (Seidel et al., 2005). In contrast, low inference coding manuals are directed at observable modes of behavior (Seidel et al., 2005). The coding instructions for low inference methods are well-defined and clear and make them easier to use than rating scales (Seidel et al., 2005). In consequence, low inference methods tend to produce more reliable scores and a higher inter-rater agreement than rating scales (Newble, 2004). However, low inference methods are criticized for trivialization because only criteria that are easy to define can be included on the coding manual at the expense of equally or more important criteria that are more difficult to define and measure (Newble, 2004). Newble (2004) warns against falling into the trap of developing detailed checklists that produce reliable scores but which do not truly reflect the examinee’s performance of the task and the underlying construct, which is the aim of the measurement. The construct might, e.g., not be captured in its complexity because the measurement is too narrow or focuses on irrelevant / not all aspects of the construct.

The comparison of low and high inference procedures in the educational and medical domain has shown that a variety of research questions can be answered more validly with high inference ratings (Regehr, MacRae, Reznick, & Szalay, 1998; Seidel et al., 2005). Regehr and colleagues (1998) showed that global scale scores have a better predictive validity and a higher inter-station reliability in OSCEs. In a study by Chesser and colleagues (2009), even excellent students did not meet all criteria on a checklist, despite performing well according to the global scores awarded by both examiners and simulated patients. In line with Chesser and colleagues, several empirical studies show that though reliable, checklists based on binary response scales cannot effectively measure differences in competencies (Hodges & McIlroy, 2003; Skillings, Porcerelli, & Markova, 2010).

Low inference methods assess only the presence or absence of a certain behavior; they neither measure its quality, nor take into account the context (Skillings et al., 2010). It is probably not simply whether persons apply a certain task, but the primary difference in competence level lies in the quality, the timing or the manner in which a task is applied

(Skillings et al., 2010). This could be the reason why global ratings are more effective in distinguishing between beginning and advanced learners (Hodges & McIlroy, 2003; Skillings et al., 2010). Thus, Skillings (2010) comes to the conclusion that high inference methods are able to evaluate competencies more effectively than low inference methods.

Newble (2004) believes that low and high inference methods may serve different purposes and argues for a balanced approach with checklists used for identifying specific elements of a competence that must be demonstrated and global ratings used for providing a measure of process aspects and a more complex picture. For assessing communication competence, he considers high inference methods as better fitted than checklists (Newble, 2004). Whether low or high inference methods are the instrument of choice also depends on the raters who are to judge the performance and the amount of training they receive. Since checklists provide clearer behavioral definitions, the amount of training required to train observers to use checklists reliably is usually lower than for rating scales (Schirmer et al., 2005). However, when intensively trained raters use rating scales, they may be as, or even more, reliable than checklists (Clausen, Reusser, & Klieme, 2003; Regehr et al., 1998; Schirmer et al., 2005). Thus, checklists might be the preferred tool when raters have less experience or cannot receive an intensive training and rating scales when raters are trained intensively (Schirmer et al., 2005). In sum, it is easier to achieve a high reliability with checklists. However, global ratings tend to capture the content of the underlying construct in a more valid way. The choice of checklists or global ratings should depend on the background of the raters, the amount of training they will receive and the purpose of the results.

5.2 Summary

The brief sketch of how simulated conversations have been used and labelled in the medical domain and of research findings regarding an adequate design of the simulated conversations, has provided essential information for the application and evaluation of simulated conversations in the educational domain summarized in the following subsection. Simulated conversations were introduced in the medical domain in 1964 by Barrows and Abrahamson in order to support clinical skills learning (Barrows & Abrahamson, 1964). Today they are used for teaching and assessing clinical and communicative competencies of (pre-service) physicians worldwide and their use has recently expanded to other domains (Association of Standardized Patient Educators, 2014). This speaks for their acceptance and benefit. Terminology with regard to simulated conversations varies with *simulated* or

standardized patients being the most frequent (Cleland et al., 2009). In this dissertation, simulated conversations are referred to as simulated unless they are fully standardized. The term *simulated conversations* is preferred over *simulated parents / patients* for two reasons. Firstly, it is neutral and can be applied to a variety of simulated encounters no matter who the conversational partner is. Secondly, it helps to avoid ambiguity since simulated patients / parents are only one component of simulated conversations besides cases, raters and coding manuals.

A considerable amount of research in the medical domain has shown that simulated conversations are well-accepted, perceived as authentic and do fulfill psychometric quality criteria if constructed properly (Barman, 2005; Cleland et al., 2009; Newble, 2004). This indicates that simulated conversations possess potential for an employment in the educational domain. Decisive for the quality of simulated conversations are the number and content of cases, the recruitment and training of the actors, the recruitment and training of the raters and the coding manual the ratings are based on. These four factors should be taken into account when transferring simulated conversations to the educational domain (cf. chapter 7). The following research findings from the medical domain provide specific information for the development of the components of the simulated parent-teacher conversations in this dissertation. Assessments based on simulated conversations should consist of several cases (Lurie et al., 2008) and cases should be selected according to their prevalence, social impact and instructional importance (Barrows, 1987). Conversational partners that match the person to portray as closely as possible should be recruited (Cleland et al., 2009) and (re-)training and quality checks should ensure that the role is displayed in a consistent way. As raters, both simulated conversational partners and external observers can be employed and provide reliable results (Barman, 2005; tEACH Assessment subgroup, 2012). Schirmer and colleagues (2005) recommend employing external raters and simulated conversational partners simultaneously as raters in order to add a second perspective and in this way increase validity. Coding manuals may consist of symbol systems, category systems or rating scales or of a combination of those. While it requires less rater training to achieve a high inter-rater agreement with low inference ratings, the validity of high inference ratings tends to be higher. The choice of low or high inference ratings should depend on the background of the raters, the amount of training they will receive, the purpose of the results and the level of the learners.

Combined with the research findings on simulated conversations in the educational domain presented in the subsequent subchapter 5.3, the findings on simulated conversations in the medical domain constitute the basis for the development, employment and evaluation of the simulated conversations in this dissertation. Special emphasis with regard to the evaluation will consequently lie on whether the development and employment was successful regarding the impact of different case vignettes, the consistent portrayal of simulated parents by the actors, the agreement of the raters, the reliability of the coding manual and the validity of the measurement with the newly-developed coding manual.

5.3 Simulated Conversations in the Educational Domain

In order to compile further evidence that delivers specific information regarding the design and evaluation of simulated conversations in teacher education required for the development of the simulated conversations, the deduction of the research questions and the corresponding evaluation of the simulated conversations in this dissertation, this subchapter outlines the state of practice and research on simulated conversations in the educational domain. Simulated conversations have not been used as an assessment in teacher education so far. However, in 2007 Dotger, Dotger, & Maher started using simulated conversations in the educational domain as a learning tool. They developed the Standardized Parent / Caregiver Conferencing Model, a learning cycle consisting of different simulated conversational partners, such as parents, students or colleagues (Dotger et al., 2010). In order to obtain additional information regarding the design of simulated conversations that target teachers instead of physicians, the design of the Standardized Parent / Caregiver Conferencing Model and its distinct components is described in this subchapter. The focus is particularly on those components that medical research has revealed to be influential regarding the reliability and validity of simulated conversations: It is outlined how the case vignettes for the simulated conversations for the teachers are crafted and how the actors are recruited and trained as simulated parents / colleagues / students. Moreover, differences and similarities between simulated conversations in the medical and educational domain are discussed that should be taken into account when designing simulated conversations for the educational domain. The chapter ends with a presentation of research findings on and via simulated conversations in the educational domain and an outlook on current needs for research that should be targeted in this dissertation and in future research.

5.3.1 *The Standardized Parent / Caregiver Conferencing Model and the Simulated Interaction Model*

The Standardized Parent / Caregiver Conferencing Model is a six-case, semester-long development intervention designed to provide pre- and in-service teachers with multiple opportunities to practice communicating with parents (Dotger & Sapon-Shevin, 2009). It is based on a simulation-reflection cycle (Dotger et al., 2008). Participants engage in a simulated case and reflect on the simulation using a video of the conversation (Dotger et al., 2008). As teachers reflect on the past simulated parent-teacher conference, they also begin preparing for the next one (Dotger & Smith, 2009) and construct an individual plan in order to improve certain aspects in the next conversation, like asking more questions (Dotger et al., 2008).

In the course of the Standardized Parent / Caregiver Conferencing Model, the simulated conversations change both in context and content and get more and more complex (Dotger, 2010). Each simulation is designed to build upon the teachers' increasing competency in parent-teacher communication (Dotger & Smith, 2009). The first case of the Standardized Parent / Caregiver Conferencing Model is, e.g., about conducting a getting-acquainted / role-defining conference, while the last two conferences deal with physical abuse in the home and a student with special needs (Dotger, 2010).

Pre-service teachers that participate in the Standardized Parent / Caregiver Conferencing Model exit the simulated conversations with digital access to their recorded data, along with written formative feedback from their conversational partners (Dotger, 2010). They have a week to review their simulation and to construct written reflections on what they said and how they said it, analyzing their verbal and nonverbal behavior and professional decisions (Dotger, 2010). One week later, their reflections are analyzed in a group session based on video excerpts (Dotger, 2010). At the end of this reflection session, participants receive the next case and can start preparing for the next conversation (Dotger, 2010).

From the Standardized Parent / Caregiver Conferencing Model, the Simulated Interaction Model evolved, which includes a broader context of teacher education problems (Dotger et al., 2010). The Simulated Interaction Model includes interactions between school leaders and simulated students, teachers and simulated students, teachers and simulated parents and teachers and simulated paraprofessionals (Dotger et al., 2010). The original focus on parent-teacher interactions has broadened to a more general focus on how teachers and school leaders engage in common simulated problems of practice (Dotger et al., 2010).

Most currently, Dotger and colleagues have designed and applied 14 simulations that focus on content-specific problems of practice (Dotger, 2015; Dotger, Masingila, Bearkland, & Dotger, 2014). These simulations bring to life particular student misconceptions or parental concerns about secondary (grades 7-12) content.

In addition to the Standardized Parent / Caregiver Conferencing Model or the Simulated Interaction Model, simulated conversations are also embedded into regular method courses or seminars of student teachers (Dotger, 2011a). The simulated conversation on a child with special needs was, e.g., employed as part of a larger course about working with students with significant disabilities (Dotger & Ashby, 2010).

Dotger (2008) considers the degree to which pre-service teachers perceive simulated conversations as authentic and meaningful as decisive for their learning benefit. The authenticity, however, hinges on the case vignettes and the persons that portray the conversational partners (Dotger et al., 2008). The design principles of the case vignettes and the recruitment and training of actors are discussed next.

5.3.2 *Cases*

Dotger and colleagues design the cases for the educational simulated conversations according to the medical design principles outlined in chapter 5.1.3.1 (e.g. Dotger et al., 2008). In order to base the cases on real-life scholastic situations, they conducted both individual and focus group interviews with parents, teachers, principals, school superintendents, and guidance counselors (Dotger et al., 2008). These interviews yielded numerous conferencing contexts that were further developed into cases (Dotger et al., 2008). Dotger and colleagues screened the cases with regard to the criteria proposed by Barrows (1987): prevalence, clinical / social impact and instructional importance (cf. chapter 5.1.3.1). With regard to prevalence, they chose cases that every teacher is likely to come across, such as a conference where the teacher is challenged to convey information on the child's academic and behavioral performances in class (Dotger et al., 2008). Regarding the clinical / social impact, cases were selected that might be rarely experienced, but that present a variable that is of great importance or has a potentially high impact if it is overlooked or mishandled by the professional, such as child abuse, special needs or bullying (Dotger et al., 2008). To regard the criteria of instructional importance, Dotger and colleagues created one very emotional conference where a parent strongly disagrees with the teachers so that teachers can practice their competence to handle emotionally difficult parent-teacher conversations

since parent-teacher conferences often are sources of anger, frustration, tension and strong emotions (Dotger et al., 2008). The first drafts of the cases were piloted and revised and then made into the final case vignettes (Dotger et al., 2008).

Since multiple individuals portray the same simulated parent role, each case profile outlines exactly who the simulated conversational partner will be during the conversation (Dotger et al., 2008). The case vignettes consist of two parts: informational content and interactional content (Dotger et al., 2008). The informational content comprises background information on the person to portray (i.e. employment history, marital status, disposition, socioeconomic status, dress etc.) (Dotger et al., 2008). The interactional content describes the information and types and degrees of emotion to be conveyed to the teacher via verbal and nonverbal behavior, e.g. the exact tone of voice, bodily-kinesthetic positions and facial expressions (Dotger et al., 2008). The interactional content also focuses on the triggers of a case (Dotger et al., 2008). Triggers are desired verbalizations that the simulated conversational partners should present to the teachers during the simulated conferences (Dotger et al., 2008). The triggers within the simulations are enacted at a certain point in the conversation or in response to a particular action on the part of the teacher (Dotger et al., 2008). Often these triggers are outlined in if-then statements (Dotger et al., 2008). In addition to the triggers, there is also verbal and non-verbal fixed interactional content that is not contingent on the teacher's actions, such as the initial posture of the simulated conversational partner or exact questions to ask or comments to make and their order and timing (Dotger, Dotger, & Tillotson, 2010).

While the simulated conversational partners are provided with very specific profiles on which to base their actions, the participating (pre-service) teachers are given a more general academic profile that describes a hypothetical student (or colleague) (Dotger & Smith, 2009). The (pre-service) teachers receive the name of their students, physical descriptions, their academic and behavioral history and a detailed rationale for why a certain student is the focus of the simulated parent-teacher conference (Dotger & Smith, 2009). As in real life, the amount of information the (pre-service) teachers receive before a conversation varies depending on how much the parent in the case has disclosed and whether the conference is teacher- or parent-initiated (Dotger, 2010). Importantly, the case vignettes for the (pre-service) teachers differ from the ones written for the simulated conversational partners in that they do not specify what decisions he or she should make regarding pedagogy or content (Dotger & Smith, 2009). Instead, participating (pre-service) teachers are

given the freedom to place themselves within the simulation and operate from their own individual, professional perspective (Dotger & Smith, 2009). In order to secure that pre-service teachers already have the necessary pedagogical content knowledge to handle the simulation, the content of the simulations is coordinated with the rest of their studies (Dotger & Ashby, 2010).

Similar to the findings in the medical domain (cf. chapter 5.1.3.1), Dotger and colleagues (2010) come to the conclusion that the number of cases is decisive for the effectiveness and quality of simulated conversations. Participants that only take part in one simulation report being busy processing the simulation process and recording technologies and hardly being able to reflect on their professional behavior (Dotger et al., 2008). In contrast, participants who took part in multiple simulations reported increased comfort with the simulation process and an increased ability to focus on the content of the case and how they conducted themselves (Dotger et al., 2008). Therefore, Dotger and colleagues (2010) state that a key lesson they have learned from conducting simulations is that simulations can only realize their full potential if multiple simulations are provided since the novelty of a participant's first simulation often negates the educational value of whatever context is simulated.

5.3.3 Recruitment and Training of Simulated Conversational Partners

The simulated conversational partners come from the database of simulated patients of the Upstate Medical University's Clinical Skills Center and are recruited according to case requirements, such as gender, age, demographics and background knowledge (Dotger et al., 2008). They receive around two-three hours of training approximately one week prior to the simulated interactions (Dotger & Ashby, 2010). The training starts with roughly 20 minutes of general introduction (Dotger & Ashby, 2010). In case there are persons that have not been simulated conversational partners before, the concept and purpose of simulated conversations as well as the responsibilities of being a simulated conversational partner are introduced (Dotger et al., 2008). The general introduction also includes an outline of the particular case to display and the intended teacher audience (Dotger & Ashby, 2010).

The remaining time is devoted to the case vignette (Dotger & Ashby, 2010). The trainer guides the simulated conversational partners through the informational and interactional information, verbally outlining the case vignette sentence by sentence (Dotger et al., 2008).

The nature, structure and order of triggers are explained (Dotger et al., 2010). Each verbal trigger is presented by the trainer and then rehearsed by the simulated conversational partners until tone, mannerisms, and verbalizations are as desired (Dotger & Ashby, 2010).

A complete standardization, as sometimes undertaken in the medical domain, is not possible in the educational domain since it is impossible to anticipate exactly what the participants will say in response to verbal triggers (Dotger & Ashby, 2010). The training serves as an opportunity to anticipate broader response patterns and the simulated conversational partners are provided with additional contingency responses that they may or may not employ in the conversation depending on the behavior of the (pre-service) teacher (Dotger, 2010). Additionally, the simulated conversational partners are given extensive background context on their character and are allowed to improvise if something unpredictable happens (Dotger, 2010). The most important point emphasized in the training is that they always stay in character in order to ensure the authenticity of the situation (B. Dotger, personal communication, April, 2014). In case the simulations are carried out for the first time, trial runs are also included in the training (Dotger & Ashby, 2010). All training and rehearsal sessions are video recorded for documentation (Dotger et al., 2010).

After the first deployment of simulated patients as simulated parents in the Standardized Parent / Caregiver Conferencing Model, twelve actors took part in a two hour reflection group on the transition from simulated patients to parents (Dotger et al., 2008). The simulated conversational partners reported that the conversations felt very realistic, but that being a simulated parent differed a lot from being a simulated patient since the characters as a simulated patient are much narrower (Dotger et al., 2008). As a simulated patient, “you have your case, you have your [medical] complaint, and these are what [the doctors] are gonna ask about and, in essence, they all ask the same things mostly. But this [standardized parent role] [...] all of a sudden, this opens wide. There’s no medical constraint” (Dotger et al., 2008, p. 342).

5.3.4 *Research Findings*

Dotger and colleagues have used simulated conversations to investigate a variety of research questions. A short overview of their research findings is given in the following.

Dotger and Sapon-Shevin (2009) saw evidence that simulated conversations improve teachers’ sensitivity to parents’ perspectives and concerns and teachers’ communication competence. Moreover, simulated conversations not only make the development of a pre-

service teacher's professional identity transparent, but also shape it (Dotger & Smith, 2009). As a result from simulated conversations, pre-service teachers start to question the nature and position of professional boundaries and reflect on gaps in their professional knowledge, thereby developing a more realistic self-concept (Dotger & Smith, 2009). Simulated conversations also hold the potential to illuminate (pre-service) teachers' dispositions towards parents, students, colleagues and the purposes of schools (Dotger, 2011a) and to train their navigation of emotions (Dotger, Harris, Maher, & Hansel, 2011). Additionally, simulated conversations raise (pre-service) teachers' awareness and increase sensitivity to socio-cultural contexts (Dotger, 2010). Simulated conversations can also yield insights into how pre-service teachers enact teacher preparation curricula (Dotger & Ashby, 2010). Dotger and Ashby (2010) showed that when inclusively trained pre-service teachers were confronted with a colleague with a different professional ideology they were often willing to defer to the contrasting philosophies, practices and / or approaches of the other professional.

5.4 Summary

Since 2007, Dotger and colleagues employ simulated conversations in teacher education (Dotger et al., 2010) as a formative model designed to foster the development of professional competencies (Dotger & Smith, 2009). Additionally, Dotger and colleagues were able to show that research with simulated conversations promotes and yields insights into pre-service teacher development and delivers information that can help fine-tune teacher education (Dotger & Ashby, 2010; Dotger & Smith, 2009). Thus, simulated conversations hold the potential to bridge the gap between teacher education and praxis (Walker & Dotger, 2012).

The successful employment of simulated conversations as a learning tool in teacher education is promising with regard to the planned implementation of simulated conversations as an assessment for (pre-service) teachers. With regard to this planned implementation, the following research findings by Dotger and colleagues might be helpful: Dotger and colleagues (2008) consider the degree to which (pre-service) teachers perceive simulated conversations as authentic and meaningful as decisive for the benefit that can be drawn from them. In order to ensure authenticity, they recommend to base simulations on real-life cases (Dotger et al., 2008) and suggest that participants take part in more than one simulation since experience shows that their attention shifts from the form of the simulated con-

versations to the content (Dotger et al., 2010). Apart from the cases, the training of the simulated conversational partners is especially critical in the educational domain since educational simulations differ from medical ones in that they are far less predictable and cannot be fully standardized (Dotger et al., 2008). Thus, actors need extensive training in order to be prepared for simulated conversations with teachers (Dotger et al., 2008).

Since the authenticity of the simulated conversations is probably equally or even more decisive when using them as an assessment, it seems recommendable to pay special attention to authenticity when developing case vignettes for the assessment, e.g. by basing the cases on real-life cases. Moreover, assessments with simulated conversations should consist of more than one case in order to ensure that participants are not distracted by the new format but can focus on the content. In order to verify if these two arrangements are effective, the perceived authenticity of the simulated conversations should be evaluated. Additionally, since medical research has shown that a consistent performance of the simulated conversational partners is a precondition for the reliability of assessments and the findings by Dotger and colleagues (2008) show that parent-teacher conversations cannot be fully standardized and are far less predictable, specific emphasis should also be put on the training of the simulated parents. In order to evaluate whether this training was successful, the consistent performance of the actors should be investigated. This proceeding is in line with a call from Dotger and colleagues for more research on simulated conversations, especially regarding the fulfillment of psychometric quality criteria:

Reliability of this pedagogy has not been measured, as no other researchers have implemented the SIM within different teacher education settings. As this body of research grows, though, there is clearly a need to assess the degree to which other researchers using SIM materials arrive at similar results. Of note, the entire pedagogy of simulated interactions hinges on multiple individuals who accurately and reliably serve as standardized parents, students, or paraprofessionals. Ensuring reliability across standardized individuals is an on-going task that is central to the success of each individual case. To date, though, no large-scale, cross-case assessment of standardized individual reliability has been conducted (Dotger et al., 2010, p. 135).

Additionally, they state that attention to and evaluation of the validity of simulated conversations in the educational domain is paramount for practice and research (Dotger et al., 2010, p. 135). Medical research has shown that apart from the case vignettes and the actors, the coding manuals and the observers who rate the simulated conversations are decisive for the reliability and validity of assessments consisting of simulated conversations (cf. chapter 5.1). Since simulated conversations have not been implemented as an assessment in teacher education so far, there is no evidence regarding the latter two components

of simulated parent-teacher conversations yet. Thus, apart from providing further evidence regarding the influence and authenticity of the case vignettes and the actors of simulated conversations in the educational domain, this dissertation focusses particularly on providing initial evidence regarding inter-rater agreement, the reliability of the newly-developed coding manual and the validity of simulated conversations with regard to diagnosing the communication competence of pre-service teachers in conversations with parents.

6 Conclusions From the Theory for this Dissertation

Chapter 1 has highlighted the relevance of parent-teacher cooperation and conversations for students. Formal conversations between a teacher and one or two parents are often considered the heart of parent-teacher cooperation, since they offer the chance to exchange relevant information about a student and, in this way, to synchronize educational processes at school and at home and to inspire further forms of parent-teacher cooperation (cf. chapter 2). However, the analysis of parent-teacher conversation practice in Germany has revealed that parents are not entirely satisfied with the quality and quantity of conversations offered by teachers (cf. chapter 2). The supply-demand model of parent-teacher conversations by Hertel and colleagues (2013) supports the conclusion that teacher education might be one key to improving parent-teacher conversation praxis. While an analysis of teacher education curricula and teacher surveys showed that parent-teacher conversations are not yet sufficiently integrated into teacher education, some progress has been made during the last years and parent-teacher conversations are increasingly becoming a part of teacher education (cf. chapter 2). This ongoing process should be monitored since teacher education has to be based on evidence in order to ensure its quality (Prenzel, 2013). Thus, now there is a need for instruments and assessments that can bridge the gap between theory and practice and provide information about the effectiveness of teacher education with regard to parent-teacher conversations, as well as about the readiness of teachers to conduct these conversations and about starting points for further teacher training. Consequently, the aim of this dissertation is to develop and evaluate an instrument for diagnosing the communication competence of pre-service teachers in conversations with parents.

The foundation for the development of measurement instruments are theoretical models of the competencies to be measured (cf. chapter 3). Theoretical models of competencies should be adequately context-specific and their structure needs to be differentiated (Hartig, 2008). Thus, chapter 3 has provided an overview of theoretical models of the competence of teachers to communicate with parents and empirical evidence supporting these models. The Munich Model of Communication Competence in Parent-Teacher Conversation was chosen as the theoretical foundation for the construction of the measurement instrument in this dissertation since it has the following advantages: While it is context-specific it still has a comparatively wide range of applicability. Moreover, its subdivision into three com-

petence facets corresponds to prior research findings and allows for a differentiated competence measurement (cf. chapter 3).

Based on theoretical models, appropriate instruments can be chosen or developed that allow for measuring a specific competence, such as the competence of teachers to communicate with parents. Chapter 4 compared the advantages and disadvantages of instruments applied to measure (teacher) competencies with regard to the diagnosis of teachers' competence to conduct conversations with parents. Simulated conversations seem particularly promising for measuring this competence since they allow observing otherwise unobservable behavior, offer measurement conditions close to reality, are performance-oriented, context-specific and well-accepted (cf. chapter 4). For these reasons, simulated conversations are employed to measure the competence of pre-service teachers to conduct conversations with parents in this dissertation.

As well as teacher training programs, new assessment methods need to be validated in order to ensure the quality of teacher education (Caspé et al., 2011). Multimethod measurements can play an important role in the validation process of instruments, such as the simulated conversations developed in this dissertation. In multimethod measurements several instruments are combined to measure one construct (Eid & Diener, 2006). The degree to which the correlational pattern and strengths between an instrument and other measurement instruments correspond to theoretical expectations is indicative of its validity (American Educational Research Association et al., 2014). In order to gather evidence for the validity of the simulated conversations developed in this dissertation, they will be complemented with other measurements located at two levels. With regard to the first level, the conversation situation, the independent observer ratings of the performance of the pre-service teachers in the simulated conversations will be complemented by ratings of the conversational partners and self-assessments of the pre-service teachers. Moreover, independent observers will rate a subsample of the simulated conversations with six different coding manuals with different forms and content in order to provide convergent and discriminant evidence for measurements with the newly-developed coding manual. The second level of the multimethod measurement goes beyond the actual conversation situation and comprises a situational judgment test on parent-teacher conversations and external criteria, which could be related to communication competence, like previous knowledge. The measurements at the second level provide further evidence for convergent and discriminant validity.

To date simulated conversations have predominantly been used in the medical domain in order to teach and assess the clinical and communication skills of (pre-service) physicians (cf. chapter 5). For the purpose of applying simulated conversations to the context of parent-teacher conversations, research findings from the medical domain on simulated conversations and on how their different components should be designed were compiled and examined in chapter 5. A considerable amount of research in the medical domain has shown that simulated conversations are well-accepted, perceived as authentic and do fulfill psychometric quality criteria if properly constructed (Barman, 2005; Cleland et al., 2009; Newble, 2004). These findings indicate that simulated conversations possess potential for incorporation into teacher education. However, medical research also shows that there are four factors that particularly influence the reliability and validity of simulated conversations: The number and content of the cases, the recruitment and training of the actors, the recruitment and training of the raters and the coding manual the ratings are based on (cf. chapter 5). These four factors should be taken into account when transferring simulated conversations to the educational domain (cf. chapter 7).

The transfer of the simulated conversations to the educational domain and the resulting development of new case vignettes for parent-teacher conversations, a corresponding actor training as well as a coding manual and a corresponding rater training for analyzing the performance in parent-teacher conversations necessitate an evaluation of the newly-developed simulated conversations in the educational domain. Guided by research findings from the medical domain this evaluation should in particular target the impact of the different case vignettes, the consistent portrayal of the simulated parents, the agreement of the raters and the reliability of the coding manual. These foci for evaluation can be expanded based on first research findings on simulated conversations in the educational domain (cf. chapter 5.3). Dotger and colleagues (2008) consider the degree to which (pre-service) teachers perceive simulated conversations as authentic and meaningful as decisive for the benefit that can be drawn from them. This finding suggests a further point for evaluation: the degree to which pre-service teachers perceive the developed simulated parent-teacher conversations as authentic and accept them as an assessment method. The question of acceptance is particularly interesting when introducing new assessment methods. Dotger and colleagues (2008) stress that the authenticity of educational simulated conversations hinges on the case vignettes and on the actors portraying the case. Since their comparison of medical and educational simulations showed that educational conversations are far less predictable than medical conversations and cannot be standardized as much, they consider

the consistency of the actors in their roles as particularly challenging and ask for more research regarding this point (Dotger et al., 2008).

The educational findings support the choice of evaluation foci for this dissertation regarding the case vignettes and the consistent portrayal of the roles by the actors deduced from medical research. Simulated conversations have not been used as an assessment in teacher education thus far. Hence, there are no findings regarding rater agreement and the reliability of coding manuals with regard to teacher education. The fact that medical research indicates that these two factors can be influential speaks for investigating these points with regard to the future integration of simulated conversations as an assessment in teacher education. Moreover, with regard to this integration, there is a need for evidence regarding the validity of the results in the simulated conversations. Thus, the final evaluation focus in this dissertation will be the validity of the results of simulated conversations. The six evaluation foci presented in this chapter are united in the following research question, which is the overall guideline for this dissertation:

To what extent are simulated conversations suited to diagnose the communication competence of pre-service teachers in parent-teacher conversations?

In part C individual research questions and hypotheses for all evaluation foci will be formulated. Moreover, these research questions are assigned to psychometric quality criteria. Psychometric quality criteria provide reference points for the development and evaluation of instruments (American Educational Research Association et al., 2014).

C Research Questions & Hypotheses

The first employment of simulated conversations as an assessment in the educational domain necessitates an evaluation of the instrument with regard to its suitability for assessing the competence of (pre-service) teachers to conduct formal conversations with one or both parents (cf. chapter 6). The evaluation in the empirical part is guided by the following overarching research question:

To what extent are simulated conversations suited to diagnose the communication competence of pre-service teachers in parent-teacher conversations?

This research question is broken down in a two-step procedure that enables investigating the evaluation foci outlined in chapter 6 by addressing the most important aspects of the corresponding psychometric quality criteria. Step one addresses the fundamental aspects of building scores, that is, checking inter-rater agreement, the factorial structure and the reliability with which the factors are measured. Step two is based on the previous analyses and inquires more deeply into important aspects of the three main quality criteria, objectivity, reliability and validity. Figure 4 provides an overview of the main psychometric quality criteria investigated in this dissertation by allocating the research questions to them. The fulfillment of ancillary quality criteria (Kubinger & Proyer, 2005) is discussed in chapter 18.

Since this is the first evaluation of simulated conversations as a measurement instrument in the educational context, it cannot be exhaustive. Instead the investigation aims at compiling first evidence for the suitability and readiness for use of simulated conversations as an instrument for measuring the competencies of pre-service teachers to communicate with parents. The results of the investigation form the basis of a cost-benefit analysis of simulated conversations in the educational domain and the development of recommendations for their successful development and integration into teacher education. The long term goal of this dissertation is to contribute to the integration of simulated conversations into teacher education and, in this way, to promote successful conversations between parents and teachers.

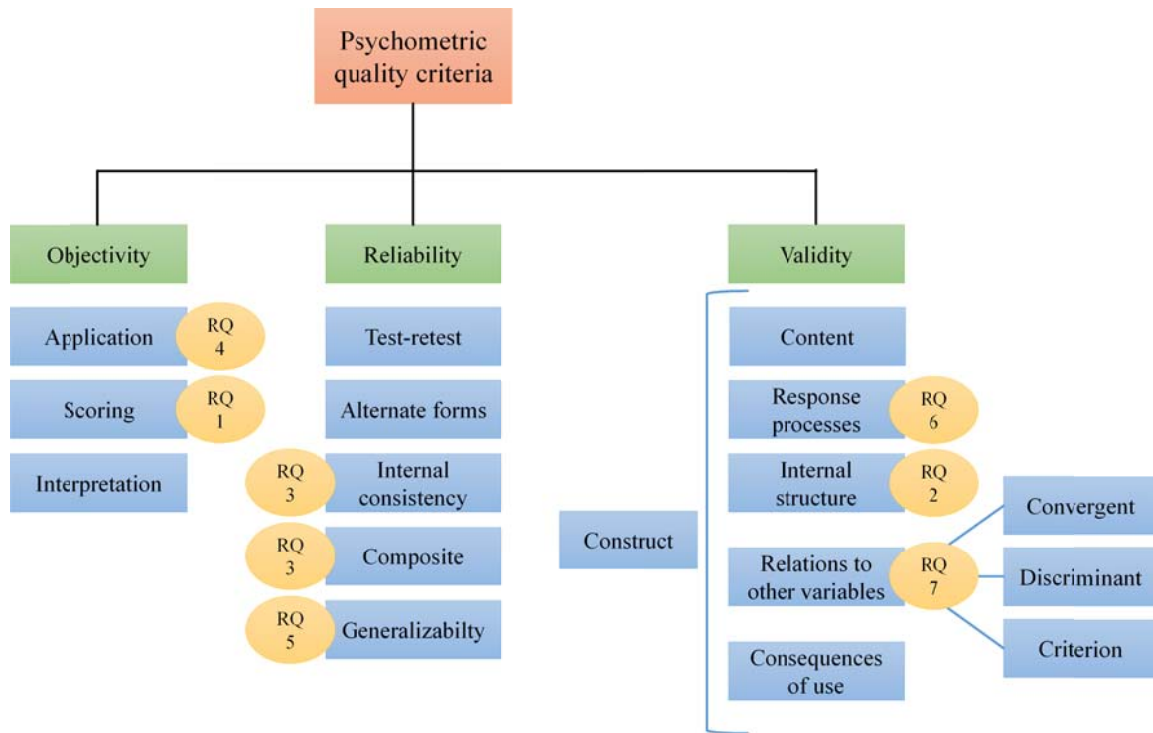


Figure 4. Research questions assigned to psychometric quality criteria (adapted from Bühner, 2011, p. 74, American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014)¹⁵

Research questions on step one, fundamental aspects of building scores:

The **first research question** concerns the inter-rater agreement of the two independent observers who rated the video-taped simulated conversations based on the newly-developed coding manual:

- (1) Do the independent observers achieve a satisfactory ($ICC \geq 0.6$) inter-rater agreement with regard to the communication competence of pre-service teachers in parent-teacher conversations?

This research question addresses the objectivity of scoring. Objectivity of scoring refers to the fact that the scoring process should not be influenced by the observer (Bühner, 2011). In order to ensure objectivity of scoring, guidelines for scoring should be set up, e.g. in a coding manual, and observers should be trained according to these guidelines (Bühner, 2011). Finally, inter-rater agreement should be calculated (Bühner, 2011).

¹⁵ The psychometric quality criteria grid is self-developed based on current standards for test development and literature on psychometric quality criteria, such as American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014), Bühner (2011), Cronbach & Meehl (1955), Frey (2014), Furr and Bacharach (2008), Kane (2001), Rammstedt (2010) and Scheibe and colleagues (2014).

The newly-developed coding manual consists of high inference items. According to the findings from the medical domain outlined in subsection 5.1.3.4, a large amount of training is needed to achieve a high inter-rater agreement when applying high inference scales. Consequently, the observers receive intensive training (cf. chapter 7.5). An inter-rater agreement of $ICC \geq 0.6$ is expected, which would indicate that scoring is adequately objective.

Research question 2 aims at providing evidence for the validity of the measurement via examining the internal structure of the data. Analyses of the internal structure can indicate the degree to which the relationships among items and instrument components conform to the theoretical construct on which the proposed instrument score interpretations are based (American Educational Research Association et al., 2014). With regard to this the following research question arises:

- (2) Does the factorial structure of the data match the theoretical construct - the Munich Model of Communication Competence in Parent-Teacher Conversation (acceptable model fit)?

The structure of the video ratings should correspond to the theoretical construct - the Munich Model of Communication Competence in Parent-Teacher Conversation - outlined in chapter 3.3. This correspondence would provide evidence for the validity of the measurement.

Research question 3 targets the reliability of the measurement. Reliability refers to the accuracy of a measurement (Bühner, 2011). It is always a matter of degree, not a yes or no feature (Furr & Bacharach, 2008). Reliability reveals to what degree differences in respondents' observed scores are consistent with differences in their true scores (Furr & Bacharach, 2008). Two different reliability coefficients, internal consistency and composite reliability, are calculated to comply with the recommendations of the AERA, APA & NCME Standards (2014) that suggest that several reliability coefficients should be calculated since they convey different information:

- (3) Are the scales of the coding manual developed for diagnosing the communication competence of pre-service teachers in parent-teacher conversations reliable ($\alpha^{16} \geq 0.6$), ($\rho \geq 0.7$)?

¹⁶ α = Cronbach's Alpha throughout the dissertation

Due to its theoretical basis in the Munich Model of Communication Competence in Parent-Teacher Conversation, it is expected that the reliability of the scale and its subscales is satisfactory.

Research questions on step two, in-depth analyses on objectivity, reliability, and validity:

Research question 4 addresses the objectivity of application. Objectivity of application refers to the fact that testing conditions should be comparable for everyone and should not influence the measurement (Rammstedt, 2010). In the simulated conversations different actors acted out the same simulated parent role. It is important that they perform consistently to provide all pre-service teachers with comparable conditions in the conversations. To verify if this is the case, research question 4 investigates whether the performance of the simulated conversational partners has an effect on the self-assessed performance or on the observer-rated performance of the pre-service teachers in the simulated conversations:

- (4) Do the different conversational partners have an effect on the self-assessed performance / observer-rated performance of the pre-service teachers in the simulated parent-teacher conversations?

With the help of the information from the case vignettes (cf. chapter 7.1) and the actor training (cf. chapter 7.2), the simulated conversational partners should perform comparably. Consequently, the performance of the pre-service teachers should not be influenced significantly by their different conversational partners. There should neither be a statistically significant effect of the conversational partner on the observer-rated nor on the self-assessed performance of the pre-service teachers in simulated parent-teacher conversations. This would be an indicator of the objectivity of application.

Research question 5 addresses the influence of the case vignettes on the performance of the pre-service teachers in the simulated conversations and in line with this the generalizability of the results of simulated conversations to other (simulated) conversations. To answer research question 5, it is investigated to what degree the results in the two simulated conversations are correlated with each other and if the two different case vignettes affect the observer-rated and self-assessed performance in simulated parent-teacher conversations.

- (5) To what degree are the results of simulated conversations generalizable to other (simulated) conversations?

Both case vignettes contain the same type of conversation, shared decision making, and target beginning learners. Thus, the results of the two conversations should positively correlate with each other. The higher the correlation is, the more generalizable the results are. Additionally, no statistically significant effect of the case vignette on the self-assessed and observer-rated performance would indicate that both cases are equally difficult.

Research question 6 investigates the response processes of the pre-service teachers. Theoretical and empirical analyses of the response processes can provide evidence of the fit between the construct and the detailed nature of the performance or response actually engaged in by test takers (American Educational Research Association et al., 2014). The corresponding research question targets the perceived authenticity of the conversations:

- (6) To what degree do the pre-service teachers perceive the simulated conversations as authentic?

If the pre-service teachers perceive the conversations as authentic, they probably behave as if in a real situation. A high perceived authenticity would speak for the validity of the results and their prognostic value regarding real parent-teacher conversations.

Research question 7 aims at providing further evidence for validity. A crucial part of the validation process is evaluating to which degree scores of an instrument show the theoretically assumed associations and relations with other variables (American Educational Research Association et al., 2014). Relations to other variables comprise convergent, discriminant and criterion evidence (American Educational Research Association et al., 2014). Convergent evidence concerns the relationships between an instrument's scores and other measures intended to assess the same or similar constructs (American Educational Research Association et al., 2014). Measures of purportedly different constructs provide discriminant evidence (American Educational Research Association et al., 2014). Criterion related evidence refers to the relation of an instrument's scores to a relevant criterion which is ascertained at the same point in time (concurrent), in the future (predictive) or in the past (retrospective) (American Educational Research Association et al., 2014).

Research question 7 examines whether the relations of the observer ratings of the performance in simulated conversations to other variables in the context of a multimethod measurement correspond to theoretical expectations. The multimethod measurement consists of two levels: Measurements that concern the situation of the simulated conversations and the performance of the pre-service teachers in this situation per se, such as the ratings of the simulated conversations by the independent, trained observers, ratings of the simulated

parents, self-assessments of the pre-service teachers and ratings of the simulated conversations based on six coding manuals adapted from the medical context (cf. chapter 5.1.3.4). The second level of the multimethod measurement also includes measurements that go beyond the performance in the actual communication situation but might be related to the construct, such as the results of the pre-service teachers in a situational judgment test on parent-teacher communication and external criteria, like previous knowledge. From this multimethod approach the following research question arises:

- (7) Do the correlations between the observer ratings of the performance of the pre-service teachers in the simulated conversations based on the newly-developed coding manual and ratings by the simulated parents, self-assessments of the pre-service teachers, ratings based on other instruments and external criteria correspond to theoretical expectations?

The independent observer ratings of the communication competence of the pre-service teachers in the simulated parent-teacher conversations based on the newly-developed coding manual should positively correlate with self-assessments of the pre-service teachers and ratings by the simulated parents since all of them aim at measuring the same construct (convergent evidence for validity). It is probable that the two external assessments (observers and simulated conversational partners) correlate higher with each other than with self-assessments (cf. chapter 4.3.1). Prior research, e.g. by Aich (2011) and Hertel (2009) (cf. chapter 4.3), shows that self-assessments of pre-service teachers have often only a low correlation with other measurements. The observer ratings should also correlate significantly with the ratings of the performance in simulated conversations with the adapted medical instruments. The expected strength and the direction of the correlations depend on the design and content of the individual instruments (cf. chapter 7.7) (convergent and discriminant evidence for validity). Lastly, it can also be assumed that the observer ratings correlate moderately with the results in the situational judgment test (convergent evidence for validity) and with external criteria (criterion related evidence for validity). Previous knowledge, training and other preconditions, like the Abitur grade, could or should have a positive impact on communication competence. Correlations with self-assessed competence (before the simulated conversations) are also assessed. However, they are expected to be low (Aich, 2011; Hertel, 2009). A correlational structure that corresponds to the theoretical expectations would be an indicator of the validity of the measurement.

D Method

At the beginning of the method part (7), the development of the instruments applied in this dissertation is depicted. The following two chapters outline data collection (8) and data analysis (9). The focus of the instrument development in chapter 7 is on the different components of the simulated conversations that have been influential with regard to their objectivity, reliability and validity in prior medical and educational research (cf. chapter 5): the case vignettes (7.1), the recruitment and training of actors (7.2), the coding manual for analyzing the conversations (7.3) and the raters and their training (7.6).

With regard to the validation process, the newly-developed simulated conversations are embedded into a multimethod measurement in order to investigate whether their relations to other measurements and external criteria correspond to theoretical expectations and provide evidence of validity (American Educational Research Association et al., 2014). The development of the instruments applied in the validation process is also described in chapter 7. An overview of the instruments that are part of the validation process and the multimethod measurement is given in the following.

The multimethod measurement consists of two levels. The first level encompasses all measurements at the conversation level. The second level ranges beyond the actual conversation level and comprises all kinds of criteria that might be stronger (convergent evidence for validity) or weaker (discriminant evidence for validity) related to the construct at stake. On the first level, in addition to the ratings of the performance of the pre-service teachers in the simulated conversations by trained observers based on the coding manual, the pre-service teachers also self-assessed their performance in questionnaires and the simulated parents rated the performance of the pre-service teachers on a rating scale. The coding manual for the independent observers, the self-assessment questionnaire and the rating scale are theoretically based on the Munich Model of Communication Competence in Parent-Teacher Conversation. Their development is described in 7.3, 7.4 and 7.5. Besides these three ratings, independent observers rated a subsample of 20 simulated conversations with six adapted coding manuals from the medical domain. The adaptation of the medical coding manuals is described in 7.7.

The second level of the multimethod measurement investigates the relations between the results of the observer ratings, the self-assessments of the pre-service teachers and the ratings of the simulated parents with external criteria that might be related to the communi-

cation competence of pre-service teachers in conversations with parents¹⁷ and the results of the pre-service teachers in a situational judgment test on parent-teacher conversation¹⁸. Figure 5 provides an overview of the different components of the multimethod measurement for the validation process of the simulated conversations. The independent observer ratings of the performance of the pre-service teachers in the simulated conversations (center of Figure 5) should correlate (at different strengths cf. part C and chapter 15.4) with the other measurements and criteria gathered at the first and the second level of the multimethod measurement.

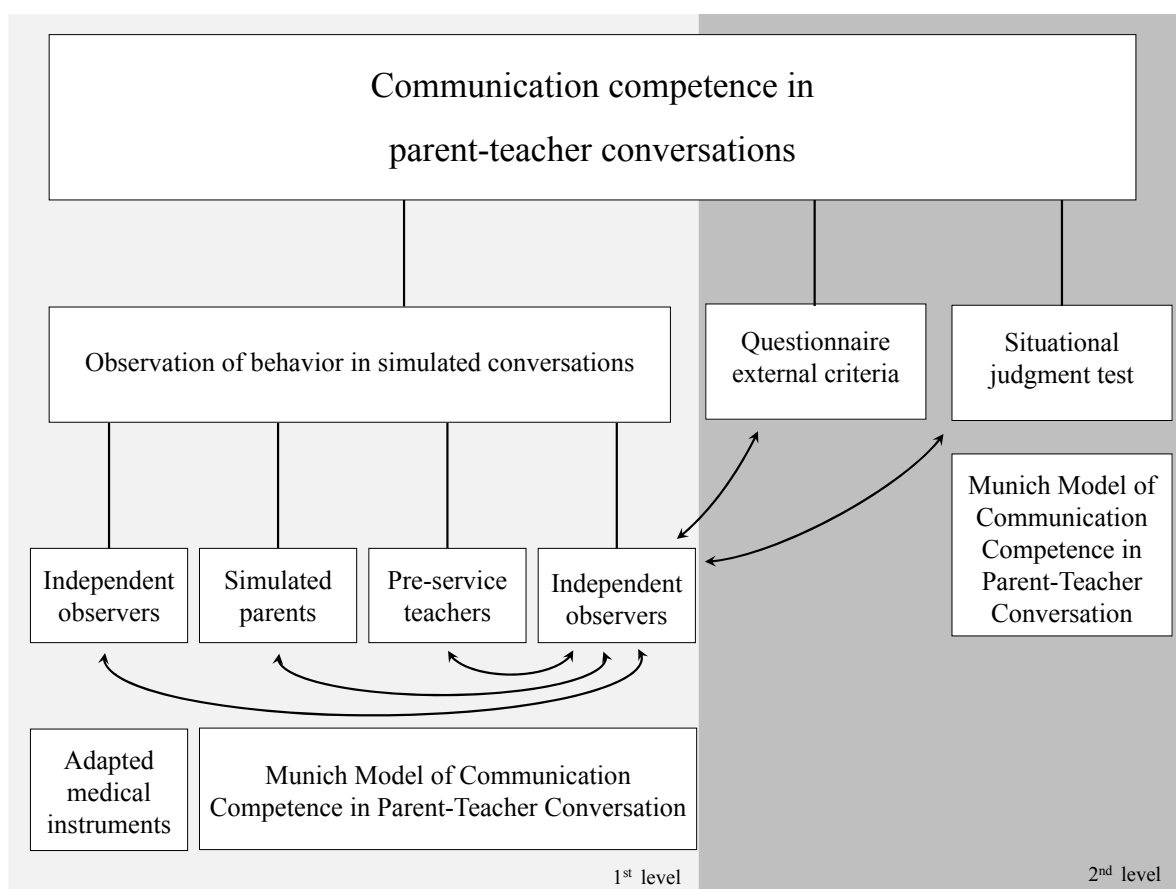


Figure 5. Components of the multimethod measurement for validating the results of the simulated conversations based on the independent observer ratings with the developed coding manual

¹⁷ For more information on the questionnaire inquiring after external criteria, see chapter 8.

¹⁸ The situational judgment test was developed by Kiessling, Gartmeier, Iblher, Karsten, Kiesewetter, Möller, Wiesbeck, Zupanic & Fischer (in preparation) and is also based on the Munich Model of Communication Competence in Parent-Teacher Conversation. Further information on its development, form, content and fulfillment of psychometric quality criteria can be found in Kiessling and colleagues (in preparation).

7 Development of the Instruments

In chapter 7 the development of the components of the simulated conversations as well as of the instruments employed for their validation is outlined in order of use. Pre-service *Gymnasium* teachers were selected as the instrument's target group, since the analysis of German parent-teacher conversation praxis in chapter 2.2 has revealed that parents are least satisfied with the conversations with *Gymnasium* teachers. Thus, the role of parent-teacher conversation in *Gymnasium* teacher education should be strengthened in particular. In the first step of the development process, the case vignettes were designed (7.1). Based on the case vignettes actors were recruited and trained (7.2). Subsequently, a coding manual for the rating of the videotaped conversations by external observers was developed (7.3). Analogously, short coding manuals for the evaluation of the simulated conversations by the conversational partners and self-assessment questionnaires for the pre-service teachers were developed for rating the performance of the pre-service teachers immediately after each simulated conversation (7.4 / 7.5). After the pre-service teachers had conducted the simulated conversations, external observers were recruited and trained to rate the videotaped conversations (7.6). In a last step, six coding manuals from the medical domain were adapted to the educational domain (7.7) and a subsample of the simulated conversations was rated with these instruments in order to investigate whether the correlational structure between the adapted medical instruments and the newly-developed coding manual corresponds to theoretical expectations.

7.1 Development of the Case Vignettes

The Munich Model of Communication Competence in Parent-Teacher Conversation distinguishes three conversation types that teachers frequently encounter: shared decision making, handling complaints and breaking bad news (cf. chapter 3.3). For the case vignettes the situation type shared decision making was selected since the target group was pre-service teachers with comparatively little previous experience and the level of difficulty was supposed to be appropriate for beginning learners. Shared decision making is supposedly less challenging for pre-service teachers than handling complaints or breaking bad news because they consider their self-efficacy regarding shared decision making conversations as higher than regarding the other two conversation types (cf. chapter 8.1.2).

With regard to the adequate number of cases, practicability and requirements regarding reliability and validity were weighed (cf. chapter 5.1.3.1). Research on the employment of simulated conversations in teacher education shows that simulations can only realize their full potential if more than one simulation is provided, since the novelty of a participant's first simulation often distracts him / her from the content of the simulation (cf. chapter 5.2.3). However, the sample in this study was comparatively large and simulated conversations require time and money. Thus, for practicability reasons, as a start two case vignettes were developed. In line with medical guidelines (cf. chapter 5), the cases were based on real-life cases extrapolated in a preceding Delphi study by Gartmeier, Bauer, Noll and Prenzel (2012). In this Delphi study Gartmeier and colleagues (2012) questioned 23 *Gymnasium* teachers, with an average professional experience of 13.1 years, to extract typical situations, challenges and strategies in parent-teacher conversations. Routine cases with multiple opportunities to demonstrate core competence facets were chosen, since medical research on simulated conversations shows that this decreases the number of cases needed for a valid measurement (cf. chapter 5.1.3.1).

In the pilot study both cases concerned students with poor school achievements, a situation likely to be experienced by teachers sooner or later during their professional career. In the main study one of these cases was substituted by another shared decision making case that concerns the counseling of parents with regard to the student's choice of the scientific or linguistic branch in order to investigate the generalizability of the results of simulated conversations to other, similar situations. The choice between the linguistic and the scientific branch is another situation most German teachers will face in their lives because most German *Gymnasium* students have to make this decision during their first two years. The choice of the linguistic or scientific branch as well as the dealing with poor marks decisively influences the future of the students. Thus, both cases comply with the three medical tenets for crafting cases. They are prevalent and are of social and instructional importance (cf. chapter 5.1.3.1).

Similar to the cases crafted by Dotger and colleagues (cf. chapter 5.2.3), the case vignettes for the simulated parents consist of informational and interactional content. The first part of a case vignette consists of background information about the parent, such as age, education level, domestic situation and relationship to the child. The next part outlines the rationale for the conversation and gives information about the student's scholastic situation. The interactional content is subdivided into fixed content and triggers in an if-then format.

Fixed interactional content as well as triggers are comprised of verbal and nonverbal behaviors, e.g. types and degree of emotions to be portrayed via facial expressions or bodily-kinesthetic positions.

The case vignettes for the pre-service teachers are more general than the ones for the simulated parents. The pre-service teachers are told to portray themselves or a future teacher version of themselves and to act according to their best professional knowledge. The teacher case vignettes describe a hypothetical student and give a rough idea about the rationale for the conversation. Moreover, they give information about the pedagogical content of the conversation, e.g. about the advantages and disadvantages of the scientific and linguistic branches. This information was given to ensure that all of the participating pre-service teachers possessed the relevant pedagogical content knowledge to conduct the conversation since study participants came from different Bavarian universities and were in different semesters. The options provided for the shared decision making process are not per se better or worse. It is up to the pre-service teacher to elicit information from the parent in order to ascertain which option fits the student best. Consequently, the focus of the assessment is on how the pre-service teachers conduct the conversations. The case vignette dealing with the choice of the scientific or linguistic branch is enclosed as an example in the appendix (cf. chapter 21.1).

7.2 Recruitment and Training of the Simulated Parents

For the pilot study four actors were recruited to portray parents. In the main study participant numbers were higher. Thus, six actors portrayed parents in the main study. Three of them had already been simulated parents in the pilot study. One pilot study actress had other appointments and was not available for all dates of the main study. In consequence, three new actors were recruited for the main study. The recruitment process took into account the four key factors for the recruitment of simulated patients proposed by Cleland and colleagues (2009): ability, suitability, conscientiousness and credibility (cf. chapter 5.1.3.2). Five of the recruited people are professional actors and the sixth works as a simulated patient, which suggests that all of them are able to serve as simulated parents (ability). During the recruitment process they were asked for their motives for participating (suitability) and had to assure that they would be available for all dates at which simulated conversations would take place (conscientiousness). Finally, applicants were screened and

selected with regard to case requirements, such as gender, age, demographics and personal background (credibility).

For the training the actors were split into two groups according to the two case vignettes. The female actors acted out one case vignette and the male actors the other¹⁹. Training for each case lasted approximately two hours. The first part of the training consisted of a general introduction to the concept and purpose of simulated conversations and the responsibilities of being a simulated parent. In the second part of the training, the case vignette and its triggers were discussed in detail and the actors asked questions. The last part of the training consisted of trial runs. In order to provide a model for the simulated parents, the workshop leader²⁰ acted out the first trial run. Each actor did at least one trial run. The simulated parents were advised to closely watch each other during the trial runs and to adjust their behavior and portrayal of the role to each other. Moreover, they were warned against cueing the pre-service teachers and becoming expert parents.

The main difference between portraying patients and parents is that for authenticity purposes, simulated conversations in the educational domain are less predictable and less standardized than in the medical domain (cf. chapter 5.2.4). Thus, in order to achieve a balance between reliability and validity, the simulated parents acted out a variety of possible responses and conversational directions. Moreover, they were encouraged to make up additional background information and to improvise if something unpredictable happened. It was stressed that the most important point was that they always stayed in character no matter what happened.

In the main study the simulated parents additionally received a short introduction to the rating scale on which they had to judge the performance of the pre-service teachers in the simulated conversations (cf. chapter 7.4). They were told to rate each conversation but were asked not to give any feedback to the study participants. Since the two simulated conversations took place immediately after each other, the quality and kind of feedback to the first conversation could have influenced the performance in the second. Moreover, the simulated parents were informed that study participation was voluntary and poor grades on the rating scale would not have negative consequences for the pre-service teachers. Main study participants received detailed feedback and access to their videos in a follow-up study supervised by the author of this dissertation (Altmann, 2014).

¹⁹ A systematic variation of gender and case vignette was not possible in this study but is planned for future research in order to single out possible effects of gender.

²⁰ Author of this dissertation

The final guideline for the training of the simulated parents is included in the appendix (cf. chapter 24.2). It is slightly more comprehensive than the training of the simulated parents described in this chapter since it was revised based on the results of the pilot study and some additional components, such as video analysis of pilot study conversations and critical phases and incidents in the conversation, were added. The refinement of the training of the simulated parents is described in chapter 13.2.

7.3 Construction of the Coding Manual

As outlined in chapter 5.1.3.4, the choice of low or high inference methods for assessing the performance in simulated conversations depends on the raters, the amount of training they receive and the purpose of the results. In this dissertation the results should provide information about the communication competence of pre-service teachers in conversations with parents. Since this requires a qualitative assessment of performance, rating scales were preferred over symbol systems or category systems. As a consequence, the raters needed a large amount of training (cf. chapter 7.5).

The development of the coding manual, the corresponding rater training and the coding of the videos followed the guidelines set up in Seidel and colleagues (2005): In a first step, a theoretical foundation for the development of the coding manual and the rating scales was selected, the Munich Model of Communication Competence in Parent-Teacher Conversation (cf. chapter 3.3). Since Franke (2005) suggests that identifying the structure and the relations between different competence facets allows diagnosing competencies in a more differentiated way, the scale communication competence in parent-teacher conversations was partitioned into three subscales: structuring the conversation, problem solving and establishing a relationship with the conversational partner according to the competence facets of the model. This proceeding is in line with the findings by Hertel (2009), discussed in chapter 3.2.3, which suggest that the communication competence of teachers in conversations with parents is three-dimensional.

Subsequently, in line with Seidel and colleagues (2005), the sampling strategy was selected (event sampling) and items for the three rating scales were deduced from the underlying theoretical model. In total 18 items were developed, six for each rating scale / competence area. Three of these items are global items targeting the entire competence facets, structuring the conversation, problem solving and establishing an interpersonal relationship. The rest of the items were deduced from a construct map that singled out sub learning goals in

the three competence areas, like coming to a concrete agreement at the end of the conversation as part of the competence facet problem solving.

For each item a description of its characteristics and its competence levels including indicators, examples and handling instructions for coding were developed, with recourse to the theoretical model and the video material, and included in the coding manual in order to increase reliability (Langer & Schulz von Thun, 2007; Seidel et al., 2005). The items comprise five competence levels which roughly correspond to German school grades (1 = very good - 5 = fail). If the coding manual is used for summative purposes, the pass / fail requirement should be that pre-service teachers score on average four or better in all of the three competence areas, since this is the minimum requirement to successfully conduct a parent-teacher conversation. The non-compensatory approach should be chosen since even an extremely well-structured conversation will not satisfy parents if no interpersonal relationship is established or vice versa.

The items were validated with recourse to video material and by comparing various expert judgments (discussion of the items in a group of four experts with 2-15 years of experience in teaching and researching on communication competence; three of them were from the educational and one from the medical domain). The coding manual for the pilot study and the revised version for the main study (cf. chapter 14) are enclosed in the appendix (cf. chapter 24.3 / 21.4).

7.4 Construction of the Rating Scale for the Simulated Parents

The rating scale for the simulated parents consists of the three global items of the competence facets from the coding manual for the external observers. The items are exactly the same apart from an adaption of the perspective, e.g. the teacher established a good relationship with *me*. The rating scale for the simulated parents is included in the appendix (cf. chapter 24.5).

It is challenging for simulated conversational partners to play a role and rate the conversation at the same time (cf. chapter 5.1). Thus, the conversational partners only rated the performance of the pre-service teachers in the main study after the pilot study analysis had shown that the actors were able to adequately portray their characters.

7.5 Construction of the Self-Assessment Questionnaire for the Pre-Service Teachers

For the pilot study seven of the 18 items of the coding manual for the external raters were adapted and made into a self-assessment questionnaire for the pre-service teachers ($\alpha = .91$). The items represent the three competence areas, structuring the conversation (2 items, $\alpha = .70$), problem solving (4 items, $\alpha = .85$) and establishing a positive relationship to the conversational partner (1 item). For the main study three more items of the coding manual were adapted and added to the self-assessment questionnaire for the pre-service teachers ($\alpha = .83$) so that the subscales for all three competence areas consisted of a similar number of items: structuring the conversation (3 items, $\alpha = .69$), problem solving (4 items, $\alpha = .65$) and establishing a positive relationship to the conversational partner (3 items, $\alpha = .73$). The pre-service teachers had to rate their performance on 4-point Likert scales, with higher values indicating better performance; e.g. I succeeded in coming to a concrete agreement with my conversational partner. The final version of the self-assessment questionnaire for the pre-service teachers is enclosed in the appendix (cf. chapter 24.5).

7.6 Recruitment and Training of the Raters

Two teachers and three psychologists in training were recruited as potential raters for the coding of the simulated conversations. All of these potential raters participated in a communication training for parent-teacher conversations in order to adjust their previous knowledge to each other. Subsequently, they took part in a two-day rater training based on video-taped simulated conversations. The first part of the rater training consisted of six components, like a discrimination training (Langer & Schulz von Thun, 2007), a concept training (Langer & Schulz von Thun, 2007; Seidel et al., 2005) or an introduction to frequent rater errors (Wirtz & Caspar, 2002). In the discrimination training, the first component of the rater training, the raters had to sort five videos with regard to the quality of the performance of the pre-service teachers in the simulated parent-teacher conversations (Langer & Schulz von Thun, 2007). The five videos comprised a very good conversation as well as a poor conversation and some mediocre conversations. This training component aimed at imparting a feeling for the range of performance to expect and the corresponding appropriate scoring (Langer & Schulz von Thun, 2007). In the concept training, the coding manual was introduced and the raters had to assign the observed behavior to items and

competence levels (Langer & Schulz von Thun, 2007; Seidel et al., 2005). In the introduction to frequent rater errors, these errors were presented and strategies for avoiding them were discussed. For an overview of the entire rater training and all six training components see appendix 24.7. Videos were rated and discussed in group until all raters, a participating expert who had also contributed in the validation process of the coding manual and the workshop leader, stated to have no more questions and a common theoretical understanding (cf. Seidel et al., 2005). During the group discussion the coding manual was slightly revised, e.g. coding rules were differentiated or introduced additionally and more examples were included where needed (cf. Seidel et al., 2005).

After the first part of the rater training, the five potential raters, the expert and the workshop leader coded ten videos in a trial run. All raters had one week to code the videos independently from each other. The video coding lasted around eight to ten hours for all ten training videos. The order of the videos was randomized for the trial run as well as for all other analyses in order to avoid primacy-recency and other sequence effects (Wirtz & Caspar, 2002). Inter-rater agreement was calculated with intraclass correlations (cf. chapter 9). After the trial run inter-rater agreement across all seven raters and all items of the coding manual was $ICC = .84$. The raters reached the set cut-off point of $ICC = .60$ (cf. chapter 9) for all but two items²¹. Table 2 shows the inter-rater agreement sorted by items.

²¹ The raters were particularly retrained for these two items in the second part of the rater training.

Table 2

Inter-rater agreement across seven raters after the trial run sorted by items

Item	ICC
Global rating	
Structuring the conversation	.79
Problem solving	.92
Establishing a relationship to the conversational partner	.66
Structuring the conversation	
Detectability of the SDM conversational phases	.86
Adequacy of the length of the SDM conversational phases	.89
Correctness of the order of the SDM conversational phases	.84
Use of metacommunication	.91
Proportion of participation of both conversational partners	.36
Problem solving	
Successful establishment of common ground	.89
Comprehensibility of the presented options	.90
Quality of cooperation in the negotiation process	.73
Coming to a concrete agreement	.88
Subject-specific performance	.95
Establishing a positive relationship with the conversational partner	
Unconditional positive regard	.76
Authenticity	.82
Empathy	.72
Conversational climate	.87
Nonverbal behavior	-.70

Note. SDM = shared decision making; ICC = intraclass correlation.

In order to choose two raters for the coding process, inter-rater agreement for all possible pairs of raters, always in combination with the expert and the workshop leader, was calculated. Table 3 displays the inter-rater agreement across all items for all possible rater combinations:

Table 3

Total inter-rater agreement for every possible rater pair and the two experts after the trial run

Rater	1&2	1&3	1&4	1&5	2&3	2&4	2&5	3&4	3&5	4&5
ICC	.80	.75	.85	.76	.73	.84	.74	.78	.69	.79

Note. ICC = intraclass correlation.

The results in Table 3 show that all possible rater pairs achieved an inter-rater agreement higher than the targeted cut-off point of ICC = .60 (cf. chapter 9). This indicates that the rater training adequately prepared the raters for applying the coding manual and rating video-taped simulated conversations. The rater pairs 1&4 and 2&4 achieved the highest inter-rater agreement. Since rater four would not have been available for the entire time period required, rater one and two with the third highest inter-rater agreement were chosen for the coding of the pilot study videos.

In the second part of the rater training, the two selected raters, the expert and the workshop leader discussed the ratings in detail and fine-tuned them based on empirical analysis. In a first step, the inter-rater agreement was targeted with particular emphasis and discussion of the two items that had an inter-rater agreement below the cut-off point (ICC = .60, cf. chapter 9) after the trial run. Subsequently, the leniency / strictness of the four raters was equalized. In the trial runs the expert ($M = 2.49$, $SD = 1.00$) and the workshop leader ($M = 2.51$, $SD = 1.03$) had been slightly stricter than the raters: rater 1 ($M = 2.09$, $SD = 1.05$), rater 2 ($M = 2.11$, $SD = 0.73$). Means for all items were discussed and an additional rule was set up for those items in which the raters had been less strict than the expert and the workshop leader: If in doubt, they should go for the stricter score. Frequency analyses showed that rater 1 did not utilize the entire scale (she never scored a five) while all others did. In consequence, video sequences for which the three other raters had scored a

five were watched again in group and criteria for scoring a five (=fail) were set up. Moreover, video ratings were graphically analyzed and extreme cases were discussed in order to single out where and why ratings diverged. In order to further calibrate the raters, ratings were also compared to sample solutions. For an overview of all contents of the second rater training see chapter 23.7. All videos that had been used for rater training were excluded from the following data analysis (Seidel et al., 2005). Inter-rater agreement was calculated again after 50% and after 100% of the videos had been scored (cf. chapter 10.1).

7.7 Adaptation of six Coding Manuals From the Medical Domain

Since simulated conversations are a widespread and frequently used method in the medical domain (cf. chapter 5.1), multiple coding manuals to rate physician-patient encounters are in use. The tEACH Assessment subgroup (2012) lists 64 worldwide. The existing instruments vary considerably regarding the targeted instrument user, the content, the design, the evidence provided for the fulfillment of psychometric quality criteria and the theoretical background. However, there are few studies that compare them and there is little agreement on the ideal one (Schirmer et al., 2005). Six frequently used medical coding manuals - SEGUE (set the stage, elicit information, give information, understand the patient's perspective, end the encounter), OPTION (observing patient involvement in decision making), Common Ground, EPSCALE (Explanation and Planning Scale), LUCAS (Liverpool Undergraduate Communication Assessment) and BGR (Berliner Global Rating) are selected for the multimethod measurement in the present study. Their use provides convergent and discriminant evidence for the validity of measurements with the newly-developed coding manual.

SEGUE and OPTION are part of the multimethod measurement because SEGUE is one of the most frequently used checklists (Makoul, 2001b) and OPTION one of the most frequently used global ratings (Edgcumbe, Silverman, & Benson, 2012). Common Ground was chosen because reaching common ground is one part of the coding manual developed to rate the simulated conversations in this dissertation. Since explaining possible options to the conversational partner is another component of the newly-developed coding manual, EPSCALE was selected. LUCAS tries to identify problematic rather than excellent communication (Huntley et al., 2012). It is included in the multimethod measurement because it approaches communication competence from a completely different, complementary

perspective and, thus, might provide discriminant evidence for validity. The BGR is finally part of the multimethod measurement because it is a very economic global rating often used in the German context.

In chapter 7.7.1 the content and design of the instruments is compared and the state of research for each instrument is discussed. Subsequently, it is outlined how the instruments were adapted to parent-teacher conversations. Lastly, hypotheses about the correlational structure between the newly-developed coding manual and the six adapted medical instruments are set up based on the content, design and research findings of the instruments.

7.7.1 Content, Design and State of Research on the Medical Instruments

The content and design of the six instruments is depicted in Table 4. Most medical instruments are either based on the Kalamazoo Consensus Statement (Makoul, 2001a) or on the Calgary-Cambridge Guides (Kurtz, Silverman, Benson, & Draper, 2003). The Kalamazoo Consensus Statement was derived from a multidisciplinary panel of experts that defined seven essential elements with evidence of their positive effect on physician-patient communication: 1) establishing rapport, 2) opening the discussion, 3) gathering information, 4) understanding the patient's perspective of illness, 5) sharing information, 6) reaching agreement on problems and plans and 7) providing closure to the conversation (Makoul, 2001a). The Calgary-Cambridge Guides are an overall framework in which to organize the numerous skills of communication (Kurtz et al., 2003). It identifies 70 core evidence-based communication process skills that fit into the framework (Kurtz et al., 2003). This skills repertoire is meant to be used as required; it is not a list to be strictly followed in every encounter (Kurtz et al., 2003). The first part of the guides deals with interviewing the patient, the second part with explanation and planning and the third one is a content guide (Schirmer et al., 2005). There is a cross-reference to the Kalamazoo Consensus Statement (Makoul, 2001a) or the Calgary-Cambridge Guides (Kurtz et al., 2003) in the respective subchapters following the overview chart if an instrument is based on one of the two. Also in the following subchapters, research findings on the different instruments are summed up focusing on inter-rater agreement / objectivity of scoring, internal consistency / reliability and different types of evidence for validity.

Table 4

Six instruments for assessing physician-patient encounters

	SEGUE	OPTION	Common Ground	EPSCALE	LUCAS	BGR
Aim	Assess communication skills	Observe patient involvement in decision making	Assess communication skills	Assess communication skills in explanation and planning	Assess whether communication meets the needs of patients	Assess communication competence
Content	Set the stage, elicit information, give information, understand the patient's perspective, end the encounter	Identifying the problem, offering and eliciting the patient's desire for shared decision making, providing and explaining options, exploring patient's expectations and concerns, securing the understanding of the conversational partner, indicating the need for decision-making and reviewing the decision	Rapport, information management, agenda setting, active listening, addressing feelings, reaching common ground, family interviewing skills and global performance	Building the relationship between patient and clinician, providing the appropriate information for the patient, aiding accurate recall and understanding, and achieving a shared understanding	General skills, respect and empathy, questions, giving information	Response to patient's feelings and needs, degree of coherence in the interview, verbal expression, non-verbal expression
Design	Checklist with a nominal (yes/no) scale	Global 4-point rating scales with 12 items	Checklist, number of occurrences & global 5-point rating scale	Global 4-point rating scale with 15 items	Global 4-point rating scale with 12 items	Global 5-point rating scale with 4 items
Language	English	English & German	English	English & German	English	English & German

SEGUE

SEGUE is a research-based checklist of medical communication tasks (Makoul, 2001b), which has been used for 15 years (Skillings et al., 2010) and adheres to the generic communication skills emphasized by the consensus statements of Toronto and Kalamazoo (cf. chapter 5.1.3.4) (Schirmer et al., 2005). It has a high degree of acceptability (Makoul, 2001b) and its easy usability is positively highlighted (Skillings et al., 2010). Due to its nominal scale and clear coding rules, inter-rater agreement for SEGUE is high when simulated patients are scoring performance immediately after live encounters and also when coders are evaluating videotaped or audiotaped encounters (Makoul, 2001b). Intra-rater agreement is high and there is evidence for its internal consistency / reliability (Makoul, 2001b). There is some evidence for validity gathered via an expert review process (Makoul, 2001b). There are also moderate correlations with related other variables, such as patient satisfaction (Makoul, 2001b), that provide some criterion-based evidence for validity. However, SEGUE is also criticized substantially with regard to evidence for validity (Skillings et al., 2010). SEGUE was unable to detect longitudinal improvement in communication scores of students while qualitative ratings from faculty members did show improvements (Skillings et al., 2010).

OPTION

OPTION is a very frequently used global scale observing patient involvement in shared decision making (Edgcumbe et al., 2012; Elwyn et al., 2005). There are several versions of the OPTION scale including two German ones (Hirsch et al., 2012; Keller et al., 2013). Inter-rater agreement with OPTION is good (Elwyn et al., 2005). The OPTION scale is internally consistent / reliable (Elwyn et al., 2005). There is also evidence for OPTION's validity since factor analyses confirmed the assumed internal structure for the English and the German version (Elwyn et al., 2005; Hirsch et al., 2012). Moreover, physicians with more expertise in shared decision making received higher ratings based on OPTION (Hirsch et al., 2012). Keller and colleagues (2013) tested a modified German version of the OPTION scale and showed that OPTION's results are correlated to expert ratings, which is an indicator of the convergent validity of OPTION.

Common Ground

Common Ground tries to combine the merits of checklists and global assessments (Lang et al., 2004). It is linked to the consensus statements of Toronto and Kalamazoo (Lang et al.,

2004). Inter-rater agreement for the total scale was high with agreement regarding the checklist being higher than for the global ratings (Lang et al., 2004). G-coefficients indicated that assessment results are consistent across a number of cases (Lang et al., 2004). There is also evidence for validity. Common Ground was sensitive to competence differences between medical students in their first-year and in their fourth-year (Lang et al., 2004). In addition, the correlation between the ratings of trained raters and a panel of independent communication experts was high (Lang et al., 2004), which is a further indicator of the validity of Common Ground.

EPSCALE

EPSCALE aims at assessing the second part of a physician-patient interview where explanations are provided to the patient and future proceeding is planned (Silverman, Archer, Gillard, Howells, & Benson, 2011). It is based on the Calgary-Cambridge Guides to consultation, which contains 20 items on explanation and planning (Silverman et al., 2011). EPSCALE is internally consistent / reliable and according to a Generalizability study its results possess generalizability (Silverman et al., 2011). Validity was supported by a consensus exercise and expert reviews (Silverman et al., 2011). Further findings also speak for the validity of EPSCALE: Factor analysis showed that EPSCALE broadly corresponds to the domains of the Calgary-Cambridge Guides (Edgcumbe et al., 2012) and Kiessling and colleagues (2013) tested a German version of EPSCALE and showed that EPSCALE's results are significantly correlated to results attained with OPTION and the Berliner Global Rating.

LUCAS

LUCAS tries to move the primary focus of examiners away from an assessment of students' enactment of certain skills to a judgment of how well students' communication meets the need of patients (Huntley et al., 2012). To do so, the instrument aims at identifying problematic rather than excellent communication (Huntley et al., 2012). Inter-rater agreement for LUCAS was good and the scale was internally consistent / reliable (Huntley et al., 2012). There is also evidence for the validity of LUCAS: LUCAS was able to discriminate between performances of students at different levels (Huntley et al., 2012). Moreover, the dimensionality of LUCAS was tested with factor analysis and corresponded to the assumed internal structure (Huntley et al., 2012) and examiner ratings correlated with ratings by the simulated patients (Huntley et al., 2012).

Berliner Global Rating

The Berliner Global Rating is based on the analytic global rating form developed by Hodges and McIlroy (2003). The original version was slightly modified and translated into German (Scheffer, 2009). Inter-rater agreement between examiners, simulated patients and experts were high (Scheffer, 2009). An indicator of the validity of the Berliner Global Rating is that its results were highly correlated with results of the Calgary-Cambridge Guides and moderately with ratings of clinical skills (Scheffer, 2009).

7.7.2 Adaptation of the Medical Instruments to the Educational Context

The medical instruments were adapted to parent-teacher conversations by 17 pre-service teachers guided by the author of this dissertation. The pre-service teachers studied at the TUM School of Education in the last semester of their Bachelors and participated in two seminars on research-based learning in the winter term 2013/2014 and the summer term 2014 led by the author of this dissertation. They worked together in groups of two-three persons. The pre-service teachers were instructed to adjust the terminology of the instruments, e.g. teachers instead of physicians, and to transform items that did not apply to the educational context. In general, changes were only slight since the pre-service teachers received the guideline to stick to the medical instruments as close as possible. If applicable, the pre-service teachers also translated the instruments from English into German. With the final versions of the instruments each group of pre-service teachers conducted a rater training with their instrument. The rater training corresponded to a shortened version of part one (and optionally part two) of the rater training presented in chapter 7.6, which had been introduced to them in class. Training videos were excluded from further analyses. After the pre-service teachers had successfully completed the rater training and had presented the results from the rater training and / or trial runs in class, they rated a randomly selected subsample of 20 video-taped simulated conversations from the main study independently from each other in the same groups that had developed the instruments and participated in training together. Each video was watched one time only and then rated. Under the guidance of the author of this dissertation, the pre-service teachers conducted intraclass correlations to calculate their inter-rater agreement. Inter-rater agreement was above the set cut-off point ($ICC \geq .60$) for all rater groups. Table 5 provides an overview of inter-rater agreement across the entire subsample of 20 videos sorted by instrument.

Table 5

Inter-rater agreement for the six adapted medical instruments

	SEGUE	OPTION	Common Ground	EPSCALE	LUCAS	BGR
Raters	3	3	3	2	3	3
ICC	.77	.80	.85	.74	.62	.83

Note. ICC = intraclass correlation.

While inter-rater agreement with all instruments was above the set cut-off point ($ICC \geq .60$), it was lower with the instrument LUCAS than with the other instruments. The reason for this might be that it is easier for raters to judge whether pre-service teachers did a good job at conducting a parent-teacher conversation than whether their behavior was “borderline” or “unacceptable” (Huntley et al., 2012). The deficit-oriented approach that Huntley and colleagues (Huntley et al., 2012) pursue, to identify inappropriate behavior, was probably new and, thus, more challenging for the pre-service teachers who rated the videotaped conversations.

7.7.3 Expected Correlational Structure Between the Adapted Medical Instruments and the Newly-Developed Coding Manual

The adapted medical instruments differ in content and design. Due to these differences, the strengths of the correlations between measurements with the newly-developed coding manual and measurements with different adapted medical instruments supposedly differ. That is, depending on content and design, the correlation between the newly-developed coding manual and instrument x is possibly stronger than between the newly-developed coding manual and instrument y. A correlational structure between the newly-developed coding manual and the six adapted medical instruments that corresponds to theoretical expectations would provide convergent and discriminant evidence for validity since some instruments are expected to correlate highly with the newly-developed instrument while others are expected to correlate only moderately or not at all (American Educational Research Association et al., 2014). The expected correlational structure is outlined in the following.

All adapted medical instruments apart from SEGUE contain high inference rating scales. The newly-developed coding manual consists also of high inference rating scales and research findings indicate that high inference rating scales are better at making qualitative assessments of competencies than checklists (cf. chapter 5.1.3.4). SEGUE is a checklist and seems to have a low discriminant validity with regard to communication competence (Skillings et al., 2010). Thus, it is probable that the correlation between SEGUE and the newly-developed coding manual is low or non-existent.

In contrast, the highest correlation is expected between Common Ground and the newly-developed coding manual because establishing common ground is a part of the newly-developed coding manual. Moreover, the content of Common Ground coincides widely with the content of the newly-developed coding manual; e.g. agenda setting, active listening and addressing feelings are considered as important features by both coding manuals.

Results of the Berliner Global Rating are probably highly correlated to results of the newly-developed coding manual since the global rating scales of the Berliner Global Rating target, analogously to the newly-developed instrument, the competencies to structure a conversation and to establish a relationship.

The correlation between OPTION and the newly-developed instrument is also expected to be medium to high. OPTION observes how patients are involved in shared decision making conversations. The shared decision making process is also an important content of the newly-developed coding manual.

A slightly lower correlation is expected between EPSCALE and the newly-developed instrument since one sub facet of the newly-developed instrument is explanation and planning, which is the main focus of EPSCALE. However, the focus of EPSCALE is far narrower than the focus of the newly-developed coding manual. Thus, the correlation is expected to be low to medium.

Between LUCAS and the newly-developed instrument a medium correlation is expected since both instruments focus on communication competence in general and feature similar content, such as respect or empathy for the conversational partner. However, since the two instruments investigate communication competence from opposite perspectives, LUCAS addresses in how far students do not meet the needs of patients, only a medium to low correlation is expected.

In sum, the following correlational structure between observer ratings with the newly-developed coding manual and the adapted medical instruments is expected:

Common Ground > BGR > OPTION > EPSCALE > LUCAS > SEGUE.

The expected strength of the correlations is depicted in Table 6.

Table 6

Expected correlational structure between observer ratings based on the newly-developed coding manual and the six adapted medical instruments

	Common Ground	BGR	OPTION	EPSCALE	LUCAS	SEGUE
Coding manual	+++	+++	++	++	++	(+)

Note. +++ = high correlation; ++ = medium correlation; (+) = weak correlation or no correlation.

Apart from the correlations between the observer ratings based on the adapted medical instruments and on the newly-developed coding manual, correlations between a) the ratings by the simulated parents and ratings with the adapted medical instruments and b) the self-assessments of the pre-service teachers and ratings with the adapted medical instruments are also investigated in order to provide further evidence for convergent and discriminant validity. In general, it is expected that the correlational pattern between the ratings of the simulated parents and the ratings with the adapted medical instruments corresponds mainly to the expected correlational pattern between the observer ratings based on the coding manual and on the adapted medical instruments. However, the correlations between observer ratings based on the newly-developed coding manual and on the six adapted medical instruments are expected to be higher than correlations between the ratings of the simulated parents and observer ratings with the six adapted medical instruments, since ratings by independent observers tend to be more reliable than ratings by conversational partners (cf. chapter 4). Lastly, correlations between ratings with the adapted medical instruments and self-assessments of the pre-service teachers are expected to be the lowest, since self-

assessments tend to be less reliable and valid than observations by other persons (cf. chapter 4).

Since the medical instruments had already been applied and evaluated in the medical context, they were not piloted again but only employed in the main study. All other instruments and components of the simulated conversations, apart from the coding manual for the simulated parents (cf. 7.4), were tested in the pilot and reemployed in the main study.

8 Data Collection

Data collection for this dissertation took place in the context of the project *ProfKom-Professionalisierung von zukünftigen ÄrztInnen und Lehrkräften im Bereich der Kommunikationskompetenz*. The interdisciplinary research project was conducted from 2009-2012 under the direction of the School of Education of the TU München in cooperation with the KiMed Centre for Medical Education, Kiel University and the Institute for Medical Education, LMU Munich. The aim of the project was to promote and assess the competence of pre-service physicians to communicate with patients and of pre-service teachers to communicate with parents, thereby evaluating whether it is possible to develop a communication training program that can promote communication competence in expert-layperson conversations across domains and singling out especially effective training components. The project was funded by the German Federal Ministry of Education and Research. For further information consult the project website: <http://profkom.medizin.uni-kiel.de> or Gartmeier and colleagues (2011; 2015). In the following, the sample and design of the pilot and the main study are described with focus on the simulated conversations in the educational domain which provided data for this dissertation.

8.1 Pilot Study

The pilot study took place in April 2011 at the TUM School of Education in Munich, Germany. Training and assessment for the participants lasted one day. Participants could apply for any of four training days. Four classrooms were equipped with cameras and camerapersons in order to videotape the simulated conversations.

8.1.1 *Research Design*

At the beginning of each day of the study, all participants filled out questionnaires covering, among other things, sociodemographic information, personal characteristics, anticipated self-efficacy regarding future parent-teacher conversations, previous knowledge, motivation and general information. The questionnaires consisted for the most part of 4-point Likert-type scales with higher values indicating better performance. Some items had dichotomous answering categories (female / male; yes / no) or specific answering categories (type of school; subjects). After the completion of the questionnaires, the participating

pre-service teachers were randomly assigned to two different communication training programs for parent-teacher conversations in which they took part before they conducted the simulated conversations as an assessment at the end of the day. All training and assessment components were timed to keep preparation times the same for all study participants.

Each pre-service teacher conducted two conversations, one with a simulated mother and one with a simulated father. Prior to the conversations the pre-service teachers received a short general introduction and 20 minutes to prepare both simulated conversations based on the two case vignettes. Group size during the preparation for the simulated conversations was four people. In order to keep testing conditions constant, the preparation for the simulated conversations was always conducted by the same person. After the preparation each pre-service teacher was assigned a conference room in which she or he received both parents successively. To preserve the anonymity of the participants, each participant held up a code for the camera at the beginning of the simulated conversation so that data could later be assigned appropriately to the questionnaire with the corresponding code. Once pre-service teachers signaled that they were ready, the cameraperson cued the simulated parents to enter. From this moment on the simulated conversations unfolded according to the case vignette of the simulated parent and the behavior of the pre-service teachers. Conversations lasted around ten minutes, which was the approximate time frame participants were given before the conversation. The cameraperson displayed the remaining time after five and nine minutes. However, the pre-service teachers did not have to stop exactly after ten minutes but were supposed to bring the conversation to an end. When the simulated parents had left, the pre-service teachers received five minutes to self-assess their performance in questionnaires (cf. chapter 7.5). Participants were separated from their peers until everybody had completed the simulated conversations.

Two trained raters (cf. chapter 7.3) rated all 98 videotaped conversations with the newly-developed coding manual. Rating for one video took around 45 minutes. The raters watched each video three times, one time for each competence facet, structuring the conversation, problem solving and establishing a relationship. The raters did not know which type of communication training program the pre-service teachers had received in order to avoid expectancy effects.

8.1.2 *Sample*

$N = 49$ pre-service *Gymnasium* teachers took part in the pilot study. 33 of them were female. The pre-service teachers were on average 23.20 years old ($SD = 3.95$) and enrolled in their 4th to 5th semester ($M = 4.79$, $SD = 3.57$) at Bavarian universities. 79.6% of the pre-service teachers studied MINT²² subjects, 12.2% social sciences and 8.2% a combination of both. All but one of the pre-service teachers who studied social sciences or a combination of MINT and social sciences had at least one language as a main subject. All of the participants were born in Germany and reported German as their first language. 45.7% reported to have no previous knowledge with regard to conducting professional conversations. 43.4% reported having little to some previous knowledge and 10.9% reported having considerable previous knowledge.

The pre-service teachers self-assessed their general communication competence prior to the communication training program as relatively high ($M = 3.06$, $SD = 0.25$). The scale general communication competence ($\alpha = .80$) of the self-assessment questionnaire consisted of subscales targeting broad competences that might be helpful for conducting conversations, such as comprehensibility (6 items, $\alpha = .62$), articulateness (7 items, $\alpha = .73$), persuasiveness (4 items, $\alpha = .61$) and empathy (7 items, $\alpha = .74$). The pre-service teachers rated items such as “I am good at empathizing with others”.

The self-concept of the pre-service teachers regarding their subjects of study was also comparatively high ($M = 3.02$, $SD = 0.43$). They answered items such as “The content of my study program does not pose difficulties for me”. The scale self-concept about subjects of study consisted of four items and had low reliability ($\alpha = .52$).

The pre-service teachers reported a medium level of anticipated self-efficacy ($M = 2.53$, $SD = 0.30$) for conducting future parent-teacher conversations. The scale anticipated self-efficacy in conducting future parent-teacher conversations consisted of 15 items and was sufficiently reliable ($\alpha = .76$). All items of the scale were introduced as follows: “How confident are you that you will be able to successfully handle the following challenges in future parent-teacher conversations?” E.g. “sympathetically deliver bad news to someone”. The anticipated self-efficacy of the pre-service teachers varied between types of conversations being the highest for shared decision making ($M = 2.62$, $SD = 0.30$), slightly lower for breaking bad news ($M = 2.55$, $SD = 0.46$) and the lowest for handling complaints

²² Comparable to STEM; MINT is an acronym for mathematics, informatics, natural sciences and technology.

($M = 2.39$, $SD = 0.45$). This finding confirms the assumption that shared decision making conversations are the easiest conversation type for pre-service teachers and, thus, a good starting point for beginning learners (cf. chapter 7.1).

Participation in the study was voluntary. Participants received a certificate of participation and a 30 Euro voucher for their participation. Since participation was time-intensive and extra-curricular, participating students may have been motivated above average. The incentives aimed at mitigating this bias. Before their participation, study participants received and signed a consent form that explained the purpose of the study and the utilization of their data, in particular, regarding the videotapes of the simulated conversations. Participants could withdraw their consent to use the data at any time, including after having participated in the study and having received the voucher.

8.2 Main Study

The main study took place from March to July 2012 at the TUM School of Education in Munich, Germany. Training and assessment for the participants lasted again one day and participants could choose any of 10 dates. Six classrooms were equipped with cameras and camerapersons in order to videotape the simulated conversations.

8.2.1 *Research Design*

Data collection in the main study took place analogously to data collection in the pilot study (cf. chapter 8.1.1) apart from the following three changes. Firstly, at the beginning of each day the participating pre-service teachers were randomly assigned to three different communication training programs and a control group. Pre-service teachers who participated in one of the three communication training programs conducted the simulated conversations after training; study participants that were assigned to the control group conducted the simulated conversations before they participated in the training program. Secondly, the pre-service teachers did the situational judgment test on parent-teacher conversations before they conducted the simulated conversations. The third change was that not only the pre-service teachers self-assessed their performance but that the simulated parents also rated the performance of the pre-service teachers immediately after each simulated conversation (cf. chapter 7.4). Figure 6 shows the research design of the main study.

Simulated parent (SP) one and three were female, simulated parent two and four were male.

As in the pilot study, two trained raters rated the 192 videotaped conversations independently from each other with the revised coding manual (cf. chapter 13.3). The rating process was blind, i.e. the raters did not know which type of communication training program the pre-service teachers had received or if they had been in the control group in order to avoid expectancy effects.

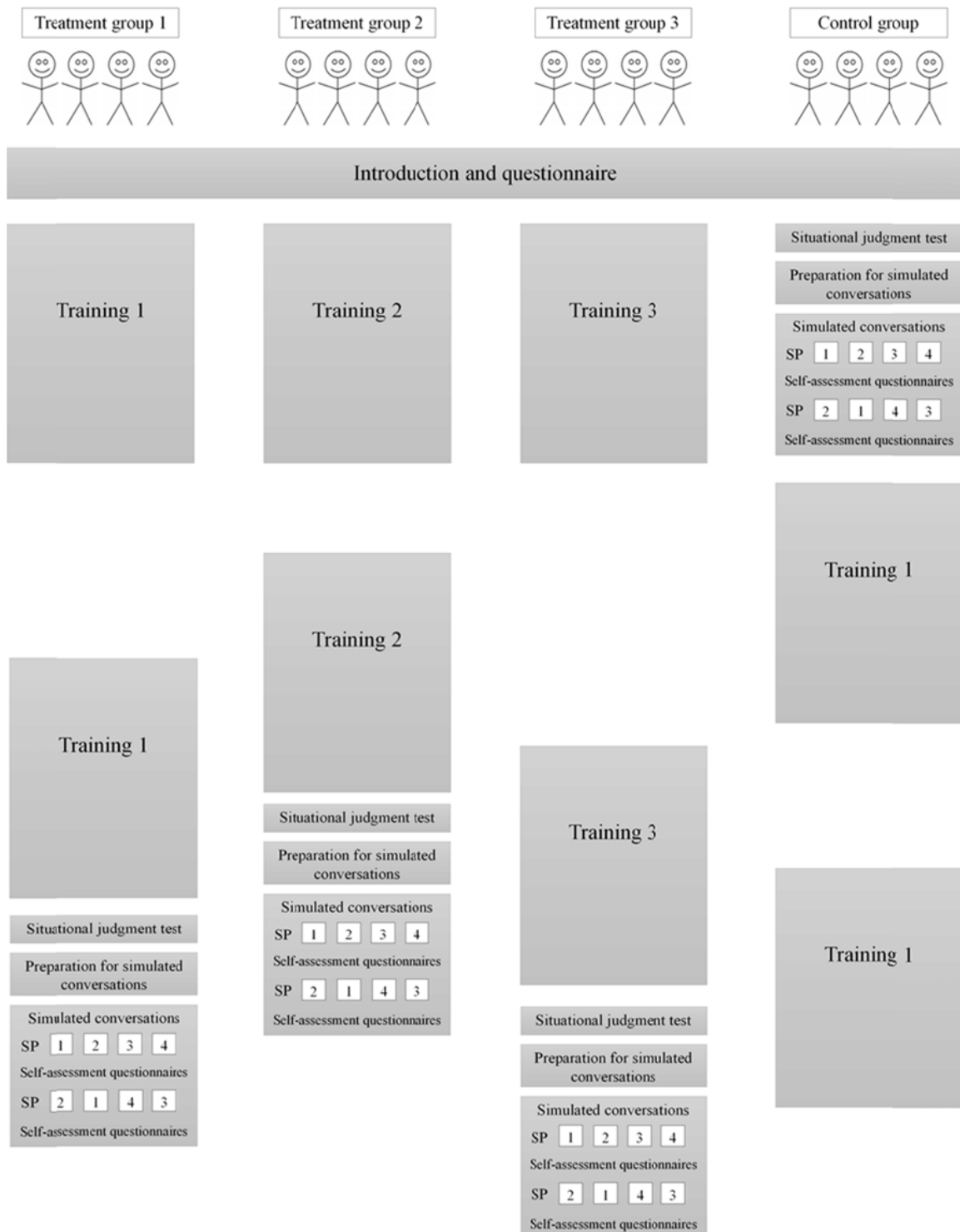


Figure 6. Research design of the main study

8.2.2 *Sample*

$N = 96$ pre-service teachers took part in the main study. 62 of them were female. The pre-service teachers were on average 22.43 years old ($SD = 3.70$) and enrolled in their 4th to 5th semester ($M = 4.62$, $SD = 2.96$) at Bavarian universities. 89 were pre-service *Gymnasium* teachers. Seven studied for other types of schools but were individually admitted due to high interest. 52.1% of the pre-service teachers studied MINT subjects, 33% social sciences and 14.9% a combination of both. All but one of the participants who studied social sciences had at least one language as a main subject. Around half of the pre-service teachers who studied a combination of MINT subjects and social sciences had one language as a main subject. All but one of the participants were born in Germany and reported German as their first language. 16.7% reported to have grown up bilingual. 54.2 % reported having no previous knowledge with regard to conducting professional conversations. 38.6% reported having little to some previous knowledge and 7.3% reported having considerable previous knowledge.

As in the pilot study, the pre-service teachers self-assessed their general communication competence as relatively high ($M = 3.06$, $SD = 0.28$) prior to the communication training program. The scale general communication competence (23 items, $\alpha = .83$) consisted of the subscales comprehensibility (6 items, $\alpha = .66$), articulateness (7 items, $\alpha = .73$), persuasiveness (3 items, $\alpha = .65$) and empathy (7 items, $\alpha = .75$). The self-concept of the pre-service teachers regarding their subjects of study was also again comparatively high ($M = 2.99$, $SD = 0.39$, 4 items, $\alpha = .52$).

As in the pilot study, participating students were possibly motivated above average since participation in the main study was again voluntary and extra-curricular. To mitigate this selection bias, participants received a certificate of participation and a 40 Euro voucher. As in the pilot study, they had to give their consent to the utilization of their data but could withdraw it anytime.

9 Data Analysis Pilot and Main Study

With the pilot study data a first evaluation of the simulated conversations was conducted, via an investigation of research questions 1-6. Moreover, the pilot study analyses provided first results regarding research question 7, the relations of the simulated conversations to other measurements and external criteria. The analyses of research questions 1-6 were replicated with data from the main study. Due to the larger sample size and additional data, the main study rendered it possible to investigate research question 7 in more depth with additional measurements and statistical procedures. If not stated otherwise, the video ratings by the independent observers form the basis for data analysis.

The scales of the coding manual, the self-assessment questionnaires and the rating scales of the simulated parents are treated as interval scales. Multilevel rating scales can be treated as interval scales if the ends of the scales are bipolar extremes of a continuum (Wirtz & Caspar, 2002), which is the case for the scales in the coding manual and the rating scales of the simulated parents. Moreover, a variety of researchers argue for a pragmatic point of view and for the usage of parametric procedures even if the equidistance of interval scale points is moderately breached as long as the results make sense content-wise (Bortz & Döring, 2006; Uebersax, 2010; Wirtz & Caspar, 2002). In case of doubt, it is recommended to complement parametric procedures with techniques for ordinally scaled variables (Bortz & Döring, 2006; Wirtz & Caspar, 2002). In line with this argumentation, the self-assessment questionnaires are also treated as interval scales and corresponding results are backed up with additional analyses for ordinally scaled variables.

Research question 1 investigates whether the independent observers achieve a satisfactory inter-rater agreement with regard to the communication competence of pre-service teachers in simulated conversations with parents (objectivity of scoring). There is little consensus about what statistical methods are best to analyze rater agreement. Uebersax (2010) argues that the choice of methods depends on the purpose of the analysis, the scaling of the data and the number of raters. For Likert-type items and interval-level data, he recommends employing the Pearson correlation coefficient for two raters and for two or more raters intraclass correlations (ICCs) (Uebersax, 2010). Intraclass correlations assess rater agreement by comparing the variability of different ratings of the same subject to the total variation across all ratings and all subjects (Uebersax, 2010). In line with the recommendation by Uebersax (2010), intraclass correlations were conducted in order to calculate the inter-

rater agreement with regard to the performance of the pre-service teachers in the simulated conversations. To enable generalization to the population of possible raters, the second class of intraclass correlations (ICC C,K) was used (Mc Graw & Wong, 1996). For interval scales absolute rater agreement is usually not required and reasonable since the underlying construct is continuous (Wirtz & Caspar, 2002). Consequently, the consistency agreement was calculated. According to existing guidelines, the cut-off point for a good inter-rater agreement was set to an ICC of .60 (Cicchetti, 1994; LeBreton & Senter, 2008). The literature on high inference ratings shows that it is difficult to achieve a high inter-rater agreement (cf. chapter 5.1.3.4). Thus, it is recommended to take into account the importance and difficulty of the measurement, the alternatives for measuring the construct at stake and further factors when judging whether a coefficient is adequate (Wirtz & Caspar, 2002).

Research question 2 aims at eliciting whether the internal structure of the data matches the structure of the theoretical construct, the Munich Model of Communication Competence in Parent-Teacher Conversation (evidence of validity). In order to investigate research question 2, confirmatory factor analyses (CFAs) were conducted. In line with the Munich Model of Communication Competence, general communication competence was modelled as a second order factor and the three competence facets, structuring the conversation, problem solving and establishing a relationship to the conversational partner, as first order factors. Robust maximum likelihood estimation was chosen because data were not normally distributed (Brown, 2006). The analyses were conducted with two data sets, one with all study participants and one without outliers (Schreiber, Nora, Stage, Barlow, & King, 2006). There were only slight differences that did not lead to substantively different conclusions. Thus, only results from the complete data set are reported. The recommended sample size for factor analyses is $N \geq 100$, the larger the sample size is, the better (Kyriakides & Charalambous, 2014). Due to the small sample size in the pilot study, factor analyses were conducted at conversation level ($N = 90$) and backed with analyses with the two conversations clustered at person level ($N = 49$). Model fit indices and factor loadings at conversation and person level were very similar and did not lead to substantively different conclusions. Thus, only results for analyses at conversation level are reported. Due to the larger sample size in the main study, confirmatory factor analyses were only calculated with the two conversations clustered at person level.

Research question 3 targets the reliability of the scales of the newly-developed coding manual. In order to investigate research question 3, two different types of reliability coeffi-

cients were calculated to determine the internal consistency of the scales (American Educational Research Association et al., 2014); firstly, Cronbachs Alpha through intraclass correlations (Mc Graw & Wong, 1996) and, secondly, composite reliability via confirmatory factor analyses (Raykov & Marcoulides, 2011). Composite reliability is a relatively new way to calculate reliability, which was introduced in the context of factor analysis (Raykov & Marcoulides, 2011). It has the advantage that it has less severe preconditions than other types of reliability calculations, like test-retest or alternate forms approaches, and is more accurate than the frequently used Cronbachs Alpha (Raykov & Marcoulides, 2011). Both reliability coefficients were calculated for the entire scale communication competence in parent-teacher conversations as well as for the three subscales, structuring the conversation, problem solving and establishing a positive relationship to the conversational partner.

Research question 4 investigates whether the different actors who portrayed parents had an effect on the self-assessed or observer-rated performance (objectivity of application). The self-assessment questionnaires respectively video ratings were divided into four (pilot study) / six (main study) groups according to the actors performing the simulated conversations. Levene tests confirmed variance homogeneity for both pilot and main study data for both overall communication competence and the three competence facets. Analyses of variances (ANOVAs) were run for communication competence and all three competence areas to check to what extent the mean self-assessed and observer-rated performance varied between groups.

Research question 5 focuses the influence of the case vignettes on the performance of the pre-service teachers and, in line with this, the generalizability of the results of simulated conversations. Thus, correlational analyses between the average results of conversation one and two were conducted. Moreover, it was investigated whether the case vignettes were equally difficult or if the different case vignettes had an effect on the self-assessed or observer-rated performance. The self-assessment questionnaires and the video ratings were each distributed into two groups according to the case vignette the simulated conversation was based on. Levene tests confirmed variance homogeneity for both pre- and main study data from the self-assessment questionnaires and the observer-ratings of the pilot study. Independent samples *t*-tests for overall communication competence and all three competence areas were conducted to evaluate if the mean self-assessed / observer-rated performance varied between groups. Variance homogeneity could not be confirmed for the main study observer-ratings. Hence, the respective analyses were repeated with the Mann-

Whitney U Test for non-parametric data. Since results were very similar and did not lead to substantively different conclusions, only the results of the *t*-tests are reported.

Research question 6 examines the response processes of the pre-service teachers (evidence of validity). The pre-service teachers rated the perceived authenticity of the simulated conversations on 4-point Likert-type items. A descriptive analysis of these items was conducted to elicit to what degree the pre-service teachers perceived the conversations as authentic and behaved as if in real situations.

Research question 7 explores the correspondence of the relations between the results of simulated conversations and other variables and measurements to theoretical expectations (evidence of validity). With the pilot study data, correlations between ratings by the independent observers, self-assessments of the pre-service teachers and external criteria were analyzed through bivariate correlational analyses. With the main study data, correlations between ratings by the independent observers, self-assessments of the pre-service teachers and, additionally, ratings by the simulated parents were examined through correlational analyses. Moreover, the correlations of these three measurements with the results in a situational judgment test on parent-teacher conversations and external criteria were calculated. The selected external criteria comprised Abitur grade, current semester, previous knowledge and self-assessed communication competence (before the simulated conversations). In a third step, the correlational pattern of the results of the independent observer ratings, the ratings by the simulated parents and the self-assessments of the pre-service teachers with the results of the six adapted medical instruments was investigated with correlational analyses (cf. chapter 7.7). In a last step, a multitrait-multimethod-matrix for the independent observer ratings, the ratings of the simulated parents and the self-assessments of the pre-service teachers was developed based on correlational analyses.

The confirmatory factor analyses were conducted with Mplus 7.1 (Muthén & Muthén, 1998-2013). All other analyses were conducted with IBM SPSS Statistics 22. There were missing data due to not readable or truncated video files or missing values in questionnaires. In the pilot study eight videos ($\leq 8\%$) and two self-assessment questionnaires ($\leq 2\%$) had to be excluded. In the main study six videos ($\leq 3.2\%$) were not readable and were excluded from data analysis. Five participants had missing values on one of the subscales of the self-assessment questionnaire ($\leq 2.6\%$). Analyses were only run for subscales that had complete values. Missing values were excluded listwise.

E Results - Pilot Study

This part presents the results of the pilot study. In the subsequent part F, the analyses of the pilot study are replicated with the main study data and results from additional analyses are outlined. According to the subdivision of the research questions, the presentation of the results of the pilot study is also subdivided into two parts: firstly, an analysis of the fundamental aspects of building scores, that is, clarifying inter-rater agreement, the factorial structure and the reliability with which the factors are measured and, secondly, in-depth analyses that are based on the fundamental analyses and inquire more deeply into important aspects of the main psychometric quality criteria. Subsequently, the pilot study results are discussed and implications for the refinement of the simulated conversations with regard to the main study are deduced.

Basic and in-depth analyses can be assigned to the research questions and psychometric quality criteria as follows. Basic analyses: The calculation of inter-rater agreement through intraclass correlations provides information about the objectivity of scoring (research question 1). A confirmatory factor analysis reveals to what degree the internal structure of the data matches the theoretical construct and provides evidence of validity (research question 2). The reliability of the coding manual is estimated with intraclass correlations, which provide information about the internal consistency of the scales and with a confirmatory factor analysis, which reveals the composite reliability of the factors (research question 3) (Raykov & Marcoulides, 2011). In-depth analyses: ANOVAs show whether the different simulated parents have an effect on the performance of the pre-service teachers and render information about the objectivity of application (research question 4). A correlational analysis shows to what degree the results in two simulated conversation are related to another and *t*-tests show whether the two case vignettes are equally difficult. The correlational analysis and the *t*-tests provide information with regard to the influence of the case vignettes and the generalizability of the results of simulated conversations (research question 5). A descriptive analysis of the perceived authenticity of the simulated conversations provides first evidence of the response processes of pre-service teachers in simulated parent-teacher conversations (research question 6). Lastly, a multimethod measurement reveals whether correlations between the observer-ratings of the performance of pre-service teachers in simulated conversations and other measurements and variables match the hypothesized correlation pattern and provides evidence of the validity of the measurement (research question 7).

10 Basic Analyses

10.1 Consistent Inter-Rater Agreement - Objectivity of Scoring

The analyses regarding the first research question investigate whether the two trained observers, who rated the $N = 90$ video-taped simulated conversations of the pilot study with the newly-developed coding manual independently from each other, achieved a satisfactory inter-rater agreement (cf. chapter 7.3 and chapter 7.6). To check the inter-rater agreement intraclass correlations were calculated after the observers had rated 50% and 100% of the videos. The coding manual consisted of 18 high inference 5-point Likert-type items. Inter-rater agreement across all 18 items was $ICC = .81$ after 50% and $ICC = .83$ after 100% of the video analysis. Inter-rater agreement was higher than the set-cut off point ($ICC = .60$) for all items. Table 7 shows the inter-rater agreement for all individual items after 50% and 100% of the video analysis.

Table 7
Inter-rater agreement pilot study

Item	ICC 50%	ICC 100%
Global rating items		
Structuring the conversation	.87	.90
Problem solving	.89	.88
Establishing a relationship to the conversational partner	.81	.81
Structuring the conversation		
Detectability of the SDM conversational phases	.87	.90
Adequacy of the length of the SDM conversational phases	.73	.79
Correctness of the order of the SDM conversational phases	.63	.79
Use of metacommunication	.66	.77
Proportion of participation of both conversational partners	.65	.68
Problem solving		
Successful establishment of common ground	.65	.69
Comprehensibility of the presented options	.85	.87
Quality of cooperation in the negotiation process	.80	.80
Coming to a concrete agreement	.84	.84
Subject-specific performance	.89	.90
Establishing a positive relationship with the conversational partner		
Unconditional positive regard	.71	.74
Authenticity	.89	.82
Empathy	.87	.89
Conversational climate	.89	.89
Nonverbal behavior	.85	.82

Note. SDM = shared decision making; ICC = intraclass correlation.

10.2 Fit of the Data to the Theoretical Construct - Internal Structure

Regarding research question 2, it was examined whether the data match the underlying theoretical construct, the Munich Model of Communication Competence in Parent-Teacher Conversation. The model assumes that communication competence is three-dimensional and comprises the competence facets, structuring the conversation, problem solving and establishing a relationship with the conversational partner (cf. chapter 3.3). Accordingly, the coding manual for rating the video-taped simulated conversations is subdivided into these three competence areas. First and second order confirmatory factor analyses based on the independent observer ratings with the coding manual were conducted to check whether the data fit the theoretical construct. Figure 7 shows the underlying measurement model for the first and second order confirmatory factor analyses. The three global rating items were not part of the analyses since they are located at a higher level of abstraction than the other items.

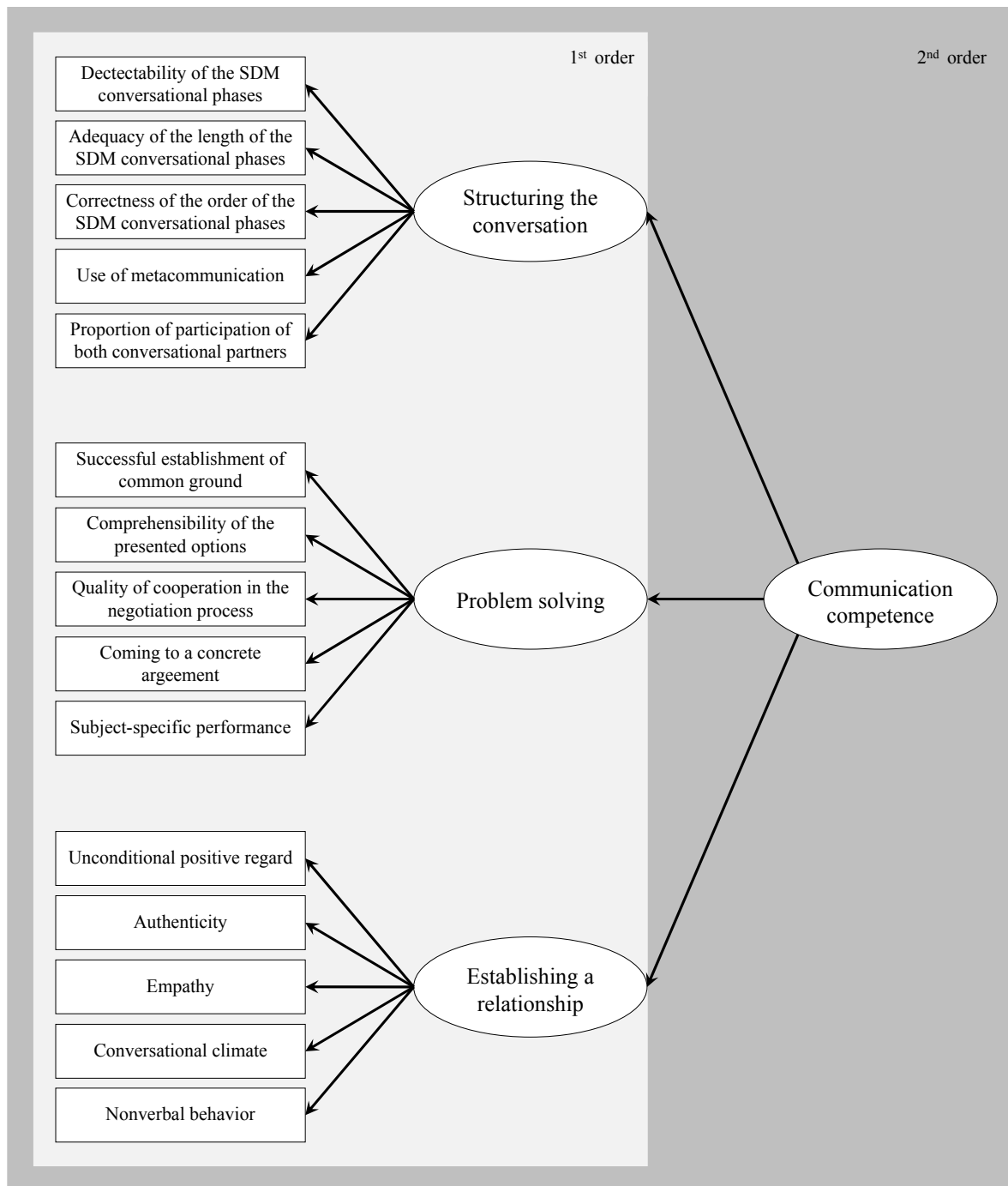


Figure 7. Measurement model for the confirmatory factor analyses of the pilot study

A first order confirmatory factor analysis with the three factors, structuring the conversation, problem solving and establishing a relationship, produced only a moderate model fit ($\chi^2(87) = 177.57, p \leq 01, RMSEA = .11, CFI = .93, SRMR = .05$). In order to improve the model fit, two items were excluded from the further analyses based on theoretical considerations and modification indices: item 1.5 “adequacy of the proportion of participation of both conversational partners” and item 2.5 “subject-specific performance of the teacher”.

Item 1.5 was the only item of the competence facet structuring the conversation that did not directly refer to the shared decision making script which was the template for the learning goals for this competence facet. Item 2.5 was removed since it emphasizes subject-specific competence more than communication competence.

A first order confirmatory factor analysis based on the remaining 13 items indicates a three-dimensional solution. A model consisting of the three competence facets, structuring the conversation, problem solving and establishing a relationship to the conversational partner, has an acceptable model fit (cf. Table 8). Since the three factors of the model correlate comparatively highly with each other, in particular the factors problem solving and establishing a relationship, the three-dimensional model was compared to a one-dimensional model with a single factor communication competence. The one-dimensional model did not fit the data adequately (cf. Table 8). This finding and the theoretical foundation of the model further support the three-dimensional solution.

Table 8

Fit indices of the one- and three-dimensional model of communication competence - pilot study

Model	χ^2	Df	RMSEA	CFI	SRMR
1-dimensional	234.37*	64	.17	.84	.06
3-dimensional	97.60*	61	.08	.97	.05

Note. χ^2 = chi-square test; df = degrees of freedom; RMSEA = root mean squared error of approximation; CFI = comparative fit index; SRMR = standardized root mean squared residual; * $p < .05$.

Based on the results of the first order confirmatory factor analyses, a second order confirmatory factor analysis was conducted with the selected 13 items. It showed that a hierarchical model with a second order factor, general communication competence, and three first order factors, structuring the conversation, problem solving and establishing a relationship to the conversational partner, has an acceptable model fit ($\chi^2(61) = 96.24, p \leq .01$, RMSEA = .08, CFI = .97, SRMR = .05). The slight difference in the model fit indices of the theoretically equivalent first and second order model is due to the fact that the factor

problem solving had a negative residual variance in the second order model and, thus, this variance was restricted to greater than zero. The negative residual variance might be due to the high correlation between the factors problem solving and establishing a relationship. The model fit indices with exception of the significant chi-square indicate that the data fit the theoretical model. Brown (2006) recommends to always regard the chi-square in combination with other fit indices since it is often criticized with regard to the fact that a variety of influence factors can compromise the statistical significance test of the model. Figure 8 provides standardized parameter estimates and residuals for the second order model of communication competence.

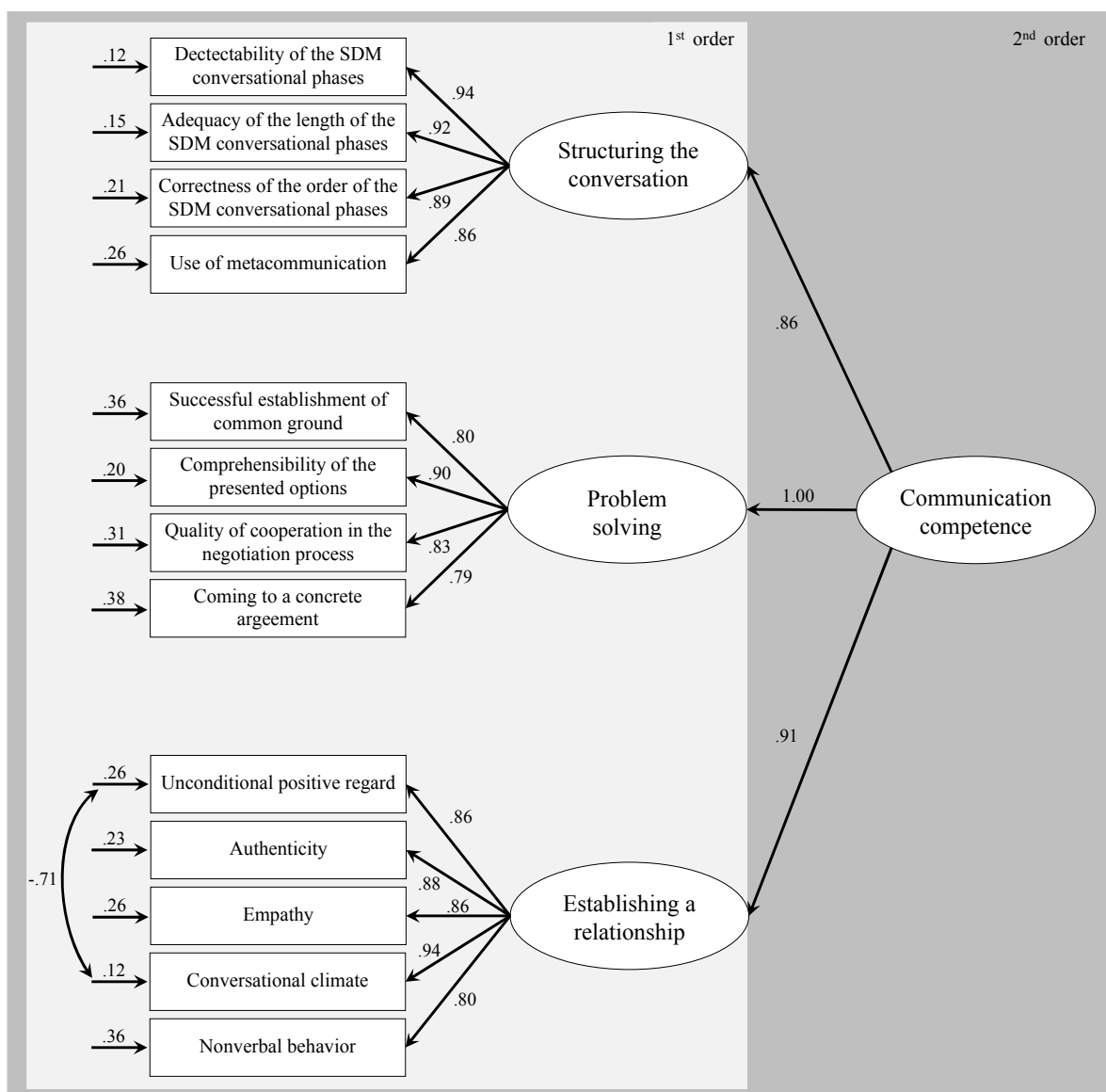


Figure 8. Results of the confirmatory factor analysis for the second order model of communication competence

The residual variances of the items unconditional positive regard and conversational climate correlate. This is theoretically plausible since unconditional positive regard is likely to be a precondition for a good conversational climate.

10.3 Reliability of the Coding Manual - Composite Reliability / Internal Consistency

The analyses in this subchapter provide results regarding research question 3. They target the reliability of the scales of the newly-developed coding manual. Two types of reliability were calculated for the coding manual: composite reliability via confirmatory factor analyses (Raykov & Marcoulides, 2011) and internal consistency in Crobach's alpha with intraclass correlations. Reliability coefficients were calculated for the scale communication competence and the three subscales, structuring the conversation, problem solving and establishing a relationship. Data basis were the observer ratings of the $N = 90$ video-taped conversations.

10.3.1 Composite Reliability

The composite reliabilities for the second order model of communication competence presented in chapter 10.2 were $\rho = .95$ for structuring the communication, $\rho = .90$ for problem solving, $\rho = .94$ for establishing a relationship and $\rho = .93$ for communication competence. These results indicate a reliable measurement of the factors.

10.3.2 Internal Consistency

The scale communication competence in parent-teacher conversations, consisting of all original 18 items ($\alpha = .98$), and the three subscales structuring (6 items, $\alpha = .96$), problem solving (6 items, $\alpha = .94$) and establishing a relationship (6 items, $\alpha = .95$) were internally consistent. All items had a positive and high item-scale correlation ($r \leq .61$). The pre-service teachers achieved a mean value of 2.23 ($SD = 0.71$), which indicates a good to satisfactory level of communication competence in conversations with parents. The mean communication competence of the pre-service teachers ranged from 1.14 to 4.33.

The in-depth analyses conducted in the next chapters are based on the scales built according to the results of the confirmatory factor analyses (cf. chapter 10.2). Thus, reliability coefficients for those scales are also depicted. The global ratings were not part of the confirmatory analyses and two items were excluded due to theoretical considerations and increased model fit. The scales built according to the confirmatory factor analysis were also internally consistent: communication competence in parent-teacher conversations (13 items, $\alpha = .96$), structuring the conversation (4 items, $\alpha = .94$), problem solving (4 items, $\alpha = .90$) and establishing a relationship (5 items, $\alpha = .93$). All items had a positive and high item-scale correlation ($r \leq .75$). Table 9 gives an overview of the characteristics of all items of the coding manual and compares the item-scale correlations of all items of the coding manual and the item-scale correlation of the 13 items selected by the confirmatory factor analyses.

Table 9

Item and scale characteristics for the pilot study coding manual

Item	<i>M</i>	<i>SD</i>	r_{it}	$r_{it(CFA)}$
Structuring the conversation				
Global rating	2.27	0.88	.93	-
Detectability of the SDM conversational phases	2.11	0.91	.92	.78
Adequacy of the length of the SDM conversational phases	2.26	0.76	.91	.84
Correctness of the order of the SDM conversational phases	1.90	0.71	.86	.75
Use of metacommunication	2.14	0.81	.84	.84
Proportion of participation of both conversational partners	2.33	0.68	.80	-
Problem solving				
Global rating	2.27	0.92	.90	-
Successful establishment of common ground	2.26	0.69	.65	.76
Comprehensibility of the presented options	2.19	0.92	.82	.85
Quality of cooperation in the negotiation process	2.27	0.87	.66	.81
Coming to a concrete agreement	1.99	0.80	.61	.75
Subject-specific performance	2.37	1.14	.76	-
Establishing a relationship				
Global rating	2.46	0.85	.95	-
Unconditional positive regard	2.16	0.74	.81	.78
Authenticity	2.52	0.95	.87	.82
Empathy	2.43	0.97	.85	.84
Conversational climate	2.42	0.99	.87	.88
Nonverbal behavior	2.38	0.85	.82	.70

Note. SDM = shared decision making; *M* = sample mean; *SD* = standard deviation; r_{it} = item-scale correlation for all items of the coding manual; $r_{it(CFA)}$ = item-scale correlation for CFA items; scaling: 4-point Likert scales, higher values indicating better performance.

11 In-Depth Analyses on Objectivity, Reliability and Validity

Based on the results of the basic analyses, in the following sections the in-depth analyses on the objectivity, reliability and validity of simulated conversations are conducted. The in-depth analyses are based on the scales elicited through the confirmatory factor analyses. The scale communication competence is calculated as a mean of the results in the three competence areas, structuring the conversation, problem solving and establishing a relationship with the three competence facets weighed coequally due to the non-compensatory approach (cf. chapter 7.3).

11.1 Consistent Performance of the Actors in the Simulated Conversations - Objectivity of Application

Due to logistic reasons, four different actors portrayed parents in the pilot study. Two of them were simulated mothers, two simulated fathers. The participating $N = 49$ pre-service teachers were randomly assigned to the actors and conducted one conversation with a simulated mother and one with a simulated father each. In order to scrutinize the objectivity of application and verify if all pre-service teachers received the same conditions, it was investigated whether the actors performed their simulated parents' roles consistently. In line with this, the analyses in 11.1.1 and 11.1.2 aim at providing evidence of research question 4 which examines whether the different simulated parents have an effect on the self-assessed and / or observer-rated performance of teachers in conversations with simulated parents.

11.1.1 *Effect of the Actor on the Self-Assessed Performance*

The $N = 96$ self-assessment questionnaires were divided into four groups according to the actor performing the conversation. One-way ANOVAs showed no statistically significant effects of the conversational partner on the self-assessed competence to communicate with parents, $F(3,92) = 1.18$, $MSE = .39$, $p = .32$, $\eta^2 = .03$. There were also no statistically significant effects in the sub competence facets, structuring the conversation, $F(3,92) = .11$, $MSE = .16$, $p = .95$, $\eta^2 = .00$, problem solving $F(3,92) = .23$, $MSE = .30$, $p = .88$, $\eta^2 = .01$ and establishing a relationship with the conversational partner, $F(3,91) = .41$, $MSE = .42$, $p = .59$, $\eta^2 = .03$. Table 10 gives an overview of per-actor sample sizes, means and standard deviations for the self-assessed performance.

Table 10

Per-actor sample sizes, means and standard deviations for the self-assessed performance of the pre-service teachers in the simulated conversations of the pilot study

	<i>n</i>	<i>M</i>	<i>SD</i>
Structuring the conversation			
Actor 1	26	3.00	0.62
Actor 2	26	2.94	0.46
Actor 3	22	3.02	0.61
Actor 4	22	2.96	0.49
Problem solving			
Actor 1	26	3.15	0.46
Actor 2	26	3.15	0.46
Actor 3	22	3.22	0.31
Actor 4	22	3.22	0.31
Establishing a relationship			
Actor 1	25	3.12	0.73
Actor 2	26	3.27	0.72
Actor 3	22	3.23	0.61
Actor 4	22	3.32	0.48

Note. *n* = number of cases in subsample; *M* = sample mean; *SD* = standard deviation; scaling: 4-point Likert scales with higher values indicating better performance.

11.1.2 *Effect of the Actor on the Observer-Rated Performance*

The video-taped conversations were divided into four groups according to the actor performing the conversation. One-way ANOVAs showed no statistically significant effects of the conversational partner on the observer-rated communication competence in parent-teacher conversations, $F(3,86) = .53$, $MSE = 4.47$, $p = .66$, $\eta^2 = .02$. There were also no statistically significant effects on the three sub competence facets, structuring the conversation, $F(3,86) = .27$, $MSE = .56$, $p = .85$, $\eta^2 = .01$, problem solving, $F(3,86) = 1.11$, $MSE = .51$, $p = .35$, $\eta^2 = .04$ and establishing a relationship with the conversational partner, $F(3,86) = .50$, $MSE = .66$, $p = .68$, $\eta^2 = .02$. Table 11 gives an overview of per-actor sample sizes, means and standard deviations for the observer-rated performance.

Table 11

Per-actor sample sizes, means and standard deviations for the observer-rated performance of the pre-service teachers in the simulated conversations of the pilot study

	<i>n</i>	<i>M</i>	<i>SD</i>
Structuring the conversation			
Actor 1	25	2.07	0.77
Actor 2	25	2.22	0.78
Actor 3	21	2.07	0.71
Actor 4	19	2.03	0.71
Problem solving			
Actor 1	25	2.32	0.71
Actor 2	25	2.28	0.71
Actor 3	21	1.97	0.74
Actor 4	19	2.11	0.71
Establishing a relationship			
Actor 1	25	2.42	0.85
Actor 2	25	2.50	0.80
Actor 3	21	2.21	0.71
Actor 4	19	2.37	0.88

Note. *n* = number of cases in subsample; *M* = sample mean; *SD* = standard deviation; scaling: 5-point Likert scales with lower values indicating better performance.

11.2 Effect of the Case Vignettes - Generalizability

The two simulated conversations each pre-service teacher conducted were based on two different case vignettes. Both conversations were shared decision making conversations regarding possibilities to improve low school achievements (cf. chapter 7.1). In order to answer research question 5 and find out if the two different case vignettes affect the performance of the pre-service teachers in the simulated conversations and, consequently, how generalizable the results in one simulated conversation are to the performance in another, the correlation of the results of the two conversations with each other was calculated and it was investigated whether the two case vignettes were equally difficult for the pre-service teachers.

11.2.1 *Correlation of the Results of the two Conversations With Each Other*

For each video a global value for communication competence in parent-teacher conversation was calculated. A correlational analysis revealed that the results of the two conversations were significantly positively correlated with each other ($r = .48$, $p \leq .01$). The strength of the correlation was medium to high (Cohen, 1988).

11.2.2 *Effect of the Case Vignette on the Self-Assessed Performance*

The $N = 96$ questionnaires in which the pre-service teachers had self-assessed their performance were divided according to the two case vignettes. Independent samples t -tests showed that there was no statistically significant effect of case vignette one ($M = 3.08$, $SD = 0.61$) and case vignette two ($M = 3.04$, $SD = 0.66$) on the self-assessed performance of communication competence in parent-teacher conversations, $t(94) = .32$, $p < .75$, $d = .06$, 95% CI [- .22, .3]. There were also no statistically significant effects of case vignette on the sub competence facets, structuring the conversation, $t(94) = -.35$, $p < .73$, $d = .04$, 95% CI [- .32, .23], problem solving, $t(94) = .25$, $p < .81$, $d = .02$, 95% CI [- .23, .29] and establishing a relationship to the conversational partner, $t(94) = .87$, $p < .39$, $d = .08$, 95% CI [- .18, .46].

Table 12 gives an overview of per-case vignette sample sizes, means and standard deviations for the self-assessed competence in the three competence facets, structuring the conversation, problem solving and establishing a relationship.

Table 12

Per-case vignette sample sizes, means and standard deviations for the self-assessed performance of the pre-service teachers in the simulated conversations of the pilot study

	<i>n</i>	<i>M</i>	<i>SD</i>
Structuring the conversation			
Case vignette 1	49	2.90	0.63
Case vignette 2	47	2.95	0.72
Problem solving			
Case vignette 1	49	3.13	0.64
Case vignette 2	47	3.10	0.63
Establishing a relationship			
Case vignette 1	49	3.22	0.77
Case vignette 2	47	3.10	0.80

Note. *n* = number of cases in subsample; *M* = sample mean; *SD* = standard deviation; scaling: 4-point Likert scales with higher values indicating better performance.

11.2.3 *Effect of the Case Vignette on the Observer-Rated Performance*

The $N = 90$ observer-ratings of the performance of the pre-service teachers in the simulated conversations were divided into two groups according to the two case vignettes. Independent samples *t*-tests showed that there was no statistically significant effect of case one ($M = 2.26$, $SD = 0.69$) and case two ($M = 2.18$, $SD = 0.71$) on the observer-rated performance of communication competence in simulated parent-teacher conversations, $t(88) = .63$, $p < .60$, $d = .12$, 95% CI [- .22, .37]. There were also no statistically significant effects of case vignette on the sub competence facets, structuring the conversation, $t(88) = .42$, $p < .68$, $d = .10$, 95% CI [- .24, .38], problem solving, $t(88) = .29$, $p < .77$, $d = .06$, 95% CI [- .26, .35] and establishing a relationship to the conversational partner, $t(88) = .73$, $p < .47$, $d = .16$, 95% CI [- .21, .46]. Table 13 gives an overview of per-case

vignette sample sizes, means and standard deviations for the observer-rated performance in the three competence facets, structuring the conversation, problem solving and establishing a relationship.

Table 13

Per-case vignette sample sizes, means and standard deviations for the observer-rated performance of the pre-service teachers in the simulated conversations of the pilot study

	<i>n</i>	<i>M</i>	<i>SD</i>
Structuring the conversation			
Case vignette 1	44	2.14	0.75
Case vignette 2	46	2.07	0.74
Problem solving			
Case vignette 1	44	2.20	0.71
Case vignette 2	46	2.16	0.74
Establishing a relationship			
Case vignette 1	44	2.45	0.83
Case vignette 2	46	2.32	0.79

Note. *n* = number of cases in subsample; *M* = sample mean; *SD* = standard deviation; scaling: 5-point Likert scales with lower values indicating better performance.

11.3 Perceived Authenticity of the Simulated Conversations - Response Processes

The analyses in 11.3 provide evidence of research question 6, which focuses on the degree to which pre-service teachers perceive simulated conversations as authentic. The pre-service teachers assessed the authenticity of the simulated conversations on a 4-point Likert-type item, “I felt like I was conducting a real parent-teacher conversation”, with higher values indicating a higher degree of authenticity. They perceived the simulated conversations as highly authentic ($M = 3.24$, $SD = 0.74$). This corresponds to the impressions of the simulated parents. In a debriefing they reported that they perceived the conversations as very authentic, too.

11.4 Multimethod Measurement - Relations of the Results of Simulated Conversations to Other Variables

The analyses in 11.4 concern research question 7 and investigate whether the relations between the results of simulated conversations and other measurements and criteria correspond to theoretical expectations (cf. part C) in order to provide evidence of the validity of the results of simulated conversations.

Table 14 gives on the one hand, an overview of the correlations between two different measurements at the conversational level: the observer ratings of the performance of the pre-service teachers in the simulated conversations and the pre-service teachers’ self-assessments of their performance (1st level of the multimethod measurement). On the other hand, it also displays results for correlations between the observer-ratings of the simulated conversations and external criteria that might be related to the construct but go beyond the actual conversational level (2nd level of the multimethod measurement). Correlational analyses were conducted at the person level (mean across both conversations and all three competence facets) for both observer ratings of the performance in simulated conversations and self-assessments of the pre-service teachers.

Table 14

Correlations of the observer ratings of the performance in simulated conversations with self-assessments of the pre-service teachers and external criteria

	Convergent Evidence of Validity	Criterion-Based Evidence of Validity		
	Pre-service teachers	Abitur grade	Previous knowledge	Semester
Independent observers	.22	.31*	.27	.23

Note. * $p = \leq .05$.

As expected, the correlation between independent observer ratings of the performance of the pre-service teachers in simulated conversations and pre-service teachers' self-assessments of their performance in is positive but low. Also corresponding to expectations, there are positive, low to medium correlations between the independent observer ratings and external criteria. However, only the correlation with the Abitur grade is significant.

12 Summary and Discussion of the Pilot Study Results

The basic analyses provided results regarding the fundamental aspects of building scores: the inter-rater agreement, the internal structure of the data and the reliability of the scales of the coding manual. Analyses regarding research question 1, showed that the inter-rater agreement across all 18 items of the coding manual and all 90 video-taped simulated conversations of the pilot study was high. Inter-rater agreement was higher than the set-cut off point ($ICC = .60$) for all 18 items of the coding manual. These results indicate that objectivity of scoring was high and that the rater training prepared the observers well for the rating of the videos. This result is particularly promising since the ratings were high inference and it tends to be challenging to achieve high inter-rater agreement with high inference rating scales (cf. chapter 5.1.3.4).

With regard to research question 2, confirmatory factor analyses revealed that a hierarchical model with a second order factor, general communication competence and three first order factors, structuring the conversation, problem solving and establishing a relationship to the conversational partner, fits the data adequately. The fit between the internal structure of the data and the Munich Model of Communication Competence in Parent-Teacher Conversation provides first evidence for the validity of the results of simulated conversations and supports the theoretical foundation of the development of the instruments. The finding is in line with the results of Hertel (2009) who considers a three-dimensional structure of the competence to communicate with parents as most likely. However, the sample size is relatively small for confirmatory factor analyses (Kyriakides & Charalambous, 2014). In order to get more reliable results, the confirmatory factor analyses need to be replicated with the larger sample of the main study.

Concerning research question 3, analyses showed that the composite reliabilities (Raykov & Marcoulides, 2011) of the presented hierarchical model were high. Moreover, the scales built based on the results of the confirmatory factor analysis were internally consistent. These findings indicate that the scales of the coding manual are sufficiently reliable.

Based on the results of the basic analyses, the in-depth analyses inquired more deeply into important aspects of the main psychometric quality criteria, such as objectivity of application, generalizability of the results of simulated conversations, authenticity of the response processes in simulated conversations and relations of the results of simulated conversations

to other variables. Analyses targeting research question 4, showed that there were no statistically significant effects of the actors on the self-assessed and observer-rated performance of pre-service teachers in simulated conversations with parents. This indicates that the actor training was successful and the actors performed their roles as parents consistently and, in this way, provided comparable conditions for all study participants and ensured objectivity of application.

Analyses concerning research question 5, showed that there is a medium to high correlation between the results of the pre-service teachers in simulated conversation one and two and no statistically significant effect of case vignette on the self-assessed and observer-rated performance. These results are indicators for the generalizability of the results of simulated conversations.

With respect to research question 6, analyses showed that the pre-service teachers perceived the simulated conversations as highly authentic. This suggests that simulated conversations provide authentic measurement conditions in which the pre-service teachers react and respond as they would in real situations. The authenticity of the response processes provides further evidence of the validity of the measurement.

Lastly, as expected, regarding research question 7, the results of simulated conversations were low to medium correlated to external criteria, such as the pre-service teachers' Abitur grade, and there was also a low correlation with the self-assessments of the pre-service teachers. The correspondence of the relations between the independent observer ratings and other variables to theoretical expectations is another indicator of the validity of the measurement. However, further measurement methods have to be included in the main study to provide sound results.

The results of the pilot study are promising regarding the diagnosis of pre-service teachers' competence to communicate with parents via simulated conversations. The objectivity of application and scoring of the simulated conversations were high, the results were reliable and there is first evidence of the validity of the interpretation of the results. Yet, the results of the pilot study also provide important information for the refinement of the simulated conversations with regard to the main study as depicted in chapter 13.

13 Implications for the Refinement of the Instrument

In the following it is described how the case vignettes, the actor training, the coding manual and the rater training were modified based on the results of the pilot study and the feedback of the simulated parents, the raters and the pre-service teachers elicited in debriefings after the pilot study.

13.1 Case Vignettes

The results of the pilot study show that the pre-service teachers perceived the content of the case vignettes as highly authentic (cf. chapter 12.3). Moreover, the two case vignettes were equally difficult (cf. chapter 12.2) and the level of difficulty of the case vignettes was adequate for beginning learners. The pre-service teachers achieved on average a good to medium value regarding their communication competence. These results show that the content and format of the case vignettes were adequate and, thus, they were only slightly refined as explained in the following.

The results of the two conversations correlated moderately - highly with each other what speaks for the generalizability of the results of simulated conversations. However, the two case vignettes in the pilot study were very similar. Both concerned the same topic, the tackling of low school achievements. In order to inquire more deeply to what degree the results of simulated parent-teacher conversations are generalizable, one of the case vignettes was substituted with a case vignette that described another frequent shared decision making conversation, the choice of the scientific or linguistic branch, for the main study (cf. chapter 7.1).

In a debriefing session the simulated parents stated that the case vignettes were relatively brief. They asked for more background information and details with regard to the main study. They reported that the discussion during the actor training had been extremely helpful to clarify the roles and determine the details and asked to include the additional information from the actor training in written form in the case vignettes. Consequently, some additional background information, details and determined triggers were additionally included in the case vignettes. As an example a final version of one of the case vignettes of the main study is included in the appendix (chapter 23.1).

13.2 Actor Training

The results of the pilot study show that the four different actors had no effect on the observer-rated and self-assessed performance of the pre-service teachers in the simulated conversations. This indicates that the actors performed their roles sufficiently consistent and that the actor training was successful. In the debriefing session the actors emphasized that the actor training per se and especially the trial runs in which they could observe their colleagues portraying the same characters were very helpful and important for them.

In a debriefing session with the raters who had coded the videos the raters commented that they had the general impression that the actors portrayed their roles as simulated parents consistently. However, they remarked that sometimes it was hard for them to rate certain items when the simulated parents cued the pre-service teachers or behaved differently from each other. Since cueing and inconsistent portrayal of triggers are frequent problems when employing simulated conversations (cf. chapter 5), the actor training was intensified and two additional components were added in order to avoid these problems. Firstly, an explanation of the phenomenon cueing and its consequences was integrated into the actor training. Typical tasks and points in the conversation where cueing occurred and should be avoided were highlighted. Secondly, important points of the conversations and expected behavior and triggers of the simulated parents were not only explained as in the pilot study but also illustrated via video sequences from the pilot study. Divergent and consistent behavior of the simulated parents in the video examples was discussed in group and the actors agreed on guidelines for behavior in future conversations. The final version of the actor training including the changes for the main study is enclosed in the appendix (cf. chapter 23.2).

13.3 Coding Manual

The results of the pilot study show that the scales of the coding manual are internally consistent and that the structure of the data matches the theoretical construct, the Munich Model of Communication Competence in Parent-Teacher Conversation. However, the results of the confirmatory factor analyses suggest some changes to the coding manual for assessing the competence of teachers to communicate with parents. For the competence facet structuring the conversation the item “adequacy of the proportion of participation of both conversational partners in the conversation” was omitted since the item did not direct-

ly address the shared decision making process as all other items of this competence facet did and the removal resulted in an improved model fit. For the competence area problem solving the item “subject-specific performance of the teacher” was modified content-wise since the original item focused a great deal more on the subject-specific performance of the pre-service teachers than on their communication competence and a removal of the item resulted in an increased model fit. The new modified item “fit of the solution to the problem and the conversational partner” also partly entails subject-specific performance. It aims at singling out whether the pre-service teacher has found a solution that fits the problem and makes sense from a pedagogical point of view. Additionally, in order to take greater account of communication processes, the item also evaluates if the solution fits the needs of the conversational partner.

The confirmatory factor analyses also showed that the competence facets problem solving and establishing a relationship were strongly correlated with each other. This is theoretically plausible since one has to collaborate well with the conversational partner in order to solve a problem. However, in order to demark the two competence facets more clearly from each other, one item that originally belonged to the competence facet problem solving “quality of cooperation in the negotiation process” and which had a high cross loading with the competence facet establishing a relationship was resorted and assigned to the competence facet establishing a relationship instead. This was done since it is likely that while the quality of cooperation in the negotiation process is decisive for the problem solving process, the cooperativeness of pre-service teachers is probably part of their competence to establish a relationship with the conversational partner.

Two further changes were made to all items of the coding manual. Firstly, the descriptions of the interstages of the items were omitted. This is recommended by Wirtz and Caspar (2002) in order to facilitate interval-scaling. Since the raters were able to handle the coding manual in the pilot study well, it is likely that they will also be able to deal with a slightly more challenging version of the coding manual with fewer explanations of the interstages in the main study. Secondly, in order to make it possible to apply the coding manual to all types of expert-layman conversations, the terminology of the items was formulated in a neutral way in order to be applicable across domains. E.g. instead of parent the more neutral term conversational partner was utilized.²³

²³ In the context of the project *ProfKom* (cf. chapter 8) the coding manual was applied to rate conversations between pre-service teachers and simulated parents as well as between pre-service physicians and simulated patients.

A lot of participants in the pilot study expressed the wish to get feedback on their performance in the simulated conversations in order to identify starting points for further learning. Since it would be supposedly difficult for pre-service teachers to interpret their results on high inference rating scales, additional behavior anchored items were included in the coding manual as a basis for feedback. These behavior anchored items target small observable units of behavior, e.g. whether the pre-service teachers greet the parents. The observers rate them process-accompanying. The behavior anchored items are straightforward and, thus, easier to understand for the pre-service teachers and probably better suited for feedback than high inference items. Moreover, a balanced approach with high inference ratings and behavior anchored items combines the merits of high inference ratings, such as a higher validity, and of low inference ratings, such as an easier manageability (cf. chapter 5.1.3.4). However, since it is not merely the presence or absence of a certain behavior but the quality of it that is informative and helpful for the pre-service teachers, the behavior anchored items are also rated on 5-point Likert scales and are, therefore, in the strict sense, no low inference items. The coding manual for the main study is enclosed in the appendix (cf. chapter 23.4).

13.4 Rater Training

The analysis of the data of the pilot study showed that inter-rater agreement was high. This indicates that the rater training had worked well to prepare the raters for handling the coding manual and coding the data in the pilot study. Thus, no changes were made to the rater training. However, since the coding rubric for the main study comprised more items than in the pilot study more time was allowed for all training components. After the trial run, inter-rater agreement for the two experts and the two raters was $ICC = .84$. As in the pilot study, items that were below the cut-off point of $ICC = .60$ were discussed in detail. Both raters were equally strict and located in between the two experts: expert 1 ($M = 2.61, SD = 1.37$), expert 2 ($M = 2.12, SD = 1.20$), rater 1 ($M = 2.43, SD = 1.34$), rater 2 ($M = 2.43, SD = 1.34$). The frequency analyses displayed in Table 15 showed that both raters and both experts utilized the entire scale.

Table 15

Frequencies of scores of the four raters in the main study trial run in percent

	Expert 1	Expert 2	Rater 1	Rater 2
Very good	0.3	5.2	7.2	9.0
Good	29.0	24.1	39.7	26.6
Satisfactory	35.2	32.1	25.5	25.5
Adequate	18.6	24.8	10.0	16.9
Fail	16.9	13.8	17.6	22.1

F Results - Main Study

Part F presents the results of the main study. Analogously to the results of the pilot study, the main study results are subdivided into basic analyses of the fundamental aspects of building scores and in-depth analyses that are based on the fundamental analyses and inquire more deeply into important aspects of the main psychometric quality criteria. All analyses from the pilot study are replicated with the main study data in order to consolidate the evidence of the research questions. The basic analyses replicate findings for research question 1-3. With regard to research question 1, the evaluation of inter-rater agreement is repeated with the main study data in order to provide further information about the objectivity of scoring. Concerning research question 2, the fit between the internal structure of the data and the theoretical construct, the Munich Model of Communication Competence in Parent-Teacher Conversation is reexamined in order to solidify evidence of validity. With respect to research question 3, the reliability of the measurement, the internal consistency of the scales of the coding manuals and the composite reliability of the factors elicited by a confirmatory factor analysis are again evaluated with the main study sample (Raykov & Marcoulides, 2011).

The in-depth analyses are aimed at supporting the findings of the pilot study regarding the research questions 4-7. Regarding research question 4 and 5, it is investigated again whether the different simulated parents consistently perform the simulated conversations, if the different case vignettes have an effect on the self-assessed performance and how generalizable the results of the simulated conversations based on a certain case vignette are. These analyses will particularly broaden the evidence of the pilot study since one of the two very similar case vignettes of the pilot study was substituted by a slightly more distinct case vignette for the main study, in order to inquire more deeply into the implementation and generalizability of simulated conversations. This change will also broaden the evidence of research question 6, which targets the perceived authenticity and response processes in simulated conversations. Lastly, with regard to research question 7, further measurements are included into the multimethod measurement and additional analyses, like a multitrait-multimethod matrix, are conducted in order to expand the evidence of validity compiled via the investigation of the relations to other variables in the pilot study.

14 Basic Analyses

14.1 Consistent Inter-Rater Agreement - Objectivity of Scoring

The analyses in 14.1 are aimed at providing further evidence of research question 1 and investigate to what degree the two trained observers achieved inter-rater agreement with respect to the performance of the pre-service teachers in the $N = 186$ video-taped simulated conversations of the main study (cf. chapter 7.6, 13.3). The revised version of the coding manual consisted of 14 high inference and 43 behavior anchored 5-point Likert-type items.

Inter-rater agreement across all items was $ICC = .72$ after 50% and $ICC = .74$ after 100% of the video analysis. However, not all items / subscales reached the set-cut off point ($ICC = .60$). Table 16 shows the inter-rater agreement for the scales consisting of the high inference items and the behavior anchored items after 100% of the video analysis.

Table 16

Inter-rater agreement for the subscales of the main study coding manual

Subscale	ICC high inference	ICC behavior anchored
Structuring the conversation	.56	.74
Problem solving	.55	.55
Establishing a relationship	.41	.77

Note. SDM = shared decision making; ICC = intraclass correlation.

14.2 Fit of the Data to the Theoretical Construct - Internal Structure

In order to undermine the findings of the pilot study regarding research question 2, the fit of the data to the underlying theoretical construct, the Munich Model of Communication Competence in Parent-Teacher Conversation, was again investigated with factor analyses. The pilot study had shown an adequate model fit for a second order model of communication competence with the three competence facets, structuring the conversation, problem solving and establishing a relationship with the conversational partner, as first order factors. In addition to the high inference items already evaluated in the pilot study, the main study coding manual contained behavior anchored items as a basis for an easier understandable feedback for the pre-service teachers. However, research findings indicate that high inference items are better suited to validly capture competencies and their internal structure (cf. chapter 5.1.3.4). A first confirmatory factor analysis based on all items did not produce an adequate model fit and confirmed this finding. Thus, further confirmatory factor analyses were calculated based on the high inference items only. Since the number of participants was higher than in the pilot study, analyses were run with the two conversations clustered at person level ($N = 92$). Figure 9 shows the underlying measurement model for the confirmatory factor analyses.

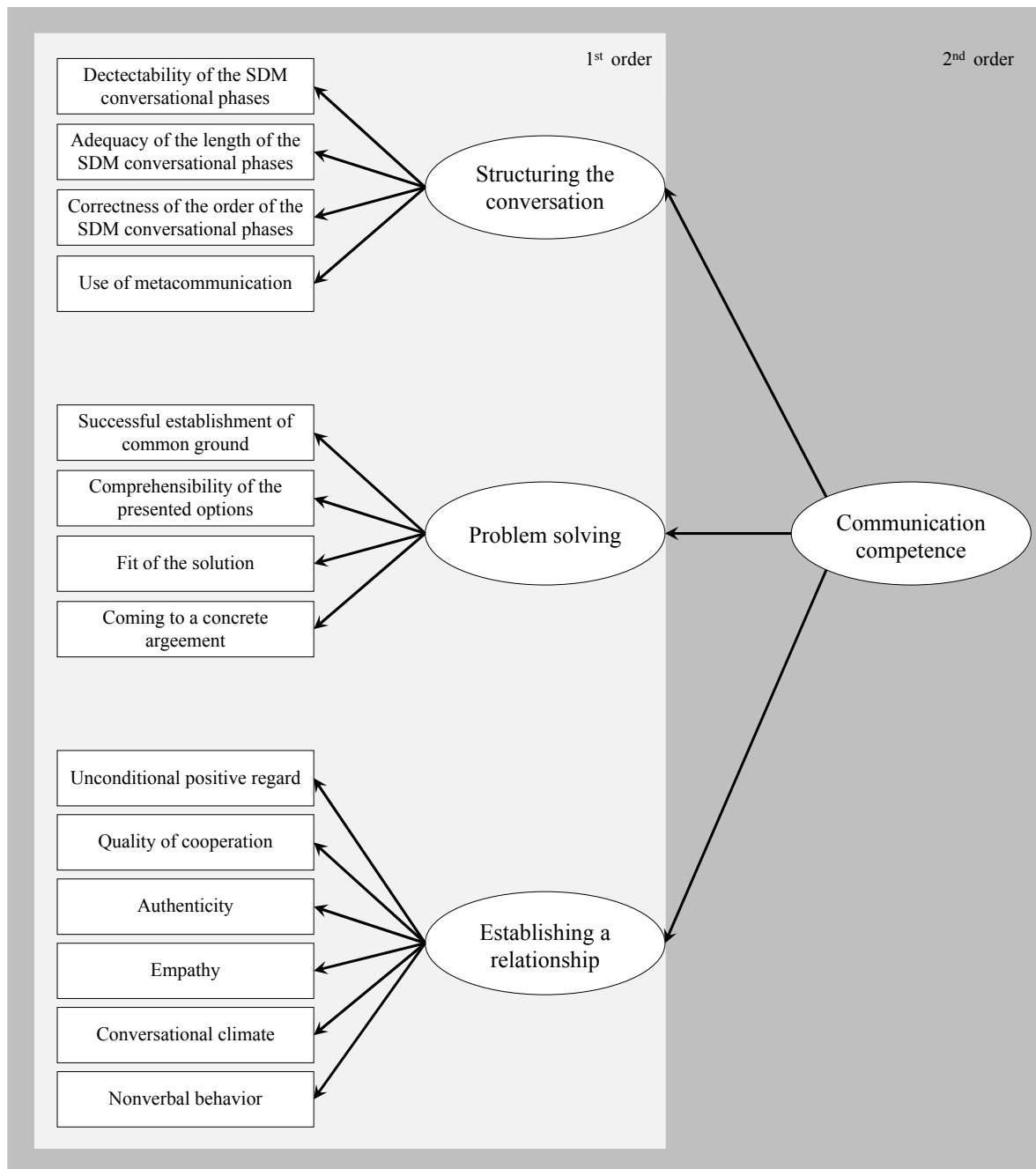


Figure 9. Measurement model of communication competence for the confirmatory factor analyses of the main study

A second order model of communication competence with the three first order factors, structuring the conversation, problem solving and establishing a relationship, based on all high inference items did not fit the data adequately. In order to improve the model fit, four items were excluded from the further analyses based on modification indices and corresponding theoretical considerations: “successful establishment of common ground at the

beginning of the conversation”, “comprehensibility of the presented options for solving the problem”, “quality of cooperation in the negotiation process” and “empathy with the conversational partner and his / her perspective”. Modification indices showed that the two excluded problem solving items had high cross loadings and did not load as high on the factor problem solving as the two remaining problem solving items, “fit of the solution” and “coming to a concrete agreement”. A theoretical explanation for this might be that the two excluded items target the first half of the conversation whereas the other two problem solving items, fit of the solution and coming to a concrete agreement, target the end of the conversation. It might be that the problem solving competencies in the two different conversational phases were too heterogeneous. The items “empathy with the conversational partner” and “quality of cooperation in the negotiation process” were excluded since the first item had a residual correlation with one problem solving item and the second item had a high cross loading on the factor problem solving. Since the later item was relocated after the pilot study from the problem solving scale, it seems as if the item captures aspects of problem solving as well as of establishing a relationship and blurs the lines between the two competence facets.

A second order confirmatory factor analysis with the second order factor, general communication competence and three first order factors, structuring the conversation, problem solving and establishing a relationship to the conversational partner, revealed an acceptable model fit ($\chi^2(32) = 41.97, p < .13, RMSEA = .06, CFI = .98, SRMR = .06$). Figure 10 provides standardized parameter estimates and residuals of this model.

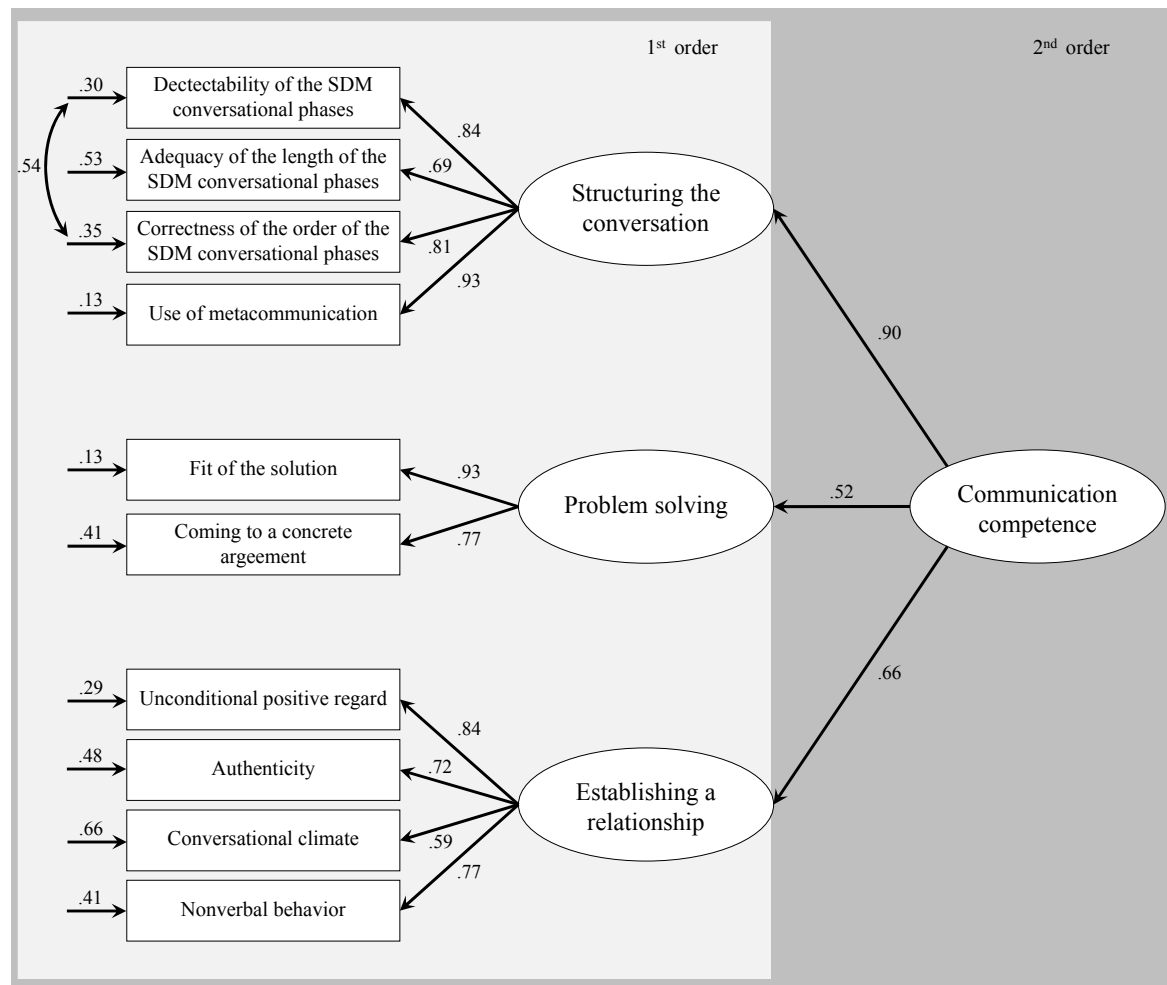


Figure 10. Results of the confirmatory factor analysis of the main study

The residual variances of the item “detectability of the shared decision making conversational phases” and the item “correctness of the order of the shared decision-making conversational phases” are correlated. This is theoretically plausible since the detectability of the conversational phases and the correctness of their order are likely interdependent.

14.3 Reliability of the Coding Manual - Internal Consistency / Composite Reliability

The analyses in this chapter provide further evidence of research question 3 since they focus the reliability of the measurement. As in the pilot study, two types of reliability were calculated for the scale communication competence in parent-teacher conversations and its subscales: composite reliability and internal consistency. Data basis were the observer ratings of the $N = 186$ video-taped conversations.

14.3.1 *Composite Reliability*

The composite reliabilities (Raykov & Marcoulides, 2011) of the hierarchical model calculated in chapter 14.2 were good; $\rho = .90$ for structuring the communication, $\rho = .85$ for problem solving, $\rho = .82$ for establishing a relationship and $\rho = .72$ for communication competence. These results indicate a reliable measurement of the factors.

14.3.2 *Internal Consistency*

The scale communication competence in parent-teacher conversations consisting of high inference and behavior anchored items (57 items, $\alpha = .90$) and the three subscales, structuring (16 items, $\alpha = .75$), problem solving (19 items, $\alpha = .83$) and establishing a relationship (22 items, $\alpha = .80$), were internally consistent. Corrected item-scale correlations varied from .04 to .75. One behavior anchored item had a negative correlation and two behavior anchored item had close-to-zero correlations. The pre-service teachers achieved a mean value of 2.51 ($SD = 0.34$) on the scale communication competence consisting of 5-point Likert-type items. The mean communication competence of the pre-service teachers ranged from 1.75 to 3.82. Table 17 gives an overview of the internal consistency of the scales consisting of the behavior anchored items, the high inference items and the high inference items selected by the confirmatory factor analysis.

Table 17

Internal consistency of the main study scales in Cronbach's Alpha

Items	Communication competence	Structuring	Problem solving	Establishing a relationship
Behavior anchored	.83	.61	.75	.64
High inference	.86	.90	.61	.83
CFA selected	.81	.90	.81	.77

Note. CFA = confirmatory factor analysis.

Table 18 shows the characteristics of the ten items and the scale which were selected / was built based on the results of the confirmatory factor analyses. All items selected by the confirmatory factor analyses had a positive and medium to high item-scale correlation.

Table 18

Item and scale characteristics for the main study scale based on the results of the confirmatory factor analyses

Item	<i>M</i>	<i>SD</i>	<i>r_{it}</i>
Structuring the conversation			
Detectability of the SDM conversational phases	1.99	0.63	.64
Adequacy of the length of the SDM conversational phases	2.33	0.67	.53
Correctness of the order of the SDM conversational phases	1.96	0.65	.60
Quality of the use of metacommunication	2.75	0.68	.72
Problem solving			
Fit of the solution	2.32	0.77	.51
Coming to a concrete agreement	2.59	0.88	.30
Establishing a positive relationship with the conversational partner			
Unconditional positive regard	1.82	0.54	.56
Authenticity	1.69	0.53	.44
Conversational climate	1.98	0.57	.22
Nonverbal behavior	1.90	0.54	.43

Note. SDM = shared decision making; *M* = sample mean; *SD* = standard deviation; *r_{it}* = item-scale correlation.

15 In-Depth Analyses on Objectivity, Reliability and Validity

Based on the results of the basic analyses, in the following section the in-depth analyses on the objectivity, reliability and validity of the pilot study are replicated with the main study data and expanded. The additional analyses concern in particular an extended multimethod measurement and include a multitrait-multimethod matrix. The multimethod measurement and the multitrait-multimethod matrix allow the investigation of whether the relations of the results of simulated conversations to other variables correspond to theoretical expectations and provide evidence of validity.

The in-depth analyses are based on the scales elicited through the confirmatory factor analyses. As in the pilot study, the scale communication competence is calculated as a mean of the results in the three competence areas, structuring the conversation, problem solving and establishing a relationship, with the three competence facets weighed coequally due to the non-compensatory approach (cf. chapter 7.3).

15.1 Consistent Performance of the Actors in the Simulated Conversations - Objectivity of Application

The analyses in chapter 15 concern research question 4 and investigate if the different actors that portray parents have an effect on the self-assessed or observer-rated performance of the pre-service teachers in simulated conversations with parents. Due to the higher number of participants in the main study, six different actors portrayed parents; two more than in the pilot study. Three actresses portrayed mothers, three actors portrayed fathers. The participating $N = 96$ pre-service teachers were randomly assigned to the actors and conducted one conversation with a simulated mother and one with a simulated father each.

15.1.1 *Effect of the Actor on the Self-Assessed Performance*

The $N = 182$ self-assessment questionnaires were divided into six groups according to the actors performing the conversation. One-way ANOVAs showed no statistically significant effects of the conversational partner on the self-assessed competence to communicate with

parents, $F(5,176) = 2.11$, $MSE = .22$, $p = .07$, $\eta^2 = .06$. There were also no statistically significant effects in the sub competence facets, structuring the conversation, $F(5,177) = 2.28$, $MSE = .38$, $p = .05$, $\eta^2 = .06$, problem solving $F(5,181) = 2.08$, $MSE = .29$, $p = .07$, $\eta^2 = .06$ and establishing a relationship with the conversational partner, $F(5,180) = 1.48$, $MSE = .31$, $p = .20$, $\eta^2 = .04$. Table 19 gives an overview of per-actor sample sizes, means and standard deviations of the self-assessed performance of the pre-service teachers.

Table 19

Per-actor sample sizes, means and standard deviations for the self-assessed performance of the pre-service teachers in the simulated conversations of the main study

	<i>n</i>	<i>M</i>	<i>SD</i>
Structuring the conversation			
Actor 1	33	2.92	0.47
Actor 2	40	2.84	0.55
Actor 3	19	2.42	0.65
Actor 4	32	2.75	0.62
Actor 5	33	2.74	0.77
Actor 6	26	2.97	0.58
Problem solving			
Actor 1	34	3.15	0.53
Actor 2	41	3.15	0.61
Actor 3	19	3.00	0.59
Actor 4	33	3.19	0.48
Actor 5	34	3.24	0.50
Actor 6	26	3.48	0.50
Establishing a relationship			
Actor 1	34	3.33	0.47
Actor 2	41	3.16	0.56
Actor 3	19	3.14	0.64
Actor 4	33	3.06	0.53
Actor 5	33	3.30	0.61
Actor 6	26	3.36	0.52

Note. *n* = number of cases in subsample; *M* = sample mean; *SD* = standard deviation; scaling: 4-point Likert scales with higher values indicating better performance.

15.1.2 *Effect of the Actor on the Observer-Rated Performance*

Analogously to the self-assessment questionnaires, the 185 video ratings were divided into six groups according to the actor performing the conversation. One-way ANOVAs showed statistically significant effects of the conversational partner on the observer-rated competence to communicate with parents, $F(5,179) = 4.09$, $MSE = .16$, $p \leq .01$, $\eta^2 = .10$. There were also statistically significant effects on the sub competence facets, structuring the conversation, $F(5,179) = 3.82$, $MSE = .31$, $p \leq .01$, $\eta^2 = .10$, problem solving $F(5,179) = 6.52$, $MSE = .51$, $p \leq .01$, $\eta^2 = .15$ and establishing a relationship with the conversational partner, $F(5,179) = 2.50$, $MSE = .17$, $p = .03$, $\eta^2 = .07$.

Since the six actors acted out two different cases, the effect of the case vignette and of the actor might be confounded. In order to extrapolate the effects of actors and case vignettes, two separate one-way ANOVAs were conducted for the three actors acting out case one (improving poor school achievements) and the three actors acting out case two (choice of the linguistic and scientific branch). For case one there was no statistically significant effect of the conversational partner on the observer-rated competence to communicate with parents, $F(2,88) = 1.26$, $MSE = .15$, $p = .29$, $\eta^2 = .03$. There were also no statistically significant effects in the sub competence facets, structuring the conversation, $F(2,88) = 0.16$, $MSE = .22$, $p = .85$, $\eta^2 \leq .01$, problem solving $F(2,88) = 1.11$, $MSE = .43$, $p = .33$, $\eta^2 = .03$ and establishing a relationship with the conversational partner, $F(2,88) = 2.45$, $MSE = .17$, $p = .09$, $\eta^2 = .05$. For case two there was also no statistically significant effect of the conversational partner on the observer-rated competence to communicate with parents, $F(2,91) = 1.39$, $MSE = .18$, $p = .25$, $\eta^2 = .03$. There were also no statistically significant effects in the sub competence facets, structuring the conversation, $F(2,91) = 1.85$, $MSE = .39$, $p = .16$, $\eta^2 = .04$ and establishing a relationship with the conversational partner, $F(2,91) = 3.13$, $MSE = .17$, $p = .05$, $\eta^2 = .06$. There was a statistically significant but small effect on the competence facet problem solving $F(2,91) = 5.05$, $MSE = .58$, $p \leq .01$, $\eta^2 = .10$. Table 20 gives an overview of per-actor sample sizes, means and standard deviations of the observer-rated performance.

Table 20

Per-actor sample sizes, means and standard deviations sorted by competence facet and case vignette for the observer-rated performance of the pre-service teachers in the simulated conversations of the main study

	<i>n</i>	<i>M</i>	<i>SD</i>		<i>n</i>	<i>M</i>	<i>SD</i>
Structuring the conversation							
Case vignette 2				Case vignette 1			
Actor 1	34	2.46	0.72	Actor 4	26	2.07	0.41
Actor 2	41	2.27	0.52	Actor 5	31	2.09	0.53
Actor 3	19	2.59	0.64	Actor 6	34	2.14	0.45
Total	94	2.41	0.63	Total	91	2.10	0.46
Problem solving							
Case vignette 2				Case vignette 1			
Actor 1	34	2.94	0.74	Actor 4	26	2.12	0.69
Actor 2	41	2.67	0.78	Actor 5	31	2.36	0.68
Actor 3	19	2.25	0.76	Actor 6	34	2.18	0.61
Total	94	2.68	0.79	Total	91	2.23	0.66
Establishing a relationship							
Case vignette 2				Case vignette 1			
Actor 1	34	1.78	0.35	Actor 4	26	1.73	0.35
Actor 2	41	1.74	0.47	Actor 5	31	1.94	0.44
Actor 3	19	2.02	0.37	Actor 6	34	1.95	0.35
Total	94	1.81	0.42	Total	91	1.88	0.38

Note. *n* = number of cases in subsample; *M* = sample mean; *SD* = standard deviation; scaling: 5-point Likert scales with lower values indicating better performance.

Since these results indicate that there is no or only a very small effect of actor (Rasch, Friese, Hofmann, & Naumann, 2006a, Rasch, Friese, Hofmann, & Naumann, 2006b) but an effect of case vignette, the effect of the case vignette on the self-assessed and observer rated performance is investigated in more detail in chapter 15.2.

15.2 Effect of the Case Vignettes - Generalizability

As in the pilot study, each pre-service teacher conducted two randomized simulated parent-teacher conversations based on two different case vignettes. Again, both conversations were shared decision making conversations. However, in order to inquire more deeply into the effects of the case vignettes and the generalizability of the results of simulated conversations and to provide further evidence of research question 5, this time the conversations contained two different topics. Half of the pre-service teachers started with a conversation on the choice of the linguistic or the scientific branch, the other half began with a conversation about a student who has to decide between different alternatives to improve his grades (cf. chapter 7.1).

The pre-service teachers rated the difficulty of the conversations on eight 4-point Likert items ($\alpha = .76$), with higher values indicating a higher level of difficulty. They answered items such as “I found it difficult to conduct the conversation”. The pre-service teachers perceived the overall level of difficulty of both case vignettes ($M = 2.2$, $SD = 0.51$) as well as of each individual case vignette as adequate: case vignette on the choice of the linguistic or the scientific branch ($M = 2.27$, $SD = 0.61$), case vignette on improving poor school achievements ($M = 2.15$, $SD = 0.65$).

15.2.1 Correlation of the Results of the two Conversations With Each Other

For each video, a global value for communication competence in parent-teacher conversation was calculated. A correlational analysis for the global results of each pre-service teacher in conversation one and conversation two showed that the results of the two conversations were significantly positively correlated with each other ($r = .35$, $p \leq .01$). The strength of the correlation was medium (Cohen, 1988).

15.2.2 *Effect of Case Vignette on the Self-Assessed Performance*

The $N = 183$ questionnaires in which the pre-service teachers had self-assessed their performance were divided according to the two case vignettes. Independent samples t -tests showed that there was no statistically significant effect of case one ($M = 3.13$, $SD = 0.49$) and case two ($M = 3.04$, $SD = 0.47$) on the self-assessed performance of communication competence in parent-teacher conversations, $t(181) = 1.33$, $p \leq .19$, $d = .09$ 95% CI [-.05, .23]. There were also no statistically significant effects of case vignette on the sub competence facets, structuring the conversation, $t(182) = .43$, $p \leq .67$, $d = .03$ 95% CI [-.14, .22] and establishing a relationship to the conversational partner, $t(185) = .12$, $p \leq .91$, $d = .01$ 95% CI [-.15, .17]. There was a statistically significant effect on the competence facet problem solving, $t(186) = 2.05$, $p \leq .04$, $d = .16$ 95% CI [.01, .32]. However, the effect size was small (Cohen, 1988). Table 21 gives an overview of per-case vignette sample sizes, means and standard deviations for the three competence facets, structuring the conversation, problem solving and establishing a relationship.

Table 21

Per-case vignette sample sizes, means and standard deviations for the self-assessed performance of the pre-service teachers in the simulated conversations of the main study

	<i>n</i>	<i>M</i>	<i>SD</i>
Structuring the conversation			
Case vignette 1	92	2.82	0.67
Case vignette 2	92	2.78	0.57
Problem solving			
Case vignette 1	94	3.29	0.51
Case vignette 2	94	3.12	0.58
Establishing a relationship			
Case vignette 1	93	3.23	0.57
Case vignette 2	94	3.22	0.55

Note. *n* = number of cases in subsample; *M* = sample mean; *SD* = standard deviation; scaling: 4-point Likert scales with higher values indicating better performance.

15.2.3 *Effect of Case Vignette on the Observer-Rated Performance*

The $N = 185$ video ratings were divided according to the two case vignettes. Independent samples *t*-tests revealed a statistically significant, medium effect of case one ($M = 2.07$, $SD = 0.38$) and case two ($M = 2.31$, $SD = 0.42$) on the observer-rated performance of communication competence in parent-teacher conversations, $t(183) = -4.05$, $p \leq .01$, $d = .60$, 95% CI [- .36, -.12] (Cohen, 1988). There were also statistically significant, medium effects of case vignette on the sub competence facets, structuring the conversation, $t(183) = -3.88$, $p \leq .01$, $d = .64$, 95% CI [- .47, -.15] and problem solving, $t(183) = -4.28$, $p \leq .01$, $d = .62$, 95% CI [- .67, -.25]. There was no statistically significant effect on the competence facet, establishing a relationship to the conversational partner, $t(183) = .84$, $p = .40$, $d = .12$, 95% CI [- .07, .17]. Table 22 gives an overview of per-case vignette sam-

ple sizes, means and standard deviations for the three competence facets, structuring the conversation, problem solving and establishing a relationship.

Table 22

Per-case vignette sample sizes, means and standard deviations for the observer-rated performance of the pre-service teachers in the simulated conversations of the main study

	<i>n</i>	<i>M</i>	<i>SD</i>
Structuring the conversation			
Case vignette 1	91	2.10	0.47
Case vignette 2	94	2.45	0.62
Problem solving			
Case vignette 1	91	2.23	0.66
Case vignette 2	94	2.68	0.79
Establishing a relationship			
Case vignette 1	91	1.87	0.42
Case vignette 2	94	1.82	0.42

Note. *n* = number of cases in subsample; *M* = sample mean; *SD* = standard deviation; scaling: 5-point Likert scales with lower values indicating better performance.

15.3 Perceived Authenticity of the Simulated Conversations - Response Processes

With regard to research question 6 and the authenticity of the response processes of pre-service teachers in simulated conversations, the pre-service teachers assessed the authenticity of the two simulated conversations on three 4-point Likert-type items, with higher values indicating higher authenticity ($\alpha = .79$). They perceived the simulated conversations as highly authentic ($M = 3.43$, $SD = .59$). 81.7% of the pre-service teachers assigned a value of three or higher than three. Only one out of 96 pre-service teachers perceived the conversations as not authentic.

15.4 Multimethod Measurement - Relations of the Results of Simulated Conversations to Other Variables

The analyses in 15.4 correspond to research question 7 and investigate to what degree correlations between the observer ratings of the simulated conversations and other measurements and external variables correspond to theoretical expectations. For the main study, the multimethod measurement in which the simulated conversations were embedded was extended. As in the pilot study, the multimethod measurement took part on two levels. The first measurement level concerned the actual conversation and the second measurement level included also measurements and variables that extended beyond the actual conversation situations, such as previous knowledge. On the first level, as in the pilot study, independent observers rated the performance of the pre-service teachers in the simulated conversations and the pre-service teachers self-assessed their performance. In addition, in the main study the simulated parents also rated the performance of the pre-service teachers in the simulated conversations and a subsample of 20 videos from 20 different pre-service teachers was coded with six adapted medical instruments in order to provide evidence of the convergent and discriminant validity of simulated conversations (cf. chapter 7.7). On the second level, the pre-service teachers did a situational judgment test on parent-teacher conversations and reported about external criteria, such as their previous knowledge, their Abitur grade or their study time in questionnaires.

Correlations with the medical instruments were calculated at conversation level since the sample was only $n = 20$. All other correlations were calculated at conversation level and at

person level based on the average result of both conversations. Only correlations at person level are reported since medical research (cf. chapter 5.1.3.1) indicates that several cases are needed in order to validly assess communication competence. The comparison of correlations at conversation level and person level confirmed this finding. Correlations at person level were a little higher and corresponded slightly better to the theoretical expected structure.

With regard to prior research, e.g. by Aich (2011) and Hertel (2009) (cf. chapter 4.3), it is probable that pre-service teachers' self-assessments of the performance in simulated conversations are only low correlated to independent observer ratings. It is also likely that there is a low or no correlation at all between the general perceptions pre-service teachers have of their communication competence prior to the simulated conversations and their performance in the simulated conversations (cf. chapter 4.3). However, the independent observer ratings should be positively correlated to the ratings of the simulated parents, external criteria, such as semester or previous knowledge, and the results of the situational judgment test on parent-teacher communication and other instruments measuring similar constructs, such as the adapted medical instruments. The strength of the correlation between the ratings with the adapted coding manual and with the adapted medical instruments depends on the contents and designs of the adapted medical instruments (cf. chapter 7.7). The correlations with the instruments Common Ground, Berliner Global Rating and OPTION are, e.g., expected to be high (cf. chapter 7.7). The lowest correlation is expected with SEGUE since it is a low inference instrument (cf. chapter 7.7).

Table 23 shows the correlations between the observer-rated performance of pre-service teachers in simulated conversations, the ratings of this performance by the simulated parents and the self-assessed performance of the pre-service teachers. Moreover, correlations of these three measurements with the results of the situational judgment test and with external criteria are depicted.

Table 23

Correlations of the observer ratings of the performance in simulated conversations with ratings of the simulated parents, self-assessments of the pre-service teachers and external criteria

	Convergent / Discriminant Evidence of Validity			Criterion-Based Evidence of Validity			
	Simulated parents	Pre-service teachers	Situational judgment test	Abitur grade	Previous knowledge	Semester	Self-assessed communication competence
Independent observers	.48**	.21**	.43**	.23*	.11	.37**	.08
Simulated parents		.36**	.26**	-.02	.21*	.47**	.27**
Pre-service teachers			.15	.08	.14	.05	.33**

Note. * $p \leq .05$; ** $p \leq .01$; self-assessed communication competence = pre-service teachers self-assessment of their general communication competence prior to communication training program and simulated conversations.

The results of the correlational analyses presented in Table 23 provide convergent, discriminant and criterion-based evidence for the validity of simulated conversations as an instrument for measuring the performance of pre-service teachers in conversations with parents. As expected, there is a medium to high correlation between the ratings of the independent observers and the ratings of the simulated parents and there are low to medium correlations between those two measurements and the self-assessments of the pre-service teachers (Cohen, 1988). Moreover, the ratings of the independent observers and the simulated parents are medium correlated to the results of a situational judgment test on parent-teacher conversations. This result also corresponds to theoretical expectations. There are also, as expected, low-medium correlations between the ratings of the independent observers and ratings of the simulated parents with external criteria (Cohen, 1988). The self-assessments of the pre-service teachers of their performance in the simulated conversations are not related to external criteria, apart from their general self-assessed communication competence. The ratings of the independent observers are not related to the general self-assessed communication competence of the pre-service teachers, while the ratings of the simulated parents are. In addition to these findings, further correlational analyses showed that previous knowledge is positively correlated to semester ($r = .20, p \leq .05$) and to general self-assessed communication competence ($r = .23, p \leq .04$).

In a second step, the correlations between the observer ratings / the ratings of the simulated parents / the self-assessments of the pre-service teachers and the results of the adapted medical instruments were investigated. Table 24 gives an overview of the corresponding correlational structure.

Table 24

Correlations of the ratings of the performance in simulated conversations by the observers, the simulated parents and the pre-service teachers with observer ratings based on six adapted medical instruments

	Convergent / Discriminant Evidence of Validity					
	SEGUE	OPTION	Common Ground	EPSCALE	LUCAS	BGR
Independent observers	.38	.53*	.70*	.41	.58**	.64**
Simulated parents	.32	.46	.62**	.48*	.53*	.49*
Pre-service teachers	.39	.17	.12	.25	.18	.11

Note. * $p \leq .05$; ** $p \leq .01$; SEGUE = Set the stage, elicit information, give information, understand the patient's perspective, end the encounter; OPTION = Observing patient involvement in decision making; EPSCALE = Explanation and Planning Scale; LUCAS = Liverpool Undergraduate Communication Assessment Scale; BGR = Berliner Global Rating.

As expected, the correlations between the independent observer-ratings based on the coding manual and the adapted medical instruments are highest (mean of $r = .55$)²⁴, followed by still comparatively high correlations between the ratings of the simulated parents and the ratings based on the adapted medical instruments (mean of $r = .49$). Lastly, correlations between the adapted medical instruments and self-assessments of the pre-service teachers are low (mean of $r = .21$) and not significant with one exception; the correlation between the self-assessments of the pre-service teachers and SEGUE is comparatively high (though not significant; this might be due to the small sample size).

The correlational structure between the independent observer ratings based on the coding manual and the adapted medical instruments mainly corresponds to theoretical expectations. As expected, there are high correlations between the observer ratings based on the coding manual and ratings with Common Ground, the Berliner Global Rating and

²⁴ The correlation coefficients were transformed to Fishers Z, averaged and then retransformed to Pearson's r for easier interpretation (Bortz, 2005).

OPTION. The correlation with LUCAS is higher than expected. Correlations with EPSCALE are lower than expected and not significant. Lastly, as expected, the correlation between observer ratings with the newly-developed coding manual and SEGUE are the lowest and not significant.

In a last step, a multitrait-multimethod matrix was developed in order to compile further evidence of validity (Bühner, 2011). According to Bühner (2011), the highest correlations in a multitrait-multimethod matrix should be between the same traits measured with the same methods, followed by the same traits measured with different methods, followed by different traits measured with the same methods and the lowest correlations should be between different traits, measured with different methods. Table 25 features a multitrait-multimethod matrix for the three competence facets, structuring the conversation, problem solving and establishing a relationship, measured with different methods: independent observer ratings, ratings of the simulated parents and self-assessments of the pre-service teachers. In this example the same traits were not measured with the same methods.

Table 25

Multitrait-Multimethod Matrix

		Independent observers			Simulated parents			Pre-service teachers		
		Structuring	Solving	Relationship	Structuring	Solving	Relationship	Structuring	Solving	Relationship
Independent observers	Structuring									
	Solving	.33**								
	Relationship	.43**	.27*							
Simulated parents	Structuring	.34**	.21*	.32**						
	Solving	.42**	.30**	.46**	.67**					
	Relationship	.32**	.15	.49*	.63**	.74**				
Pre-service teachers	Structuring	.15	.02	.08	.26*	.19	.22*			
	Solving	.13	.26*	.05	.32**	.24*	.18	.51**		
	Relationship	.17	.16	.16	.31**	.32**	.35**	.48**	.72**	

Note. * $p \leq .05$; ** $p \leq .01$; convergent validity: same trait, different methods; discriminant validity: different traits, same method / different traits, different methods; structuring = competence facet structuring the conversation; solving = competence facet problem solving; relationship = competence facet establishing a relationship with the conversational partner.

As expected, the correlation between different traits measured with different methods is the lowest (mean of $r = .23$)²⁵. However, in contrast to the theoretical expectations, the correlations between the different traits measured with the same method are higher (mean of $r = .54$) than the correlations of the same traits measured with different methods (mean of $r = .29$). Possible explanations for these results that are contradictory to expectations are discussed in chapter 17 and 18.

²⁵ The correlation coefficients were again transformed to Fishers Z , averaged and then retransformed to Pearsons r for easier interpretation (Bortz, 2005).

16 Ancillary Psychometric Quality Criteria

In this chapter it is discussed to what degree simulated conversations fulfill ancillary psychometric quality criteria. The main psychometric quality criteria discussed in part E and F provide fundamental information about the quality of an instrument and its readiness for use. Ancillary quality criteria are additional indicators for a good psychometric instrument (Bühner, 2011). The ancillary quality criteria an instrument should adhere to fulfill comprise fairness, scaling, normalization, economy, utility, reasonableness, non-susceptibility and comparability (Bühner, 2011). Main psychometric quality criteria and ancillary psychometric quality criteria are mostly independent from each other. For example, an instrument can fulfill certain ancillary psychometric quality criteria and be economic and non-susceptible but still, interpretations of results may not be valid. Thus, all quality criteria should be taken into account when developing and evaluating instruments. While the main psychometric quality criteria are essential for instruments and should be at the core of an instrument's evaluation, instruments should also adhere to fulfill ancillary quality criteria. To which degree individual quality criteria have to be fulfilled, depends on the purpose of the instrument's use. The focus in the following is on selected criteria which are particularly important and informative regarding simulated conversations: utility, non-susceptibility, fairness, reasonableness and economy.

Utility stands for the idea that an instrument should measure a characteristic that is required by and relevant for the praxis (Bühner, 2011). In addition, the instrument should either aim at measuring a construct for which no instruments are available so far, or provide an added value regarding the targeted information (Bühner, 2011). Simulated conversations are utile. They measure the communication competence of pre-service teachers in conversations with parents, a competence required by and relevant for the praxis. They also provide an added value in comparison with other instruments that have been employed so far to measure this competence, such as self-assessments or role-plays, since they are performance-based and provide highly authentic measurement conditions (cf. chapter 15.3). Due to the high authenticity of simulated conversations, it is also likely that they are non-susceptible. *Non-susceptibility* refers to the fact that a person should neither be able to falsify a score on purpose nor unintentionally in order to derive a benefit from it (Bühner, 2011). It would be very difficult for pre-service teachers to falsify their results upwards in order to derive a benefit from it.

Fairness is the third ancillary quality criterion discussed regarding the usage of simulated conversations. A fair test or instrument does not advantage or disadvantage some individuals because of characteristics that are irrelevant to the construct (American Educational Research Association et al., 2014). With regard to fairness, it can be hypothesized that simulated conversations are comparatively fair because they resemble reality closely and, thus, probably capture mostly only construct-relevant characteristics. However, future research should investigate the fairness of simulated conversations in more detail, e.g. via an investigation of differential item functioning since the characteristics of actors or specific details of the case vignettes might advantage or disadvantage specific subgroups of pre-service teachers, e.g. because of migration background or gender.

Since simulated conversations are a new method in the educational domain, their reasonableness and acceptance is particularly important. *Reasonableness* concerns the ethical aspects of measurement. An instrument should go easy on a test person's time, psyche and body (Bühner, 2011). Negative consequences for the test persons have to be carefully weighed against positive consequences of an instrument's use (2014). In a follow-up study by Altmann (2014), supervised by the author of this dissertation, with a subsample of the pre-service teachers of the main study ($n = 10$), the pre-service teachers reported that they perceived the simulated conversations in addition to the conversation training they received as highly beneficial ($M = 3.90$, $SD = 3.20$, 4-point Likert-type item, higher values indicating higher degree of benefit). All pre-service teachers scored a three or higher. This finding concords with qualitative data from the main study. At the end of each study day, the pre-service teacher had open space for comments. The pre-service teachers highlighted that they found the simulated conversations very beneficial. The high degree of acceptance of the conversations speaks for their reasonableness.

Another important question for the introduction of a new method concerns its *economy*. Economy refers to the fact that an instrument should be as time- and money-efficient and easy to handle as possible (Bühner, 2011). However, the cost-benefit calculation has to take into account the importance of the research question and alternatively available instruments (Bühner, 2011). It requires a certain amount of time and money to conceptualize simulated conversations so that they can diagnose competencies reliably and validly. However, regarding the validity of their results and their consequences of use, simulated conversations have certain advantages over other instruments commonly employed to assess

teachers' competencies. Since the cost-benefit analysis of simulated conversations is central for their future employment, it is discussed in more detail in part G.

17 Summary and Discussion of the Main Study Results

The basic analyses of the main study provided results regarding the fundamental aspects of building scores: inter-rater agreement, the internal structure of the data and the reliability of the scales of the coding manual. Analyses regarding research question 1, showed that the inter-rater agreement across all 74 behavior anchored and high inference items of the coding manual and all 186 video-taped simulated conversations of the main study was good. However, inter-rater agreement was not above the set-cut off point ($ICC = .60$) for all items of the coding manual. It tends to be challenging to achieve high inter-rater agreement with high inference rating scales in general (cf. chapter 5.1.3.4). Hence, Wirtz and Caspar (2002) recommend not to focus too much on coefficients but to take into account the difficulty of the measurement process, the importance of the results and alternatively available instruments. In sum, the results of the main study regarding inter-rater agreement are satisfactory since the ratings were high inference and the total inter-rater agreement was above the set cut-off point.

With regard to research question 2, the fit of the data to the underlying theoretical construct was investigated. A hierarchical model with a second order factor, general communication competence and three first order factors, structuring the conversation, problem solving and establishing a relationship to the conversational partner, fit the data adequately. The behavior anchored items of the main study coding manual were not part of the model since in line with findings in medical research the high inference items captured the construct better (cf. chapter 5.1.3.4). The fit between the internal structure of the data and the Munich Model of Communication Competence in Parent-Teacher Conversation with the main study data provides further evidence for the validity of the interpretation of the results of simulated conversations and supports the theoretical foundation of the development of the instruments. Moreover, the finding is in line with the results of prior research indicating a three-dimensional structure of the competence to communicate with parents (Hertel, 2009).

Analyses targeting research question 3, showed that the composite reliabilities (Raykov & Marcoulides, 2011) of the hierarchical model of communication competence were good. The scales built based on the results of the confirmatory factor analysis were internally consistent. These findings provide further evidence for a sufficient reliability of the scales of the coding manual.

Based on the results of the basic analyses, the in-depth analyses inquired more deeply into important aspects of the main psychometric quality criteria, such as objectivity of application (research question 4), effects of the case vignettes and generalizability of the results (research question 5), response processes (research question 6) and fit of the correlations between observer ratings of the performance in simulated conversations and other variables to theoretical expectations (research question 7). The in-depth analyses replicated the pilot study in-depth analyses and, additionally, extended the investigation of research question 7, the correlational structure of the results in simulated conversations to other variables.

Analyses regarding research question 4, revealed that there was no statistically significant effect of the different actors on the self-assessed performance of pre-service teachers in simulated conversations with parents. However, there was a statistically significant effect of the actor on the observer-rated performance. The inclusion of the case vignette into the analysis showed that there was no statistically significant effect of actor for case one (improving poor school achievements) and a statistically significant yet small effect of the actor on the competence facet problem solving for case two (choice of the linguistic or the scientific branch). There were no statistically significant effects in the other sub competence areas or on overall communication competence for case two. These findings indicate that there was only a very slight effect of actor but an effect of case vignette. These results indicate that the actor training was overall successful and the actors mainly performed their roles as parents consistently and, in this way, provided comparable conditions for all study participants and ensured objectivity of application.

In order to investigate more deeply into the effects of the case vignettes and the generalizability of the results of simulated conversations (research question 5), one of the two case vignettes of the pilot study (both on shared decision making regarding how to promote students with low school achievements) was substituted with a similar case vignette on a shared decision making conversation regarding the choice of the linguistic or the scientific branch. The correlation between the results of the pre-service teachers in those two shared decision making conversations was medium. There was a statistically significant yet small effect of case vignette on the self-assessed performance in the competence area problem solving and statistically significant, medium effects on the observer-rated performance of communication competence in the two sub competence facets, structuring the conversation and problem solving. These results are in line with findings from the medical domain in

that they show that communication competence is context-dependent and that the validity of simulated conversations depends on the content and number of the case vignettes (cf. chapter 5.1.3.1).

Analyses targeting research question 6, the response processes in simulated conversations, investigated the perceived authenticity of the simulated conversations. The pre-service teachers perceived the simulated conversations as highly authentic. This finding indicates that simulated conversations provide authentic measurement conditions in which the pre-service teachers react and respond as they would in real situations and provides further evidence for the validity of the measurement. The high perceived authenticity is particularly promising since Dotger and colleagues (2008) consider the degree to which pre-service teachers perceive simulated conversations as authentic and meaningful as decisive for their benefit.

Lastly, analyses regarding research question 7, showed that the correlation matrix of the multimethod measurement corresponded to theoretical expectations. The two external ratings of the performance of the pre-service teachers in simulated conversations by independent observers and the simulated parents were correlated higher with each other than with self-assessments of the pre-service teachers. There were also medium correlations between these two external ratings and the results in a situational judgment test on parent-teacher conversations and external criteria, such as the study time of the pre-service teachers. The self-assessments of the pre-service teachers were not related to external criteria, apart from their self-assessed general communication competence. It is interesting that the ratings of the independent observers are not related to the general self-assessed communication competence of the pre-service teachers, while the ratings of the simulated parents are. A possible explanation is that the simulated parents might judge pre-service teachers with more self-confidence as better. A further interesting result of the correlational analysis is that previous knowledge is positively correlated to semester and to general self-assessed communication competence. This indicates that during the course of their studies, pre-service teachers become more proficient in conducting conversations with parents and that a higher amount of previous knowledge might be a means to improve the perceived competence and self-efficacy of teachers.

Correlations between the independent observer ratings / ratings of the simulated parents / self-assessments of the pre-service teachers and measurements with the six adapted medical instruments also mainly corresponded to the expected theoretical structure and provid-

ed discriminant and convergent evidence of the measurement. There were low to high correlations between the independent observer ratings and the six adapted medical instruments and the ratings of the simulated parents and the six adapted medical instruments with the first being slightly higher. Correlations between the independent observer ratings and the ratings of the simulated parents with the only low inference instrument were the lowest and correlations with the high inference instruments with similar content the highest. This finding supports again the claim that high inference instruments measure something different than low inference instruments and are probably better suited to assess complex constructs. The correlations between the ratings of the independent observers and the simulated parents and the instrument LUCAS were higher than expected. A medium correlation was expected since the newly-developed coding manual assesses to what degree the pre-service teachers show appropriate behavior and LUCAS assesses to what degree pre-service physicians / teachers show inappropriate behavior. The high correlation indicates that this change of perspective plays only a minor role. Correlations between observer ratings and simulated parents' ratings and the instrument EPSCALE were lower than expected and not significant. This might be due to the fact that while explaining and planning were part of the originally developed coding manual, the two problem solving items that concerned explanation and planning were excluded due to the results of the confirmatory factor analyses. As expected, there were no significant correlations between the self-assessments of the pre-service teachers and the results of the adapted medical instruments. Interestingly, the correlation between the self-assessments of the pre-service teachers and the only low inference instrument, SEGUE, is comparatively high / the highest (though not significant). This might be an indicator of the fact that pre-service teachers have a hard time coming to a holistic, qualitative assessment of their competencies and rather focus on isolated skills.

Finally, a multitrait-multimethod matrix revealed that in line with expectations, the correlations between different traits measured with different methods were the lowest. However, in contrast to the theoretical expectations, the correlations between different traits measured with the same method were higher than the correlations of the same traits measured with different methods. This last result might be due to two reasons. Firstly, the three competence facets, structuring the conversation, problem solving and establishing a relationship, are all sub competencies of communication competence. They might be too similar to represent different traits. This hypothesis is also supported by the results of the confirmato-

ry factor analyses, which revealed that the latent variables, structuring the conversation, problem solving and establishing a relationship, are correlated. If this hypothesis applies, the corresponding correlations would fall into the category of same traits measured with the same methods and the high correlation would then correspond to theoretical expectations. Secondly, there might be an effect of method. Self-assessments and ratings by the conversational partners are no objective accounts (cf. chapter 4.3). Thus, the pre-service teachers and the simulated parents might judge traits differently and not as differentiated as independent, trained observers. This hypothesis is supported by the fact that the correlations within the self-assessments and the ratings by the simulated parents are higher than the correlations between the different observer ratings. With regard to these results it could be interesting for future multitrait-multimethod matrices to a) include constructs, which have a relationship to the communication competence of pre-service teachers but might be better distinguishable than the three competence facets, like content knowledge relevant for parent-teacher conversations, and b) to include further measurement methods, which have a presumably smaller method bias than self-assessments. Altogether, the empirically validated relations to other variables are a further indicator of the validity of the measurement.

Taken in sum, most results of the main study are promising regarding the diagnosis of pre-service teachers' competence to communicate with parents via simulated conversations. The objectivity of application and scoring of the simulated conversations was acceptable, the results were reliable and there is evidence for the validity of the interpretation of the results. However, the results of the main study also show that, in particular, the content and number of the case vignettes and the content and design of the coding manual are decisive for the degree to which simulated conversations fulfill psychometric quality criteria.

G Discussion

Parent-teacher conversations are a core task for teachers and can positively influence the development and academic success of pupils (cf. chapter 1 and 2). However, German teachers are often not sufficiently prepared for conducting conversations with parents and, thus, during the last years parent-teacher conversations have been more strongly institutionalized (cf. chapter 2.1) and integrated into teacher education in Germany (cf. chapter 2.3). The latter ongoing process has to be monitored since teacher education should be based on evidence in order to ensure its future quality (Prenzel, 2013). In consequence, now there is a need for instruments and assessments that can bridge the gap between theory and practice, provide information about the effectiveness of teacher education regarding parent-teacher conversations, as well as about the readiness of teachers to conduct parent-teacher conversations and about starting points for further teacher education. Consequently, the aim of this dissertation was to develop and evaluate an instrument for assessing the communication competence of pre-service teachers in conversations with parents.

Simulated conversations were chosen as the main instrument for this dissertation since they are a promising approach to measuring communication competence: they are performance-oriented and context-related, offer authentic measurement conditions and allow observing otherwise unobservable behavior (cf. chapter 5). Moreover, they are better accepted, more motivating and provide more valid results than other measurement methods, such as role-plays or self-assessment questionnaires (cf. chapter 4). Additionally, simulated conversations provide an authentic (learning) experience for pre-service teachers. This is particularly important since a common research finding is that pre-service teachers are not able to realistically self-assess their performance in conversations with parents (Aich, 2011; Hertel, 2009). This might be due to a lack of necessary practical experience in teacher education that allows sorting learning content and competencies with regard to practical relevance (Hertel, 2009). Simulated conversations hold not only the potential to bridge this gap between theory and practice but also to provide empirical evidence of the quality of teacher education and the teachers' level of readiness regarding future parent-teacher conversations.

Due to their advantages and positive research findings on their objectivity, reliability and validity, simulated conversations have become a wide-spread method to train and assess competencies in the medical domain (cf. chapter 5.1). However, medical research has also identified four main foci that should be investigated when employing simulated conversa-

tions because they influence the degree to which simulated conversations fulfill psychometric quality criteria: the number and content of cases, the recruitment and training of actors, the recruitment and training of raters and the coding manual for rating the performance (cf. chapter 5.1).

Since 2007 simulated conversations have also been employed in the educational domain, as a learning tool (cf. chapter 5.3). First research regarding this new area of application provides further information about necessary evaluation foci for the simulated conversations developed in this dissertation. The findings reveal e.g. that conversations in the educational domain are far less predictable than those in the medical domain (Dotger et al., 2008). Thus, it is crucial to achieve an adequate balance between standardization and authenticity when employing simulated conversations to assess teachers' competencies. In consequence, the consistent portrayal of the parents' roles is particularly challenging in the educational domain and have been one evaluation focus for the simulated conversations developed in this dissertation.

A further finding on simulated conversations in the educational domain is that their degree of authenticity is decisive for their acceptance and the benefit students can draw from them (Dotger et al., 2008). Therefore, another focus for the evaluation of the simulated conversations in this dissertation was their perceived authenticity. In addition to the portrayal of the parents, the perceived authenticity of the conversations hinges on the case vignettes the conversations are based on (Dotger et al., 2008). Thus, a third evaluation focus for the simulated conversations developed in this dissertation was the case vignettes. The educational findings presented thus far reinforce the medical findings and speak for the corresponding choice of evaluation foci.

With regard to the importance of coding manuals and raters for the objectivity, reliability and validity of simulated conversations, there were no findings from the educational domain since simulated conversations had not been previously used as an assessment for teacher competencies. However, medical research indicates that the coding manual (Barman, 2005; Chesser et al., 2009; Iramaneerat et al., 2008) and the raters (Chesser et al., 2009; Lurie et al., 2008) influence the objectivity, reliability and validity of simulated conversations when employed as an assessment and simulated conversations were used for the first time to assess teachers' competencies in this dissertation. Thus, the coding manual and the raters were chosen as two further evaluation foci. Furthermore, when developing measurement instruments the validity of the results should always be evaluated (American

Educational Research Association et al., 2014) and was, thus, chosen as the final evaluation focus.

The transfer of simulated conversations to the educational domain as an instrument for assessing the communication competence of pre-service teachers to communicate with parents necessitated an evaluation of the instrument with regard to this purpose, concentrating on the evaluation foci deduced from prior research in the medical and educational domain. The main research findings regarding these evaluation foci are summarized and assigned to psychometric quality criteria in chapter 18. Moreover, based on these main findings recommendations for the future employment of simulated conversations are developed.

18 Main Findings and Practice Implications

Research question 1 investigated the objectivity of scoring and provided information regarding the evaluation foci coding manual and case vignettes. The analyses of the pre- and main study data showed that overall inter-rater agreement was higher than the set cut-off point for both studies. This finding indicates that the developed rater training sufficiently prepared the raters for using the newly-developed coding manual. However, inter-rater agreement in the pilot study was much higher than inter-rater agreement in the main study. Moreover, in the pilot study the raters reached the set cut-off point for every individual item, which was not the case in the main study. While most researchers agree that it is difficult to achieve high inter-rater agreement with high inference items and interpretation of inter-rater agreement coefficients should take this into account and not solely stick to cut-off levels (Seidel et al., 2005; Wirtz & Caspar, 2002), the results of the pilot study show that it is possible to achieve a high inter-rater agreement with a high inference coding manual.

The lower inter-rater agreement in the main study in comparison to the pilot study might be due to the following reasons. Firstly, due to the interdisciplinary approach of the research project *ProfKom* (cf. chapter 8), the observers had to rate expert-layman conversations from two different areas in the main study: doctor-patient and parent-teacher conversations. This proceeding could have increased the difficulty of the rating process. Secondly, the coding manual of the pilot study was modified with regard to the main study. The two main differences between the two coding manuals were that, in the main study coding manual, the written explanations of the item interstages were omitted to facilitate interval scale level (Wirtz & Caspar, 2002) and additional behavior anchored items were included as a basis for an easier interpretable feedback for the participants. The omission of the interstages in combination with the inclusion of further behavior anchored items seems to have produced a cognitive overload for the raters in the main study. On top of that, Altmann (2014) compared the satisfaction with feedback based on the behavior anchored items and on the high inference items and found no difference. Since the supplementary behavior anchored items do not seem to produce the anticipated additional value for the participants but instead overstrain the raters, whereas the omitted interstages seem to have been helpful for the raters, a slightly modified version of the pilot study coding manual is recommended for future applications. The modifications concern the items “ade-

quacy of the proportion of participation of both conversational partners in the conversation”, “quality of cooperation” and “subject-specific performance”. The first two items should be omitted since data analyses from the pre- and main study showed that the item “adequacy of the proportion of participation of both conversational partners in the conversation” does not seem to be part of the construct communication competence in parent-teacher conversations and the item “quality of cooperation” seems to blur the lines between the competence facets establishing a relationship and problem solving since it captures aspects from both competence facets. The third item, “subject-specific performance”, should be substituted with the item “fit of the solution to the problem” since this modified version of the item was evaluated as better in the main study.

The rater training, which was employed in the pilot and the main study, can be recommended as is for the future training of observers for rating high inference items with regard to simulated conversations or other types of observation of behavior. Since the main study results show how difficult it is to keep a high inter-rater agreement, the inter-rater agreement should be regularly reassessed when coding behavior over a certain period of time or number of instances.

Research question 2 investigated whether the internal structure of the data matched the theoretical model, which was the basis for the instrument development: the Munich Model of Communication Competence in Parent-Teacher Conversation (cf. chapter 3). The investigation of research question 2 provided further information about the coding manual and first evidence for the validity of the measurement. Regarding the coding manual the main study analyses showed that the high inference items captured the construct better than the behavior anchored items. This finding is consistent with prior research (Regehr et al., 1998; Seidel et al., 2005) and speaks also for the use of the pilot study coding manual in the future. The fit between the high inference items of the pilot study and the theoretical model that could be replicated with the (slightly modified²⁶) set of high inference items of the main study is an indicator of the validity of the measurement (American Educational Research Association et al., 2014) and supports the theoretical foundations for instrument development. The findings are consistent with prior research in so far as the Munich Model of Communication Competence in Parent-Teacher Conversation conceptualizes communication competence as a three-dimensional construct (Hertel, 2009). Additionally, Gartmeier and colleagues (2015) showed that the structure of the communication competence of a

²⁶ The models of the pilot study and of the main study differed only with regard to the exclusion of two problem solving and one relationship item in the main study.

sample of pre-service teachers and pre-service physicians also corresponded to the model (cf. chapter 3). The findings furthermore indicate that the model captures important aspects of the competence of (pre-service) teachers to communicate with parents and is context-specific enough while still being widely applicable, e.g. also to other types of expert-layperson conversations (Gartmeier et al., 2015).

Research question 3 targeted the reliability of the scales of the newly-developed coding manual and provided further information regarding this evaluation focus. All scales of the pilot study and the main study coding manual were reliable. Though the scales were slightly edited between the pilot study and the main study due to the results of the confirmatory factor analyses, these findings speak for the structure and the content of the coding manuals and their future usage.

Based on the results of the basic analyses regarding research questions 1-3, several in-depth analyses were conducted that target further evaluation foci and corresponding psychometric quality criteria.

Research question 4 investigated whether the actors consistently portrayed the parents. As explained above, this objectivity of application is considered particularly challenging for simulated conversations in the educational domain. The pilot study results showed no statistically significant effects of actor on the self-assessed or observer-rated performance of the pre-service teachers in the simulated conversations, which indicated a sufficiently consistent portrayal. Regarding the main study the consistent portrayal was on the whole confirmed. However, for one case vignette, there was a statistically significant yet small effect of actor on the observer-rated performance of the pre-service teachers in the competence facet problem solving. This result indicates that there can be effects of actors on the (self-assessed) performance as well as interaction effects between actors and case vignettes.

From these results follow implications regarding the future construction of case vignettes and the design of the actor training program. In general the format of the case vignettes and the subdivision into background information, interactional knowledge and triggers seems to be appropriate since they provided the actors most of the times with enough information to consistently portray the parents. For future employment - if possible - even more fixed content should be included to avoid interaction effects between case vignettes and actors. The second implication is that an extensive actor training including trial runs is important for every single case vignette in order to ensure a consistent portrayal. It is notable that the effect of the actor is not on the competence facet establishing a relationship with the con-

versational partner but on the competence facet problem solving. One possible explanation is that the three actors portraying this case did not simply differ with regard to their kindness; but instead it is likely that the degree to which the actors either helped the pre-service teachers to solve the problem, e.g. through cueing, or hindered them differed. In order to prevent incidents like this, the actors should be cautioned against cueing and, if possible, inconsistent portrayals should be exemplified, e.g. with selected video sequences from trial runs or a pilot study. Lastly, the consistent portrayal of the parents should be evaluated on a regular basis since their performance is likely to change over the course of time (Adamo, 2003). For this kind of consistent reevaluation, Scheiderer (2013) has developed a category system under the supervision of the author of this dissertation.

Research question 5 investigated the effects of the case vignettes on the self-assessed and observer-rated performance of the pre-service teachers and the generalizability of the results from one simulated conversation to another. In the pilot study there was no statistically significant effect of the case vignettes on the self-assessed or observer-rated performance and the results of the two conversations based on two different case vignettes were significantly positively correlated with each other. The main study investigated the evaluation focus case vignette in greater depth. One of the two very similar case vignettes of the pilot study was substituted with another case vignette, which still concerned shared decision making but had slightly different content. Even though the two case vignettes in the main study were still comparatively similar, the correlation between them was only medium and there was a statistically significant effect of case vignette on the self-assessed and observer-rated performance of the pre-service teachers. These main study findings confirm prior research from the medical and educational domain, which shows that the case vignettes are extremely important for the objectivity, reliability, validity and acceptance of simulated conversations (Barman, 2005; Dotger et al., 2008; Guiton et al., 2004; Iramaneerat et al., 2008; Newble, 2004). Since the content and number of the case vignettes seems to be critical for the validity of an assessment consisting of simulated conversations, it is recommended for the future employment of simulated conversations to thoroughly select the content for the case vignettes, e.g. according to medical guidelines, and to construct assessments that consist of multiple cases. The fewer cases included, the more important it is to provide students with routine cases and multiple opportunities to show a certain behavior (Lang et al., 2004).

The findings regarding the influence of the case vignettes' content on the performance in simulated conversations have also certain implications regarding the theoretical model of communication competence underlying the instrument development. As depicted in chapter 3, the Munich Model of Communication Competence distinguishes three sub competencies needed for successfully conducting conversations with parents: the competence to structure a conversation, to solve a problem and to establish a relationship. Competencies allow solving complex problems and are, thus, multidimensional, i.e. consist of a variety of learnable cognitive abilities and associated motivational, volitional and social capabilities (cf. chapter 3). Content knowledge seems to be either a precondition for or one part of the competence to conduct parent-teacher conversations. Future research should investigate the relation between content knowledge and communication competence and / or identify further components of the three sub competence facets, which are influential with regard to the requirements in different types of parent-teacher conversations. Figure 11 depicts content knowledge as a precondition for the communication competence of teachers to communicate with parents (left) or as one part of it (right), which is particularly important with regard to the sub competence facet problem solving.

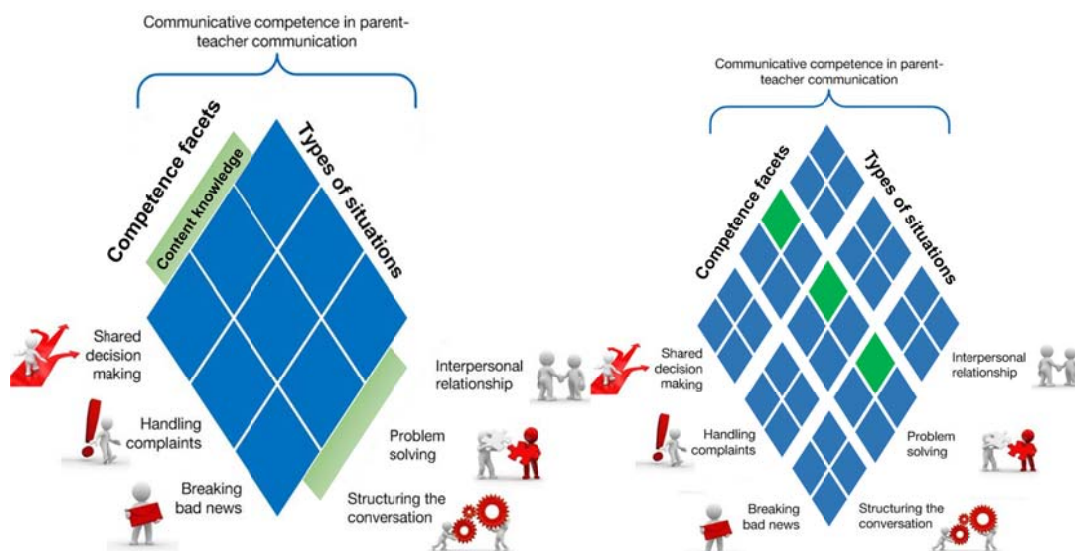


Figure 11. Content knowledge and the Munich Model of Communication Competence in Parent-Teacher Conversation (Adapted from Gartmeier et al., 2011, p. 414)

One of the advantages of the Munich Model of Communication Competence in Parent-Teacher Conversations is its wide applicability while it is still context-specific, i.e. tailored to parent-teacher conversations. In order to maintain this wide applicability, content

knowledge is only included in general and not narrowed down to specific kinds of situations, like the models of Hertel (Hertel, 2009) and Bruder and colleagues (2010), which target conversations regarding learning. In this way, the modified model highlights the importance of content knowledge needed to successfully conduct parent-teacher conversations and is still applicable to a variety of different parent-teacher conversation situations.

Research question 6 targeted the evaluation focus authenticity of the conversations. The corresponding analyses provide information about the response processes of the pre-service teachers and, in line with this, about the validity of the results. The pre-service teachers perceived the conversations as highly authentic in both the pre- and the main study. From this result it can be concluded that the response processes of the pre-service teachers were also authentic, which speaks for the validity of the measurement. Additionally, these results further support the finding that the content of the case vignettes and the portrayal of the characters by the actors were adequate and realistic.

Research question 7 investigated to what degree the relations between the results of the simulated conversations and other measurements and criteria corresponded to theoretical expectations in order to compile further evidence of the validity of the measurement. This multimethod validation approach comprised measurements and criteria compiled at two levels: in the conversation situation per se (perspectives of different raters) and beyond the actual conversation situation (external variables). Correlations between the observer ratings of the performance of the pre-service teachers in the simulated conversations and the other measurements and criteria mainly corresponded to theoretical expectations (for an in-depth discussion of all individual results see chapter 17). At the conversation level ratings of the trained observers and the simulated parents were, e.g., as expected, higher correlated with each other than with self-assessments of the pre-service teachers. Also in line with expectations, the two external assessments were higher related than self-assessments with measurements at the second level, such as a situational judgment test on parent-teacher conversations and external criteria. Correlations between observer ratings of the simulated conversations with the newly-developed coding manual and six adapted medical instruments mainly corresponded to theoretical expectations (see chapter 17 for possible explanations of results which did not correspond to a priori hypotheses). The lowest correlation was, as expected, between ratings with the newly-developed coding manual and ratings with the only low inference instrument. This finding is in line with prior research and indicates that low inference instruments measure something different than high inference in-

struments (Chesser et al., 2009; Newble, 2004; Regehr et al., 1998; Seidel et al., 2005; Skillings et al., 2010). The self-assessments of the pre-service teachers were not related to ratings with the medical instruments what also corresponded to expectations. This finding is again indicative of the fact that it is difficult for pre-service teachers to come to a realistic self-assessment of their competencies to conduct conversations with parents. The highest, though still not significant, correlation was between the self-assessments of the pre-service teachers and the only low inference medical instrument. This finding suggests that the pre-service teachers rather focus on isolated skills when judging their performance and might not be able to come to a holistic qualitative assessment of their competencies.

The results of a multitrait-multimethod matrix for the three competence facets, structuring the conversation, problem solving and establishing a relationship, corresponded only partly to theoretical expectations. As expected, the correlations between different traits measured with different methods were the lowest. However, contrary to expectations, the correlations between different traits measured with the same methods were higher than the correlations between the same traits measured with different methods. This last result might be due to two reasons. Firstly, the three competence facets, structuring the conversation, problem solving and establishing a relationship, are all sub competencies of the competence of teachers to communicate with parents. They might be too similar to represent different traits. If this applies, the corresponding correlations would fall into the category same traits measured with the same methods and the high correlation would then correspond to theoretical expectations (Bühner, 2011). Secondly, there might be an effect of method. Self-assessments and ratings by the conversational partners are no objective accounts (cf. chapter 4.3). Thus, the pre-service teachers and the simulated parents might judge traits differently and not as differentiated as independent, trained observers. This hypothesis is supported by the fact that the correlations within the self-assessments of the pre-service teachers and the ratings by the simulated parents are higher than the correlations between the different observer ratings. This result indicates that while the simulated parents are *per se* potential raters (Barman, 2005; Blake et al., 2006), they do need a more intense rater training than the minimal treatment they received in this dissertation, which corresponds to prior research findings in the medical domain (Cleland et al., 2009).

In sum, the correspondence of the empirically validated relations between the results of the simulated conversations and other variables to theoretical expectations is another indicator for the validity of the results of the simulated conversations (American Educational Re-

search Association et al., 2014). Further support for the validity of the simulated conversations comes from a recent study by Gartmeier and colleagues (2015). They could show that the simulated conversations could differentiate between pre-service teachers that had received communication training on parent-teacher conversations and a control group (Gartmeier et al., 2015). Moreover, the simulated conversations also allowed to differentiate between groups that had received different types of conversation training (Gartmeier et al., 2015) and, in this way, could deliver information regarding the effectiveness of different types of teacher education training components. Based on the results of the simulated conversations, points for refinements of teacher training programs regarding parent-teacher conversations could be deduced (Gartmeier et al., 2015).

To answer the overall research question, the consistency of the results of the pre- and the main study indicates that simulated conversations are suited to diagnose the communication competence of pre-service teachers in parent-teacher conversations. If constructed properly, they deliver objective, reliable and valid results. In comparison to other instruments (cf. chapter 4) simulated conversations provide an added value since they are performance-based, context-related and provide comparatively authentic measurement conditions. Moreover, due to the high degree of authenticity of the measurement conditions, simulated conversations are probably non-susceptible by the participants and supposedly comparatively fair since they mostly capture construct relevant characteristics (cf. chapter 16). However, the results presented in this dissertation also show that the different components of simulated conversations, such as the case vignettes or the coding manual, influence the degree to which simulated conversations do fulfill psychometric quality criteria and capture competencies adequately. Thus, the different components of simulated conversations have to be carefully designed and evaluated in order to establish and ensure an adequate balance between authenticity and standardization. The design and evaluation process requires a considerable amount of expertise, time and money. Thus, the costs and benefits of simulated conversations should be carefully weighed with regard to the purposes and framework conditions of their employment.

19 Limitations and Future Research Areas

Since this study was a first evaluation of simulated conversations as an assessment for teachers' competencies, it could not be exhaustive but merely aimed at compiling first evidence of their suitability and readiness for use. While the study could show that simulated conversations - if constructed properly - are suited to diagnose the competencies of pre-service teachers to communicate with parents, it has also several limitations. First of all, the sample consisted only of pre-service *Gymnasium* teachers in the first phase of their teacher education²⁷. Consequently, results cannot be generalized to teachers in-service since they might react differently to simulated conversations. In contrast to pre-service teachers, in-service teachers have already conducted conversations with real parents and might judge the authenticity of simulated conversations differently. Thus, future research should evaluate to what degree simulated conversations are also suited to diagnose the communication competence of in-service teachers and pre-service teachers in the second, practical phase of teacher education. Experiences from the medical domain show that simulated conversations are not only well-accepted by students but also by physicians and can be used to train and assess them (Beullens et al., 1997; Rethans et al., 2007). This finding is promising regarding a future employment of simulated conversations with in-service teachers. Associated to this research area is the question for good points in time during teacher education and advanced teacher training at which simulated conversations provide a high benefit and should, thus, be applied; they might, e.g., be better suited for pre- than for in-service teachers due to different acceptance levels or they might not be suited for pre-service teachers at the very beginning of their education since these might lack necessary content knowledge. With regard to this, the prognostic validity of simulated conversations is also of interest. Future research should investigate to what degree results of simulated conversations of teachers at different educational and professional stages are related to the quality of real parent-teacher conversations (Stokoe, 2013). A last, long-term point for this line of research is, if and how the employment of simulated conversations and the improved capacity of teachers to communicate with parents affect students.

²⁷ Teacher education in Germany is divided into two parts. During the first phase pre-service teachers acquire theoretical knowledge at a university or teacher training college in combination with practical phases at schools. The first phase concludes with the first state examination. In the second, practical phase trainee teachers work in schools accompanied by theoretical sessions. This phase concludes with the second state examination.

A second limitation of this study is that the results are only to a limited extent generalizable to pre-service teachers studying for other types of school since at least one of the case vignettes was tailored especially to the needs of *Gymnasium* teachers and the analyses conducted in this dissertation showed that the case vignettes are comparatively influential. In the future simulated conversations for (pre-service) teachers of other school types should be developed and evaluated for the following reason: several studies hint to the fact that especially students with low socio-economic or migration background benefit from parent-teacher cooperation and that parent-teacher cooperation might be a means to provide more equal chances for all students (Hertel, 2009; Hill et al., 2004). In Germany a disproportional high number of students with low socio-economic and / or migration background attend other types of school than the *Gymnasium*.

A further limitation of the study is that the simulated conversations targeted a very specific type of communication competence: the competence of (pre-service) teachers to conduct formal conversations with one or two parents. Future research should investigate to what extent this specific type of communication competence, depicted in the Munich Model of Communication Competence in Parent-Teacher Conversations, can be transferred to other types of conversations / conversational partners. It is one of the advantages of the model that it is applicable to a variety of contexts and it can be assumed that communication competence as conceptualized in the model might be helpful in other types of conversations or with other conversational partners, like students or colleagues, as well: one also needs to structure the conversation, solve a problem and establish a relationship to the conversational partner. However, this assumption needs to be empirically tested.

Lastly, due to the practical relevance of parent-teacher conversations, the potential of simulated conversations for the training and assessment of (pre-service) teachers for parent-teacher conversations and the comparatively few preceding research findings in this area, future research on parent-teacher conversations and simulated conversations is needed. With regard to simulated conversations, the evaluation of factors that influence their objectivity, reliability and validity should be continued. Since the slight differences in the design of the components of the simulated conversations for the pre- and main study have turned out to be influential and since there are also complex interaction effects between the different components, an analysis of simulated conversations via Generalizability Theory seems promising. Generalizability Theory allows singling out and quantifying the contribution of different sources of error variance, which critically influence the objectivity, reliability and

validity of an assessment, as well as interaction effects between the different error variance sources (Cardinet, Johnson, & Pini, 2010). Furthermore, with the help of decision studies, it is also possible to single out points of improvement for an assessment and, in this way, to refine simulated conversations (Cardinet et al., 2010). The application of Generalizability and Decision studies requires the development of a design, which systematically varies all potential sources of error (Cardinet et al., 2010). This was not possible in this dissertation but should be done in future research based on the potential sources of error identified in this study.

20 Outlook

The simulated conversations, which were presented and discussed in this dissertation, were developed and evaluated in a complex validation study that was part of a research project (cf. chapter 8). The results of the dissertation show that simulated conversations are a valuable instrument for research on teacher education, which can deliver empirical evidence of its effectiveness, and that has potential for further usage in research on teacher training. As a next step, the suitability of simulated conversations for daily use in teacher education is of the utmost relevance since teachers as well as parents demand more teacher education regarding parent-teacher conversations (cf. chapter 2). The developments in the medical domain show that over the course of 50 years simulated conversations have become a worldwide and frequently used method for the training and assessment of medical students (Association of Standardized Patient Educators, 2014; Barrows & Abrahamson, 1964) due to their decisive advantages in comparison with other instruments, such as their high authenticity, acceptance and their fulfillment of psychometric quality criteria (Barman, 2005; Cleland et al., 2009; Newble, 2004). Today, simulated conversations are used as a training method or a formative assessment in medical education in order to give group or individual feedback (Cleland et al., 2009; Görlitz et al., 2014; May et al., 2009) as well as in high-stakes final exams in order to assess whether the competencies of individual medical students are at an adequate level to start their profession (Adamo, 2003; United States Medical Licensing Examination, 2015). With awareness of the developments in the medical domain, it is conceivable that simulated conversations will be integrated into teacher education as both training tools and / or formative assessments during the course of studies in order to identify starting points for future training or as a summative, final exam at the end of teacher education in order to assess whether teachers are adequately prepared for conversations with parents.

With regard to an application as summative assessments, simulated conversations must fulfill a variety of (psychometric quality) criteria. The findings in the medical domain (cf. chapter 5) and this dissertation illustrate that in order to use simulated conversations in high-stakes exams much time, effort, research and money are necessary and that in the educational domain simulated conversations are still at the first development level. However, the results of this dissertation are promising in so far as they indicate that, if constructed and evaluated properly, simulated conversations also possess potential to be

integrated as summative assessments in the educational domain, e.g. in order to assess the competences of pre-service teachers to conduct conversations with parents at the end of their study program.

The research findings presented in this dissertation indicate, furthermore, that simulated conversations could be integrated into teacher education as formative assessments at their current development stage. They are well-accepted, offer authentic learning experiences and are sufficiently objective, reliable and valid. The usage of simulated conversations as formative assessments is less time and money intensive than as summative assessments. With regard to formative purposes, it is, e.g., sufficient if the simulated conversational partners rate the performance in the conversations instead of independent raters (Cleland et al., 2009). This finding from the medical domain is supported by the comparatively high correlation between ratings by the independent raters and simulated parents found in this dissertation despite the fact that the simulated parents were only minimally trained for rating the conversations (cf. chapter 15.4). However, if simulated conversational partners (only) rate the conversations, they should be trained intensively for this task (Cleland et al., 2009). The results from the multi-trait multimethod matrix presented in this dissertation indicate e.g. that the ratings of the trained observers were more differentiated (the method effect was smaller) than those of the simulated parents. Moreover, descriptive analyses showed that the simulated parents were not as strict as the independent observers.

A further possibility to design simulated conversations in a more economical way is to train students from higher semesters as raters or simulated parents. This possibility is sometimes used in medical education (S. Harris, personal communication, April 08, 2014). However, this approach has the danger that students portraying patients and participating students know each other and the simulated conversations end up being role-plays and lose some of their authenticity.

A last possibility to save time and money when employing simulated conversations concerns the videotaping process. It is not absolutely necessary to videotape the conversations if simulated parents or independent observers rate the performance of the pre-service teachers during or immediately after the conversations. Not videotaping conversations saves money, e.g. with regard to the equipment and the salaries of the camerapersons. Another possibility - if the quality of the videos is not decisive and the videos are not used for summative assessments - is to have students videotape their own performance or those

of peers with their phones or computers since today a lot of those devices possess integrated cameras.

One point that is particularly important when using simulated conversations as a formative assessment is feedback. Feedback can significantly enhance learning (Hattie & Timperley, 2007) and was considered very important by the participants in this and other studies in which simulated conversations (B. Dotger, personal communication, March 07, 2014) or role-plays (Hertel, 2009) were used to train and assess the competencies of pre-service teachers to communicate with parents. In the follow-up study by Altmann (2014), in which a subsample of the pre-service teachers who had conducted simulated conversations in this dissertation received feedback on their performance, the participating pre-service teachers reported that they perceived the feedback in addition to the simulated conversations as very helpful. Feedback that allows participants to compare self-assessments to external assessments is particularly beneficial since it helps pre-service teachers to become better at realistically judging their competencies (Hertel, 2009; Kluger & DeNisi, 1996) and, in this way, contributes to bridging the gap between teacher education and practice. In order to give external feedback, representative and up-to-date comparative samples are needed, which allow interpreting and classifying individual results (American Educational Research Association et al., 2014). If simulated conversations are videotaped, pre-service teachers should also receive access to their videos since videos can not only be used when giving feedback, but also inspire self-reflection and enhance the learning effect further (Kleinknecht & Schneider, 2013).

A last point that should be considered when integrating simulated conversations into teacher education is that this study has shown that the content of simulated conversations is particularly influential. Thus, the design of simulated conversations should be based on content from regular courses taught in teacher education when embedding them into teacher education (Dotger, 2011a). This allows pre-service teachers to put the theoretically acquired knowledge promptly into practice and ensures that they possess the pedagogical content knowledge to conduct the conversations. This close connection between theory and practice probably also has a motivating effect on pre-service teachers since they long for a more practice-oriented teacher education that prepares them for the every-day professional life of teachers (Putnam & Borko, 2000; Terhart, 2009). Lastly, the connection between regular method courses and simulated conversations would make simulated conversations an integral, instead of an additional, part of teacher education. When connecting regular

method courses with simulated conversations, the idea of simulated parents can also be broadened to simulated students, school leaders or colleagues (Dotger, 2011b). An ensemble of simulated conversations with different conversational partners, addressing different content knowledge acquired during teacher education and comprising different types of conversations, will provide a comprehensive picture of a pre-service teacher's communication competence and single out starting points for further training.

In sum, this dissertation has provided evidence that simulated conversations possess potential for research on teacher education and for teacher education practice, for formative as well as summative purposes. With regard to research, simulated conversations can provide empirical evidence for the effectiveness of teacher education regarding parent-teacher conversations and about the readiness of (pre-service) teachers for conducting future parent-teacher conversations. Beyond this, simulated conversations should be integrated into teacher education and advanced teacher training since they allow promoting and assessing the competencies of (pre-service) teachers to communicate with parents and, in this way, foster relations between parents and teachers and contribute to the student success and development.

H Appendix

21 List of Figures

FIGURE 1. PREPARATION FOR PARENT-TEACHER CONVERSATIONS BY TEACHER EDUCATION PROGRAMS (ADAPTED FROM HERTEL, 2014, P. 1)	16
FIGURE 2. MUNICH MODEL OF COMMUNICATION COMPETENCE IN PARENT-TEACHER CONVERSATION (GARTMEIER ET AL., 2011, P. 414).....	34
FIGURE 3. CLASSIFICATION OF COMPETENCE ASSESSMENT METHODS ORDERED ALONG A CONTINUUM OF DIRECTNESS (ADAPTED FROM SHERNOFF AND KRATOCHWILL, 2004, P. 371)	52
FIGURE 4. RESEARCH QUESTIONS ASSIGNED TO PSYCHOMETRIC QUALITY CRITERIA (ADAPTED FROM BÜHNER, 2011, P. 74, AMERICAN EDUCATIONAL RESEARCH ASSOCIATION, AMERICAN PSYCHOLOGICAL ASSOCIATION, & NATIONAL COUNCIL ON MEASUREMENT IN EDUCATION, 2014)	81
FIGURE 5. COMPONENTS OF THE MULTIMETHOD MEASUREMENT FOR VALIDATING THE RESULTS OF THE SIMULATED CONVERSATIONS BASED ON THE INDEPENDENT OBSERVER RATINGS WITH THE DEVELOPED CODING MANUAL	87
FIGURE 6. RESEARCH DESIGN OF THE MAIN STUDY.....	113
FIGURE 7. MEASUREMENT MODEL FOR THE CONFIRMATORY FACTOR ANALYSES OF THE PILOT STUDY	123
FIGURE 8. RESULTS OF THE CONFIRMATORY FACTOR ANALYSIS FOR THE SECOND ORDER MODEL OF COMMUNICATION COMPETENCE.....	125
FIGURE 9. MEASUREMENT MODEL OF COMMUNICATION COMPETENCE FOR THE CONFIRMATORY FACTOR ANALYSES OF THE MAIN STUDY.....	149
FIGURE 10. RESULTS OF THE CONFIRMATORY FACTOR ANALYSIS OF THE MAIN STUDY.....	151
FIGURE 11. CONTENT KNOWLEDGE AND THE MUNICH MODEL OF COMMUNICATION COMPETENCE IN PARENT-TEACHER CONVERSATION (ADAPTED FROM GARTMEIER ET AL., 2011, P. 414) ..	186

22 List of Tables

TABLE 1 <i>ALLOCATION OF THE COMPETENCE FACETS OF THE MODELS BY HERTEL (2009) AND BRUDER (2011) TO THE COMPETENCE FACETS OF THE MUNICH MODEL OF COMMUNICATION COMPETENCE IN PARENT-TEACHER CONVERSATIONS.....</i>	37
TABLE 2 <i>INTER-RATER AGREEMENT ACROSS SEVEN RATERS AFTER THE TRIAL RUN SORTED BY ITEMS</i>	96
TABLE 3 <i>TOTAL INTER-RATER AGREEMENT FOR EVERY POSSIBLE RATER PAIR AND THE TWO EXPERTS AFTER THE TRIAL RUN</i>	97
TABLE 4 <i>SIX INSTRUMENTS FOR ASSESSING PHYSICIAN-PATIENT ENCOUNTERS.....</i>	100
TABLE 5 <i>INTER-RATER AGREEMENT FOR THE SIX ADAPTED MEDICAL INSTRUMENTS.....</i>	104
TABLE 6 <i>EXPECTED CORRELATIONAL STRUCTURE BETWEEN OBSERVER RATINGS BASED ON THE NEWLY-DEVELOPED CODING MANUAL AND THE SIX ADAPTED MEDICAL INSTRUMENTS</i>	106
TABLE 7 <i>INTER-RATER AGREEMENT PILOT STUDY</i>	121
TABLE 8 <i>FIT INDICES OF THE ONE- AND THREE-DIMENSIONAL MODEL OF COMMUNICATION COMPETENCE - PILOT STUDY.....</i>	124
TABLE 9 <i>ITEM AND SCALE CHARACTERISTICS FOR THE PILOT STUDY CODING MANUAL.....</i>	128
TABLE 10 <i>PER-ACTOR SAMPLE SIZES, MEANS AND STANDARD DEVIATIONS FOR THE SELF-ASSESSED PERFORMANCE OF THE PRE-SERVICE TEACHERS IN THE SIMULATED CONVERSATIONS OF THE PILOT STUDY</i>	131
TABLE 11 <i>PER-ACTOR SAMPLE SIZES, MEANS AND STANDARD DEVIATIONS FOR THE OBSERVER-RATED PERFORMANCE OF THE PRE-SERVICE TEACHERS IN THE SIMULATED CONVERSATIONS OF THE PILOT STUDY</i>	133
TABLE 12 <i>PER-CASE VIGNETTE SAMPLE SIZES, MEANS AND STANDARD DEVIATIONS FOR THE SELF-ASSESSED PERFORMANCE OF THE PRE-SERVICE TEACHERS IN THE SIMULATED CONVERSATIONS OF THE PILOT STUDY</i>	135
TABLE 13 <i>PER-CASE VIGNETTE SAMPLE SIZES, MEANS AND STANDARD DEVIATIONS FOR THE OBSERVER-RATED PERFORMANCE OF THE PRE-SERVICE TEACHERS IN THE SIMULATED CONVERSATIONS OF THE PILOT STUDY.....</i>	136
TABLE 14 <i>CORRELATIONS OF THE OBSERVER RATINGS OF THE PERFORMANCE IN SIMULATED CONVERSATIONS WITH SELF-ASSESSMENTS OF THE PRE-SERVICE TEACHERS AND EXTERNAL CRITERIA.....</i>	138
TABLE 15 <i>FREQUENCIES OF SCORES OF THE FOUR RATERS IN THE MAIN STUDY TRIAL RUN IN PERCENT</i>	145
TABLE 16 <i>INTER-RATER AGREEMENT FOR THE SUBSCALES OF THE MAIN STUDY CODING MANUAL</i>	147
TABLE 17 <i>INTERNAL CONSISTENCY OF THE MAIN STUDY SCALES IN CRONBACH'S ALPHA</i>	152
TABLE 18 <i>ITEM AND SCALE CHARACTERISTICS FOR THE MAIN STUDY SCALE BASED ON THE RESULTS OF THE CONFIRMATORY FACTOR ANALYSES</i>	153

TABLE 19 <i>PER-ACTOR SAMPLE SIZES, MEANS AND STANDARD DEVIATIONS FOR THE SELF-ASSESSED PERFORMANCE OF THE PRE-SERVICE TEACHERS IN THE SIMULATED CONVERSATIONS OF THE MAIN STUDY</i>	156
TABLE 20 <i>PER-ACTOR SAMPLE SIZES, MEANS AND STANDARD DEVIATIONS SORTED BY COMPETENCE FACET AND CASE VIGNETTE FOR THE OBSERVER-RATED PERFORMANCE OF THE PRE-SERVICE TEACHERS IN THE SIMULATED CONVERSATIONS OF THE MAIN STUDY</i>	158
TABLE 21 <i>PER-CASE VIGNETTE SAMPLE SIZES, MEANS AND STANDARD DEVIATIONS FOR THE SELF-ASSESSED PERFORMANCE OF THE PRE-SERVICE TEACHERS IN THE SIMULATED CONVERSATIONS OF THE MAIN STUDY</i>	161
TABLE 22 <i>PER-CASE VIGNETTE SAMPLE SIZES, MEANS AND STANDARD DEVIATIONS FOR THE OBSERVER-RATED PERFORMANCE OF THE PRE-SERVICE TEACHERS IN THE SIMULATED CONVERSATIONS OF THE MAIN STUDY</i>	162
TABLE 23 <i>CORRELATIONS OF THE OBSERVER RATINGS OF THE PERFORMANCE IN SIMULATED CONVERSATIONS WITH RATINGS OF THE SIMULATED PARENTS, SELF-ASSESSMENTS OF THE PRE-SERVICE TEACHERS AND EXTERNAL CRITERIA</i>	165
TABLE 24 <i>CORRELATIONS OF THE RATINGS OF THE PERFORMANCE IN SIMULATED CONVERSATIONS BY THE OBSERVERS, THE SIMULATED PARENTS AND THE PRE-SERVICE TEACHERS WITH OBSERVER RATINGS BASED ON SIX ADAPTED MEDICAL INSTRUMENTS</i>	167
TABLE 25 <i>MULTITRAIT-MULTIMETHOD MATRIX</i>	169

23 List of Abbreviations

ANOVA	analysis of variance
CFA	confirmatory factor analysis
cf.	compare
e.g.	for example / <i>exempli gratia</i>
i.e.	<i>id est</i>
ICC	intraclass correlation
SDM	shared decision making

24 Materials

All materials were developed in the context of the research project *ProfKom* funded by the German Federal Ministry of Education and Research. I would like to especially thank all project partners, colleagues, student assistants and the simulated parents for their cooperation and support in particular with regard to the development of the simulated conversations. I developed the coding manual for the main study in cooperation with Dr. Martin Gartmeier. The self-assessment questionnaire is part of a questionnaire for evaluating the entire project developed by all project partners (Bauer et al., 2013). The original materials are in German. Since the dissertation is in English, I translated all materials into English. For space reasons only the English version of the materials are enclosed.

24.1 Case Vignette

“Christiane and the Choice of the Second Foreign Language“

Case Vignette for Simulated Father

Name: Mr. Dieter Knapphelm (father of **Christiane Knapphelm**, fifth grade at secondary school)

Age: around 40 years

Background information

- **Qualification and job:** You work as a department leader at the local bank. This is a save and good position. You have great plans for the future of your daughter. You want her to make something of her life and to have success at school.
- **Family situation:** You are married and have two children. Your older daughter, Christiane, is 11 years old. She is in fifth grade at secondary school. Your son, Hans-Peter, is 9 years old and goes to elementary school.
- **Relationship to Christiane:** You have a good relationship with your daughter. Christiane speaks with you openly about school matters, when / if you are free after work. The primary reference person of your daughter is your wife. Christiane told your wife that she has not made up her mind about the choice of the second foreign language, yet. More specifically, she does not know whether she should attend the scientific or the linguistic branch.

Setting and aim of the conversation

Today you are going to the office hour of the class teacher. You have phoned the teacher beforehand and made an appointment. On the phone you have told the teacher that you would like to have his / her counsel with regard to the choice of the linguistic or the scientific branch. You know that after fifth grade Christiane has to decide whether she will take the linguistic branch (with Latin) or the scientific branch (with French). You mentioned to the teacher that Christiane has not made up her mind, yet. You personally slightly prefer the scientific branch but, like your daughter, you are uncertain and hope for further information today. Important for you is that your daughter comes to a good decision concerning her future, which will contribute to Christiane's interest in school and her enthusiasm about learning.

Christiane's school situation

Christiane is a diligent pupil. Success at school is very important for her. In general, she is a rather sensitive and shy girl. You have the impression that she feels fine in her class.

Christiane is enthusiastic about scientific subjects as well as about English. You know that she has told her English teacher very proudly that she reads Harry Potter in English and that English is her favorite subject at the moment. Then again, in the context of the open all-day school program, she participates voluntarily in the project group „Discover the world with natural sciences!“ in the afternoon. The teacher and Christiane herself told you that she enjoys working in this group and that she does very well there. However, most of the participants are boys and they teased Christiane two weeks ago. One of them told her that only boys comprehend natural sciences and that girls should do other things. In consequence, Christiane was very angry and sad and did not want to go to the project group last week.

Most of Christiane's friends are girls who plan to attend the linguistic branch with Latin as second and French or Spanish as third foreign language. Christiane herself is undecided between the linguistic and the scientific branch with French as the second foreign language and a subsequent specialization in a scientific subject. On the one hand, she enjoys the language classes and likes to attend classes together with her friends. Moreover, last year Christiane spent one week in England in a guest family. She was very excited about her stay. Since that time she would like to go abroad for a year to attend a school in England. On the other hand, she finds the natural sciences very exciting and has expressed the wish to become an engineer like her uncle, who constructs robots. In the context of a cooperation project between her school and companies she was allowed to accompany her uncle at work.

The decision between the linguistic and the scientific branch is uppermost in Christiane's mind. Your wife told you that, because of the upcoming decision, Christiane is thoughtful and pores over her future. Since the incident in the natural sciences project group, she fears to be separated from her friends. Christiane's thoughtfulness and worries concern you and it means a lot to you to come to a good decision for your daughter.

Your behavior in the conversation

In general:

You are looking forward to the conversation. You are glad that the class teacher takes time for discussing the different options with you. However, if you feel not treated well, you show this to the teacher through distinct (but not exaggerated) body language: look to the ground, contract your eyebrows, cross your arms, face away your body axis etc. The teacher should notice this and react. If he or she does not react, stay in your posture and become more and more monosyllabic.

The teacher offers shared decision making:

- You are a little confused. Actually, you have hoped for the teacher's concrete advice. Why should you decide together?
- If the teacher explains why you should participate in the decision-making process (involving your and your child's needs and life conditions), you are satisfied and agree.

- If the teacher does not explain why you should participate in the decision-making process, you inquire more vehemently.

Your response to the teacher's suggestions

- When the teacher addresses the scientific branch, you look very worried. You are still concerned about the incident that happened two weeks ago and you worry if Christiane will feel fine in the new group. What does the teacher think?
- When the linguistic branch is addressed, you are very interested and nod. Certainly, Christiane would be happy to be in a class with her friends, but is this the right decision for her future?

Your point of view

- You are intent on deciding soon, so that Christiane can stop poring over the decision and start to be happy again.
- You think Christiane is capable of achieving good grades in the linguistic branch as well as in the scientific branch.
- In general you do have the impression that Christiane would find her position in a new group (also amid boys). Are there any other girls, too? Still you wonder whether the linguistic branch would be the easier way for her, because she could stay together with her girls, what is really important for her.
- In contrast, you are well aware of Christiane's enthusiasm for the natural sciences and you know that she is serious about becoming an engineer. And when your daughter sets her mind on something...
- That is why you are in the middle of a difficult decision-making process. You are feeling very unsure about choosing either the one or the other possibility.
- If you are directly asked for your preference, you choose the scientific branch, but in the same breath: "But I think that my wife prefers that Christiane stays together with her friends...." "I am really in a quandary... What would you do if you were me?" "What should I do?"

Coming to a decision

- When you reach the decision-making point, you say: "I'm really not sure." You ask for the teacher's opinion / advice once more. Then you tend to the option which is more plausible after the conversation.
- If the teacher suggests something (no matter what), you are hard to convince and keep hesitating (1x).
- You refuse to postpone the decision ("I like to decide now. I would like to know what I can recommend my daughter, otherwise she will keep on worrying...")

You would agree

- If the two of you came to a preliminary decision that corresponds to the course of the conversation (either linguistic or scientific). You tend to the natural sciences because your daughter is always very delighted after the project group. After reflecting the situation you think, that she will find her position in a new group and that her dream of constructing robots seems to be her greatest wish. Still, if other arguments that convince you have been brought up by the teacher, you would also go for the linguistic branch. No matter which decision, you would definitely talk to Christiane once again.
- If the teacher suggests to make an appointment with Christiane.
- At the end you should accept the solution that seems best after the course of the conversation - so that the teacher can come to an agreement with you.

You wait, whether the teacher makes a suggestion with regard to the further proceeding. If the teacher does not present concrete next steps, you inquire for them: "What's next now?" You wait for the teacher to close the conversation.

You are a teacher at a secondary school!

Your name is Mrs. / Mr. Schuster und you work as a teacher at a secondary school in your subjects of study. You have your office hour today and you expect several parents of your pupils.

At the moment you are waiting for **Mr. Knapphelm. Christiane**, his daughter, is in fifth grade. You have been her class teacher for nearly one year now. Christiane has to decide soon whether she will take the linguistic or the scientific branch after fifth grade. Mr. Knapphelm has made an appointment by phone. He would like to have your counsel on the upcoming decision because Christiane and her parents have not made up her mind, yet.

In the following you find important background information for the conversation with Mr. Knapphelm:

Christiane participates actively in your lessons. You have the impression that success at school is very important for her. In general, you think that Christiane is a rather sensitive and shy girl. Nevertheless, you have the impression that she has found her position in class and is popular with her classmates.

Christiane is enthusiastic about scientific subjects as well as about languages. In the context of the open all-day school afternoon program she participates voluntarily in the project group „Discover the world with natural sciences!“. The teacher who leads the project group has told you that she enjoys working on natural sciences a lot.

Most of Christiane`s friends are going to attend the linguistic branch that includes Latin as second and French or Spanish as third foreign language. You do not know what Christiane herself prefers. However, you believe that her friends will have a huge impact on her decision.

You want to discuss the following two alternatives with Mr. Knapphelm:

- 1.) The linguistic branch:** (English from the fifth grade on, Latin from the sixth grade on, French or Spanish from the eighth grade on)

Advantages:

- Latin is a good basis for acquiring other languages (especially the Romance languages).
- If Christiane chooses the linguistic branch, she will attain the Latinum, which is required for several university careers. It is expensive and time-consuming to make up for the Latinum in university.
- Spanish and French are world languages. A good knowledge of foreign languages is indispensable for a lot of professions.

- Most of Christiane's friends are girls who plan to attend the linguistic branch. At Christiane's age social inclusion is particularly important for pupils.

Disadvantages:

- A lot of pupils that have taken the linguistic branch in school and choose studies of science in university complain about a lack of preparation in comparison with students that chose the scientific branch in school. However, there are possibilities to catch up on the advance in knowledge, for example, by taking preparatory courses or enrolling in a "Studium Naturale", a one year course that prepares students for scientific studies and helps them in making their decision on which major to choose.
- Informatics is often not attended in the linguistic branch. Sometimes it can be chosen as an elective course in upper classes. However, good computer skills are important.
- You have to learn a lot of vocabularies in the linguistic branch.

2.) The scientific branch (English from the fifth grade on, French from the sixth grade on, a scientific specialization, informatics and chemistry from the eighth grade on)

Advantages:

- The pupils have a great deal of scientific courses, which is a good foundation for studies of science. This branch is particularly suited for pupils who are enthusiastic about natural sciences.
- It involves the acquisition of two foreign languages (English and French).
- Informatics is part of the curriculum and good computer skills are important for the entire life.

Disadvantages:

- Though languages are taught and acquired in the scientific branch, in comparison to the linguistic branch far less time is dedicated to them.
- Often later in life former pupils from the scientific branch have to expand their language skills with regard to professional requirements. This can be expensive and time-consuming. However, it is possible, for example in the context of a stay abroad.
- Furthermore, only two languages are acquired and the Latinum, which is an admission requirement for some university careers, for example for teacher training and for languages, is not acquired.

Both options are theoretically equally good. It depends on Christiane, her talents, her interests, her aims and social factors which option suits her best.

24.2 Training of the Simulated Parents

1. Theoretical introduction and analysis of video excerpts
(45 minutes)
2. Trial runs
(60 minutes)

Take Home Messages

- **Always try to act out your role as alike to the other simulated parents and your own previous performance as possible!**
- **Don't cue the pre-service teachers! It is their task to structure the conversation, to find an adequate solution for the problem and to establish a positive conversation climate!**
- **Try to keep an appropriate level of difficulty for the pre-service teachers. Neither be too cooperative nor give them an extremely hard time!**
- **Always stay in character!**
- **Do not give feedback to the pre-service teachers but judge their performance on the global ratings for us!**

Thank you and have a good time!

Assessment With Simulated Conversations

Instruction for the Actors

The project *ProfKom* aims at developing a training program for the communication competence of pre-service teachers in parent-teacher conversations. In order to evaluate the effectiveness of this newly-developed training program, we need an assessment that can provide information about the learning progress of the participants. The most important component of this assessment is you! 😊

You conduct simulated parent-teacher conversations with pre-service teachers. These conversations are video-taped and will be evaluated later on. In order to make the assessment as reliable and informative as possible, we have compiled certain pieces of advice in the following:

1. Try to always act out your role the same way. We compare the performance of the pre-service teachers in the conversations. Thus, it is crucial that every participant receives the same terms and conditions (*1st video analysis: model for the desired acting out of the role from the pilot study*).
2. Your colleagues portray the same parent. Pay attention to how they enact their roles during the trial runs and in the video sequences. We also compare the performance of the pre-service teachers in the simulated conversations across simulated parents, i.e. we compare the performance of a pre-service teacher who conducts a conversation with you to the performance of a pre-service teacher who conducts a simulated conversation with your colleague. Thus, it is important that all actors portray their characters as similar to each other as possible in order to offer as equal terms and conditions as possible to all participants independent of their conversational partner (*2nd video analysis: video excerpts of the realization of triggers are compared, video excerpts were selected based on an analysis of where the simulated parents differed*).
3. The assessment aims at diagnosing the competence of pre-service teachers to conduct parent-teacher conversations in a structured and solution-oriented way. In order to make this possible, it is important that you as the simulated parent are neither too active nor too passive. On the one hand, it is important that you realistically portray parents who want to communicate with the teacher and who are interested in exchanging information about the child with the teacher. With regard to this you should talk about and provide information about your child and participate actively in the conversation. On the other hand, it is important that you do neither cue the pre-service teachers nor direct the conversation. In the following you find typical sequences of (simulated) parent-teacher conversations, which are taught in the project *ProfKom*. It is the task of the teacher to **direct the conversation** through these typical conversational phases. It is important that you do not guide the conversation and lead over from one conversational phase to the next. For in detail advice for certain conversational phases have a look at the following table.

Conversational phases and content

The conversations consist roughly of the following conversational phases:

1. Establishing common ground / a shared understanding of the situation
2. Presentation of options / possible solutions by the teacher
3. Discussion of options and decision-finding (may partly take place in phase 2)
4. Coming to an agreement

	Your performance in the particular conversational phases
Common ground	<ul style="list-style-type: none"> • Talk about your child and how it is doing. If you are asked for certain information, give the teacher the information that is in your case vignette. Otherwise only talk about the bare necessities. • You rather stay at the surface of the problem. However, if the teacher asks or provides more information, you are o.k. to look at it in more detail. • Do not cue the teacher about possible solutions to the problem. • You hope that the teacher describes his perspective on your child and the rationale for the conversation. If this does not happen and the teacher offers you to come to a shared decision or starts to present possible solutions to you, you should ask for the teacher's perspective on the problem.
Presentation of options	<ul style="list-style-type: none"> • With regard to the options your main focus is on pedagogical aspects. Ask e.g. for the following: What is the (pedagogical) advantage of option x? How will my child benefit from this option? What is the short-term / long-term benefit? • You are less / only a little interested in technical aspects, such as money or organizational aspects. (E.g. How often does option x take place?) • When the teacher presents the options, try to understand how your child will benefit from the different options and which options suits its needs best. However, do not direct the conversation to a decision.
Discussion & decision-finding	<ul style="list-style-type: none"> • At the end of the conversation you are undecided and weigh the options. Let the teacher figure out and suggest what suits the needs of your child best. • Talk about your preferences (see case vignette) only when asked for it.
Agreement	<ul style="list-style-type: none"> • At the end of the conversation the teacher will come to a shared decision with you and make concrete agreements with regard to the further proceeding. Let the teacher take the initiative! • If the teacher does not come to a concrete agreement with you and starts to say goodbye and you are unclear about the further proceeding, ask for it!

4. The pre-service teachers should not only structure the conversations but also establish a positive **interpersonal relationship** to you, their conversational partner. Thus, it is important that while you in general have a positive attitude to the conversation, you are a little reserved and cautious in the beginning. (You remember your own school days with mixed feelings and wonder what you have to expect during this conversation.) If the teacher gives you a warm welcome, responds to you and your needs and you feel like you can trust him / her, you unbend gradually.
5. The third task of the pre-service teachers is that, in cooperation with you, they should find a **solution to your problem** that meets the needs of your child and that, finally, they come to a shared decision with you. The pre-service teacher should explain the proceeding in the conversation, e.g. by giving an advanced organizer like "I would like to come to a shared decision with you today". Of course you are interested why the teacher aims at finding a decision with you instead of just giving you an expert advice. Thus, you ask him why he / she proceeds this way. You should try to neither make the teacher's task too easy nor to complicate it too much. If you are asked for certain information about your child, you are happy to provide it. If not, you only talk about the bare necessities. You comment on the provided options but you do not direct the conversation towards a decision.
6. Please, always stay in character also in between the conversations! Do not introduce yourself or talk privately / longer to the pre-service teachers if you happen to meet someone in between conversations. However, you can greet the participants.
7. All pre-service teachers conduct two conversations. Please, do not give feedback to the participants after the conversations, since this might influence their performance in the next conversation or the performance of other participants.
8. Please, fill in the attached rating sheet after each conversation and give all rating sheets to the cameraperson at the end of each day.
9. All participants take place in the training and the assessment voluntarily. The grades you assign are not relevant for pass / fail of their studies.
10. Please, come to the organizers in room X at the end of each day for a short debriefing and feedback about your day.

24.3 Coding Manual Pilot Study

Coding Manual for the Rating of Video-Taped Shared Decision Making Conversations Between Parents and Teachers

Facets of the Munich Model of Communication Competence in Parent-Teacher Conversation - assigned observation areas and items		
Competence facet 1: Structuring the conversation	Competence facet 2: Problem solving	Competence facet 3: Establishing a relationship
0. Global rating items for the three competence facets		
1.1 Detectability of the shared decision making conversational phases	2.1 Successful establishment of common ground at the beginning of the conversation	3.1 Unconditional positive regard and respect for the conversational partner
1.2 Adequacy of the length of the shared decision making conversational phases	2.2 Comprehensibility of the presented options for solving the problem	3.2 Authenticity of the teacher
1.3 Correctness of the order of the shared decision making conversational phases	2.3 Quality of cooperation in the negotiation process	3.3 Empathy of the teacher with the conversational partner and his perspective
1.4 Quality of the use of meta-communication for structuring the conversation	2.4 Coming to a concrete agreement	3.4 Constructiveness of the conversational climate
1.5 Adequacy of the proportion of participation of both conversational partners in the conversation	2.5 Subject-specific performance of the teacher	3.5 Adequacy of the nonverbal behavior of the teacher

CODING INSTRUCTION

The raters can stop and rewatch the video as required. The levels of the scale 1 to 5 correspond roughly to German school grades: 1 (very good), 2 (good), 3 (satisfactory), 4 (adequate), 5 (fail).

1. **After the raters have watched the video for the first time:** Rating of the scale 0 and the items 1.1 - 1.5
2. **After the raters have watched the video for the second time:** Rating of the items 2.1-2.5
3. **After the raters have watched the video for the third time:** Rating of the items 3.1-3.5

O. GLOBAL RATING OF THE COMPETENCE FACETS

Judge the global ratings according to the overall impression you have of the teacher's performance in the three competence areas after watching the video for the first time.

1. STRUCTURING THE CONVERSATION

	1	2	3	4	5
The teacher succeeded in structuring the conversation (according to the shared decision making model) and in adequately shaping the length of the conversational phases, the transitions between the conversational phases and the proportion of participation of both conversational partners in the conversation.					
1 = totally applies / 2 = rather applies / 3 = partially applies / 4 = does rather not apply / 5 = does not apply at all					

2. PROBLEM SOLVING

	1	2	3	4	5
The teacher successfully established common ground with regard to the rationale of the conversation, developed possible solutions in cooperation with the parent and came to a concrete agreement regarding the further proceeding which satisfies the parent and his needs.					
1 = totally applies / 2 = rather applies / 3 = partially applies / 4 = does rather not apply / 5 = does not apply at all					

3. ESTABLISHING A RELATIONSHIP

	1	2	3	4	5
The teacher succeeded in establishing a positive interpersonal relationship with the parent that was beneficial for the course of the conversation by showing positive regard and acting congruently and constructively.					
1 = totally applies / 2 = rather applies / 3 = partially applies / 4 = does rather not apply / 5 = does not apply at all					

1. FACET: STRUCTURING THE CONVERSATION

1.1 DETECTABILITY OF THE SHARED DECISION MAKING CONVERSATIONAL PHASES

Question: Are the fundamental phases of a shared decision making conversation (greeting, establishing common ground with regard to the problem, offering to come to a shared decision, presenting possible solutions, negotiating a solution, coming to a concrete agreement, saying goodbye) clearly detectable?

All fundamental shared decision making conversational phases are clearly detectable.	1
The conversational phases are mainly detectable.	2
The conversational phases are only partly distinguishable or some conversational phases are clearly distinguishable, others not.	3
Only few conversational phases (1-2) are detectable.	4
The conversational phases (apart from greeting and saying goodbye) are not distinguishable.	5

TYPICAL EXAMPLES / KEY PHRASES

Shared decision making offer:

Teacher: "I have thought about possible solutions as well. I would suggest that we make use of the conversation today to JOINTLY find a solution for the problem."

NOTICE:

The conversational phases "presenting possible solutions" and "negotiating a solution" often overlap. It is not always necessary to assign a lower score for that as long as you have the impression that the course of the conversation is still well structured.

1.2 ADEQUACY OF THE LENGTH OF THE SHARED DECISION MAKING CONVERSATIONAL PHASES

Question: Are the conversational phases adequately long?

The shared decision making conversational phases are adequately long.	1
The length of most conversational phases is adequately long. 1-2 phases are a little too long or short.	2
Approximately half of the conversational phases is adequately long. The others are rather too long or too short.	3
Only few conversational phases are adequately long. Most are either too long or too short.	4
Some conversational phases do not or hardly exist. The rest is either too long or too short.	5

TYPICAL EXAMPLES / KEY PHRASES

-

NOTICE:

If conversational phases are missing, rate them as too short.

1.3 CORRECTNESS OF THE ORDER OF THE SHARED DECISION MAKING CONVERSATIONAL PHASES

Question: Is the order of the shared decision making conversational phases (greeting, establishing common ground with regard to the problem, offering to come to a shared decision, presenting possible solutions, negotiating a solution, coming to a concrete agreement, saying goodbye) correct?

All shared decision making conversational phases are in correct order.	1
One conversational phase takes place at the wrong time.	2
More than one conversational phase take place at the wrong time. Some conversational phases take place in correct order.	3
The majority of the conversational phases takes place at the wrong time.	4
The order of the shared decision making process is completely incorrect. Conversational phases take either place in the wrong order or are mixed with each other.	5

TYPICAL EXAMPLES / KEY PHRASES

-

NOTICE:

If conversational phases are missing, rate them as in incorrect order.

1.4 QUALITY OF THE USE OF METACOMMUNICATION FOR STRUCTURING THE CONVERSATION

Question: How well does the teacher structure the conversation by using metacommunication?

Criteria: Possible types of metacommunication are structuring phrases, summaries, transitions, inquiries (e.g. whether the conversational partner has understood the presented option). Decisive points in the conversations where the teacher should use metacommunication are the beginning and the end of the conversation, the offer to come to a shared decision, transitions between different conversational phases and unclear situations, e.g. misunderstandings.

The teacher uses metacommunication adequately in all relevant situations.	1
The teacher mostly uses metacommunication adequately in most relevant situations.	2
The teacher uses metacommunication only seldom and / or content-wise only partly adequate.	3
The teacher uses metacommunication only seldom and / or mostly not adequately.	4
The teacher does not use metacommunication with regard to his / her proceeding in the conversation.	5

TYPICAL EXAMPLES / KEY PHRASES

Advanced organizer: *"I would like to present three possible solutions to you. Subsequently, we can discuss which one is best for your child. I am sure we will come to a good decision today."*

Transition: *"This was the first possibility. A second possibility would be that...."*

"O.k. I think we agree so far about xyz. Now, I would like to discuss..."

Summaries: *"So if I understand you correctly, you think that..."*

1.5 ADEQUACY OF THE PROPORTION OF PARTICIPATION OF BOTH CONVERSATIONAL PARTNERS IN THE CONVERSATION

Question: Does the teacher succeed in providing both conversational partners with an adequate proportion of participation in the conversation?

Criteria: How much participation of the conversational partners is adequate depends on the conversational phase. During the presentation of possible solutions the teacher is likely to talk more than the parent. When negotiating a possible solution or establishing common ground it is more important that the teacher involves the parent and actively listens to his perspective.

The participation of both conversational partners is adequately proportioned throughout the entire conversation.	1
The participation of both conversational partners is mainly adequately proportioned throughout the conversation.	2
The participation of both conversational partners is only at times adequately proportioned.	3
The participation of both conversational partners is only in few situations adequately proportioned.	4
The teacher either talks far too much or not enough during the conversation.	5

TYPICAL EXAMPLES / KEY PHRASES

-

2. FACET: PROBLEM SOLVING

2.1. SUCCESSFUL ESTABLISHMENT OF COMMON GROUND AT THE BEGINNING OF THE CONVERSATION

Question: To what extent does the teacher succeed in establishing common ground about the rationale of the conversation at the beginning of the conversation / to what extent serves the common ground as a basis for the following conversation and the problem solving process?

Criteria: Profundity of common ground with regard to the rationale of the conversation, symptom (e.g. 5 in the last exam), problem (e.g. has trouble doing his homework) and underlying reasons (e.g. parents got divorced).

At the beginning of the conversation the teacher establishes common ground with regard to the rationale of the conversation.	1
At the beginning of the conversation the teacher mostly establishes common ground with regard to the rationale of the conversation.	2
At the beginning of the conversation the teacher partly establishes common ground with regard to the rationale of the conversation.	3
At the beginning of the conversation it remains rather unclear whether the two conversational partners have a common understanding about the rationale of the conversation.	4
At the beginning of the conversation it remains unclear whether the two conversational partners have a common understanding about the rationale of the conversation.	5

TYPICAL EXAMPLES / KEY PHRASES

“Do you have any idea why your daughter...?”

“Did you notice anything with regard to xyz during the last weeks?”

“If I understand you correctly, you have the impression / you think that the reason for xyz is ...”

NOTICE:

“1” can only be awarded if symptoms, problem and underlying reasons are clarified.

“2” should be awarded if symptoms and the problem are clarified but no underlying reasons.

“3” is the best possible rating if only the symptoms but neither the problem nor the underlying reasons are clarified.

2.2 COMPREHENSIBILITY OF THE PRESENTED OPTIONS FOR SOLVING THE PROBLEM

Question: How comprehensible does the teacher present possible solutions?

Criteria: Comprehensibility and conciseness of the presentation of possible solutions and their advantages and disadvantages.

The teacher presents possible solutions in a very comprehensible way.	1
The teacher presents possible solutions in a comprehensible way.	2
The teacher partly presents possible solutions in a comprehensible way. Sometimes it is hard to follow.	3
Only few of the teacher's explanations about the possible solution are easy comprehensible. A lot of explanations are hard to understand.	4
It is very hard to impossible to understand the possible solutions and their advantages and disadvantages.	5

TYPICAL EXAMPLES / KEY PHRASES

"I think these three options are promising and would like to discuss them with you."

"Do you have any more questions?"

"Do you know what xyz means?"

2.3 QUALITY OF COOPERATION IN THE NEGOTIATION PROCESS

Question: How cooperative is the teacher with regard to negotiating possible solutions?

Criteria: This item concerns the conversational phase “negotiating a solution”. It is about the direction of the negotiation process to a solution that suits the needs of the parent and the pupil. The teacher can ensure this by replying thoroughly to objections and inquiring for perspectives of the parent.

The conversational partners cooperate very well with each other. The teacher replies adequately to all objections and asks very often for the opinion of the parent.	1
The conversational partners cooperate well with each other. The teacher mostly replies to objections and often asks for the opinion of the parent.	2
The conversational partners cooperate partly well with each other. The teacher sometimes replies to objections and sometimes asks for the opinion of the parent.	3
The teacher hardly cooperates with the conversational partner. The teacher hardly replies to objections and scarcely asks for the opinion of the parent.	4
The conversational partners do not cooperate. The teacher does not take into account the perspective or objections of the parent.	5

TYPICAL EXAMPLES / KEY PHRASES

“What is the best solution for your child in your opinion?”

“About what can we decide after the conversation today?”

“Do you think this would be a good solution for your son?”

“Is this a possibility for you?”

“I would like to jointly.....”

“You say you have made a bad experience with private tutoring. What happened exactly?”

2.4 COMING TO A CONCRETE AGREEMENT

Question: To what degree does the teacher come to a concrete agreement with the parent about a further proceeding, which can solve the problem, at the end of the conversation?

Criteria: Clarity with regard to further proceeding

At the end of the conversation the teacher comes to a concrete agreement with the parent and discusses the further proceeding in detail.	1
At the end of the conversation the teacher comes to a relatively concrete agreement with the parent and discusses the further proceeding.	2
At the end of the conversation the teacher comes to a relatively concrete agreement with the parent and discusses the further proceeding vaguely.	3
The agreement is only vague. Teacher and parent do not discuss the further proceeding.	4
The teacher and the parent do not come to a concrete agreement.	5

TYPICAL EXAMPLES / KEY PHRASES

“What shall we do, then?”

“Then you take your time to talk about everything with your wife and daughter at home and I phone you on Monday.”

“Is it o.k. for you if I write down the fixed dates and give them to your child?”

“Then I will call you next Tuesday around 7 pm on your mobile.”

“I would like to get in contact with you again in three weeks to tell you how things have changed.”

Negative example:

“You know I do have other parent-teacher conversations today, too. Let’s just leave it like that.”

NOTICE:

Additional rule: If the parent forestalls the teacher in proposing an agreement, rate only the quality of the agreement. Ignore who took the initiative for coming to an agreement.

2.5 SUBJECT-SPECIFIC PERFORMANCE OF THE TEACHER

Question: How competent seems the teacher regarding the subject-specific content of the conversation?

The teacher is a subject-specific competent contact person.	1
The teacher is generally a competent subject-specific contact person.	2
The teacher is partly a competent subject-specific contact person, partly the teacher seems a little inept with regard to subject-specific content.	3
The teacher is sometimes a competent subject-specific contact person, more often the teacher seems a little inept with regard to subject-specific content.	4
The teacher seems mostly inept with regard to subject-specific content.	5

TYPICAL EXAMPLES / KEY PHRASES

-

3. FACET: ESTABLISHING A RELATIONSHIP

3.1 UNCONDITIONAL POSITIVE REGARD AND RESPECT FOR THE CONVERSATIONAL PARTNER

Question: To what degree is the behavior of the teacher, i.e. his statements and conduct towards the parent, appreciative?

Criteria: Friendliness / respecting the parent and his attitudes / let the parent finish speaking

The teacher is very attentive, appreciative, friendly and polite towards the parent.	1
The teacher is mostly attentive, appreciative, friendly and polite towards the parent.	2
The teacher is partly attentive, appreciative, friendly and polite towards the parent.	3
The teacher is only sometimes attentive and appreciative towards the parent. In some situations he seems rather reserved and / or impolite.	4
The teacher is rarely attentive or appreciative. He seems reserved and impolite.	5

TYPICAL EXAMPLES / KEY PHRASES

"I appreciate it a lot that you look after xyz and have come to talk with me today."

"Thank you for coming!"

"Thank you for taking your time to talk with me today!"

Negative example:

Sarcasm

Making fun of someone / something

3.2 AUTHENTICITY OF THE TEACHER

Question: Does the teacher seem authentic during the conversation? I.e. do you have the feeling that he takes not only part in the conversation as a “professional” but also as an individual that shows his feelings and thoughts openly.

Criteria: Openness with which the teacher shares his experience of the conversation (thoughts / feelings) in the conversation. No distanced hiding behind a professional role.

The teacher shares thoughts and feelings openly in the conversation.	1
In a lot of situations the teacher shares thoughts and feelings openly.	2
The teacher shares thoughts and feelings only sometimes.	3
The teacher shares thoughts and feelings only in few situations. Sometimes he seems distanced or as if hiding behind the professional role.	4
The teacher seems distanced or as if hiding behind the professional role.	5

TYPICAL EXAMPLES / KEY PHRASES

“For me it is also often difficult to detect...”

“Maybe I also realized a little too late that...”

“As far as I am concerned I try to...”

“With regard to this point I believe that ...”

“I often feel that...”

3.3 EMPATHY OF THE TEACHER WITH THE CONVERSATIONAL PARTNER AND HIS PERSPECTIVE

Question: Is the teacher interested in the parent and his perspective on the rationale of the conversation? Or does he seem rather disinterested / not involved?

Criteria: Asking for the parent’s perspective / attentive and active listening

It is of greatest concern for the teacher to empathize with the parent and to understand his perspective of the rationale for the conversation.	1
It is mostly of greatest concern for the teacher to empathize with the parent and to understand his perspective of the rationale for the conversation.	2
The teacher partly empathizes with the parent and tries to understand his perspective of the rationale for the conversation.	3
Only in few situations the teacher empathizes with the parent and tries to understand his perspective of the rationale for the conversation. In some situations he seems rather indifferent regarding the parent and his perspective.	4
The teacher seems most of the time indifferent regarding the parent and his perspective on the rationale for the conversation.	5

TYPICAL EXAMPLES / KEY PHRASES

“I can understand your worries / concerns very well.”

“Yes, I understand that...”

“I understand that it is very difficult for you. You are giving your very best but feel like you are not getting through to your son.”

Negative example:

Trivial or dishonest encouragement: *“Well, eventually your son is going to figure it out somehow.”*

3.4 CONSTRUCTIVENESS OF THE CONVERSATIONAL CLIMATE

Question: To what degree does the teacher create a positive, constructive conversational climate?

Criteria: Constructive, positive conversational climate

The conversational climate is open and constructive. The teacher signals that he believes in a positive outcome of the conversation.	1
The conversational climate is most of the time open and constructive. The teacher mainly signals that he believes in a positive outcome of the conversation.	2
The conversational climate is only partly open and constructive. The teacher only partly signals that he believes in a positive outcome of the conversation.	3
Only in few situations the conversational climate is constructive. Sometimes it is tense and rather confrontational. The teacher signals that a solution of the problem will not be reached easily.	4
The conversational climate is tense and rather confrontational. The teacher signals that he does not believe a solution for the problem will be found.	5

TYPICAL EXAMPLES / KEY PHRASES

"Your daughter can certainly achieve xyz."

"I am very glad we made such a good decision and I am convinced that if we continue working on xyz together we are going to make it in the end."

"I am glad we reached an agreement."

"If all of us act in concert, we are going to succeed in..."

Negative example:

"Someone has to help you."

"I do not see a perspective for your son with regard to this problem."

"Your daughter is a lazybones."

3.5 ADEQUACY OF THE NONVERBAL BEHAVIOR OF THE TEACHER

Question: To what extent is the nonverbal behavior of the teacher appreciative and constructive?

Criteria: Approachable body posture / kind facial expression / responsive facial expression and gestures

The teacher conveys approachableness and attentiveness through his facial expressions and body postures.	1
The teacher conveys mostly approachableness and attentiveness through his facial expressions and body postures.	2
The teacher conveys partly approachableness and attentiveness through his facial expressions and body postures.	3
The teacher only seldom conveys approachableness and attentiveness through his facial expressions and body postures. Sometimes he appears to be rather distanced and disinterested.	4
The teacher hardly conveys approachableness and attentiveness through his facial expressions and body postures. Most of the time he appears to be rather distanced and disinterested.	5

TYPICAL EXAMPLES / KEY PHRASES

Teacher faces toward parent

The teacher directly looks the parent into the eyes (does not stare!)

Expressive facial expressions

Accompanying gestures

Negative examples:

Teacher leans back or turns away from parent

Crossed arms

Bored facial expression

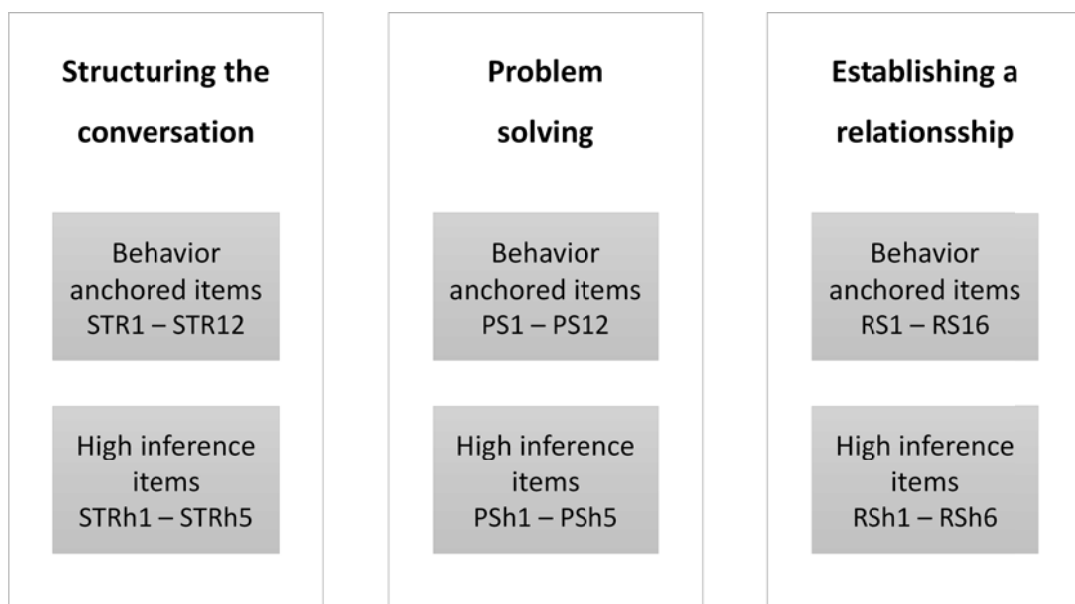
Facial expressions and tone of voice that reveal that the teacher is not convinced of what he says, e.g. "I do believe your son is gonna make it."

24.4 Coding Manual Main Study

Coding Manual for Rating Video-Taped Shared Decision Making Conversations

This coding manual is a guideline for the rating of video-taped shared decision making conversations. It is structured according to the Munich Model of Communication Competence in Parent-Teacher Conversation and distinguishes the three competence facets structuring the conversation, problem solving and establishing a relationship to the conversational partner. The coding manual contains two types of items:

1. Items that target small observable units of the student teacher's behavior (behavior anchored items)
2. Items that evaluate the performance at a higher level (high inference items)



CODING INSTRUCTION

The raters watch each video three times for coding it. During each run they stop the video according to the event sampling plan and rate the behavior anchored items. Additionally, the raters can stop and rewatch sequences of the video as required. They assign the levels of the scale 1 to 5 like German school grades: 1 (very good), 2 (good), 3 (satisfactory), 4 (adequate), 5 (fail). The rating process is as follows:

1st video run: rating of the behavior anchored items for structuring the conversation

After the raters have watched the video for the first time: rating of the high inference items for structuring the conversation

2nd video run: rating of the behavior anchored items for problem solving

After the raters have watched the video for the second time: rating of the high inference items for problem solving

3rd video run: rating of the behavior anchored items for establishing a relationship

After the raters have watched the video for the third time: rating of the high inference for establishing a relationship

OVERVIEW OF THE ITEMS

1. Competence Facet: Structuring the Conversation

Event sampling plan	Behavior anchored items
Welcome: Initial rapport (before the problem-oriented conversation starts)	STR1 The student refers to the background of the conversation or previous contacts (like a phone call).
Clarifying the problem (problem-oriented conversation before the first solution is presented)	STR2 The student states the rationale for the conversation and starts the conversation.
	STR3 The student sums up the shared perception of the problem.
	STR4 The student inquires if the parent has further questions.
Presenting possible solutions (starts with the shared decision making offer or with the presentation of the first option)	STR5 The student offers shared decision making and highlights the communality of the decision-making.
	GS6 The student gives an advanced organizer over the different options.
	STR7 The student seeks the approval of the conversational partner for his proceeding.
	STR8 The student inquires whether the conversational partner has understood the options.
Negotiating solutions / decision-making	STR9 The student leads over from one option to the next.
	STR10 The student sums up.
Coming to an agreement	STR11 The student asks if the conversational partner already favors one of the options.
Farewell	STR12 The student asks about the contentedness with the conversation (e.g. <i>"Did you benefit from the conversation?"</i>).

High inference items	STRh1 Detectability of the shared decision making conversational phases
	STRh2 Adequacy of the length of the shared decision making conversational phases
	STRh3 Correctness of the order of the shared decision making conversational phases
	STRh4 Quality of the use of metacommunication for structuring the conversation

2. Competence Facet: Problem Solving

Event sampling plan	Behavior anchored items
Welcome: Initial rapport (before the problem-oriented conversation starts)	-
Clarifying the problem (problem-oriented conversation before the first solution is presented)	PS1 The conversational partner can outline his perspective on the problem in detail early in the conversation.
	PS2 The student listens actively.
	PS3 The student poses adequate questions.
	PS4 The student adequately presents his perception of the situation.
	PS5 The student determines a goal for the conversation.
Presenting possible solutions (starts with the shared decision making offer or with the presentation of the first option)	PS6 The student presents the options in a comprehensible way.
	PS7 The student explains advantages and disadvantages of the options.
	PS8 The student expresses his opinion adequately (does not try to steer the conversational partner in a direction).
Negotiation solutions / decision-making	PS9 The student weighs advantages and disadvantages.
	PS10 The student relates advantages and disadvantages to the needs of the conversational partner.
	PS11 The student guides the negotiation process and avoids divagations.
Coming to an agreement	PS12 The agreement comprises concrete actions (<i>What will be done?</i>).
	PS13 The student and the conversational partner agree on a time frame (<i>Until when will it be done?</i>).
	PS14 The student and the conversational partner determine the responsibilities for different parts of the agreement (<i>Who has to do what?</i>).
Farewell	PS15 The student gives a realistic but encouraging prognosis.

High inference items	PS1 Successful establishment of common ground at the beginning of the conversation
	PS2 Comprehensibility of the presented options for solving the problem
	PS3 Fit of the solution to the problem and the conversational partner
	PS4 Coming to a concrete agreement

3. Competence Facet: Establishing a Relationship

Event sampling plan	Behavior anchored items
Welcome: Initial rapport (before the problem-oriented conversation starts)	RS1 The student greets the conversational partner friendly, offers a chair and takes a seat after the conversational partner.
	RS2 The student introduces himself and his professional role.
	RS3 The student makes a personal remark, which is not related to the rationale of the conversation (e.g. <i>"I am glad that you are here!"</i> , <i>"I am glad to meet you"</i>).
Clarifying the problem (problem-oriented conversation before the first solution is presented)	RS4 The student manages a positive start of the problem-oriented discussion (not: <i>"The situation is very difficult."</i> / <i>"Things look black..."</i>).
	RS5 The student does not interrupt the conversational partner.
	RS6 The student seizes remarks of the conversational partner.
	RS7 The student reflects the facial expression, gestures and voice of tone of the conversational partner.
	RS8 The student respects attitudes of the conversational partner.
	RS9 The student formulates his opinions and statements adequately.
Presenting possible solutions (starts with the shared decision making offer or with the presentation of the first option)	RS10 The student reacts to verbal and nonverbal signals of understanding from the conversational partner (<i>"I have the impression that you are skeptical - is it possible that you dislike this option?"</i>).
Negotiating solutions / decision-making	RS11 The student addresses objections constructively.
	RS12 The student inquires after the perspective of the conversational partner.
	RS13 The student takes requests of the conversational partner into account.
	RS14 The student makes encouraging remarks (<i>"I think we will get it about right."</i> / <i>"I think we will be able to help."</i>).
Coming to an agreement	RS15 The student is forthcoming and makes concrete suggestions how he can support the problem solving process.
Farewell	RS16 The student makes a personal remark, which is not related to the content of the conversation (e.g. <i>"I am glad we got to speak with each other today!"</i>).

High inference items	RSh1 Unconditional positive regard and respect for the conversational partner
	RSh2 Quality of cooperation in the negotiation process
	RSh3 Authenticity of the student
	RSh4 Empathy with the conversational partner and his perspective
	RSh5 Constructiveness of the conversational climate
	RSh6 Adequacy of the nonverbal behavior of the student

DESCRIPTION OF THE HIGH INFERENCE ITEMS

1. COMPETENCE FACET: STRUCTURING A CONVERSATION (STR)

STRh1. DETECTABILITY OF THE SHARED DECISION MAKING CONVERSATIONAL PHASES

Question: Are the fundamental phases of a shared decision making conversation (greeting, establishing common ground with regard to the problem, offering to come to a shared decision, presenting possible solutions, negotiating a solution, coming to a concrete agreement, saying goodbye) clearly detectable?

All fundamental shared decision making conversational phases are clearly detectable.	1
	2
	3
	4
The conversational phases (apart from greeting and saying goodbye) are not distinguishable.	5

TYPICAL EXAMPLES / KEY PHRASES

Shared decision making offer, student should point out that all options are of equal value:

“There are different options with different advantages and disadvantages. Together with you I would like to single out the one that matches your needs and expectations best.”

NOTICE:

The conversational phases “presenting possible solutions” and “negotiating a solution” often overlap. It is not always necessary to assign a lower score for that as long as you have the impression that the course of the conversation is still well structured (e.g. via transitions or explicit metacommunication).

STRh2. ADEQUACY OF THE LENGTH OF THE SHARED DECISION MAKING CONVERSATIONAL PHASES

Question: Are the conversational phases adequately long?

The shared decision making conversational phases are adequately long.	1
	2
	3
	4
Some conversational phases do not or hardly exist. The rest is either too long or too short.	5

TYPICAL EXAMPLES / KEY PHRASES

-

NOTICE:

If conversational phases are missing, rate them as too short.

If students take more than five minutes too long, score one level less.

STRh3. CORRECTNESS OF THE ORDER OF THE SHARED DECISION MAKING CONVERSATIONAL PHASES

Question: Is the order of the shared decision making conversational phases (greeting, establishing common ground with regard to the problem, offering to come to a shared decision, presenting possible solutions, negotiating a solution, coming to a concrete agreement, saying goodbye) correct?

All shared decision making conversational phases are in correct order.	1
	2
	3
	4
The order of the shared decision making process is completely incorrect. Conversational phases take either place in the wrong order or are mixed with each other.	5

TYPICAL EXAMPLES / KEY PHRASES

-

NOTICE:

If conversational phases are missing, rate them as in incorrect order.

STRh4. QUALITY OF THE USE OF METACOMMUNICATION FOR STRUCTURING THE CONVERSATION

Question: How well does the student make the sense and goal of individual conversational phases / transitions transparent and the course of the entire conversation comprehensible?

Criteria: Possible types of metacommunication are structuring phrases, summaries, transitions, inquiries (e.g. whether the conversational partner has understood the presented option). Decisive points in the conversations where the student should use metacommunication are the beginning and the end of the conversation, the offer to come to a shared decision, transitions between different conversational phases und unclear situations, e.g. misunderstandings.

The student uses metacommunication adequately in all relevant situations.	1
	2
	3
	4
The student does not use metacommunication with regard to his proceeding in the conversation.	5

TYPICAL EXAMPLES / KEY PHRASES

Advanced organizer: *“I would like to present three possible solutions to you. Subsequently, we can discuss which one is best for your child. I am sure we will come to a good decision today.”*

Transition: *“This was the first possibility. A second possibility would be that...”*

“O.k. I think we agree so far about xyz. Now, I would like to discuss...”

Summaries: *“So if I understand you correctly, you think that...”*

“Are there any more questions?”

“So the two of us agree that...?”

“O.k. in conclusion I would like to give you the following advice...”

2. COMPETENCE FACET: PROBLEM SOLVING (PS)

PSh1. SUCCESSFUL ESTABLISHMENT OF COMMON GROUND AT THE BEGINNING OF THE CONVERSATION

Question: To what extent does the student succeed in establishing common ground about the rationale of the conversation at the beginning of the conversation / to what degree serves the common ground as a basis for the following conversation and the problem solving process?

Criteria: Comparison of the different perspectives (e.g. perceptions and interpretations of the behavior of the student, type and underlying reasons for the problem), clarifying and comprehension questions, reassurance if everything was understood correctly, determination of concrete, shared conversation goals.

At the beginning of the conversation the student establishes common ground with regard to the rationale of the conversation.	1
	2
	3
	4
At the beginning of the conversation it remains unclear whether the two conversational partners have a common understanding about the rationale of the conversation.	5

TYPICAL EXAMPLES / KEY PHRASES

“So you expect that in the conversation today we will....”

“If I understand you correctly, you want to find a solution for XXX today.”

“You have made an appointment to talk about XXX. Is that correct?”

“If I understand you correctly, you have the impression ...”

NOTICE:

Profundity of common ground with regard to the rationale of the conversation is decisive: symptom (e.g. 5 in the last exam), problem (e.g. has trouble doing his homework) and underlying reasons (e.g. parents got divorced).

- Nothing better than a “2” should be awarded if symptoms and the problem are clarified but no underlying reasons.
- If only the symptoms but neither the problem nor the underlying reasons are clarified, the best possible rating is “3”.

PSh2. COMPREHENSIBILITY OF THE PRESENTED OPTIONS FOR SOLVING THE PROBLEM

Question: How comprehensible does the student present the different options? Does the student ensure that the conversational partner has understood the most important information?

Criteria: Comprehensibility and conciseness of the presentation of the options, adequate choice of words, reassurance that everything was understood correctly, purpose of the option is understandable for the conversational partner, options are presented correctly with relevant advantages and disadvantages in a neutral way

The student presents possible solutions in a very comprehensible and concise way and always ensures the understanding of the conversational partner.	1
	2
	3
	4
It is very hard to impossible to understand the possible solutions and their advantages and disadvantages. The student does not ensure understanding.	5

TYPICAL EXAMPLES / KEY PHRASES

"I think these three options are promising and I would like to discuss them with you."

"This was a lot of information. What is the most important point for you?"

"Do you have any more questions?"

"Which points are not clear, yet?"

Comprehensibility:

- Explaining specialist terms and avoiding unnecessary specialist terms
- Short, clearly structured sentences (no long, complicated sentences)
- Comprehensible, simple contexts of meaning
- Illustration of utterances through drawings, gestures and lively examples

PSH3. FIT OF THE SOLUTION TO THE PROBLEM AND THE CONVERSATIONAL PARTNER

Question: To what extent does the student succeed in identifying a solution that fits the problem and the needs of the conversational partner?

Criteria: Development of a solution in cooperation with the conversational partner, professional knowledge is adapted to the problem and the perspective and the background of the advice seeker, deviations from the topic are avoided, the final solution / result of the conversation fits the wishes, perspectives and preferences of the conversational partner optimally

A solution is identified which is very promising for solving the problem and which fits the needs of the conversational partner (respectively the pupil) very well.	1
	2
	3
	4
It remains unclear how well the solution fits the problem or the conversational partner / pupil or the solution does not fit.	5

TYPICAL EXAMPLES / KEY PHRASES

-

NOTICE: The focus is on the final solution; if the needs and requests of the conversational partner have remained unclear during the conversation, assign a lower score.

PSh4. COMING TO A CONCRETE AGREEMENT

Question: To what degree does the student come to a concrete agreement with the conversational partner about the further proceeding at the end of the conversation?

Criteria: Clarity with regard to further proceeding / concrete appointments / actions / contacts

At the end of the conversation the student comes to a concrete agreement with the conversational partner and discusses the further proceeding.	1
	2
	3
	4
The student and the conversational partner do not come to a concrete agreement.	5

TYPICAL EXAMPLES / KEY PHRASES

“What shall we do exactly, then?”

“Then you take your time to talk about everything with your wife and daughter at home and I phone you on Monday.”

“O.k. then we make another appointment for next Tuesday.”

Negative example:

“You know I do have other conversations today, too. Let’s just leave it like that.”

NOTICE:

Additional rule: If the conversational partner forestalls the student in proposing an agreement, rate only the quality of the agreement. Ignore who took the initiative for coming to an agreement.

3. COMPETENCE FACET: ESTABLISHING A RELATIONSHIP (RS)

RSh1. UNCONDITIONAL POSITIVE REGARD AND RESPECT FOR THE CONVERSATIONAL PARTNER

Question: To what degree is the behavior of the student, i.e. his statements and conduct towards the conversational partner, appreciative?

Criteria: Quality of the relationship between the student and the conversational partner, friendliness / respecting the conversational partner, his attitudes / points of views (without necessary approving or sharing them), letting the conversational partner finish speaking, differences in opinion are resolved in a way that does not interfere with the relation between the conversational partners

The student is very attentive, appreciative, friendly and polite towards the conversational partner.	1
	2
	3
	4
The student is rarely attentive or appreciative. He seems reserved and impolite.	5

TYPICAL EXAMPLES / KEY PHRASES

"Good to see you! I am glad you could come today!"

"I appreciate it a lot that you look after xyz and have come to talk with me today."

Unconditional positive regard: non-judgmental attitude towards the other person

Attentive: Reacting to verbal and nonverbal clues (and worries)

Making sure the conversational partner is well seated

Negative example:

Sarcasm

Making fun of someone / something

RSh2. QUALITY OF COOPERATION IN THE NEGOTIATION PROCESS

Question: To what degree does the student cooperate with the conversational partner?

Criteria: Cooperative partnership / equality of the conversational partners / discussion on eye level / involving the conversational partner, his attitudes and wishes / inquiring for perspectives and preferences of the conversational partner / thorough reaction to objections / guiding the negotiation process to a solution that fits the situation of the conversational partner

The student cooperates very well with the conversational partner.	1
	2
	3
	4
The student cooperates insufficiently with the conversational partner.	5

TYPICAL EXAMPLES / KEY PHRASES

“Can you tell me why your son...?”

“What is the best solution for your child in your opinion?”

“About what can we decide after the conversation today?”

“When you hear about these different options, what comes to your mind?”

“Do you already favor one of the options?”

“What is particularly important for you with regard to choosing one of the options?”

Negative example:

Patronizing advice or lecture that raises the competence of the conversational partner to question.

NOTICE:

In contrast to PSh3 the focus is not on the result but on the negotiation process.

RSh3. AUTHENTICITY OF THE STUDENT

Question: Does the student seem authentic during the conversation? I.e. do you have the feeling that he takes not only part in the conversation as a “professional” but also as an individual who shows feelings and thoughts openly.

Criteria: Student adequately shares point of views, emotions or impressions without attacking the conversational partner, no distanced hiding behind a professional role, congruency of verbal and nonverbal behavior

The student’s verbal and nonverbal behavior is congruent. The student shares thoughts and feelings openly in the conversation.	1
	2
	3
	4
The student’s verbal and nonverbal behavior is congruent. The student seems distanced or as if hiding behind the professional role.	5

TYPICAL EXAMPLES / KEY PHRASES

-

RSh4. EMPATHY WITH THE CONVERSATIONAL PARTNER AND HIS PERSPECTIVES

Question: Is the student interested in the conversational partner and his perspective on the rationale of the conversation?

Criteria: Seizes the position, emotions and attitudes of the conversational partner tries to understand the conversational partner, e.g. by listening attentively / inquires for perspective of the conversational partner / active listening / responds adequately to verbal and nonverbal clues, to emotions and worries

It is of greatest concern for the student to empathize with the conversational partner and to understand his perspective of the rationale for the conversation. The student inquires actively after emotions and worries of the conversational partner and reacts to nonverbal cues.	1
	2
	3
	4
The student is not able to show empathy for the conversational partner or is not interested in his perspectives. The student does not discern verbal and nonverbal clues to emotions or reacts in a counterproductive way.	5

TYPICAL EXAMPLES / KEY PHRASES

"I can understand your worries / concerns very well."

"Yes, I understand that..."

Negative example:

Ignore facial expressions that express a lack of understanding or worries

RSh5. CONSTRUCTIVENESS OF THE CONVERSATIONAL CLIMATE

Question: To what degree does the student create a constructive conversational climate?

Criteria: Constructive, positive conversational climate

The conversational climate is open and constructive. The student signals that he believes in a positive outcome of the conversation.	1
	2
	3
	4
The conversational climate is tense and rather confrontational. The student signals that he does not believe a solution for the problem will be found.	5

TYPICAL EXAMPLES / KEY PHRASES

"Your daughter can certainly achieve xyz."

"I am very glad we made such a good decision and I am convinced that if we continue working on xyz together we are going to make it in the end."

"I am glad we reached an agreement."

"If all of us act in concert, we are going to succeed in..."

Negative example:

"I do not see a perspective for your son with regard to this problem." / "Your daughter is a lazybones."

RSh6. ADEQUACY OF THE NONVERBAL BEHAVIOR OF THE STUDENT

Question: To what degree is the nonverbal behavior of the student appreciative and constructive?

Criteria: Approachable body posture / kind facial expression / responsive facial expression and gestures

The student conveys approachableness and attentiveness through his facial expressions and body postures.	1
	2
	3
	4
The student hardly conveys approachableness and attentiveness through his facial expressions and body postures. Most of the time he appears to be rather distanced and disinterested.	5

TYPICAL EXAMPLES / KEY PHRASES

Student faces toward parent

Student directly looks the parent into the eyes (does not stare!)

Expressive facial expressions

Accompanying gestures

Negative examples:

Student leans back or turns away from parent

Crossed arms

Bored facial expression

24.5 Rating Scale for the Simulated Parents

1. STRUCTURING THE CONVERSATION

	1	2	3	4	5
The teacher succeeded in structuring the conversation (according to the shared decision making model) and in adequately shaping the length of the conversational phases, the transitions between the conversational phases and the proportion of participation of both conversational partners in the conversation.					
1 = totally applies / 2 = rather applies / 3 = partially applies / 4 = does rather not apply / 5 = does not apply at all					

2. PROBLEM SOLVING

	1	2	3	4	5
The teacher successfully established common ground with regard to the rationale of the conversation, developed possible solutions in cooperation with me and came to a concrete agreement regarding the further proceeding which satisfies my needs.					
1 = totally applies / 2 = rather applies / 3 = partially applies / 4 = does rather not apply / 5 = does not apply at all					

3. ESTABLISHING A RELATIONSHIP

	1	2	3	4	5
The teacher succeeded in establishing a positive interpersonal relationship with me that was beneficial for the course of the conversation by showing positive regard and acting congruently and constructively.					
1 = totally applies / 2 = rather applies / 3 = partially applies / 4 = does rather not apply / 5 = does not apply at all					

24.6 Self-Assessment Questionnaire Pre-Service Teachers

Please, remember the two simulated parent-teacher conversations that you have just conducted. Rate to what extent you succeeded in achieving the following goals in the two conversations:

☒	In the first conversation				In the second conversation			
	Does not apply	Rather not applies	Rather applies	Totally applies	Does not apply	Rather not applies	Rather applies	Totally applies
Establishing an appreciative relation to my conversational partner	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Establishing common ground with regard to the rationale for the conversation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Structuring the conversation according to the shared decision making model	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Empathizing with my conversational partner	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Weighing the advantages and disadvantages of possible solutions in cooperation with my conversational partner	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cooperating constructively with my conversational partner	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Keeping the length of the conversational phases adequate	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Reaching a decision that optimally suits the needs of my conversational partner	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Shaping the transitions between different conversational phases adequately	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Coming to a concrete agreement with my conversational partner	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

24.7 Rater Training

Rater Training 1

Time	Learning goals	Form of work	To keep in mind
Before the training	Acquisition of knowledge about parent-teacher conversations, setting up a level of expectations with regard to the performance of the pre-service teachers in the simulated conversations	All raters participate in a communication training program for parent-teacher cooperation	
Before the training	Raters should become familiar with the content of the simulated conversations	Raters read the case vignettes	
60 minutes	Raters should get a feeling for the range of the performance to expect and the corresponding appropriate score	Discrimination training: Raters have to individually sort five videos with regard to the quality of the performance of the pre-service teachers. Subsequently, group discussion to verify the results	The five videos should comprise the entire range of performance to expect including best case and worst case
20 minutes	Raters should get a feeling for typical good or bad modes of behavior / patterns of communication / utterances	Creation training: the mediocre video is converted into a good / bad version in a group discussion	
60 minutes	Raters should get to know the coding manual and learn to assign observed behavior to certain items and appropriate scores	Concept training: Raters watch one video and list all observed modes of behavior they consider important on meta plan cards. Subsequently, raters receive and read the coding manual. They assign the observed behavior to one of the competence facets and an associated item. Their allocation is discussed in group. Subsequently, the raters rate all items. Their rating is again discussed in group.	White board with a printed version of the coding manual, meta plan cards for the raters to assign the observed behavior to items
60 minutes	Raters should learn to focus on and differentiate the three competence facets	One video is watched three times. During each run raters individually rate one competence facet. After each run the rating of the competence facet is discussed in group.	
As required	Raters should develop a common understanding of the items and get more experienced and reliable with regard to their ratings	At least two videos are rated. During the rating raters note questions and difficulties. Subsequently, ratings and remarks are discussed in group and, if necessary, additional coding rules are set up and the coding manual is adapted.	
20 minutes	Raters should learn to avoid frequent observer / rating errors	Trainer presents and explains frequent observer errors to the raters	

Rater Training 2

Rater training 1 is followed by a trial run in which 6-8 videos are rated. Video ratings are statistically analyzed (e.g. inter-rater agreement, means, frequencies) and graphically depicted. The statistical results form the basis for rater training 2.

30 minutes	Each rater should examine his agreement with the other raters and finetune his ratings	Inter-rater agreement for all items is discussed in group with particular emphasis on items which are below the targeted inter-rater agreement	Rater training 2 should take place soon enough after the coding of the trial videos so raters can still recall the reasons for their scorings
30 minutes	Equalization and calibration of the leniency / strictness of the raters and the experts	Total means and means for all items are discussed and additional rules are set up, in particular, for those items in which leniency / strictness differs between raters	
15 minutes	Raters should improve their feeling for the Likert-scale	Frequencies of ratings are discussed in group	
45 minutes	Reasons for insufficient inter-rater agreement are singled out and eliminated	Raters graphically analyze extreme cases where ratings diverge and rewatch selected video sequences as a basis for group discussion on why ratings diverge	
30 minutes	Raters are calibrated to expert ratings	Ratings are compared to expert sample solutions with explanations	Expert sample solutions for randomly selected videos, check: are differences irregular or systematic?
Optional / as required	Raters should develop a common understanding of the items and get more experienced and reliable with regard to their ratings	Further videos are rated if required. During the rating raters note remaining questions and difficulties. Subsequently, ratings and remarks are discussed in group	

The rater training is completed when all raters and experts agree on a common theoretical understanding, all questions are clarified and the discussion of the codings indicates a sufficient inter-rater agreement. Inter-rater agreement should be calculated again after 50% and after 100% of the videos are scored. All videos that are used for rater training should be excluded from the following data analysis (Seidel et al., 2005).

I Literature

- Adamo, G. (2003). Simulated and standardized patients in OSCEs: achievements and challenges 1992-2003. *Medical Teacher*, 25(3), 262–270.
- Aich, G. (2011). *Professionalisierung von Lehrenden im Eltern-Lehrer-Gespräch: Entwicklung und Evaluation eines Trainingsprogramms*. Zugl.: Schwäbisch-Gmünd, Pädagog. Hochsch. und Tübingen, Univ., Diss. *Schul- und Unterrichtsforschung: Vol. 14*. Baltmannsweiler: Schneider-Verl. Hohengehren.
- Allhoff, D. W., & Allhoff, W. (2006). *Rhetorik & Kommunikation: Ein Lehr- und Übungsbuch*. München: Reinhardt.
- Altmann, R. (2014). *Entwicklung und Evaluierung von zwei verschiedenen Feedback-Varianten zur Gesprächsführungskompetenz angehender Lehrpersonen in simulierten Lehrer-Eltern-Gesprächen* (Bachelorarbeit). Technische Universität München, München.
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association.
- Anderson, K. J., & Minke, K. M. (2007). Parent Involvement in Education: Toward an Understanding of Parents' Decision Making. *Journal of Educational Research*, 100(5), 311–323.
- Association of Standardized Patient Educators. (2014). *About ASPE*. Retrieved from <http://www.aspeducators.org/>
- Barman, A. (2005). Critiques on the Objective Structured Clinical Examination. *Annals Academy of Medicine Singapore*, 34(8), 478–482.
- Barrows, H. S. (1987). *Simulated (standardized) patients and other human simulations: A comprehensive guide to their training and use in teaching and evaluation*. Chapel Hill, NC: Health Sciences Consortium.
- Barrows, H. S., & Abrahamson, S. (1964). The programmed patient: a technique for appraising student performance in clinical neurology. *Journal of Medical Education*, 39(8), 802–805.
- Bauer, J., Gartmeier, M., Gröschl, B., Fischer, M. R., Hoppe-Seyler, T., Iblher, P., Karsten, G., Möller, G., Noll, A., Töpfer, J., Zupanic, M., & Prenzel, M. (2013). *Professionalisierung von zukünftigen Ärztinnen, Ärzten und Lehrkräften im Bereich der Kommunikationskompetenz: Skalendokumentation*. Unveröffentlichtes Projektdokument. TU München, CAU Kiel, LMU München.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Baumert, J., & Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In J. Baumert (Ed.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 321–410). Opladen: Leske & Budrich.
- Bayerische Staatsregierung. (1946). *Verfassung des Freistaates Bayern*. Retrieved from <http://www.verfassungen.de/de/by/bayern46-index.htm>
- Bayerische Staatsregierung. (2000). *Bayerisches Gesetz über das Erziehungs- und Unterrichtswesen: BayEUG*. Retrieved from <http://www.gesetze-bayern.de/jportal/portal/page/bsbayprod.psml?showdoccase=1&doc.id=jlr-EUGBY2000V30IVZ>

- Bayerische Staatsregierung. (2007). *Schulordnung für die Gymnasien in Bayern: Gymnasialschulordnung - GSO*. Retrieved from <http://www.gesetze-bayern.de/jportal/portal/page/bsbayprod.psml?showdoccase=1&st=null&doc.id=jlr-GymSchulOBY2007rahmen&doc.part=X&doc.origin=bs>
- Bayerisches Staatsministerium für Unterricht und Kultus. (2014). *Schule und Familie. Verantwortung gemeinsam wahrnehmen. Rechte und Aufgaben der Eltern und Elternvertretung in der Schule*. Retrieved from http://www.isb.bayern.de/download/15462/schule_und_familie_band_2.pdf
- Bayerisches Staatsministerium für Unterricht, Kultus, Wissenschaft und Kunst. (2008). *Dienstordnung für Lehrkräfte an staatlichen Schulen in Bayern: Lehrerdienstordnung - LDO*. Retrieved from <http://www.luitpold-gymnasium.org/documents/service/LDO-2008.pdf>
- Behr, M. (2005). Differentielle Effekte von empathischen und authentischen Eltern-Lehrer-Gesprächen im Rollenspielexperiment. *Empirische Pädagogik*, 19(3), 244–264.
- Bentele, G., Brosius, H.-B., & Jarren, O. (Eds.). (2006). *Lexikon Kommunikations- und Medienwissenschaft*. Wiesbaden: VS Verl. für Sozialwiss.
- Berger, C. R. (2008). Interpersonal Communication. In W. Donsbach (Ed.), *The international encyclopedia of communication* (pp. 2473–2486). Malden, MA: Blackwell Pub.
- Berne, E. (2002). *Spiele der Erwachsenen: Psychologie der menschlichen Beziehungen*. Reinbek bei Hamburg: Rowohlt.
- Beullens, J., Rethans, J.-J., Goedhuys, J., & Buntinx, F. (1997). The use of standardized patients in research in general practice. *Family Practice*, 14(1), 58–62.
- Bieber, C. (2007). *Patientenbeteiligung bei medizinischen Entscheidungen: Manual zur Partizipativen Entscheidungsfindung (Shared Decision-Making). Patient als Partner*. Freiburg: Universitätsklinikum.
- Blake, K. D., Gusella, J., Greaven, S., & Wakefield, S. (2006). The risks and benefits of being a young female adolescent standardised patient. *Medical Education*, 40(1), 26–35. doi:10.1111/j.1365-2929.2005.02343.x
- Bloom, B. S. (1982). *All our children learning: A primer for parents, teacher, and other educators*. New York: McGraw-Hill.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler: Mit 242 Tabellen* (6th ed.). Springer-Lehrbuch. Berlin, Heidelberg, New York: Springer.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (4th ed.). Heidelberg: Springer.
- Bromme, R., & Rambow, R. (2001). Experten-Laien-Kommunikation als Gegenstand der Expertiseforschung: Für eine Erweiterung des psychologischen Bildes von Experten. In R. K. Silbereisen & M. Reitzle (Eds.), *Psychologie 2000. Bericht über den 42. Kongress der Deutschen Gesellschaft für Psychologie in Jena 2000* (pp. 541–550). Lengerich: Pabst Science Publishers.
- Bromme, R., Jucks, R., & Rambow, R. (2003). Wissenskommunikation über Fächergrenzen: ein Trainingsprogramm. *Wirtschaftspsychologie*, 5(3), 94–102.
- Bromme, R., Jucks, R., & Rambow, R. (2004). Experten-Laien-Kommunikation im Wissensmanagement. In G. Reinmann & H. Mandl (Eds.), *Der Mensch im Wissensmanagement: Psychologische Konzepte zum besseren Verständnis und Umgang mit Wissen* (pp. 176–188). Göttingen: Hogrefe.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research. Methodology in the Social Sciences*. New York: The Guilford Press.

- Bruder, S. (2011). *Lernberatung in der Schule. Ein zentraler Bereich professionellen Lehrerhandelns* (Dissertation). Technische Universität Darmstadt, Darmstadt.
- Bruder, S., Keller, S., Klug, J., & Schmitz, B. (2011). Ein Vergleich situativer Methoden zur Erfassung der Beratungskompetenz von Lehrkräften. *Unterrichtswissenschaft*, 39(2), 123–137.
- Bruder, S., Klug, J., Hertel, S., & Schmitz, B. (2010). Modellierung der Beratungskompetenz von Lehrkräften: Projekt Beratungskompetenz. In E. Klieme, D. Leutner, & M. Kenk (Eds.), *Zeitschrift für Pädagogik. Beiheft: Vol. 56. Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes* (pp. 274–285). Weinheim [u.a.]: Beltz.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3rd ed.). *Psychologie*. München, Boston [u.a.]: Pearson Studium.
- Bundesministerium der Justiz und für Verbraucherschutz. (1949). *Basic Law for the Federal Republic of Germany*. Retrieved from http://www.gesetze-im-internet.de/englisch_gg/englisch_gg.html#p0034
- Bundesministerium für Familie, Senioren, Frauen und Jugend. (2010). *Familienreport 2010: Leistungen, Wirkungen, Trends*. Retrieved from <http://www.bmfsfj.de/RedaktionBMFSFJ/Broschuerenstelle/Pdf-Anlagen/familienreport-2010,property=pdf,bereich=bmfsfj,sprache=de,rwb=true.pdf>
- Bundesverfassungsgericht. (1972). *Az 1 BvR 230/70, 1 BvR 95/71*.
- Busse, S., & Helsper, W. (2007). Familie und Schule. In J. Ecarus (Ed.), *Handbuch Familie* (pp. 322–341). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Cardinet, J., Johnson, S., & Pini, G. (2010). *Applying generalizability theory using EduG. Quantitative methodology series*. New York [u.a.]: Routledge.
- Caspe, M., Lopez, M. E., Chu, A., & Weiss, H. B. (2011). *Teaching the Teacher: Preparing Educators to Engage Families for Student Achievement* (Harvard Family Research Project Issue Brief). Retrieved from <http://www.hfrp.org/publications-resources/browse-our-publications/teaching-the-teachers-preparing-educators-to-engage-families-for-student-achievement>
- Chesser, A., Cameron, H., Evans, P., Cleland, J., Boursicot, K., & Mires, G. (2009). Sources of variation in performance on a shared OSCE station across four UK medical schools. *Medical Education*, 43(6), 526–532. doi:10.1111/j.1365-2923.2009.03370.x
- Chomsky, N. (2006). *Language and mind* (3rd ed.). Cambridge, New York: Cambridge University Press.
- Chrispeels, J., & Coleman, P. (1996). Improving Schools Through Better Home-School Partnerships. *School Effectiveness and School Improvement*, 7(4), 291–296.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290.
- Clark, H. H. (1996). *Using language*. Cambridge, MA: Cambridge University Press.
- Clausen, M., Reusser, K., & Klieme, E. (2003). Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen: Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31(2), 122–141.
- Cleland, J. A., Abe, K., & Rethans, J.-J. (2009). The use of simulated patients in medical education: AMEE Guide No 421. *Medical Teacher*, 31(6), 477–486. doi:10.1080/01421590903002821

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: US Government Printing Office.
- Collins, J. P., & Harden, R. M. (1998). AMEE Medical Education Guide No. 13: real patients, simulated patients and simulators in clinical examinations. *Medical Teacher*, 20(6), 508–521.
- Cox, D. D. (2005). Evidence based interventions using home school collaboration. *School Psychology Quarterly*, 20(4), 473–497.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.
- Department for children, schools and families. (2008). *The Impact of Parental Involvement on Children's Education*. Retrieved from http://webarchive.nationalarchives.gov.uk/20130401151715/https://www.education.gov.uk/publications/eOrderingDownload/DCSF-Parental_Involvement.pdf
- Desforges, C. & Abouchaar, A. (2003). *The Impact of Parental Involvement, Parental Support and Family Education on Pupil Achievement and Adjustment: A Literature Review* (DfES Research Report No. 433). Retrieved from bgfl.org/bgfl/custom/files_uploaded/uploaded_resources/18617/Desforges.pdf
- Djakovic, S.-K., & Hertel, S. (2013, March). *Wie gut können Lehrkräfte im Rahmen von Fortbildungen erworbene Beratungskompetenzen im Schulalltag umsetzen?* 1. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF), Kiel.
- Donsbach, W. (Ed.). (2008). *The international encyclopedia of communication*. Malden, MA: Blackwell Pub.
- Dotger, B. H. (2010). "I had no idea": Developing dispositional awareness and sensitivity through a cross-professional pedagogy. *Teaching and Teacher Education*, 26(4), 805–812. doi:10.1016/j.tate.2009.10.017
- Dotger, B. H. (2011a). From know how to do now: Instructional applications of simulated interactions within teacher education. *Teacher Education and Practice*, 24(2), 132–148.
- Dotger, B. H. (2011b). The School Leader Communication Model: An emerging method for bridging school leader preparation and practice. *Journal of School Leadership*, 21(6), 871–892.
- Dotger, B. H. (2015). Core Pedagogy: Individual Uncertainty, Shared Practice, Formative Ethos. *Journal of Teacher Education*. doi:10.1177/0022487115570093
- Dotger, B. H., & Ashby, C. (2010). Exposing conditional inclusive ideologies through simulated interactions. *Teacher Education and Special Education: The Journal of the Teacher Education Division of the Council for Exceptional Children*, 33(2), 114–130. doi:10.1177/0888406409357541
- Dotger, B. H., & Sapon-Shevin, M. (2009). "But what do I say?": New teachers learn to talk with parents. *Educational Leadership*, 66(9). Retrieved from <http://www.ascd.org/publications/educational-leadership/summer09/vol66/num09/But-What-Do-I-Say.aspx>
- Dotger, B. H., & Smith, M. J. (2009). "Where's the line?" - Negotiating simulated experiences to define teacher identity. *The New Educator*, 5(2), 161–180.

- Dotger, B. H., Dotger, S. C., & Maher, M. J. (2010). From Medicine to Teaching: The Evolution of the Simulated Interaction Model. *Innovative Higher Education*, 35(3), 129–141. doi:10.1007/s10755-009-9128-x
- Dotger, B. H., Harris, S., & Hansel, A. (2008). Emerging authenticity: the crafting of simulated parent–teacher candidate conferences. *Teaching Education*, 19(4), 337–349.
- Dotger, B. H., Harris, S., Maher, M., & Hansel, A. (2011). Exploring the emotional geographies of parent-teacher candidate interactions: An emerging signature pedagogy. *The Teacher Educator*, 46(3), 208–230. doi:10.1080/08878730.2011.581882
- Dotger, B., Masingila, J., Bearkland, M., & Dotger, S. (2014). Exploring iconic interpretation and mathematics teacher development through clinical simulations. *Journal of Mathematics Teacher Education*. doi:10.1007/s10857-014-9290-7
- Dotger, S., Dotger, B. H., & Tillotson, J. (2010). Examining how preservice science teachers navigate simulated parent-teacher conversations on evolution and intelligent design. *Science Education*, 94(3), 552–570. doi:10.1002/sce.20375
- Dudenverlag. (2014). *Duden*. Retrieved from <http://www.duden.de/>
- Eccles, J. S., & Harold, R. D. (1993). Parent-school involvement during the early adolescent years. *Teacher College Record*, 94(3), 568–587.
- Edgcumbe, D. P., Silverman, J., & Benson, J. (2012). An examination of the validity of EPSCALE using factor analysis. *Patient Education and Counseling*, 87(1), 120–124. doi:10.1016/j.pec.2011.07.011
- Eid, M., & Diener, E. (Eds.). (2006). *Handbook of Multimethod Measurement in Psychology*. Washington, DC: American Psychological Association.
- Elwyn, G., Hutchings, H., Edwards, A., Rapport, F., Wensing, M., Cheung, W.-Y., & Grol, R. (2005). The OPTION scale: measuring the extent that clinicians involve patients in decision-making tasks. *Health Expectations*, 8(1), 34–42.
- Epstein, J. L. (1990). School and Family Connections: Theory, Research, and Implications for Integrating Sociologies of Education and Families. *Marriage & Family Review*, 15(1-2), 99–126. doi:10.1300/J002v15n01_06
- Epstein, J. L., & Sheldon, S. (2006). Moving Forward: Ideas for Research on School, Family, and Community Partnerships. In C. F. Conrad & R. C. Serlin (Eds.), *Handbook for research in education. Engaging ideas and enriching inquiry* (pp. 117–138). Thousand Oaks, CA: Sage Publications.
- Ericsson, K. A., Charness, N., Feltovich, P. J., & Hoffman, R. R. (Eds.). (2006). *The Cambridge Handbook of Expertise and Expert Performance (Cambridge Handbooks in Psychology)*. Cambridge, New York: Cambridge University Press.
- Erlendsdóttir, G. (2010). *Effects of Parental Involvement in Education. A Case Study in Namibia* (M.Ed. Thesis). University of Iceland, Reykjavík. Retrieved from http://skemman.is/stream/get/1946/6925/18121/1/MEd_thesis-Gu%C3%B0laug.pdf
- Erpenbeck, J., & von Rosenstiel, L. (2003). *Handbuch Kompetenzmessung: Erkennen, Verstehen und Bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis*. Stuttgart: Schäffer-Poeschel.
- Fan, W., & Williams, C. M. (2010). The effects of parental involvement on students' academic self-efficacy, engagement and intrinsic motivation. *Educational Psychology*, 30(1), 53–74. Retrieved from 10.1080/01443410903353302
- Flanigan, C. B. (2005). *Partnering With Parents and Communities: Are Preservice Teachers Adequately Prepared?* (Harvard Family Research Project). Retrieved from

- <http://www.hfrp.org/publications-resources/browse-our-publications/partnering-with-parents-and-communities-are-preservice-teachers-adequately-prepared>
- Franke, G. (2005). *Facetten der Kompetenzentwicklung. Schriftenreihe des Bundesinstituts für Berufsbildung Bonn*. Bielefeld: Bertelsmann.
- Frey, A. (2014, March). *Validität: Internationaler Forschungsstand und Umsetzung in Deutschland*. 2. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF), Frankfurt am Main.
- Frey, A., & Balzer, L. (2005). Der Beurteilungsbogen smk: Ein Messverfahren für die Diagnose von sozialen und methodischen Fähigkeitskonzepten. In A. Frey, R. S. Jäger, & U. Renold (Eds.), *Kompetenzdiagnostik. Theorien und Methoden zur Erfassung und Bewertung von beruflichen Kompetenzen* (2nd ed., pp. 31–56). Landau: Verlag Empirische Pädagogik; Verl. Empirische Pädag.
- Freyaldenhoven, I. (2005). *Schule in der Krise? Psychologische Beratung als Antwort: Theoretische und praktische Hinweise für eine gelingende Schulberatung - lösungsorientiert und individuell*. Stuttgart: Ibidem-Verlag.
- Friedman, I. A. (2003). Self-efficacy and burnout in teaching: the importance of interpersonal-relations efficacy. *Social Psychology of Education*, 6(3), 191–215.
- Friedrich, G. (2002). Gesprächsführung - Ausbildungsziel der Lehrerqualifikation. In G. Brüner, R. Fiehler, & W. Kindt (Eds.), *Methoden und Anwendungsgebiete: Vol. 2. Angewandte Diskursforschung* (pp. 126–147). Radolfzell: Verlag für Gesprächsforschung.
- Fuchs, L. S., & Fuchs, D. (2000). Analogue assessment of academic skills: Curriculum-based measurement and performance assessment. In E. S. Shapiro & T. R. Kratochwill (Eds.), *Behavioral assessment in schools* (2nd ed., pp. 168–201). New York.
- Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An introduction*. Los Angeles: Sage Publications.
- Gaba, D. M. (2004). The future vision of simulation in health care. *Quality and Safety in Health Care*, 13(suppl_1), i2–10. doi:10.1136/qshc.2004.009878
- Gallagher, M. S., & Hargie, O. D. W. (1989). An investigation into the validity of role play as a procedure for counsellor skill assessment. *British Journal of Guidance & Counselling*, 17(2), 155–165. doi:10.1080/03069888908258050
- Gartmeier, M., Bauer, J., Fischer, M. R., Karsten, G., & Prenzel, M. (2011). Modellierung und Assessment professioneller Gesprächsführungskompetenz von Lehrpersonen im Lehrer-Elterngespräch. In O. Zlatkin-Troitschanskaia (Ed.), *Stationen Empirischer Bildungsforschung. Traditionslinien und Perspektiven* (pp. 412–426). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gartmeier, M., Bauer, J., Fischer, M., Hoppe-Seyler, T., Karsten, G., Kiessling, C., Möller, G., Wiesbeck, A., & Prenzel, M. (2015). Fostering professional communication skills of future physicians and teachers: effects of e-learning with video cases and role-play. *Instructional Science*, 43(4), 443–462. doi:10.1007/s11251-014-9341-6
- Gartmeier, M., Bauer, J., Noll, A., & Prenzel, M. (2012). Welchen Problemen begegnen Lehrkräfte beim Führen von Elterngesprächen? Und welche Schlussfolgerungen ergeben sich daraus für die Vermittlung von Gesprächsführungskompetenz? *Die Deutsche Schule*, 104(4), 374–382.
- Gartmeier, M., Gebhardt, M., & Dotger, B. H. (in preparation). Teachers' skills of communicating with parents: Latent classes and relationships to workplace behaviors.
- Görlitz, A., Bachmann, C., Blum, K., Höfer, S., Peters, T., Preusche, I., Raski, B., Rüttermann, S., Wagner Menghin, M., & Kiessling, C. (2014, September). *Lehren und Prüfen kommunikativer Kompetenzen im Medizinstudium - Ergebnisse einer Umfrage im*

- deutschsprachigen Raum*. Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA), Hamburg.
- Greene, J. O., & Burlison, B. R. (Eds.). (2003). *Handbook of Communication and Social Interaction Skills*. Mahwah: Lawrence Erlbaum Associates.
- Grewe, N. (2005). Gesprächsführung und Leitlinien der Beratung. In N. Grewe (Ed.), *Praxishilfen Schule. Praxishandbuch Beratung in der Schule. Grundlagen, Aufgaben und Fallbeispiele* (pp. 13–34). Neuwied: Luchterhand.
- Guiton, G., Hodgson, C. S., Delandshere, G., & Wilkerson, L. (2004). Communication skills in standardized patient assessment of final-year medical students: a psychometric study. *Advances in Health Sciences Education*, 9(3), 179–187. doi:10.1023/B:AHSE.0000038174.87790.7b
- Guttormsen, S., Beyeler, C., Bonvin, R., Feller, S., Schirlo, C., Schnabel, K., Schurter, T., & Berendonk, C. (2013). The new licencing examination for human medicine: from concept to implementation. *Swiss Medical Weekly*, 143:w13897. doi:10.4414/smw.2013.13897
- Habermas, J. (1984). *The theory of communicative action*. Cambridge, Boston: Polity; Beacon Press.
- Harden, R. M., Stevenson, M., Downie, W. W., & Wilson, G. M. (1975). Assessment of clinical competence using objective structured examination. *British Medical Journal*, 1(5955), 447–451. doi:10.1136/bmj.1.5955.447
- Harper, D. (2014). *Online Etymology Dictionary*. Retrieved from <http://etymonline.com/>
- Hartig, J. (2008). Kompetenzen als Ergebnisse von Bildungsprozessen. In N. Jude, J. Hartig, & E. Klieme (Eds.), *Kompetenzerfassung in pädagogischen Handlungsfeldern. Theorien, Konzepte und Methoden. Bildungsforschung Band 26* (pp. 15–25). Bonn, Berlin.
- Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Ed.), *Leistung und Leistungsdiagnostik* (pp. 127–143). Heidelberg: Springer.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487
- Henderson, A. T. & Mapp, K. L. (2002). *A New Wave of Evidence: The Impact of School, Family, and Community Connections on Student Achievement*. Annual Synthesis National Center for Family & Community Connections with Schools. Austin, Texas. Retrieved from <http://www.sedl.org/connections/resources/evidence.pdf>
- Hertel, S. (2009). *Beratungskompetenz von Lehrern: Kompetenzdiagnostik, Kompetenzförderung, Kompetenzmodellierung. Pädagogische Psychologie und Entwicklungspsychologie: Vol. 74*. Münster: Waxmann.
- Hertel, S., Bruder, S., & Schmitz, B. (2009). Beratungs- und Gesprächsführungskompetenz von Lehrkräften. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus, & R. Mulder (Eds.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (pp. 117–128). Weinheim [u.a.]: Beltz.
- Hertel, S., Bruder, S., Jude, N., & Steinert, B. (2013). Elternberatung an Schulen im Sekundarbereich. Schulische Rahmenbedingungen, Beratungsangebote der Lehrkräfte und Nutzung von Beratung durch die Eltern. In N. Jude & E. Klieme (Eds.), *PISA 2009 - Impulse für die Schul- und Unterrichtsforschung. Zeitschrift für Pädagogik. 59. Beiheft* (pp. 40–62). Weinheim u.a.: Beltz.
- Hertel, S., Hemker, N., Larcher, S., Kerwer, M., Djakovic, S.-K., & Abrie-Kuhn, M. (2014, May). *Projekt elbe: Besonderheiten in der Beratung von Eltern mit Migrations-*

- hintergrund im Schulalltag. Lehrer-Schüler-Eltern-Gespräche. Beiträge qualitativ-empirischer Forschung, Koblenz.*
- Hill, N. E., Castellino, D. R., Lansford, J. E., Nowlin, P., Dodge, K. A., Bates, J. E., & Pettit, G. S. (2004). Parent Academic Involvement as Related to School Behavior, Achievement, and Aspirations: Demographic Variations Across Adolescence. *Child Development, 75*(5), 1491–1509. doi:10.1111/j.1467-8624.2004.00753.x
- Hirsch, O., Keller, H., Müller-Engelmann, M., Heinzl-Gutenbrunner, M., Krones, T., & Donner-Banzhoff, N. (2012). Reliability and validity of the German version of the OPTION scale. *Health Expectations, 15*(4), 379–388. doi:10.1111/j.1369-7625.2011.00689.x
- Hitziger, H. (1987). *Soziale Kompetenz des Lehrers in der Elternarbeit: Analyse und Bewertung eines problemorientierten gruppendynamischen Lehrertrainings zur Vermittlung sozialer Kompetenz in der Lehrerfortbildung*. Frankfurt a. M: Haag & Herchen.
- Hodges, B., & McIlroy, J. H. (2003). Analytic global OSCE ratings are sensitive to level of training. *Medical Education, 37*(11), 1012–1016. doi:10.1046/j.1365-2923.2003.01674.x
- Honal, W. H., & Schlegel, H. (2002). Axiome der Schulberatung – Wissenschaftliche Vorgaben für die Beratung in der Schule. In W. H. Honal (Ed.), *Handbuch der Schulberatung* (pp. 1–4). Landsberg: mvg-Verlag.
- Hui, L., & Nirmala, R. (2000). Parental influences on Chinese literacy development: A comparison of preschoolers in Beijing, Hong Kong, and Singapore. *International Journal of Behavioral Development, 24*(1), 82–90. doi:10.1080/016502500383502
- Huntley, C. D., Salmon, P., Fisher, P. L., Fletcher, I., & Young, B. (2012). LUCAS: a theoretically informed instrument to assess clinical communication in objective structured clinical examinations. *Medical Education, 46*(3), 267–276. doi:10.1111/j.1365-2923.2011.04162.x
- Iramaneerat, C., Yudkowsky, R., Myford, C. M., & Downing, S. M. (2008). Quality control of an OSCE using generalizability theory and many-faceted Rasch measurement. *Advances in Health Sciences Education, 13*(4), 479–493. doi:10.1007/s10459-007-9060-8
- Jäger-Flor, D. & Jäger, R. S. (2009). *Bildungsbarometer zur Kooperation Elternhaus-Schule: Ergebnisse, Bewertungen und Perspektiven*. Retrieved from http://vep-landau.de/Bildungsbarometer/Bildungsbarometer_2009_4.pdf
- Jahn, G. (2014). *Studien zur Überprüfung der Validität eines Instruments zur Erfassung professioneller Unterrichtswahrnehmung von Lehramtsstudierenden* (Dissertation). Technische Universität München, München.
- Jensen, E., & Jensen, H. (2008). *Dialog mit Eltern: Gelungene Lehrer-Elterngespräche*. München: Voelchert.
- Jeynes, W. (2011). *Parental Involvement and Academic Success*. New York: Routledge.
- Jeynes, W. H. (2007). The Relationship Between Parental Involvement and Urban Secondary School Student Academic Achievement: A Meta-Analysis. *Urban Education, 42*(1), 82–110. doi:10.1177/0042085906293818
- Johns, K. M. (1992). Lowering beginning teacher anxiety about parent-teacher-conferences through role-playing. *School Counselor, 40*(2), 146–152.
- Jucks, R. (2001). *Was verstehen Laien? Die Verständlichkeit von Fachtexten aus der Sicht von Computer-Experten. Internationale Hochschulschriften: Vol. 377*. Münster: Waxmann.

- Jurczyk, K., & Klinkhardt, J. (2014). *Vater, Mutter, Kind? Acht Trends in Familien, die Politik heute kennen sollte*. Gütersloh: Bertelsmann Stiftung.
- Kane, M. T. (2001). Current Concerns in Validity Theory. *Journal of Educational Measurement*, 38(4), 319–342. doi:10.1111/j.1745-3984.2001.tb01130.x
- Kauffeld, S. (2005). Fachliche und überfachliche Weiterbildung: Welche Investitionen zahlen sich für die berufliche Handlungskompetenz aus? In A. Frey, R. S. Jäger, & U. Renold (Eds.), *Kompetenzdiagnostik. Theorien und Methoden zur Erfassung und Bewertung von beruflichen Kompetenzen* (2nd ed., pp. 57–75). Landau: Verlag Empirische Pädagogik; Verl. Empirische Pädag.
- Keller, H., Hirsch, O., Müller-Engelmann, M., Heinzl-Gutenbrunner, M., Krones, T., & Donner-Banzhoff, N. (2013). Trying to Optimise the German Version of the OPTION Scale Regarding the Dyadic Aspect of Shared Decision Making. *Methods of Information in Medicine*, 52(6), 514–521. doi:10.3414/ME13-01-0011
- Kempen, A. C. (2008). *Gespräche zwischen Lehrern und Eltern. Eine Einschätzung von Voraussetzungen, Schwierigkeiten und Chancen*. (Unveröffentlichte Diplomarbeit). Westfälische Wilhelms-Universität Münster, Münster.
- Kiessling, C., Bachmann, C., Pruskil, S., & Fischer, M. R. (2013, September). *Übersetzung und erste Validierung der deutschsprachigen Version der "Explanation and Planning Scale"*. Jahrestagung der Gesellschaft für Medizinische Ausbildung (GMA), Graz, Österreich.
- Killus, D., & Tillmann, K.-J. (Eds.). (2012). *Eltern ziehen Bilanz: Ein Trendbericht zu Schule und Bildungspolitik in Deutschland*. 2. JAKO-O Bildungsstudie. Münster: Waxmann.
- Kleinknecht, M., & Schneider, J. (2013). What do teachers think and feel when analyzing videos of themselves and other teachers teaching? *Teaching and Teacher Education*, 33, 13–23. doi:10.1016/j.tate.2013.02.002
- Klieme, E., & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In M. Prenzel, I. Gogolin, & H.-H. Krüger (Eds.), *Kompetenzdiagnostik* (pp. 11–29). VS Verlag für Sozialwissenschaften.
- Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In E. Klieme, D. Leutner, & J. Hartig (Eds.), *Assessment of competencies in educational contexts* (pp. 3–22). Toronto: Hogrefe & Huber Publishers.
- Kluger, A. N., & DeNisi, A. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, 119(2), 254–284.
- Koepfen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current Issues in Competence Modeling and Assessment. *Zeitschrift für Psychologie / Journal of Psychology*, 216(2), 61–73. doi:10.1027/0044-3409.216.2.61
- Kohl, G. O., Lengua, L. J., & McMahon, R. J. (2000). Parent Involvement in School Conceptualizing Multiple Dimensions and Their Relations with Family and Demographic Risk Factors. *Journal of School Psychology*, 38(6), 501–523.
- Kreider, H., Caspe, M., Kennedy, S., & Weiss, H. (2007). *Family involvement in middle and high school students' education* (Harvard Family Research Project: Family involvement makes a difference No. 3). Retrieved from <http://www.hfrp.org/publications-resources/publications-series/family-involvement-makes-a-difference/family-involvement-in-middle-and-high-school-students-education>
- Krumm, V. (1996). Über die Vernachlässigung der Eltern durch die Erziehungswissenschaft: Plädoyer für eine veränderte Rolle der Lehrer bei der Erziehung der Kinder. In

- A. Leschinsky (Ed.), *Die Institutionalisierung von Lehren und Lernen. Beiträge zu einer Theorie der Schule* (pp. 119–137). Weinheim: Beltz.
- Kubinger, K. D., & Proyer, R. (2005). Gütekriterien. In K. Westhoff, L. J. Hellfritsch, L. F. Hornke, K. D. Kubinger, F. Lang, H. Moosbrugger, A. Püschel, & G. Reimann (Eds.), *Grundwissen für die berufsbezogene Eignungsbeurteilung nach DIN 33430* (2nd ed., pp. 191–199). Lengerich [u.a.]: Pabst Science Publishers; Pabst.
- Kultusministerkonferenz (2003). *Erziehung als Auftrag von Elternhaus und Schule. Informationen der Länder über die Zusammenarbeit von Eltern und Schule*. Retrieved from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2003/2003_12_04-Elternhaus-Schule.pdf
- Kultusministerkonferenz. (2004). *Standards für die Lehrerbildung: Bildungswissenschaften*. Retrieved from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf
- Kunter, M., & Klusmann, U. (2010). Kompetenzmessung von Lehrkräften - Methodische Herausforderungen. *Unterrichtswissenschaft*, 38(1), 68–86.
- Kurtz, S., Silverman, J., Benson, J., & Draper, J. (2003). Marrying Content and Process in Clinical Method Teaching: Enhancing the Calgary-Cambridge Guides. *Academic Medicine*, 78(8), 802–809.
- Kyriakides, L., & Charalambous, C. (2014, July). *Using Structural Equation Modeling Techniques in Educational Research*. Workshop at the 18th Conference of the Junior Researchers of EARLI, Nicosia, Cyprus.
- Lane, C., & Rollnick, S. (2007). The use of simulated patients and role-play in communication skills training: A review of the literature to August 2005. *Patient Education and Counseling*, 67(1-2), 13–20. doi:10.1016/j.pec.2007.02.011
- Lang, F., McCord, R., Harvill, L., & Anderson, D. S. (2004). Communication Assessment Using the Common Ground Instrument: Psychometric Properties. *Family Medicine*, 36(3), 189–198.
- Langer, I., & Schulz von Thun, F. (2007). Messung komplexer Merkmale in Psychologie und Pädagogik: Ratingverfahren. In D. H. Rost (Ed.), *Standardwerke aus Psychologie und Pädagogik - Reprints* (Vol. 4). Münster: Waxmann.
- Lawrence-Lightfoot, S. (2004). *The essential conversation: What parents and teachers can learn from each other*. New York: Random House Publishing Group; Random House Pub. Group.
- LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 Questions About Interrater Reliability and Interrater Agreement. *Organizational Research Methods*, 11(4), 815–852. doi:10.1177/1094428106296642
- Lurie, S. J., Mooney, C. J., Nofziger, A. C., Meldrum, S. C., & Epstein, R. M. (2008). Further challenges in measuring communication skills: accounting for actor effects in standardised patient assessments. *Medical Education*, 42(7), 662–668. doi:10.1111/j.1365-2923.2008.03080.x
- Maag Merki, K., & Werner, S. (2011). Erfassung und Bewertung professioneller Kompetenz von Lehrpersonen. In E. Terhart, H. Bennewitz, & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (pp. 573–591). Münster: Waxmann.
- Makoul, G. (2001a). Essential Elements of Communication in Medical Encounters: The Kalamazoo Consensus Statement. *Academic Medicine*, 76(4), 390–393.

- Makoul, G. (2001b). The SEGUE Framework for teaching and assessing communication skills. *Patient Education and Counseling*, 45(1), 23–34. doi:10.1016/S0738-3991(01)00136-7
- Manz, P. H., Fantuzzo, J. W., & Power, T. J. (2004). Multidimensional assessment of family involvement among urban elementary students. *Journal of School Psychology*, 42(6), 461–475. doi:10.1016/j.jsp.2004.08.002
- Markow, D. & Pieters, A. (2010). *The Metlife Survey of the American Teacher: Collaborating for Student Success*. New York. Retrieved from https://www.metlife.com/assets/cao/contributions/foundation/american-teacher/MetLife_Teacher_Survey_2009.pdf
- May, W., Park, J. H., & Lee, J. P. (2009). A ten-year review of the literature on the use of standardized patients in teaching and learning: 1996–2005. *Medical Teacher*, 31(6), 487–492. doi:10.1080/01421590802530898
- Mc Graw, K. O., & Wong, S. (1996). Forming Inferences About Some Intraclass Correlation Coefficients. *Psychological Methods*, 1(1), 30–46.
- McLeod, J. (2003). *An Introduction to Counselling* (3rd ed.). Berkshire: Open University Press.
- Merriam-Webster. (2014). *Merriam-Webster Dictionary*. Retrieved from <http://www.merriam-webster.com/dictionary>
- Merten, K. (1977). *Kommunikation: Eine Begriffs- und Prozeßanalyse. Studien zur Sozialwissenschaft: Vol. 35*. Opladen: Westdt. Verlag.
- Müller, K., & Ehmke, T. (2013). Soziale Herkunft als Bedingung der Kompetenzentwicklung. In M. Prenzel, C. Sälzer, E. Klieme, & O. Köller (Eds.), *PISA 2012. Fortschritte und Herausforderungen in Deutschland* (pp. 245–274). Münster.
- Muthén, L. K. & Muthén, B. O. (1998-2013). *Mplus Version 7.11*. Los Angeles, CA: Muthén & Muthén.
- Newble, D. (2004). Techniques for measuring clinical competence: objective structured clinical examinations. *Medical Education*, 38(2), 199–203. doi:10.1111/j.1365-2923.2004.01755.x
- Ortwein, H., Fröhmel, A., & Burger, W. (2006). Einsatz von Simulationspatienten als Lehr-, Lern- und Prüfungsform. *PPmP - Psychotherapie · Psychosomatik · Medizinische Psychologie*, 56(1), 23–29. doi:10.1055/s-2005-867058
- Papadakis, M. A., Croughan-Minihane, M., Fromm, L. J., Wilkie, H. A., & Ernster, V. L. (1997). A Comparison of Two Methods to Teach Smoking-cessation Techniques to Medical Students. *Academic Medicine*, 72(8), 725–727.
- Prenzel, M. (2009). Von der Unterrichtsforschung zur Exzellenz in der Lehrerbildung. *Beiträge zur Lehrerbildung*, 27(3), 327–345.
- Prenzel, M. (2013). Initiativen und Perspektiven zur Weiterentwicklung der Lehrerbildung. In W. Benz, J. Kohler, & K. Landfried (Eds.), *Handbuch Qualität in Studium und Lehre. Evaluation nutzen - Akkreditierung sichern - Profile schärfen* (pp. 1–21). Berlin: Raabe Verlag.
- Prenzel, M., Artelt, C., Baumert, J., Blum, W., Hammann, M., Klieme, E., & Pekrun, R. (Eds.). (2007). *PISA 2006: Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Prenzel, M., Reiss, K., & Seidel, T. (2011). Lehrerbildung an der TUM School of Education. *Erziehungswissenschaft*, 22(43), 47–56.

- Prochaska, J. O., & Norcross, J. C. (2010). *Systems of psychotherapy: A transtheoretical analysis* (7th ed., International student ed.). Belmont: Brooks/Cole, Cengage Learning.
- Putnam, R. T., & Borko, H. (2000). What Do New Views of Knowledge and Thinking Have to Say About Research on Teacher Learning? *Educational Researcher*, 29(1), 4–15. doi:10.3102/0013189X029001004
- Rammstedt, B. (2010). Reliabilität, Validität, Objektivität. In C. Wolf & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 239–258). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2006a). *Quantitative Methoden 2: Einführung in die Statistik* (2nd ed.). Heidelberg: Springer.
- Rasch, B., Friese, M., Hofmann, W., & Naumann, E. (2006b). *Quantitative Methoden. Band 2: SPSS-Ergänzungen*. Retrieved from http://quantitative-methoden.de/Dateien/Auflage2/BdII_Dateien/Kapitel_5_SPSS_Ergaenzungen.pdf
- Raykov, T., & Marcoulides, G. A. (2011). *Introduction to Psychometric Theory*. New York: Routledge.
- Reder, L., Anderson, J. R., & Simon, H. A. (1996). Situated Learning and Education. *Educational Researcher*, 25(4), 5–11. doi:10.3102/0013189X025004005
- Regehr, G., MacRae, H., Reznick, R. K., & Szalay, D. (1998). Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine*, 73(9), 993–997.
- Rethans, J.-J., Gorter, S., Bokken, L., & Morrison, L. (2007). Unannounced standardised patients in real practice: a systematic literature review. *Medical Education*, 41(6), 537–549. doi:10.1111/j.1365-2929.2006.02689.x
- Rogers. (1951). *Client-centered Therapy: Its Current Practice, Implications and Theory*. London: Constable.
- Ruthemann, U. (2004). Die Psychologie in der Lehrerbildung zwischen Berufsfeld- und Wissenschaftsorientierung. *Beiträge zur Lehrerbildung*, 22(3), 353–361.
- Sacher, W. (2009). Elternarbeit - Partnerschaft zwischen Schule und Familie. In S. Blömeke, T. Bohl, L. Haag, G. Lang-Wojtasik, & W. Sacher (Eds.), *UTB Schulpädagogik: Vol. 8392. Handbuch Schule. Theorie - Organisation - Entwicklung* (pp. 519–526). Bad Heilbrunn: Klinkhardt [u.a.].
- Sanson-Fisher, R. W., & Poole, A. D. (1980). Simulated patients and the assessment of medical students' interpersonal skills. *Medical Education*, 14(4), 249–253. doi:10.1111/j.1365-2923.1980.tb02269.x
- Schäfer, K.-H. (2005). *Kommunikation und Interaktion: Grundbegriffe einer Pädagogik des Pragmatismus*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Scheffer, S. (2009). *Validierung des „Berliner Global Rating“ (BGR) - ein Instrument zur Prüfung kommunikativer Kompetenzen Medizinstudierender im Rahmen klinisch-praktischer Prüfungen (OSCE)* (Dissertation). Charité - Universitätsmedizin Berlin, Berlin.
- Scheibe, A.-K., Trittel, M., Klug, J., & Schmitz, B. (2014). Forschungsmethoden. In T. Seidel & A. Krapp (Eds.), *Pädagogische Psychologie. Mit Online-Materialien zum Download* (6th ed., pp. 83–106). Weinheim [u.a.]: Beltz.
- Scheiderer, C. (2013). *Messung der Gesprächsführungskompetenz angehender Lehrkräfte in Elterngesprächen durch simulierte Gesprächspartner. Erstellung und Überprüfung*

- der Reliabilität eines Kategoriensystems zur Messung der konstanten Umsetzung der Fallvignette durch simulierte Gesprächspartner* (Bachelorarbeit). TU München, München.
- Schirmer, J. M., Mauksch, L., Lang, F., Marvel, M. K., Zoppi, K., Epstein, R. M., Brock, D., & Pryzbylski, M. (2005). Assessing Communication Competence: A Review of Current Tools. *Family Medicine*, 37(3), 184–192.
- Schmitt, M. (2006). Conceptual, theoretical, and historical foundations of multimethod assessment. In M. Eid & E. Diener (Eds.), *Handbook of Multimethod Measurement in Psychology* (pp. 9–26). Washington, DC: American Psychological Association.
- Schnebel, S. (2007). *Professionell beraten: Beratungskompetenz in der Schule. Studententexte für das Lehramt: Vol. 20*. Weinheim; Basel: Beltz.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review. *Journal of Educational Research*, 99(6), 323–338.
- Schulz von Thun, F. (1998). *Miteinander Reden I: Störungen und Klärungen. Allgemeine Psychologie der Kommunikation*. Reinbek: Rowohlt.
- Schuwirth, L. W. T., & van der Vleuten, C. P. M. (2011). Programmatic assessment: From assessment of learning to assessment for learning. *Medical Teacher*, 33(6), 478–485. doi:10.3109/0142159X.2011.565828
- Schwarzer, C., & Buchwald, P. (2006). Beratung in Familie, Schule und Beruf. In A. Krapp & B. Weidenmann (Eds.), *Pädagogische Psychologie. Ein Lehrbuch* (5th ed., pp. 575–612). Weinheim [u.a.]: Beltz.
- Scofield, M. E., & Yoxtheimer, L. L. (1983). Psychometric Issues in the Assessment of Clinical Competencies. *Journal of Counseling Psychology*, 30(3), 413–420.
- Searle, J. R. (1962). *Speech acts: An essay in the philosophy of language*. New York: Cambridge University Press.
- Seidel, T., Prenzel, M., & Kobarg, M. (Eds.). (2005). *How to run a video study: Technical report of the IPN Video Study*. Münster: Waxmann.
- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. Urbana, Chicago, London: University of Illinois Press.
- Shartrand, A. M., Weiss, H. B., Kreider, H. M., & Lopez, M. E. (1997). *New skills for new schools: Preparing teachers in family involvement* (Harvard Family Research Project). Cambridge, MA. Retrieved from <http://www.hfrp.org/publications-resources/browse-our-publications/new-skills-for-new-schools-preparing-teachers-in-family-involvement>
- Shavelson, R. J. (2012). Assessing business-planning competence using the collegiate learning assessment as a prototype. *Empirical research in vocational education and training*, 4(1), 77–90.
- Shavelson, R. J. (2013). On an Approach to Testing and Modeling Competence. *Educational Psychologist*, 48(2), 73–86. doi:10.1080/00461520.2013.779483
- Shernoff, E. S., & Kratochwill, T. R. (2004). The Application of Behavioral Assessment Methodologies in Educational Settings. In S. N. Haynes & E. M. Heiby (Eds.), *Comprehensive handbook of psychological assessment. Volume 3. Behavioral assessment* (pp. 365–385). Hoboken, New Jersey: John Wiley & Sons.
- Silverman, J., Archer, J., Gillard, S., Howells, R., & Benson, J. (2011). Initial evaluation of EPSCALE, a rating scale that assesses the process of explanation and planning in the medical interview. *Patient Education and Counseling*, 82(1), 89–93. doi:10.1016/j.pec.2010.02.022

- Skillings, J. L., Porcerelli, J. H., & Markova, T. (2010). Contextualizing SEGUE: Evaluating Residents' Communication Skills Within the Framework of a Structured Medical Interview. *Journal of Graduate Medical Education*, 2(1), 102–107. doi:10.4300/JGME-D-09-00030.1
- Spielberg, L. (2011). Successful Family Engagement in the Classroom: What teachers need to know and be able to do to engage families in raising student achievement. *Family Involvement Network of Educators (FINE) Newsletter*, 3(1). Retrieved from http://flamboyanfoundation.org/wp/wp-content/uploads/2011/06/FINE-Flamboyan_Article-March-2011.pdf
- Spitzberg, B. H. (2008). Interpersonal Communication Competence and Social Skills. In W. Donsbach (Ed.), *The international encyclopedia of communication* (pp. 2486–2492). Malden, MA: Blackwell Pub.
- Stadtler, M. (2009). Die Psychologie des Konfliktmanagements in schulischen Beratungsgesprächen mit Eltern. *Beruf: Schulleitung*, 3(1), 19–20.
- Stokoe, E. (2013). The (In) Authenticity of Simulated Talk: Comparing Role-Played and Actual Interaction and the Implications for Communication Training. *Research on Language and Social Interaction*, 46(2), 165–185. doi:10.1080/08351813.2013.780341
- Strasser, J., & Gruber, H. (2008). Kompetenz von Beratungslehrern im Vergleich. In M. Gläser-Zikuda & J. Seifried (Eds.), *Lehrerexpertise. Analyse und Bedeutung unterrichtlichen Handelns* (pp. 237–259). Münster: Waxmann.
- Tacke, M. (1997). Haben die Belastungen für ältere Lehrerinnen zugenommen? *Die Grundschule*, (6), 37–46.
- tEACH Assessment subgroup. (2012). *Assessment Tools for Communication Skills*. Utrecht: EACH.
- Terhart, E. (2009). Erste Phase: Lehrerbildung an der Universität. In O. Zlatkin-Troitschanskaia, K. Beck, D. Sembill, R. Nickolaus, & R. Mulder (Eds.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (pp. 425–438). Weinheim [u.a.]: Beltz.
- Textor, M. R. (2009). *Bildungs- und Erziehungspartnerschaft in der Schule: Gründe, Ziele, Formen*. Norderstedt: Books on Demand.
- Uebersax, J. (2010). *Statistical Methods for Rater and Diagnostic Agreement*. Retrieved from <http://www.john-uebersax.com/stat/agree.htm>
- United States Medical Licensing Examination. (2015). *Step 2 CS*. Retrieved from <http://www.usmle.org/step-2-cs/>
- Unterbrink, T., Zimmermann, L., Pfeifer, R., Wirsching, M., Brähler, E., & Bauer, J. (2008). Parameters influencing health variables in a sample of 949 German teachers. *International Archives of Occupational and Environmental Health*, 82(1), 117–123. doi:10.1007/s00420-008-0336-y
- van der Vleuten, C. P., Schuwirth, L. W., Scheele, F., Driessen, E. W., Hodges, B., Currie, R., & Currie, E. (2010). The assessment of professional competence: building blocks for theory development. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 24(6), 703–719. doi:10.1016/j.bpobgyn.2010.04.001
- Walker, J. M., & Dotger, B. H. (2012). Because Wisdom Can't Be Told: Using Comparison of Simulated Parent-Teacher Conferences to Assess Teacher Candidates' Readiness for Family-School Partnership. *Journal of Teacher Education*, 63(1), 62–75. doi:10.1177/0022487111419300
- Watzlawick, P., Beavin, J. H., & Jackson, D. D. (2011). *Menschliche Kommunikation: Formen Störungen Paradoxien* (12th ed.). Bern: Hans Huber.

- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and Selecting Key Competencies. Theoretical and Conceptual Foundations* (pp. 45–65). Göttingen: Hogrefe.
- Weiss, H. B., Lopez, M. E., & Rosenberg, H. (2010). *Beyond Random Acts: Family, School, and Community Engagement as an Integral Part of Education Reform* (Harvard Family Research Project). Retrieved from <http://www.hfrp.org/publications-resources/browse-our-publications/beyond-random-acts-family-school-and-community-engagement-as-an-integral-part-of-education-reform>
- Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review*, *19*(3), 188–202. doi:10.1016/j.hrmr.2009.03.007
- Wild, E. (2003). Einbeziehung des Elternhauses durch Lehrer: Art, Ausmaß und Bedingungen der Elternpartizipation aus Sicht von Gymnasiallehrern. *Zeitschrift für Pädagogik*, *49*(4), 513–533.
- Wippermann, K., Wippermann, C., & Kirchner, A. (2013). *Eltern - Lehrer - Schulerfolg: Wahrnehmungen und Erfahrungen im Schulalltag von Eltern und Lehrern. Eine sozialwissenschaftliche Untersuchung der Katholischen Stiftungshochschule Benediktbeuern für die Konrad-Adenauer-Stiftung und das Bundesministerium für Familie, Senioren, Frauen und Jugend*. Stuttgart: Lucius & Lucius.
- Wirtz, M. A., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität: Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen [u.a.]: Hogrefe.
- Wissenschaftsrat. (2006). *Empfehlungen zur künftigen Rolle der Universitäten im Wissenschaftssystem*. Köln. Retrieved from http://www.wissenschaftsrat.de/download/archiv/rolle_universitaeten.pdf
- Wollert, A. (1997). Intergenerative Kompetenzbilanz. In Arbeitsgemeinschaft Qualifikations-Entwicklungs-Management Berlin (Ed.), *Kompetenzentwicklung '97. Berufliche Weiterbildung in der Transformation - Fakten und Visionen* (Vol. 2, pp. 317–362). Münster [u.a.]: Waxmann.