Dissertation

Testing the Usability of Role-Plays to Assess Performance-Based Competences in Higher Education under Consideration of FAIR Data Management

von
submitted by

Kristina Walz, M.A.

zur Erlangung des akademischen Grades Dr. phil.
for the degree of Dr. phil.

im Fach Erziehungswissenschaften
to the Department of Education Science

am Fachbereich Sozial- und Kulturwissenschaften
at the faculty of Social Sciences and Cultural Studies

der Justus-Liebig-Universität Gießen
at Justus-Liebig-University Gießen

Eingereicht am: 02.12.2022
Submission:

Verteidigt am: 29.08.2023
Thesis defense:

Gutachten von:
Referees:

1. Prof. Dr. Edith Braun, Justus-Liebig-Universität Gießen
2. Prof. Dr. Ludwig Stecher, Justus-Liebig-Universität Gießen
3. Prof. Dr. Auli Toom, University of Helsinki

ACKNOWLEDGEMENTS

This dissertation is about communication and performance competence, about performance-based assessment methods and about the sustainable handling of valuable research data. It emphasizes the importance of generic performance competence that enables individuals to react to different and changing situations. This dissertation is furthermore the result of a nearly four year long journey, which included several changing situations itself: first research experiences and international scientific exchange, a sudden fallback to complete offline working ending with a sudden U-turn towards digital and online. Along this journey I did acquire several new competences myself.

I am very happy to not have faced this journey alone. On each step I could count on the guidance, encouragement and support of friends, family members, and colleagues. First and foremost there is my doctoral supervisor Prof. Edith Braun, who did not only give me the opportunity to do research in a thriving and innovative area, but has furthermore been a steady source of advice, fruitful collaboration and a great help improving my understanding of the area significantly. Thank you, for your support, for proposing interesting problems to study, and for introducing me to the fantastic work of the KomPrü research team. Furthermore, I thank my colleagues Julia Fecke, Anastasia Falkenstern, Katharina Lohberger, and Julia Mertens for an ever-fruitful exchange of ideas and thoughts – concerning professional topics as well as cooking recipes. I will definitively miss our joint lunch breaks and digital coffee-dates. I would particularly like to thank Ulrike Schwabe for our intensive collaboration and your advice during all steps of research, presentation, and publication of our joint results.

I also want to express my gratitude to my coauthors Edith Braun, Anastasia Falkenstern, Julia Osada, and Ulrike Schwabe. The contributions of this dissertation are the result of close collaboration with these coauthors.

This dissertation has been supported by the German research program, ‘The Giessen Campaign: Teacher Education 2.0 (GOL)’ as part of the German Federal Ministry of Education and Research (BMBF), under Grant 01JA1929.

ZUSAMMENFASSUNG

Die evidenz-basierte Untersuchung von Lern- und Kompetenzergebnissen im Hochschulbereich ist ein wichtiges Forschungsthema, in dem nicht nur innovative Lern- und Prüfungsmethoden entwickelt, sondern darüber hinaus auch Forschungsdaten in großer Menge produziert werden. Die vorliegende Dissertation verbindet zwei dabei relevante Themen: innovativen performanz-basierten Kompetenzprüfungen und FAIR[1]es Forschungsdaten-management. Damit hat sie zum Ziel, die Nutzungsmöglichkeiten von Rollenspielen bei der Prüfung von performanz-basierten Kompetenzen an Hochschulen besser zu verstehen, während durch FAIRes Management und Dokumentation der verwendeten Forschungsdaten Transparenz der Ergebnisse und Wiederverwendbarkeit der Daten ermöglicht wird.

Performanz-basierte Kompetenzen werden als wichtiges Ergebnis von Hochschulbildung herausgearbeitet, für deren Prüfung angemessene Methoden wie Rollenspiele nötig sind. Durch tiefgehende Voranalysen wird die Anwendbarkeit eines bestimmten Rollenspielinstruments gezeigt, das 2016 von der Forschungsgruppe „Kompetenzorientierte Prüfung kommunikativer Fähigkeiten von Studierenden“ (KomPrü) entwickelt wurde (Braun et al. 2016; Braun et al. 2018). Drei publizierte Artikel, sowie ein Artikel im Begutachtungsverfahren adressieren die Dissertationsziele weiter, indem

- die individuelle und soziale Relevanz von Hochschulergebnissen betont wird (Braun et al. 2020),
- Rollenspiele als Methode vorgestellt werden, die theoretisch holistische Ansprüche mit empirischer Messbarkeit von performanz-basierten Kompetenzen vereint werden (Falkenstern et al. 2020),
- Diese Verknüpfung von Theorie und Empirie genutzt wird, um ein Kompetenzstufenmodell mit qualitativen Deskriptoren entwickelt werden (Walz & Braun 2022) und
- Ergebnisse des innovativen Rollenspielinstruments mit denen von etablierten Wissenstests an Hochschulen verglichen werden (Walz et al. under review).

[1] Das Akronym FAIR bezieht sich auf vier zentrale Prinzipien von Forschungsdatenmanagement: Auffindbarkeit (Findability), Zugänglichkeit (Accessibility), Interoperabilität (Interoperability) und Wiederverwendbarkeit (Reusability) (Wilkinson et al. 2016).

Für die empirischen Analysen wurden verschiedene Teilgruppen einer Stichprobe von 546 Studierenden aus deutschen Lehramtsstudiengängen und wirtschaftswissenschaftlichen Studiengängen verwendet. Diese Datenbasis wurde unter Beachtung der FAIR-Prinzipien für Forschungsdatenmanagement dokumentiert und publiziert (Walz et al. 2022). Die Ergebnisse, Stärken und Einschränkungen dieses Ansatzes werden diskutiert und im Kontext möglicher zukünftiger Forschung eingeordnet.

ABSTRACT

Evidence-based research on learning and competence outcomes of higher education is a thriving field that does not only develops innovative learning and assessment methods, but furthermore produces an abundance of valuable research data. This dissertation combines the two relevant topics of innovative performance-based assessment of competences and FAIR[2] management of research data. Hence, it aims at a better understanding of the usability of role-plays in higher education competence assessment, while FAIR management of applied research data facilitates transparency of the results and reuse of data.

Performance-based competence is emphasized as a relevant outcome of higher education requiring appropriate assessment methods like role-plays. In-depth pre-analysis constitutes the usability of a particular role-play instrument developed by the research group "Performance-based Assessment of Students' Communication Skills" (KomPrü) (Braun et al. 2016; Braun et al. 2018). Three published articles and one article currently under review cover this dissertations scope further by

- highlighting the personal and social importance of higher education outcomes (Braun et al. 2020),
- emphasizing role-plays as a method that can merge theoretical holistic requirements with empirical measurability of performance-based competence (Falkenstern et al. 2020),
- using this connection between theory and application to derive a competence level model with qualitative descriptors (Walz & Braun 2022), and

[2] The acronym FAIR includes four central principles of research data management: Findability, Accessibility, Interoperability, and Reusability (Wilkinson et al. 2016).

- comparing the results of the innovative role-play instrument with established knowledge tests (Walz et al. under review).

The empirical analysis employed different subsamples of 546 German higher education students studying teacher education or economics. This data has been documented and published under consideration of FAIR data management principles (Walz et al. 2022). Results and limitations of this approach are discussed and connected to directions of further research.

Affiliations:

- Justus-Liebig-Universität, Gießen (kristina.walz@erziehung.uni-giessen.de)
- Universität Bielefeld (kristina.walz@uni-bielefeld.de)
- ORCID: 0000-0001-9606-4831

INHALT

List of Figures 1
List of Tables 1
List of Abbreviations 1
1. Introduction 3
1.1 The Potential of Role-Plays to Assess Performance-Based Competence in Higher Education 4
1.2 FAIR Principles and Documentation of Scientific Data 7
2. Theoretical Framework 9
2.1 Learning Outcomes of Higher Education: Knowledge, Skill or Competence? 9
2.2 The Concept of Communication Skills and Communication Competence 11
2.3 Assessment of Knowledge, Skills, and Competences in Higher Education 13
2.4 Role-Play Methods 15
3. The Komprü-Test-Instrument and Empirical Sample 21
3.1 Sample 22
3.2 Test Procedure and Multi-Matrix Design 24
3.3 Management and Documentation of Data 26
3.4 Analysis of Gender, Age and Experience of Rater and Confederates 30
3.5 Empirical Methods for Further Examination of the Observation Data 40
4. Discussion of Research Goals and Cumulative Contributions 44
4.1 Discussion of the Potential of Role-Plays to Assess Performance-Based Competence in Higher Education 45
4.2 Discussion of FAIR Principles and Documentation of Scientific Data of the KomPrü Instrument and Data Sample 50
4.3 Limitations 52
4.4 Directions for Further Research 54
4.5 Reflection 57
5. References 60
6. Appendix 73
6.1 First Paper (Published) 73
6.2 Second Paper (Published) 92
6.3 Third Paper (Published) 107
6.4 Fourth Paper (currently under review) 133
Eigenständigkeitserklärung 168

LIST OF FIGURES

Figure 1. Testing Procedure (derived from Braun 2021) 25
Figure 2. Example of Variable Description in the Codebook 30

LIST OF TABLES

Table 1. Description of Participants 23
Table 2. Description of Confederates and Rater 23
Table 3. Distribution of Role-Play Combinations (in %) 26
Table 4. Differences in Ranksum for Female and Male Rater 35
Table 5. Differences in Ranksum for Female and Male Confederates 37
Table 6. Correlation Between Role-Play Results and Age 38
Table 7. Correlation between Role-Play Results and Experience 39

LIST OF ABBREVIATIONS

AIC	Aikakes Information Criterion
AHELO	Assessment of Learning Outcomes in Higher Education
ACER	Australian Council for Educational Research
BIC	Bayesian Information Criterion
CHEERS	Careers after Graduation: a European Research Study
CCC	Category-Characteristic-Curve
4C	Collaboration, Communication, Creativity and Critical Thinking
CRediT	Contributor Roles Taxonomy
ENQA	European Association for Quality Assurance in Higher Education
EARLI	European Association for Research on Learning and Instruction
NEPS	German National Educational Panel Study
GRM	Graded Response Model
GOS-L	Graduate Outcomes Survey - Longitudinal
ICT	Information and communication technology
IRT	Item Response Theory
ICC	Item-Characteristic-Curve
GPK	Knowledge Test: "General Pedagogical Knowledge"
KMK	Kultusministerkonferenz
LEGACY	Learning and Employability Gain Assessment CommunitY
M	Mean value of a variable
N	Number of valid observations in a sample
OLS	Ordinary Least Squares
OECD	Organisation for Economic Co-operation and Development
r	Pearson‘s correlation coefficient
FAIR	Principles of Findability, Accessiblitiy, Interoperability and Reusability of Data
p	Probability of the test result under the assumption that the null hypothesis is correct
RDM	Research Data Management
HEGESCO	Research initiative: "Higher Education as a Generator of Strategic Competences"

KoKoHs	Research Initiative: "Modeling and Measuring Competencies in Higher Education – Validation and Methodological Innovations"
REFLEX	Research into Employment and Professional FLEXibility
KomPrü	Research project: "Performance-based Assessment of Students' Communication Skills"
CALOHEE	Reserach initiative: "Measuring and Comparing Achievemehts of Learning Outcomes in Higher Education in Europe"
STEM	Science, Technology, Engineering and Mathematics
sd	Standard deviation of a variable
SABER-TE	Systems Approach for Better Education Results
TIF	Test-Information-Function
UK	United Kingdom
VALUE	Valid Assessment of Learning in Undergraduate Education
U	Whitney-U test-statistic
WM	Wright Map
z	z-Standardized Coefficient

1. INTRODUCTION

Change and changing requirements for competences can be regarded as a key characteristic of modern society (Berezovska et al., 2020; Penprase, 2018). Especially the vast increase of technology and data in the last decades has transformed the society towards a globalized and knowledge based one (Allen et al., 2013). State-of-the-art technologies and procedures keep changing so that specific knowledge and skills that are bound to a specific technology or procedure might be outdated quickly after mastering. This has implications on a human level (e.g. in development of generic competences) as well as on a technical level (e.g. in scientific data management).

On a human level, employees need to learn a different set of skills to work with new technology and to perform tasks in a new environment. As a result, the 21st century society is characterized by the necessity of lifelong learning (Medel-Anonuevo et al., 2001). This learning includes but is not completed by mastering of specialized knowledge and sets of skills for specific tasks. Furthermore, it is inevitable to accomplish a more general competence that helps applying specific knowledge and specific skills to different and new situations (Ahmed et al., 2021; Braun et al., 2020; Pellegrino & Hilton, 2012). In this sense, performance-based competences serve as a mediator between knowledge and action.

Accordingly, on the technical level, technical solutions and data management produce new information, data, and methods. Advances in data analysis and computer science facilitate new and innovative computation which increases the amount generated and processed data. Furthermore, it becomes inevitable, to manage information and research data in such ways that it can be re-used in changing technical systems and methods. The principles of findability, accessibility, interoperability and reusability of research data, summarized with the acronym FAIR, address the application of sustainable research data management (Wilkinson et al., 2016).

On a human level, the increasing importance of general competences are summarized in the discourse of 21st century skills. On the technical level of scientific data management, we can observe a vivid discourse regarding data reusability and accessibility. Both levels, require mediators that are capable of processing new knowledge and information and transform it towards action.

Communication competence can be regarded as such a mediator. Communication links knowledge or information to active behavior within a social situation. Effective

communication is widely regarded as contributing to personal as well as professional contexts (Hargie, 2022; McConville et al., 2017). Communication, as a form of skilled behavior, is considered as learned behavior (Hargie, 2022). Hence, it is possible, to develop improve communication competence (Verderber et al., 2014). This dissertation understands communication competence as a central performance-based competence that can transform knowledge into action. Role-plays are a method with a strong focus on interaction and communication in a simulated environment. This dissertation aims at testing the usability of role-plays to assess performance-based competences in higher education under consideration of FAIR data management. It focusses on the examination of a particular role-play instrument. This instrument has been developed by the research group "Performance-based Assessment of Students' Communication Skills" (KomPrü). The instrument contains ten role-plays where participants interact with trained actors (confederates) and are observed by trained rater (Braun et al., 2016).

Furthermore, this dissertation contributes on both, the human and the technical level of scientific data management. On the human level, it ties in with the discourse of 21st century skills as a learnable outcome of higher education and examines the usability of role-plays to assess a specific performance-based competence, namely communication competence. On the technical level, it ties in with the discourse of open science and reuse of data and provides a full documentation of a complex set of data.

1.1 The Potential of Role-Plays to Assess Performance-Based Competence in Higher Education

Scientific and political actors often summarize general competences as 21st century skills. They use this term to declare specific sets of skills and link them to means of teaching, learning and application. Mostly, these sets include collaboration, communication, ICT literacy and social and cultural skills or refer to a set of "4Cs" that focusses on collaboration, communication, creativity and critical thinking (Voogt & Roblin, 2012). It is generally assumed, that these skills can be learned and should therefore be part of educational curricula and learning outcomes (Argyle, 1993; Wallace & Becker, 2019). In the first contribution of this dissertation, two colleagues and me examined learning outcomes for the case of higher education institutions (Braun et al., 2020). We identified three main actors that initiate, support, and conduct research within this field: These are scientific actors, policy makers and economic stakeholder.

The interest of researchers and policy makers in 21st century skills refers especially to assessing and enhancing these skills, for example during vocational and higher education. For the European higher education area implementations and transformation within the so-called Bologna-Process aim not only on at a better comparability between different countries but also on a higher competence orientation of higher education with impacts on individual, economical, educational and societal level (Curaj et al., 2018). Economic and governmental initiatives fund major projects that aim at developing assessment methods and gather internationally comparable data. One example is the German program "Modeling and Measuring Competencies in Higher Education – Validation and Methodological Innovations (KoKoHs)" which was funded by the German Federal Ministry of Education and Research. This research initiative developed and validated a variety of new models and tools for the valid assessment of student competencies in higher education (Zlatkin-Troitschanskaia et al., 2020). Other examples are the pan-European initiatives HEGESCO and CALOHEE. While HEGESCO focusses on the transition between higher education and labor market and provides higher education institutions with empirical evidence to enhance curricula and teaching strategies (HEGESCO, 2022), CALOHEE set out to define competence based frameworks of higher education outcomes on subject-level and to supplement these framework with international applicable and comparable tests (Wagenaar, 2021).

Still, there is some reluctance to this economic and policy driven quantification of general competences and skills. Most skills or competences that may be regarded as a 21st century skill, are characterized by a their complexity. They are often based on performance instead of mere cognition and require a deep theoretical understanding for conceptualization and assessment. Traditional quantitative assessment and analyzing methods are capable to capture the complexity of latent traits to a certain extent for the case of cognitive dispositions. Still, they require a strictly outlined definition of an assessed trait and its potential factors. The more complex a skill or competence is, the harder it gets to conceptualize it as an assessable trait. From a pure theoretical point of view, most assessment tools only capture a small part of a generic skill or competence and are therefore inadequate to describe complex 21st century skills. However, in the second contribution of this dissertation, three colleagues and me aim to bridge this alleged gap between complex theoretical understanding of performance skills and its empirical conceptualization and assessment (Falkenstern et al., 2020). We consider the example of communication skills,

which appear in nearly all main frameworks for 21st century skills and constitute its complexity and learnability in a holistic transformative learning environment before evaluating the approach of the KomPrü role-play instrument.

To describe a person's competence more detailed and to compare between different persons or points of assessment, different levels of skills need a description. These can be found in so-called competence level models. In the third contribution of this dissertation, one colleague and me apply the idea of competence level models to the assessment results of the KomPrü role-play instrument (Walz & Braun, 2022). We used data collected with the simulation-based tool and examined if the tasks of the simulation can be distinguished by difficulty in a way that individuals that are more competent succeed at more difficult tasks than individuals that are less competent. Based on these findings, we developed a competence level model that can be used to describe an individual's communication competence as assessed with the role-play instrument (Walz & Braun, 2022).

Most 21st century skills can be categorized as performance-based. They enable individuals to apply specialized knowledge to a variety of contexts and allow complex decision-making. This links performance skills to the practical area of knowledge application. Traditionally, higher education trains students in a theory driven way and is therefore distinguished from, for example, vocational education and training (Wolter & Kerst, 2015). Still, as 21st century skills are regarded as a learning outcome of higher education it is research topic of its own to implement and foster performance skills within the context of higher education curricula. Frameworks for qualification specify learning goals to be achieved in an institution of higher education and universities start to anchor performance skills to the scope of their curriculum. In the fourth contribution of this dissertation, two colleagues and me examine this assumed conjunction of theoretical knowledge and applicable skills for the case of teacher education students (Walz et al., under review). As teachers are not only citizens of the modern society but also responsible for preparing and teaching necessary skills to the next generation, their training in particualar should include performance skills. In the contribution, we compare the impacts of teacher training on traditional knowledge and performance skills and examine the share of both in teacher training. Again, we can build on the work of the previous contributions of this dissertation that strengthen the theoretical background of the role-play test and provide a competence level model for interpreting the results of the role-play test.

The previous thoughts contribute to the first goal of this dissertation: I seek to contribute to the thriving research field of assessing higher education learning outcomes by gaining a deeper understanding of role-plays as a method to assess performance based-competences in institutions of higher education.

Chapter 2 sets the theoretical foundation for this goal. It contains definitions of central terms and introduces core concepts such as the two types of communication or the method of role-play in competence assessment. Chapters 3.1 gives a general overview of the applied empirical data. Chapter 3.4 provides further analysis and validation of the particular role-play instrument. Chapter 6.1 to Chapter 6.4 provide theoretical as well as empirical results to the field of learning and assessment of communication competence. A final discussion of this first research goal that combines results of pre-analysis and all four contributions is given in Chapter 4.1.

1.2 FAIR Principles and Documentation of Scientific Data

When describing the vast changes of the 21st century one cannot ignore the high increase of data. Just in this short introduction, I described and introduced new fields of research, empirical based research projects and assessment tests to examine learning outcomes. All these fields, projects, and tests produce a worldwide abundance of data available for scientific use. While most data is collected for a specific research aim, it furthermore contains information beyond this primary research aim. Using this available data and information in a secondary data analysis is therefore a promising method of research (Gerecht & Kminek, 2018; Johnston, 2017; Smith, 2008). The use of available data in a secondary analysis holds several benefits for researcher, participants of tests and surveys, and methodological development. It enhances scientific quality indicators such as transparency, replicability, or comparability (Gerecht & Kminek, 2018). For example, researchers can use available data instead of collecting resource-intensive new data. Furthermore, they may achieve a scale of data that they could not replicate on their own (Smith, 2008). Once data of a certain topic is collected, a variety of research questions can be addressed without surveying or assessing the same population more than once. This can save time and resources and avoids overuse of participants' time and motivation (Gerecht & Kminek, 2018). Furthermore, existing data can be (re-)used with new methodological approaches or better software (Smith, 2008). Still, facilitating secondary data analysis has requirements. First, data sharing is supported by including a structured data management starting as early in the research process as possible. This can include a data management

plan, developed during the planning phase of a research project, as well as continuous documentation of all data-related tasks. After sharing, the data must be available and licensed for further research and needs a complete documentation that contains a description of variables and scales, time and circumstances of data collection, and operationalization of constructs. Data and documentation must be available and findable for researches so that they are able to assess existing data and confirm that it fits their research aims, which may be very different from the original research aim (Smith, 2008). Furthermore – as most data is stored digital – the data needs to be interoperable, which means the data is stored in a format that can be processed by (present and future statistical) software. For these purposes, the data needs to be archived and administrated. The guidelines of the FAIR (Findable, Accessible, Interoperable, and Reusable) data principles are established by a consortium of scientists and organizations and entail proposals and processes to meet these requirements (Gerecht & Kminek, 2018; Wilkinson et al., 2016). FAIR and comparable principles form a regulatory connection between different actors and platforms engaged in the data administration process. Key components of these principles have become mandatory for public funded research projects in Europe (Putnings et al., 2021).

The growing interest in data management, storage and reuse establishes a new area of interdisciplinary scientific work where researcher focus on developing guidelines and methods for planning, collecting and managing data. They furthermore collaborate with libraries and datacenters to provide IT-infrastructures and prepare standards of (meta-)data, manuals and codebooks. Accordingly, the publication of data, metadata, data manuals, and codebooks in data repositories has become the areas main form of publication. Other than archives for long-term storage of research data as suggested by guidelines of good scientific practice, data repositories are a platform to license and share research data in the sense of the FAIR principles. Analogous to published articles in journals, contributions or monographies, documented and published data can be used as a citable source in further research.

This dissertation and its contributions tie in closely with the work done by the KomPrü-research group between the projects phase from 2014 to 2017 (Braun et al., 2016; Mainz, 2022) but add new and considerable thoughts and results in a secondary data analysis. Previous research provided a theoretical frame to understand communication competence (Braun, Athanassiou, et al., 2018) as well as a performance-based assessment

and first empirical results (Braun, 2021; Braun, Schwabe, & Klein, 2018). During this research, the KomPrü-research group collected a large sample that contains performance-based information, self-assessments, extensive background information and a subject-specific knowledge assessment for teacher education and economics students. The data was collected at eleven different German higher education institutions (Braun et al., 2016). The focus on performance-based assessment and its validity for the German higher education area make this sample unique and allow for a variety of research questions to find a suitable answer. Accordingly, there remain several potentials that have not been investigated in-depth by previous research.

Following these thoughts, it is the second goal of this dissertation, to improve the reusability of the KomPrü role-play instrument as well as the KomPrü data that has been collected in 2016 under consideration of the FAIR principles. In Chapter 3.3 I introduce a full documentation of the data, containing administrated files of data, commentated survey sheets as well as a comprehensive codebook. The single documents and files have been published at the repository of the Justus-Liebig-Universität Gießen's library (Walz et al., 2022). This documentation has been developed in close cooperation with researchers from the KomPrü group and closes gaps and inconsistencies within the raw data. In an pre-analysis of raters' and confederates' gender, age and experience I illustrate, that the role-play instrument is robust against these unconscious elements of communication and evaluation of behavior (see Chapter 3.4). This enhances the instruments overall quality. With the development of a competence level model (Walz & Braun, 2022) I seek to enhance the instruments reusability by providing important means for interpretation of the test results. A final discussion including results of contributions, documentation, and further results in light of this second research aim is given in Chapter 4.2.

2. THEORETICAL FRAMEWORK

2.1 Learning Outcomes of Higher Education: Knowledge, Skill or Competence?

When it comes to the definition of qualification frameworks or the assessment of higher education learning outcomes, there are several concepts used, to describe and categorize these qualifications and outcomes (Dudek et al., 2019; Glaesser, 2019; Sanagavarapu et al., 2019; Schwabe & Braun, 2021; Sysoieva & Mospan, 2018). Three commonly used terms for these concepts are "knowledge", "skill", and "competence" (Allen et al., 2005; Weinert, 1999). While these three terms refer to partly overlapping

concepts it is inevitable to define and use them in a distinctive way that implies qualitative and meaningful differences and explicitly states what exactly is described, assessed or analyzed (Allen et al., 2005; Braun et al., 2020; Dietrich & Olson, 2010).

In educational science and research, the term knowledge often refers to cognitive processes of understanding, remembering, and reproducing/reciting facts or content information of a certain topic (Atkins et al., 1993; Mayhew et al., 2016) including knowledge that may be gained through learning or experience (Brockmann et al., 2011; Winterton et al., 2006). The knowledge a person has, can be described as the "body of understood information possessed" by this person (Weinert, 1999, p. 35). The contributions of this dissertation follow this content-based definition of knowledge. I refer to knowledge as the set of facts and information a person can reproduce.

The terms "skill" and "competence" differ from this cognitive and information-based concept of knowledge by adding some sort of application that goes beyond a mere replication of information (Blömeke, 2019; Chernikova et al., 2020; Walz et al., under review). Instead, a person uses the knowledge available to actively resolve a given situation. This means the knowledge available is transformed into action. This knowledge-based, (solution-oriented) action can be referred to as performance. The performance-based transformation of knowledge in a situation is referred to as "skill" or "competence". At this point, both terms are used to describe very similar and overlapping concepts (Brockmann et al., 2011). Skills, for example are referred to as "(a) an acquired aptitude; (b) an ability to perform complex motor and/or cognitive acts with ease, precision, and adaptability to changing conditions" (Weinert, 1999, p. 35), or simply as a combination of cognitive and physical abilities (Winterton et al., 2006). Competence sometimes specifies a more complex, higher level concept that might contain several skills (Allen et al., 2005; Weinert, 1999). Competences can be considered as context-neutral in a sense that a specific competence has the same meaning across different contexts (Allen et al., 2005) and is furthermore applicable across different contexts (Braun & Mishra, 2016). In this sense, "competences" fulfill central aspects of the 21st century skills: They are widely applicable and enable individuals to master changing situations. Hence, it becomes clear that some features of the concept "competence" of the above-mentioned features of "skills" are similar. For example, Klieme and Hartig (2008) state that the usage of the term "competence" implies a focus on a given situation or task that is to be solved but also on the ability to autonomously solve a variety of similar situations and tasks.

In this dissertation, I would like to keep the separation between knowledge as a theory-based body of information and skills/competences as an action-oriented ability of performance. In this sense, the two features of competence mentioned by Klieme and Hartig (2008) above are used to indicate characteristics of performance-based skills and competences. Performance refers to available knowledge in order to actively solve a given situation. Furthermore, performance-based competence includes the capability to consciously transfer this solution to other situations and contexts to produce successful action (Cameron, 2000). I consider performance-based competence as a latent trait, that can not be observed directly but that may be proxied by the observation of a person's action within an observed situation.

Outcomes of higher education include knowledge outcomes as well as skill outcomes. Higher education institutions train students in subject-specific skills and competences as well as in generic skills and competences. In this sense communication competence can be regarded as a performance-based competence outcome of higher education. The next chapter describes a theoretical concept of communication competence that is used in the KomPrü instrument and the empirical analysis.

2.2 The Concept of Communication Skills and Communication Competence

Whenever two or more individuals communicate, they interact not only with words, but with their whole behavior and presence. Verbal communication can transfer information between two or more individuals, para- and nonverbal communication supplies these pieces of information with an interindividual context and meaning, the general impression of a person can influence the course of communication (Degner et al., 2009). This means, that communication is not only about what is (not) said, but also about how and by whom it is (not) said. Paul Watzlawick et al. (1967) refer to these two features of communication as information and meta-information or content and relational level respectively. In communication, both levels intertwine and facilitate interindividual dialogue. Everything that is part of or happens during a communicational setting or situation is therefore part of the communication. This persistency of communication sets this skill apart from other 21st century skills and emphasizes its importance for successful life in the global knowledge society (Ellis, 2009; Hargie, 2022). Based on these thoughts I regard communication as a performance-based competence that contains knowledge about the process and strategy of communication and further on builds on a set of action skills that are applied autonomously to a situation.

As we will see in Chapter 2.3, performance-based skills and competences require a performance-oriented test, namely a simulation, for assessment. In order to develop such a simulation adequately, it is necessary to also conceptualize a skill or competence in a performance-oriented way. A theory-driven conceptualization of communication competence that is the basis for a simulation instrument has the potential to adequately assess performance-oriented communication competence in its complexity (Falkenstern et al., 2020). Such a conceptualization of communicative performance has been derived from the theory of communicative action by Habermas (1984) by the KomPrü research group (Braun, Athanassiou, et al., 2018). The theory of communicative action understands communication as interaction between individuals who communicate in a language and about matters that are existent for and understood by all participants. Furthermore, communication is characterized by two standards of truth: no participant uses the communication to deceive the other participants and all interaction has to take place within the scope of accepted social behavior. Habermas calls a communication that follows these claims as oriented to a common goal of understanding and nonviolent agreement. This ideal type of social interaction is therefore called "understanding-oriented action" (Braun, Athanassiou, et al., 2018; Frindte & Geschke, 2019; Habermas, 1984). If one of the communication partners does not aim at this common agreement based on total honesty and openness but instead exerts influence in order to achieve a certain reaction/goal within or after the communication Habermas considers this ideal violated. He then refers to the communication as "strategic action". Braun, Athanassiou, et al. (2018) go beyond this normative distinction and accept both types of communicative action as equally valid strategies in their theoretical framework. Following this definition, the contributions of this dissertation differentiate between two types of communication: a strategic and an understanding-oriented type of communication. The term "type of communication" refers to the fundamental difference in performance in the communicative situation. Furthermore, as Braun, Athanassiou, et al. (2018) postulate, a competent person makes use of both types of communication depending on whether a situation requires a common agreement or the achievement of a certain goal. In each situation, this decision and the following performance can be observed and assessed with the help of specific facets of communication that are also anchored in communication theory. As stated above, communication consists of content and relational elements that include everything that is said by participants of the conversation, that happens during or that is otherwise part of a communicational situation. On a mere content level a communication can be characterized

and observed by its content goal as well as by the clarity, quantity, relevance and quality (Grice, 1975) of the communicated information. On a relational level, a communication can be characterized by the self-disclosure (Hargie, 2010) of a person as well as by the way of dealing with social circumstances. These characterizations of communication are established parts of communication theory and lead to six facets that are used to define the observable behavior during a communication:

- content conversational goals,
- relational conversational goals,
- clarity and quantity of a conversation,
- relevance and quality of a conversation,
- self-disclosure, and
- symmetry or complementarity of a conversation.

The six facets of communication and their observable indicators are an important basis for describing the competence level of a person qualitatively. Therefore, each facet will be described in the theory section of the third contribution of this dissertation (Walz & Braun, 2022). In general, the contributions of this dissertation use this theoretical frame of two types and six facets of communication for further analysis. The results show that (1) this framework allows a holistic assessment of performance-based communication competence and that the established simulation can entail reflection (Falkenstern et al., 2020), and that (2) the chosen categories for observing the performance are valid and useful for describing differences in competence (Walz & Braun, 2022).

2.3 Assessment of Knowledge, Skills, and Competences in Higher Education

Assessment of knowledge, skills and competences needs adequate testing procedures. Hence, it is necessary to understand in which context a qualification or outcome can be reliably observed and assessed. For the case of higher education, written or oral exams are examples to assess the knowledge of a person (Pereira et al., 2016). By asking students to reproduce their gained knowledge in standardized way, it becomes possible to assess the amount of knowledge a single student has and to compare the results between students.

Both, skills, and competences built on a person's knowledge by applying it into performance and real-world solutions (see Chapter 2.3). In educational research, there exist several approaches that can be considered adequate for assessing performance-based skills and competences (Braun & Mishra, 2016). These approaches can be information-based or

performance-based. Information-based approaches operationalize skills and competences with either self-reports or an analysis of required activities (e.g. job-requirement or student engagement approaches) (e.g. Felstead et al., 2019; Hu et al., 2008). These methods use the information about a skill or competence to assess and compare it. Assessment tests go further. They are still based on information given in a survey but instead of asking participants about their skills, assessment tests contain tasks/exercises that require at least a written application of knowledge and therefore a skill (Döring & Bortz, 2015, p. 453). Furthermore, tasks/exercises of an assessment test can contain written descriptions or videos of a scenario that requires skill-based action. The participant is asked to report which knowledge would be applied in which way to solve the scenario appropriately. Within the KomPrü project, two assessment tests are applied to supplement the performance-based approach of the role-play: the general pedagogical knowledge test, developed by König and Blömeke (2009) and a test assessing the subject specific competencies students of economics developed by Lauterbach (2015). Both tests entail a pool of tasks that require teaching and economics competence for correct solutions (for further description see also Chapter 3). In contrast, performance-based approaches focus on the observation of this skill-based action by using different simulation methods (e.g. Chernikova et al., 2020; Hyytinen & Toom, 2019). A simulation allows an approximation of real-world scenarios, in which students can interactively use their skills (Chernikova et al., 2020; Cook et al., 2013; Grossman et al., 2009). They offer a secure environment of learning and assessing skills and competences required by higher education graduates in general and for specific professions (e.g. teaching, nursing or medical professions) in particular. In this sense, simulations can supplement real-world-experience like practical training by providing ethically credible learning environments as well as the opportunity to practice and prepare for real-world situations that are complex but rare (Chernikova et al., 2020; Kettula & Berghäll, 2013). Furthermore, simulations can be used to deepen the understanding and relationships between different concepts. The approximation of practice in a standardized way allows to control and alter key features of this simulated situation experimentally. In a simulation, the student interacts with either a real or virtual object, device or person. These interactions are furthermore crucial for the flow and result of the simulated situation. In a meta-analysis Chernikova et al. (2020) give a broad definition of the term simulation. They consider anything that includes skill-based interaction – either with other people or with electronic/digital devices – a simulation (Chernikova et al., 2020). This skill-based interaction can be observed and reflected in two ways: (1) One can observe/reflect how the

participant performed on the task of the specific simulation scenario. This assessment would than allow training and evaluation of the participants skill. (2) One can observe/reflect with standardized criteria that are derived from a deeper theoretical framework and that aim not only at the simulated scenario but to an underlying latent competence. In this sense, performance in a simulation serves as an indicator for performance-based competence. Current research examines the use of simulations for teaching and learning purposes (Chernikova et al., 2020). Simulations can offer a great opportunity to give feedback and induce learning processes (Gartmeier et al., 2015; Kilgour et al., 2015; Moreno-Guerrero et al., 2020). Using simulations for learning purposes allows practice-oriented training and rehearsing of real-world scenarios that require professional performance. However, simulations can furthermore be useful for practice-oriented testing and assessment of skills and competence. In this sense, simulations allow to grade performance-based competence. These grades could then be included visibly within certificates or grade point average. As performance-based competences are an important learning outcome of higher education, students' results and achieved levels of competence could be part of their certificate, transcript of records or grade point average. In the fourth contribution of this dissertation, two colleagues and me did not find performance-based competence as a visible part of grade-point average for German teacher education students yet (Walz et al., under review).

Simulations require interaction and can be used for training and assessment of a broad variety of performance-based skills and competences (see Chapter 2.4 for examples). This holds especially for communication competence, which is basically a performance-based competence consisting of interaction between two or more individuals.

2.4 Role-Play Methods

Following the previous thoughts and definitions, role-plays are characterized as simulations that take place in an environment approximated to the real world. The environment entails a task or problem that motivates the participants to interact solution-oriented with this environment (Crow & Nelson, 2015). While there are applications using digital environments, a role-play can also be applied without any technical support relying only on face-to-face interaction instead. As such, role-plays are considered as a simulation method, that is appropriate for training but also for assessing of performance skills (Braun, Athanassiou, et al., 2018; Braun & Mishra, 2016; Gulikers et al., 2004). Within the context of higher education, role-plays are applied as learning and assessment method in a variety

of subjects. Role-plays are a common and long established method for different fields of medical education, preparing future nurses, physicians or pharmacists for their future working fields (e.g. Bosse et al., 2012; Charlton, 1993; Joyner & Young, 2006; Nemec et al., 2021; Nestel & Tierney, 2007; Rao, 2011). In light of the rising importance of performance skills, the application of role-plays spread to further subjects like economics and management courses (e.g. Barrera et al., 2021; Kettula & Berghäll, 2013; McConville et al., 2017; Westrup & Planander, 2013), engineering (e.g. Andersson & Andersson, 2010; Martin et al., 2019), law and counseling (e.g. Phillips, 2012; Stewart-Spencer et al., 2019), politics (e.g. Duchatelet et al., 2021), or teacher education (e.g. Crow & Nelson, 2015; Gartmeier et al., 2015). Furthermore, role-plays in higher education focus on different aspects of professional skills and performance competence and, therefore, contribute to and assess different learning outcomes. While some role-plays address students' communication skills (e.g. Bosse et al., 2012; Gartmeier et al., 2015; Lane & Rollnick, 2007), others focus on intercultural competences (e.g. Romijn et al., 2021; Schnurr et al., 2014), diagnostic skills (e.g. Fischer & Opitz, 2022), critical thinking (e.g. Grose-Fifer, 2017; Nemec et al., 2021), or decision-making and problem-solving (e.g. McConville et al., 2017; Nicolaides, 2012; Vermeiren et al., 2022). The role-play instrument developed by the KomPrü- research group that serves in this dissertation as a particular example for testing the usability of role-plays to assess performance-based competences in higher education, focusses on the assessment of communication skills of teacher education and economics students in German higher education institutions (Braun et al., 2016).

The role-play applications deployed in the various and different contexts of higher education share on the one hand some common elements, marking them all as a role-play method. On the other hand, some clear differences distinguish different applications of role-playw resulting in different learning environments and experiences that emphasize the broad scope and potential of role-plays in higher education.

Role-plays in the context of higher education entail some approximation of reality. As it is impossible to simulate all aspects of reality, the focus of this approximation depends on the goal of the role-play application. Furthermore, the approximation should facilitate an immersive experience as well as the option to render/alter parts of the setting comparable to experimental manipulation (Lane & Rollnick, 2007). Students who participate in a role-play interact according to this approximated reality and within a previously defined role. They aim at accomplishing a task by applying professional knowledge and skills within

their enactment. Participants receive information of setting (i.e. the approximated reality), role and task before the role-play. Most role-plays in the context of higher education aim at the enhancement of performance skills and competences that are relevant for students' future professions. Hence, the approximated parts of reality as well as the roles and their required skills focus on workplace-related scenarios (e.g. Cook et al., 2013; Crow & Nelson, 2015; Lane & Rollnick, 2007; Martin et al., 2019; Nicolaides, 2012). To assure for this enhancement, role-plays entail assessment and feedback criteria (Crow & Nelson, 2015). If role-plays are used for learning purposes, this could be a video recording or observation of the performance that is discussed and reflected (e.g. Gartmeier et al., 2015; Lane & Rollnick, 2007). An in-depth discussion that is followed by a repeated but slightly altered turn of role-play can be part of the learning process and role-play application (e.g. Crow & Nelson, 2015). If role-plays are used for assessment purposes, it could be a standardized observation form or summarized record that is used to rank/grade the participants performance (e.g. Phillips, 2012). The KomPrü role-play instrument aims at approximating realistic scenarios of communication that appear in teaching and economics contexts. Within each scenario, participating students receive context information and a goal to achieve. To ensure immersive and realistic scenarios the research group employed workshops with experts of both fields (Braun et al., 2016). The performance of participants is assessed with a standardized observation survey that is based on a theoretical framework.

Apart from these common elements, applications of a role-play can be based on different characteristics. Rao and Stupans (2012) developed a typology of role-plays based on qualitative interviews about role-play applications. They distinguish three categories of role-play that are increasingly immersive. The three categories are named (1) role-switching, (2) acting, and (3) almost real life. Role-plays aiming at a role-switch foster the students ability to understand actions and positions of other people by taking their point of view (e.g. Crow & Nelson, 2015; Martin et al., 2019; McConville et al., 2017). Role-plays with a focus on acting entail a specific skill that students practice by enacting in small group scenarios (e.g. Charlton, 1993; Joyner & Young, 2006; Phillips, 2012; Westrup & Planander, 2013). The category of almost real life role-plays approximates real scenarios as close as possible and facilitate an environment for students to apply their skills (Duchatelet et al., 2021; e.g. Rao & Stupans, 2012). Furthermore, role-play applications can be distinguished by methodological relevant characteristics such as number and type of participants of the role-play, length of the role-play, assessment and feedback criteria

and procedure, attendance of the participants, or the background of the role-play scenarios. When developing or reusing a role-play application these characteristics need to be tailored to fit the desired learning situation and outcome. Seven of such characteristics are introduced in the following along with some examples of role-play applications that include that characteristic in their design. The presented characteristics entail relevant information to denote the features of the KomPrü role-play instrument compared to other role-play applications. Furthermore, they give a general impression of the broad scope of usability of role-play methods in higher education. Characteristics and examples have been identified based on an explorative literature research using the JustFind and GoogleScholar search engines.

First, each role-play needs at least two participants that can interact. They can be peer students that enact different roles within a role-play scenario. In some applications, all roles in a scenario require professional performance-based skills to achieve a goal (e.g. Duchatelet et al., 2021; Phillips, 2012). In other applications, only one role requires professional performance-based skills to achieve a goal, while peer students enact students in a classroom, clients, or patients (e.g. Charlton, 1993; Joyner & Young, 2006; Nestel & Tierney, 2007; Rao, 2011). These applications still provide a learning environment for students outside the professional role in the sense of a role-switch (e.g. Charlton, 1993; Rao & Stupans, 2012). However, the approximation of a real workplace experience can be further enhanced by interaction with a trained and confederate actor (e.g. Bosse et al., 2012; Fecke & Müller, 2022; Gartmeier et al., 2015; Stewart-Spencer et al., 2019). In an application with peer-interaction, the student in the professional role has previous knowledge about the interaction partners and may also know the description of the student/client/patient roles. Involving a trained actor enhances uncertainty for the student in the professional role as neither the interaction partner nor his:her role description are known. A review of role-play applications with simulated patients in medical education revealed, that students experience the role-play as less artificial and more realistic if it involves trained actors (Lane & Rollnick, 2007). In the KomPrü role-play scenarios one participating student interacts with one confederate who is a trained actor.

Second, a role-play instruction can entail a pre-defined role in which the students are asked to project their thoughts (e.g. Crow & Nelson, 2015; Joyner & Young, 2006; Martin et al., 2019; McConville et al., 2017). These role-plays are often used in the training and assessment of intercultural competence (e.g. Romijn et al., 2021; Schnurr et al., 2014). In

contrast, other role-play instructions entail only a simple briefing of the simulated situation and asks the participating students to mainly enact themselves and apply their performance skills, as they would do in a professional context (e.g. Phillips, 2012). The KomPrü role-play scenarios entail a simple briefing with main characteristics of the simulated situation and students are asked to enact as their professional selves.

Third, role-plays can contain short and single situations of enactment (e.g. Charlton, 1993; Joyner & Young, 2006; Rao, 2011). Participating students receive information about a clearly defined situation of their future work environment that entails a distinct task. Role-playing within this situation takes usually only a short amount of time (e.g. Nestel & Tierney, 2007). There can be several clearly defined situations that are enacted after each other. Additionally, each situation can be followed by feedback (e.g. Charlton, 1993; Crow & Nelson, 2015; Nestel & Tierney, 2007). Planning games, in contrast, describe a type of role-play that supplies a more complex environment that approximates reality in a more mechanical/rule-based gameplay (e.g. McConville et al., 2017). Achieving goals may take longer, involving longer periods of role-playing and interaction with this setting and other participants. Other than role-plays of single situations, a planning game can take a longer amount of time to conduct (e.g. Vermeiren et al., 2022; Westrup & Planander, 2013). Simulation games are prevalent in politics and economics courses (e.g. Duchatelet et al., 2021; Schnurr et al., 2014). The KomPrü role-play scenarios contain ten single situations that take about ten minutes to enact.

Fourth, there are differences considering the preparation of role-play scenarios and role-play itself. Scenarios and roles can be part of a pre-defined role-play instrument (e.g. Charlton, 1993; Crow & Nelson, 2015; Martin et al., 2019). In contrast, the development of scenarios and roles can be part of the method, where students and teachers jointly decide which situations they want to enact (e.g. Kettula & Berghäll, 2013). This is connected to the amount of time a participant is allowed to prepare for a role-play. Some applications allow for a longer preparation time before the actual role-play (e.g. Crow & Nelson, 2015; Kettula & Berghäll, 2013), others allow only a short preparation immediately before the role-play is conducted (e.g. Nestel & Tierney, 2007). Similar to the difference between peer-interaction and interaction with a confederate actor, an immediate preparation to an unknown situation can add to the realistic uncertainty of a situation. The KomPrü role-play instrument contains predefined scenarios and roles each presented to the participants immediately before enacting the scenario with only a few minutes of preparation.

Fifth, students' performance can be assessed with a prepared and standardized observation survey that contains important evaluation criteria (e.g. Joyner & Young, 2006). Alternatively, the performance can be assessed with a more open approach of general debrieving, observation and feedback (e.g. Kettula & Berghäll, 2013). Furthermore, different groups of observers can conduct the assessment. Observers may be peer-students (e.g. Charlton, 1993; Joyner & Young, 2006; Rao, 2011), teachers (e.g. McConville et al., 2017; Nicolaides, 2012), or external and trained observers (e.g. Rao, 2011) that evaluate the performance without prior knowledge of the participant. The performance of participants in the KomPrü role-plays is assessed with a standardized observation survey that is completed by a trained observer.

Sixth, role-play applications can differ in forms of attendance. While traditional role-plays require a face-to-face setting with physical presence of all participants (e.g. Charlton, 1993; Crow & Nelson, 2015; McConville et al., 2017), there exist innovative applications with online settings where the interaction is supported by electronic communication media (e.g. Cook et al., 2013; e.g. Fitó-Bertran et al., 2014; Russell & Shepherd, 2010). The KomPrü role-play instrument has been developed for a face-to-face deployment. Nevertheless, there are first empirical attempts of online applications (Fecke & Müller, 2022).

Seventh, settings and tasks of a role-play can be derived from either practical or theoretical backgrounds. A role-play that aims at the performance of a specific professional skill or situation is likely to contain specific scenarios like a parent-teacher conference or a clinical anamnesis. Participants' performance and its assessment is linked to correct application of professional knowledge and skills within this context (e.g. Crow & Nelson, 2015; Kettula & Berghäll, 2013; McConville et al., 2017). For example, the performance in a simulated parent-teacher conference could be assessed according to practice-oriented goals like the number of topics discussed or the overall success of the interaction. If a role-play application is based on a theoretical framework, this can be different (e.g. Gartmeier et al., 2015). The single settings will still approximate typical real-world situations of professional routine, but the goals and the assessment go beyond a mere interpretation of performance in context of the setting. Instead, the setting, tasks and ultimately the performance is regarded as indicator of a theory-based and latent competence trait. For example, within the setting of parent-teacher conference the participants self-disclosure could be assessed as facet of a theory-based concept. The scenarios of the theory driven

KomPrü role-play instrument simulate typical situations of professional life but assess the performance based on a theoretical framework of communicative competence.

3. THE KOMPRÜ-TEST-INSTRUMENT AND EMPIRICAL SAMPLE

The KomPrü-research group developed a role-play instrument for assessment of communication skills of teacher education and economics students in German higher education based on a theoretical framework of communication competence. The test consists of ten different situations that require a specific communicative behavior to solve. The situations are settled within job-related contexts of teaching and economics. Further on, the situations are characterized by parameters of communication theory like social positions, content, or relational goals of the conversation (see Chapter 2.2; Braun, Athanassiou, et al., 2018; Walz & Braun, 2022).

The development of the role-plays included extensive preparation on two levels: theoretical and content wise. Theoretically, the role-plays draw on an extensive framework that has been published by Braun, Athanassiou, et al. (2018). A fundamental part of the theoretical frame goes back to Habermas theory of communicative action and differentiates between a strategic and understanding-oriented type of communicative action. Differences in intention and goals of conversation partners characterize both types and each type leads to different observable behavior within a communicative situation. The theoretical frame includes further established theories of communication, which describe the observable differences in strategic and understanding-oriented communicative action. The third article of this dissertation (Walz & Braun, 2022) uses this framework to develop aspects of communication skills that are then empirically examined and described for different levels of competence.

Content wise, the role-play method draws on an assorted selection of interaction scenarios that supplement the theory-driven framework of the test instrument with realistic and typical forms of interaction in teaching and economics contexts. These scenarios are the result of a literature review and a nationwide survey among experts that identified typical interaction situations. A first selection of scenarios has then been validated by expert groups for both backgrounds (Braun et al., 2016). All ten resulting scenarios are applicable for the teaching and the economics background. For example the scenario “Participation in professional development training” describes a scenario in which the participant has to convince a supervisor to approve a further training program the participant wants to take part in. For the background of teaching, the topic of the training program has a teaching

context. For the background of economics, the topic of the training program is related to management. The scenario includes similar obstacles for both backgrounds like a certain reluctance of the supervisor towards the training programs topic. This results in similar descriptions of the scenarios that are only modified in specific details that anchor the scenario in a specific background.

Because of this close integration of theoretical components and realistic scenarios, the resulting assessment instrument has the potential to simulate complex and authentic situations but also allows a standardized observation and interpretation. Furthermore, it seems plausible, to apply the instrument to other disciplines by modifying the anchor-details for further backgrounds.

3.1 Sample

The instruments has been implemented in 2016 at 11 German higher education institutions. These institutions have been randomly picked out of all higher education institutions, that offered courses either in economics or teacher education or both. The participants were addressed with an open call using flyers. A total of 546 students took part in the role-plays: 273 students of economics and 273 students of teacher education. Table 1 presents descriptive numbers of this sample. On average the students are 23.17 (sd 3.99) years (economics: 23.01 (sd 3.24) years; teacher education 23.32 (sd 4.63) years) old and studied for 6.61 (sd 4.90) semesters (economics: 6.2 (sd 4.2) semesters; teacher education: 7.03 (sd 5.52) semesters) in higher education. 300 students associated themselves with female (economics: 122; teacher education: 178), 233 with male (economics: 150; teacher education: 83) gender. This distribution of gender reflects the general distribution for German higher education students for both disciplines (Statistisches Bundesamt, 2017, 2019). There are 251 undergraduate students (economics: 162; teacher education: 89) and 126 graduate students (economics: 79; teacher education: 47) in the sample. Traditionally, teacher education programs lead to a state examination and 123 teacher education students in our sample aim at this degree.

Besides the role-plays and a background survey, 260 students (economics: 155; teacher education: 105) took part in a test that assessed subject specific knowledge. Further on, 58 students (economics: 32; teacher education: 26) were tested a second time two weeks after the first assessment.

Table 1. Description of Participants

	N			M (sd)		
	T	***E***	***Te***	***T***	***E***	***Te***
Age	531	269	262	23.17 (3.99)	23.01 (3.24)	23.32 (4.63)
Semester	529	269	260	6.61 (4.90)	6.20 (4.20)	7.03 (5.52)
Gender	546	273	273			
Female	300	122	178			
Male	233	150	83			
Other	13	1	12			
Program	546	273	273			
Undergraduate	251	162	89			
Graduate	126	79	47			
State Examination	123	-	123			
Other	46	32	14			
Knowledge-Test	260	155	105			
Retest	58	32	26			

Legend: N – number of observations, M – mean value, sd – standard deviation, T – total number of observations, E – students of economics, TE – teacher education students

During the role-play scenarios the students interacted with a trained confederate actor. A trained rater observed and assessed the performance. Table 2 presents some descriptive numbers of these two groups. There have been 14 rater and 12 confederates. On average, they have been 39.3 years (rater; sd = 12.3), and 39.0 years (confederates; sd = 4.99) old. Raters observed an average of 41.5 (sd = 23.6) role-plays and confederates interacted with the participants in 49.2 (sd = 42.6) role-plays on average. Regarding the single role-plays, raters observed each role-play on average between 15.5 (role-play 6; sd = 8.88) and 17.4 (role-play 1; sd = 10.6) times, while confederates interacted with participants on average between 18.3 (role-play 6; sd = 15.9) and 20.5 (role-play 1; sd = 17.6) times. There are high standard deviations, because some rater and confederates participated in very few, others in nearly all institutions of higher education. Gender is distributed less equal for rater and confederates. There are 12 female rater who observed N = 505 role-plays, while only two male rater observed N= 35 role-plays. The five female confederates interacted with participants in N = 159 role-plays, while the seven male confederates had been assigned to N = 382 role-plays. Hence, there will be small and different sized groups within the gender comparisons, which are analyzed in Chapter 3.4. Results in these cases are largely dependent on very few people. This should be considered, when interpreting the findings in Chapter 3.4.3.

Table 2. Description of Confederates and Rater

	Rater		Confederates	
	N	**M (sd)**	**N**	**M (sd)**
Age	290	39.2 (12.3)	535	39.0 (4.99)

Experience	540	41.5 (23.6)	541	49.2 (42.6)
Role-Play 1	226	17.4 (10.6)	226	20.5 (17.6)
Role-Play 2	223	17.2 (10.9)	223	20.3 (16.8)
Role-Play 3	204	15.7 (9.38)	205	18.6 (16.2)
Role-Play 4	207	15.9 (9.19)	207	18.8 (17.5)
Role-Play 5	215	16.5 (8.98)	216	19.6 (17.2)
Role-Play 6	201	15.5 (8.88)	201	18.3 (15.9)
Role-Play 7	216	16.6 (9.64)	217	19.7 (15.8)
Role-Play 8	219	16.8 (10.9)	220	20.0 (20.0)
Role-Play 9	216	16.6 (9.20)	216	19.6 (16.5)
Role-Play 10	222	17.1 (10.4)	222	20.2 (17.3)
Gender	540		541	
Female	505		159	
Male	35		382	
Other	-		-	

Legend: N – number of observations, M – mean value, sd – standard deviation

3.2 Test Procedure and Multi-Matrix Design

Each assessment took approximately one hour without completing the knowledge test and one and a half hours when a participant completed the knowledge test. The testing procedure is presented in Figure 1. At the beginning of the assessment, four role-plays are alloted to each student: always two strategic and two understanding-oriented. After a preparation of 5 minutes during which the students read the instructions for the first role-play, the students interact with a trained confederate for about 10 minutes. A trained rater observed the interaction and filled out an observation sheet afterwards, while the student prepared for the next role-play. This procedure is repeated until the student performed all four of the alloted role-plays. Afterwards, a subsample of students continued by answering a knowledge-test. At the end of the assessment, each student filled out a background

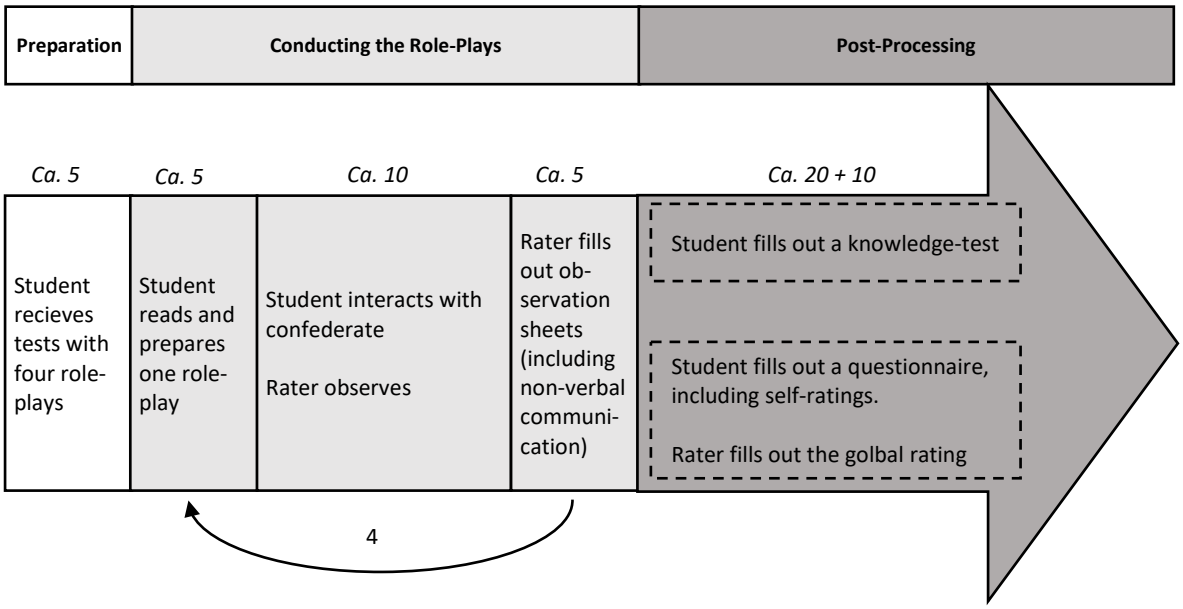

Figure 1. Testing Procedure (derived from Braun 2021)

questionnaire that included self-assessments of subject-specific knowledge and communication competence as well as socio-demographic information.

The procedure with four of ten alloted role-plays per student leads to a multi-matrix design with 40 percent observed data and 60 percent data missing completely at random. This has some implications for the further analysis of the data as we will see in Chapter 3.5 and two contributions (Walz et al., under review; Walz & Braun, 2022). Table 3 presents the distribution of role-play combinations for the strategic and understanding-oriented type. The cells present the share of each combination while the marginal distributions count the share for each single role-play. As each student performed two strategic respectively two understanding-oriented role-plays, these marginal relative distributions sum up to 200 per communication type. Due to the random allocation of role-plays, we would expect each role-play to appear in 40 of these 200 cases and each combination to appear in ten percent of the cases. Most of the combinations vary within one standard deviation (strategic: sd = 1.14 (total sample), sd = 2.28 (economics), sd = 0.31 (teacher education); understanding-oriented: sd = 1.5 (total sample), sd = 1.28 (economics), sd = 1.84 (teacher education)) around ten percent.

The combinations of the strategic role-play 1 and 3, concerning all samples (total sample: 11.7%; economics: 12.9%; teacher education: 10.4%) as well as role-play 2 and 5 concerning the total (11.5%) and economics (12.9%) sample, and role-play 3 and 4 concerning the teacher education sample (10.4%) are in this sense overrepresented.

Concerning the total and the economics sample, the combinations of the strategic role-plays 1 and 4 (total sample: 8.74%; economics: 7.38%) and 3 and 5 (total sample: 8.01%; economics: 5.9%) are in this sense underrepresented as well as the combination of 2 and 4 concerning the teacher education sample (9,35%).

The combinations of the understanding-oriented role-plays 7 and 9, concerning all samples (total sample: 12.3%; economics: 12.2%; teacher education: 12.5%), as well as 9 and 10, concerning the total (11.8%) and the economics (11.8%) sample, are in this sense overrepresented. The combinations of the understanding-oriented role-plays 6 and 7, concerning the economics sample (8.49%), as well as 6 and 9, concerning the total (7.9%) and the teacher education (6.96%) sample, and 8 and 9, concerning the total (8.46%) and the economics (8.49%) sample, are in this sense underrepresented.

However, even in case of over- or underrepresentation of a combination, each single role-play appears frequently within other combinations. Therefore, I do not expect an influence of these outlier combinations.

Table 3. Distribution of Role-Play Combinations (in %)

SC \ UOC	6			7			8			9			10			Total		
	T	***E***	***TE***	***T***	***E***	***TE***	***T***	***E***	***TE***	***T***	***E***	***TE***	***T***	***E***	***TE***	***T***	***E***	***TE***
1	8.82	8.49	9.16	11	10.3	11.7	7.9	8.86	6.96	9.38	9.59	9.16	37.1	37.3	37			
2	10.4	10.7	10.1	9.93	9.59	10.3	12.3	12.2	12.5	9.19	9.96	8.42	40.3	40.2	40.3			
3	11.7	12.9	10.4	9.65	9.23	10.1	8.46	8.49	8.42	11.2	10.7	11.7	40.6	39.1	42.1			
4	8.74	7.38	10.1	9.11	8.86	9.35	9.47	8.49	10.4	11.8	11.8	11.7	40.4	41.3	39.6			
5	10.6	11.4	9.71	11.5	12.9	10.1	8.01	5.9	10.1	10.9	12.2	9.71	41.5	42.1	41			
Total	41.3	42.4	40.3	40.6	41.7	39.6	38.8	36.5	41	38.3	36.9	39.6	41	42.4	39.6	200	200	200

Legend: SC – strategic communication; UOC – understanding-oriented communication; T – total sample; E – subsample of education students; TE – subsample of teacher education students. 1-10 – number of the role-play

3.3 Management and Documentation of Data

Transparency, replicability and comparability are traditionally among the most important quality criteria for scientific research (Döring & Bortz, 2015). More recently, the paradigm of open science postulates the managing and sharing of research data as a further criterium of good scientific practice. As sharing of data can reduce the economic costs of

future research, managing and sharing research data has become a common requirement for research funding (Gerecht & Kminek, 2018). Following an established framework, research data management and sharing can adhere to guidelines to make data findable, accessible, interoperable and re-usable – summarized with the acronym FAIR (Wilkinson et al., 2016). Research data management (RDM) can be considered as a key component to achieve not only traditional criteria of scientific practice but also requirements of a FAIR sharing of data and results. Furthermore, as RDM relates to the entire empirical research process it benefits the efficiency of the research process as well as data protection (Koltay, 2022). Models of a data life cycle have been established analogous to research process frameworks (e.g. Michener, 2015). They commonly start with planning research and include clarification about necessary data reuse or collection to fulfill the research goals. A data management plan supports specification of the need for data and the necessary steps in obtaining and managing them. Furthermore, following a data management plan can reduce the amount of effort caused by RDM, which is otherwise a commonly phrased downside of FAIR RDM (Oßwald, 2021). The next steps in the data life cycle include the search for available data or the collecting of data if no appropriate data could be found or if secondary use of data is otherwise no option for achieving the research goal. In both cases, the quality of used data needs to be secured to ensure the findability and interoperability of the data as well as the quality of the own analysis. This can be done by using standard formats and enriching the data with meaningful meta-data. An extensive documentation of data ensures the transparency and replicability of the results and improves the re-usability of the data as well as overall data quality. Documentation and meta-data are key elements for sharing the data at a research data repository at the end of the research process. Research data repositories collect and license research data and thus facilitate their accessibility (Dierkes, 2021; Michener, 2015).

While standardized meta-data makes data machine readable and increases findability, other parts of the documentation can follow less standardized procedures that aim only at human readability. Documentation can include protocols of data collection and processing as well as descriptions of including variables (e.g., a codebook) or a summary of the applied instruments (e.g., a collection of surveys or scales). This further documentation can be stored in the research data repository along with the actual data and the standardized meta-data.

The KomPrü research sample that has been described in Chapter 3.1 and was used for the empirical pre-analysis in Chapter 3.4 as well as for empirical analysis within most of the contributions of this dissertation, contains data, collected with the newly developed performance-based instrument as well as further established scales. Furthermore, there is information from a comprehensive background-survey available, including students' self-assessment on several competences, as well as data on a knowledge test for subject specific knowledge. Taken together the data gathered by the researchers of the KomPrü project can be considered as a complex pool of research data that contains potentials exceeding the original research group's goals. Still, there is no documentation of the applied instruments and the collected data available yet. In the following I seek to address this gap by applying RDM principles to the available data and instruments to provide a usable documentation that supports the transparency of the presented contributions under consideration of the guidelines of FAIR data management.

To support transparency the documentation entails a detailed description of the testing procedure and data collection, comparable to Chapter 3.1. The original data has been collected with manually filled printed surveys which were then captured by a combination of software and manual corrections. A first step of data management has been to process this digital raw data by adjusting inconsistencies, missing or false information. Some inconsistencies revealed itself during data examination, others have been detected by thoroughly reviewing the available data and surveys. Missing data has been categorized according to its origins (see Chapter 3.5). These steps aimed at improving the data's overall quality and re-usability. To support interoperability all data has been transformed to a comma separated values format which is a non-proprietary format that can be read and processed by all current statistical software but does not contain further information like value labels or annotations. Instead, this information is provided within an associated codebook. This codebook furthermore entails information about distribution and number of valid observations and can be used to gain a first impression of the data. Furthermore, a collection of surveys and scales supplements the information of the codebook. This solution does not support a machine-readable documentation of value labels but supports the planning and collection steps of future research and data life cycles. To ensure findability and accessibility the data and the associated information have been enriched with meta-data during the publication process at the repository of the Justus-Liebig-University Gießen. The published data has been licensed for reuse under CC-BY-SA conditions

(Creative Commons, no year; Walz et al., 2022). These allow reuse of the data for non-commercial purposes, when the original source is cited and the results of the reuse are shared under the same license.

The final documentation of the KomPrü data consists of three parts. First, the edited data that is stored in several files in a comma separated values format. The files are structured contentwise: Each file contains the participants' ID and variables emphasizing a specific part of the role-play instrument, the background survey or the knowledge tests. Furthermore, some variables like the context of the role-play, the participants age or gender are part of more than one file as these variables are of general importance for most research questions. This structure supports the use of single files of data for examining specific questions, while simultaneously merging of two or more files can be done by using the unique id-values as key variable.

Furthermore, the handling of item batteries is supported by a structured scheme of variable names. Names of variables that are part of a battery (e.g. the CSRS by Spitzberg and Adams (2007)) start with a short 'speaking' part (e.g., in case of CSRS the items start with np for "non- and paraverbal communication"), followed by two numbers indicating the role-play and the item within the battery. These three parts are separated with an underscore (e.g., the first item of the CSRS for role-play 1 is named 'np_1_1'). Variables, that are not part of an item battery (e.g. participants age) are named with short but `speaking` names derived from the English meaning of the variable (e.g. the variable containing the participants age is named 'age'). Variables of the General Pedagogical Knowledge (GPK) test by König and Blömeke (2009) are named based on the original coding scheme provided by the authors. The documentation differentiates between different types of missing values, which are indicated by the same numeric value throughout all documents and files. The types of missing values are missing specification (indicated with a -9), unclear assignment (indicated with a -8) and missing by design (indicated with a -7). For the background survey and the knowledge tests, which have been completed by the participants, a missing specification is further distinguished in skipped questions with valid answers afterwards (similar to a missing specification indicated with a -9) and the termination of the survey/test, when there are no valid answers afterwards (indicated with a -10).

Second, all variables are explained in a codebook. The codebook links the variable names to the original question, refers to the file in which the variable is stored, explains

manifestations, displays their frequencies and provides further annotations if needed. For some variables an English translation of the original question and the manifestations meanings is provided. In case of string variables, the codebook provides examples of manifestations. In case of metric variables, the codebook provides the variables minimum, maximum, median, mean and standard deviation as well as the frequencies of different types of missing values. In case of the knowledge tests, manifestations are labelled 'Option 1' to 'Option 4'. The content of these options can be found within a collection of surveys that is the third part of the documentation.

sc1 **Type:** Numeric, Ordinal **File:** communication_glo.csv

Original Question:
Die Testperson zeigt in den Gesprächssituationen insgesamt gute kommunikative Fähigkeiten.

Translated Question:

Freq.	Value	Original Label; *(Translation)*
4	-9	Keine Angabe; *(Not Specified)*
2	-8	Keine eindeutige Zuordnung möglich; *(No Clear Assignment)*
34	1	Trifft nicht zu; *(Not Applicable)*
152	2	Trifft eher nicht zu; *(Rather Not Applicable)*
229	3	Trifft eher zu; *(Rather Applicable)*
127	4	Trifft zu; *(Applicable)*

Annotation:
This variable refers to the general assessment of strategic communication.

Figure 2. Example of Variable Description in the Codebook

Third, a collection that contains all surveys and tests used. The single questions within the surveys are supplemented with a reference to the name and storage location of the according variable. In case of the knowledge tests, the surveys include a comment on the manifestations of single test items.

The full documentation with information and data on all applied surveys and tests is available for scientific use at the chair of Professor Edith Braun at the University of Giessen. Furthermore, a shortened version of the documentation has been published at the repository of the Justus-Liebig-Universität Gießen's library (Walz et al., 2022). This version does not include data of the knowledge tests and the information about rater's and confederates' identity pseudonymized as well as information about the institutions of higher education.

3.4 Analysis of Gender, Age and Experience of Rater and Confederates

Participants of a communicative situation interact by interpreting and answering not only with verbal contributions. Instead, they interact with their whole behavior and presence. For example, in a face-to-face interaction non- and paraverbal behavior

accompanies the verbal contribution and provides further clues on how to interpret a message. Both, verbal and para- or nonverbal communication can be a conscious and goal-oriented part of communication in the sense of communication competence as defined in Chapter 2.2. Furthermore, the communication can be influenced by individual characteristics beyond behavior, like the gender or age of the communication partners (Hargie, 2022; Vermeiren et al., 2022). Other than verbal, para-, and nonverbal communication, characteristics like gender or age cannot be a conscious and goal-oriented part of communication in the sense that gender and age can not be altered, learned and applied to performance in the same way as verbal, para-, and nonverbal communication can be learned. Instead, their influence within a communicative situation is based on mainly unconscious and automatic processes of categorization and activation of stereotypes. When encountering another person, this person is very likely to be categorized based on simple and external characteristics like gender, age or race. Attributes that are associated with these categories can influence the behavior towards and the evaluation of a person (Degner et al., 2009).

3.4.1 Age, Gender and Experience in the KomPrü Instrument

For the case of the KomPrü role-play instrument, there are three groups of people involved. First, the participant, who's communicative competence is assessed. Second, the trained confederate, who the participant interacts with. Third, the rater who observes the role-play and assesses the participant's performance with a standardized survey. The rater does not take active part in the communication. Still, the rater observes and assesses the participants behavior and therefore needs to process the observed behavior. During the processing and rating, the same stereotypical and social categorizations might be activated and influence the result of the test. In this sense, the previous considerations about the influence of stereotypes on behavior apply for raters as well as for the two active partners of the communication, which are participants and confederates.

While characteristics like gender and age can have an influence on performance and process of a communication, they are not part of the theoretical framework communication competence, or the role-play instrument grounded on this framework. Consequently, neither the role-play instructions nor the observation sheets of the KomPrü instrument capture gender or age-based evaluation or behavior and consequently gender or age of the communication participants should not have an impact on the assessment. This assumption has not been tested empirically so far. I want to address this gap by systematically testing

gender and age effects of rater and confederates on the result of the assessment. Furthermore, I include a third characteristic that can lead to a change in evaluation and behavior of rater and confederate but is not associated to the participants' communication competence. This is the raters and confederates experience with the role-play test instrument. During their involvement in the role-play rater and confederates may establish situation specific categorizations and stereotypes. Furthermore, rater might assess the participants behavior in comparison with previous participants behavior. Again, situation or instrument specific categorizations, stereotypes or comparisons between different participants are not part of the theoretical framework.

3.4.2 Empirical Framework and Operationalization

This analysis includes four concepts which are (1) gender and (2) age of rater and confederate, (3) experience of rater and confederate, and (4) the result of the assessment.

Information about the concepts of (1) gender and (2) age could be retrieved for some rater (gender: all rater; age: 9 of 14 rater) and confederates (gender: all confederates; age: 10 of 12 confederates) in close cooperation with researchers from the KomPrü research group.

I operationalized the concept of (3) experience by drawing on participants ID. Each participant was allocated an ID-number starting with 1 for the first person participating in the test, 2 for the second and so forth. Furthermore, the rater and confederate for each role-play performance was documented and can be linked to the participant's ID. Rising ID-numbers of participants that have been observed by the same observer or interacted with the same confederate can therefore be used to count the number of assignments of a particular rater or confederate and indicate how many role-plays this rater or confederate has attended at a given assignment. This number of attended role-plays is used as an indicator for a rater's or confederate's experience with each role-play and with the role-play instrument.

To estimate the (4) results of the assessment on the level of role-plays as well as on the level of the two types of communication, I computed six values for each individual student of the sample. Four of these values are the mean values of the variables corresponding to each role-play, the student participated in. Two further values are based on the instruments structure that the total set of all ten role-plays is partitioned in the two types of strategic and understanding-oriented communication: For each such type I

computed one value that is the mean value of all variables corresponding to the union of the role-plays of that type.

To analyze a potential influence of rater's or confederates' gender on the result of the assessment I tested, whether the results of the role-play test did show any significant differences, when grouped by (1) rater's or confederates' gender, (2) participants gender, when they are observed by or interacted with a female or with a male rater or confederate, (3) same gender and mixed-gender parings of rater or confederate and participant. I applied a Mann-Whitney-U test. This test uses sums of ranks to determine differences between the mean rank of two groups. Hence, each observation is assigned with a rank, where the smallest observed value is assigned with the rank 1, the second-smallest observed value is assigned with the rank 2, and so on until the highest observed value is assigned with the highest rank. Ranks of both groups are than summed up and the test statistic U and a z-value are computed to determine, whether an observed difference in the sum of ranks is statistically significant (Döring & Bortz, 2015; Kraska-Miller, 2014). Furthermore, the probability that the rank of a random observation from the first group is larger than the rank of a random observation from the second group, indicates direction and effect size of an observed difference (StataCorp, 2017). If this probability equals 0.5, there is no or – in case of small deviations around 0.5 – only a small difference between the two groups. If the probability is significantly higher than 0.5, this means that the ranks of the first group are expected to be higher than the ranks of the second group. If the probability is significantly lower than 0.5, this means that the ranks for the first group are expected to be lower than the ranks of the second group. The Mann-Whitney-U test can be applied for groups and variables with a non-normal distribution or small sets of data (Kraska-Miller, 2014). The grouped data of the role-play sample is both, non-normal distributed and contains small sample sizes due to the complex design and different numbers of female and male rater and confederates. U- and z-values, are reported in the results of this analysis as well as the probability that the rank of a random observation from the first group is larger than the rank of a random observation from the second group. In the case of the role-play instrument a probability > 0.5 indicates, that the first group is assessed with a higher competence than the second group on average, while a probability < 0.5 indicates, that the first group is assessed with a lower competence than the second group.

Furthermore, I tested whether the age or experience of the rater or confederate is associated with the result of the role-play assessment by using Pearson correlation. For the

case of rater's and confederates' experience, I tested the influence of experience on the results separately for their experience with the single role-plays and with the role-play instrument in general.

I used the total KomPrü sample as described in Chapter 3.1 for this analysis.

3.4.3 Results

Influence of Raters and Confederates Gender

Results for the examined differences based on raters and confederates' gender are presented in Table 4 and Table 5. The tables present four results for both types of communication as well as for each role-play scenario. These results are structured by the comparisons between groups with a female or a male rater or confederate (line 1); between groups with a female or a male participant, where the rater or confederate is female (line 2) or male (line 3); and between same-gender and mixed-gender pairings (line 4). Overall, most results did not indicate significant differences between the tested groups. Few significant results can be found for the rater's gender in understanding-oriented communication and the role-plays 1, 2, 8, 9, and 10, as well as for confederates' gender in strategic communication and role-plays 3 and 8.

Rater's gender is observed to have a significant influence for the case of understanding-oriented communication, and the role-plays 1, 2, 8, 9, and 10 (see Table 4). As for all of these cases male rater rate only a minority of observations, those results should be interpreted with care. For the case of understanding-oriented communication, a significant difference between the assessment of male and female raters is observed ($U_{N_{female\ rater} = 503;\ N_{male\ rater} = 35} = 6908.5$; $z = -2.130$; $p < 0.1$). The probability for a female rater to assess the participants communication competence higher than a male rater is 39.2%. For the case of role-plays 1, 2, 8, 9, and 10, significant differences between the assessment of male and female raters is observed (roleplay 1: $U_{N_{female\ rater} = 205;\ N_{male\ rater} = 21} = 1461$; $z = -2.432$; $p < 0.05$; role-play 2: $U_{N_{female\ rater} = 212;\ N_{male\ rater} = 11} = 514.5$; $z = -3.152$; $p < 0.05$; role-play 8: $U_{N_{female\ rater} = 205;\ N_{male\ rater} = 14} = 1045.5$; $z = -1.7$; $p < 0.1$; roleplay 9: $U_{N_{female\ rater} = 200;N_{male\ rater} = 16} = 776$; $z = -3.432$; $p < 0.01$; roleplay 10: $U_{N_{female\ rater} = 205;\ N_{male\ rater} = 17} = 1211.5$; $z = -2.1$; $p < 0.05$). For all of those role-plays it is less likely for a female rater to give a higher assessment than a male rater (role-play 1: 33.8%, role-play 2: 22.1%; role-play 8: 36.4%; role-play 8: 24.2%; role-play 10: 34.8%).

Table 4. Differences in Ranksum for Female and Male Rater

		N(f)	U	z	P(f > m)
Strategic	*Rat.*	537 (502)	7524.5	-1.420	0.428
	Rat. f.	492 (278)	28013.5	1.108	0.529
	Rat. m.	34 (21)	126	0.372	0.538
	=	526 (291)	32630	0.902	0.523
Role-Play 1	*Rat.*	226 (205)	1461.5	-2.432**	0.339
	Rat. f.	198 (108)	4670	0.475	0.520
	Rat. m.	20 (15)	34.5	0.267	0.540
	=	218 (113)	5880	0.113	0.504
Role-Play 2	*Rat.*	223 (212)	514.5	-3.152**	0.221
	Rat. f.	210 (123)	5109	-0.562	0.477
	Rat. m.	11 (6)	15	0	0.500
	=	221 (128)	5607.5	-0.741	0.471
Role-Play 3	*Rat.*	204 (190)	1155	-0.822	0.434
	Rat. f.	182 (99)	3596	1.449	0.562
	Rat. m.	14 (9)	21.5	0.134	0.478
	=	196 (104)	4226.5	1.408	0.558
Role-Play 4	*Rat.*	207 (193)	1044.5	1.418	0.613
	Rat. f.	191 (109)	3887.5	1.540	0.565
	Rat. m.	13 (7)	16	-0.718	0.381
	=	204 (115)	4408.5	1.698	0.569
Role-Play 5	*Rat.*	215 (205)	824.5	-1.048	0.402
	Rat. f.	204 (118)	4804	-0.651	0.473
	Rat. m.	10 (5)	10	0.532	0.600
	=	214 (123)	5248	-0.781	0.469
Understanding-oriented	*Rat.*	538 (503)	6908.5	-2.130*	0.392
	Rat. f.	493 (278)	28632.5	-0.799	0.479
	Rat. m.	34 (21)	118	-0.656	0.432
	=	527 (291)	32893.5	-0.831	0.479
Role-Play 6	*Rat.*	201 (191)	951	-0.022	0.498
	Rat. f.	191 (112)	4136.5	-0.766	0.468
	Rat. m.	10 (7)	8	-0.577	0.381
	=	201 (115)	4706	-0.587	0.476
Role-Play 7	*Rat.*	216 (203)	1256	0.291	0.524
	Rat. f.	201 (105)	4818.5	0.538	0.552
	Rat. m.	12 (7)	14.5	-0.490	0.414
	=	213 (110)	5402	0.586	0.523
Role-Play 8	*Rat.*	219 (205)	1045.5	-1.700*	0.364
	Rat. f.	197 (122)	3862.5	-1.835*	0.422
	Rat. m.	14 (10)	18.5	0.213	0.537
	=	211 (126)	4446.5	-2.091	0.415
Role-Play 9	*Rat.*	216 (200)	776	-3.432***	0.242
	Rat. f.	197 (108)	4578.5	0.572	0.524
	Rat. m.	15 (7)	21.5	0.757	0.616
	=	212 (116)	5383.5	0.416	0.517
Role-Play 10	*Rat.*	222 (205)	1211.5	-2.100**	0.348
	Rat. f.	198 (111)	4798.5	-0.075	0.497
	Rat. m.	17 (11)	28	0.525	0.576
	=	215 (117)	5537	-0.434	0.483

*Legend: Rat. – Rater; f. – female; m. – male; N – number of observations; (f) – Numer of female observations; U – Whitney-U-statistic, z – value of z; P(f > m) – probability that a value from the female group is larger than a value of the male group; * – p < 0.1; ** – p < 0.05; *** – p < 0.01*

Furthermore, for the case of role-play 8, there is a difference observed for female rater, depending whether a female or a male participant is assessed ($U_{N_{female\ participant} = 122;\ N_{male\ participant} = 75} = 3862.5$; $z = -1.835$; $p < 0.1$). The probability of a female rater, to assess a female participant with a higher competence than a male participant is 42.2%.

Confederates' gender is observed to have a significant influence for the case of strategic communication, and the role-plays 3, and 8 (see Table 5). For the case of strategic communication, female participants who interact with female confederates are assessed with higher communicative competence than male participants who interact with female confederates ($U_{N_{female\ participant} = 90;\ N_{male\ participant} = 63} = 2350.5$; $z = 1.796$; $p < 0.1$). The probability for a female participant of being assessed with higher competence than a male participant is 58.5%. For the case of role-play 3, a significant difference between the assessment of participants based on the gender of the confederate is observed ($U_{N_{female\ confederate} = 65;\ N_{male\ confederate} = 140} = 3293$; $z = 3.183$; $p < 0.05$). The probability of a higher assessment of competence for a participant interacting with a female confederate is 63.8% and therefore more likely than a higher assessment of competence for a participant interacting with a male confederate. Furthermore, for the role-plays 3 and 8, a difference is observed between the assessment of participants interacting with a female confederate based on the participants gender (role-play 3: $U_{N_{female\ participant} = 32;\ N_{male\ participant} = 30} = 348.5$; $z = 1.875$; $p < 0.1$; role-play 8: $U_{N_{female\ participant} = 35;\ N_{male\ participant} = 17} = 208$; $z = -1.747$, $p < 0.1$). The probability for a higher assessment of competence for a female participant interacting with a female confederate a male participant interacting with a female confederate is 63.7%. For role-play 3 it is more likely for a female participant to be assessed higher than a male participant when interacting with a female confederate (63.7%). For role-play 8 it is less likely for a female participant to be assessed with a higher competence than a male participant when interacting with a female confederate (35%).

Table 5. Differences in Ranksum for Female and Male Confederates

		N(f)	U	z	P(f > m)
Strategic	*Con.*	538 (157)	27927	1.209	0.533
	Con. f.	153 (90)	2350.5*	1.796*	0.585
	Con. m.	374 (209)	17105	0.132	0.504
	=	527 (255)	32938.5	0.997	0.525
Role-Play 1	*Con.*	226 (68)	5227.5	-0.322	0.487
	Con. f.	64 (40)	371.5	1.511	0.613
	Con. m.	154 (83)	2898.5	-0.157	0.492
	=	218 (111)	5469	1.013	0.540
Role-Play 2	*Con.*	223 (64)	5046	0.097	0.504
	Con. f.	64 (37)	435.5	0.880	0.564
	Con. m.	157 (92)	2646	-1.237	0.442
	=	221 (102)	5365.5	1.498	0.558
Role-Play 3	*Con.*	205 (65)	3293**	3.183**	0.638
	Con. f.	62 (32)	348.5*	1.875*	0.637
	Con. m.	135 (76)	2129	0.502	0.525
	=	197 (91)	4560.5	0.659	0..527
Role-Play 4	*Con.*	207 (60)	4376.5	0.086	0.504
	Con. f.	59 (37)	361	0.722	0.557
	Con. m.	145 (79)	2294.5	1.242	0.560
	=	204 (103)	4938	-0.626	0.475
Role-Play 5	*Con.*	216 (58)	4201	-0.940	0.458
	Con. f.	58 (35)	397	-0.088	0.493
	Con. m.	257 (88)	2864.5	-0.609	0.472
	=	215 (104)	5605	0.368	0.514
Understanding-oriented	*Con.*	539 (157)	29538.5	-0.273	0.493
	Con. f.	153 (90)	2758	0.285	0.514
	Con. m.	375 (209)	16181	-1.118	0.466
	=	528 (256)	33119.5	0.968	0.524
Role-Play 6	*Con.*	201 (63)	4066	-0.737	0.468
	Con. f.	63 (42)	399	-0.614	0.452
	Con. m.	138 (77)	2225.5	-0.529	0.474
	=	201 (103)	5040	-0.017	0.499
Role-Play 7	*Con.*	217 (71)	4816.5	-0.845	0.465
	Con. f.	70 (39)	497.5	1.266	0.589
	Con. m.	144 (73)	2559.5	-0.128	0.494
	=	214 (110)	5432.5	0.636	0.525
Role-Play 8	*Con.*	220 (55)	4326	-0.518	0.477
	Con. f.	52 (35)	208*	-1.747*	0.350
	Con. m.	160 (97)	2816.5	-0.836	0.461
	=	212 (98)	5465	-0.272	0.489
Role-Play 9	*Con.*	216 (59)	4148	1.184	0.552
	Con. f.	58 (29)	362	0.912	0.570
	Con. m.	154 (86)	2908	0.058	0.503
	=	212 (97)	5346.5	0.520	0.521
Role-Play 10	*Con.*	222 (67)	4915.5	-0.635	0.473
	Con. f.	64 (37)	496	-0.048	0.496
	Con. m.	151 (85)	2743.5	0.232	0.511
	=	215 (103)	5661	-0.236	0.491

*Legend: Con. – Confederate; f. – female; m. – male; N – number of observations; (f) – Number of female observations; U – Whitney-U-statistic, z – value of z; P(f > m) – probability that a value from the female group is larger than a value of the male group; * – p < 0.1; ** – p < 0.05; *** – p < 0.01.*

Influence of Raters and Confederates Age

Results for the influence of raters and confederates age on the assessment of the participants' communication competence are presented in Table 6. There are only very few significant results. For the case of rater's age, the assessment of understanding-oriented communication competence is lower, the older the rater is ($r = -0.18$, $p < 0.05$). The same shows for the role-plays 6 and 10 (role-play 6: $r = -0.27$, $p < 0.05$; role-play 10: $r = -0.18$, $p < 0.05$). For the case of confederates' age, the assessment of role-play 8 is lower, the older the confederate is ($r = -0.12$, $p < 0.1$). All significant correlation coefficients are small ($r = -0.12$ to $r = -0.27$) and indicate, therefore, no strong association between raters and confederates age and the competence assessment.

Table 6. Correlation Between Role-Play Results and Age

	Rater		Confederate	
	N	**r**	**N**	**r**
Strategic	289	- 0.06	532	0.04
Role-Play 1	121	- 0.13	223	- 0.10
Role-Play 2	124	- 0.14	222	0.09
Role-Play 3	101	0.09	201	0.06
Role-Play 4	117	0.13	206	0.11
Role-Play 5	116	- 0.05	213	- 0.02
Understanding-Oriented	289	- 0.18 **	535	0.02
Role-Play 6	105	- 0.27 **	198	0.06
Role-Play 7	125	- 0.15	214	0.01
Role-Play 8	102	- 0.03	219	- 0.12 *
Role-Play 9	123	- 0.13	215	0.06
Role-Play 10	122	- 0.18 **	218	0.04

*Legend: N – number of observations; r – correlation coefficient; * – p < 0.1; ** – p < 0.05; *** – p < 0.01.*

Influence of Raters and Confederates Experience

Results for the influence of raters and confederates experience on the assessment of the participants' communication competence are presented in Table 7. The results differentiate between the raters and confederates experience with the role-play instrument in general and with the ten different role-plays explicitly. As indicator for experience I use the number of role-plays a rater or confederate has observed or interacted within so far. There are only few significant results. For the case of rater's experience, the assessment of participants communication in role-play 1 is assessed lower, the more experience the rater has ($r_{instrument} = -0.11$, $p < 0.1$). This effect is slightly stronger for explicit experience with this role-play ($r_{role\text{-}play} = -0.12$, $p < 0.1$). Furthermore, rater's explicit experience with role-play 5 is associated with a higher assessment of the participants communication ($r_{role\text{-}play} = 0.12$, $p < 0.1$). For the case of confederates' experience, the assessment of participants communication in role-plays 4 is lower, the more experience the confederate has

($r_{instrument}$ = -0.22, $p < 0.05$), and the assessment of role-play 5 is higher, the more experience the confederate has ($r_{instrument}$ = 0.18; $p < 0.05$). The effect for role-play 4 is even slightly weaker for explicit experience with this role-play ($r_{role\text{-}play}$ = -0.21, $p < 0.05$), while the effect for role-play 5 is slightly stronger for explicit experience with this role-play ($r_{role\text{-}play}$ = 0.19, $p < 0.05$). Again, all significant correlation coefficients are small (r = -0.11 to r = -0.22) and indicate, therefore, no strong association between raters and confederates experience and the competence assessment.

Table 7. Correlation between Role-Play Results and Experience

	Rater			Confederate		
	N	**$r_{instrument}$**	**$r_{role\text{-}play}$**	**N**	**$r_{instrument}$**	**$r_{role\text{-}play}$**
Strategic	543	- 0.07		543	- 0.05	
Role-Play 1	228	- 0.11 *	-0.12 *	228	0.01	0.01
Role-Play 2	224	- 0.10	-0.10	224	- 0.03	-0.03
Role-Play 3	206	- 0.05	-0.10	206	- 0.04	-0.04
Role-Play 4	209	- 0.11	-0.10	209	- 0.22 **	-0.21**
Role-Play 5	219	0.11	0.12 *	219	0.18 **	0.19**
Understanding-Oriented	543	- 0.04		543	- 0.00	
Role-Play 6	202	- 0.01	-0.03	202	0.01	0.05
Role-Play 7	218	- 0.02	-0.00	218	0.00	-0.01
Role-Play 8	221	- 0.03	-0.05	221	- 0.01	-0.01
Role-Play 9	219	0.02	0.05	219	0.03	0.06
Role-Play 10	224	- 0.12 *	-0.11	224	0.01	0.00

Legend: N – Number of observations; $r_{instrument}$ – correlation coefficient between raters/confederates experience with the role-play instrument; $r_{role\text{-}play}$ – correlation coefficient based on raters/confederates experience with a single role-play scenario.

3.4.4 Discussion

Rater and Confederates are important features of the KomPrü role-play instrument. Before they take part in the assessment of students' communication competence, they receive a training based on video recordings of the role-plays to familiarize them with the tasks, how rater should use the observation survey, and which action is required by the confederates (Braun et al., 2016). Hence, rater and confederates should be able to observe and participate in the communication during the role-play within the competence framework of the instrument. Still, as a simulation like a role-play is supposed to be authentic (Chernikova et al., 2020), external and instrument specific characteristics like age, gender or experience that may shape communicative behavior cannot be excluded from the interaction. From a theoretical point of view, these characteristics are not part of the competence framework and therefore not part of the assessment of competence. The previous analysis tested this theoretical assumption empirically.

Overall, neither age, nor gender, nor experience of rater and confederates seemed to be structurally associated with the assessment of the participants communication. There have been some significant effects but these appeared unsystematic and coefficients appeared to be rather small. Hence, it seems plausible, that these effects appeared randomly as the analysis consisted of several tests. Furthermore, most of the effects appeared only on a rather liberal significance level ($p < 0.1$). Applying a great amount of tests together with a liberal significance level can lead to an increase in significant results without empirical validity (Döring & Bortz, 2015).

The previous analysis is a secondary data analysis. When originally gathered, effects of rater or confederates have not been part of the original project. Therefore, the data did not entail information about raters or confederates age or gender, neither an assessment of raters or confederates experience. Raters and confederates age and gender could be added afterwards but with some missing values resulting in rather small groups. Experience has been operationalized relatively crudely by using the ascending order of participants' identification numbers but without including exact times, accumulations and intervals between participating in the test. Furthermore, when testing effects of rater's or confederates' gender, some tested groups have been rather small, due to the complex design of the instrument and the detailed testing of gender effects. Likewise, more sophisticated models like path analysis that entail effects and interactions of more than one characteristic could not be conducted due to the complexity of the assessment and data. However, under these constraints, all results indicate an empirical validation of the assumption that characteristics like age, gender or experience do not influence the assessment. This can be interpreted as a strength and further proof of the instruments validity in the sense that it assesses communication competence as defined in the underlying theoretical framework by Braun, Athanassiou, et al. (2018). This proof of validity enhances the usability of the role-play instrument to assess performance-based competence in the German context of higher education.

In future applications of the instrument, rater and confederate effects could be monitored more closely to test the current results on a more sophisticated level.

3.5 Empirical Methods for Further Examination of the Observation Data

Knowledge and assessment tests consist of a pool of tasks that test the extend, depth or focus of the assessed knowledge or competence. These tasks are designed to assess a variety of different levels of knowledge or competence. Accordingly, some tasks are rather

easy to solve and can be successfully achieved by participants with low knowledge or competence. Other tasks are, in comparison, difficult to solve and can only be achieved by a participant with a high level of knowledge or competence. Consequently, knowledge and assessment tests are associated with two types of competence variables: First, the relative difficulty of the single tasks in a test and second, the knowledge or competence of the person taking the test. Both types of variables can be related to each other (Hartig & Goldhammer, 2010). A task, which can only be solved by a more competent person would be considered as difficult, while a person who can solve difficult tasks would be considered as more competent than a person who cannot solve difficult tasks. Consequently, both types of variables must be examined, when designing and applying a test and assessing a person's competence.

Furthermore, data derived from a knowledge or assessment test can contain missing values because of at least three different reasons. First, values can be missing because of missing knowledge or competence. Second, values can be missing because the participant did not have enough time to complete all tasks. When applying a test, the participants are usually given a fixed amount of time to complete the test. If a person does not manage to reach and try to solve all tasks in the given time, this leads to missing values. These two types of missing values can be caused by the participants' competence, other personal characteristics, or external influences. Hence, it is important, to document the reason of a missing value as exact as possible, e.g., by offering an "I don't know"-option for each task to identify values missing because of missing knowledge or competence. Third, values can be missing because of a multi-matrix design. In this design not all tasks that are part of a test instrument are presented to each participant. Instead, these tests are designed in a booklet design where each booklet contains a subset of the available tasks. This design is called multi-matrix design (Graham et al., 2006). Even though some tasks are included in more than one booklet and presented to many or all participants, the multi-matrix design leads to a certain extend of missing information. When the assignment of participants to the different booklets is random, this missing information can be regarded as missing at random, i.e., the missing is not influenced by personal characteristic or external influences (Hartig & Goldhammer, 2010).

Item response theory (IRT) is a family of quantitative empirical methods that allow the examination of the relation between task and participant (Hartig & Frey, 2013). Furthermore, IRT-models are sensitive to missing values if the data is missing completely

at random. In an IRT-model each task is assigned with a value indicating the task's difficulty and each participant is assigned with a value, indicating the participant's competence. "If an item and an individual are both assigned with the same value, then the probability that this individual solves the corresponding item is at least 0.5. Accordingly, both, individual competence and item-difficulty parameter, are mapped on a joint scale, called theta scale (θ)" (Walz & Braun, 2022). The mean of theta is 0 and the standard deviation is 1 (Raykov & Marcoulides, 2018). Further estimation that is based on the item-difficulty parameter allows a further description of a test and its psychometric characteristics. For example, graphical post-estimations like the Item- or Category-Characteristic-Curves (ICC or CCC), Test-Information-Function (TIF) or Wright-Maps (WM) illustrate the difficulty and distribution function of a single item, the overall quality of the test for different areas of competence and the distribution of items along the theta-scale (Boone et al., 2014; Raykov & Marcoulides, 2018). For example, an ICC with a steeper increase can be regarded as more distinctive than an ICC with a less steep increase. A WM reveals which areas of the theta-scale are covered by the test's items (i.e. an area of the theta-scale is covered, if there are items assigned to that area) as well as which seem to be neglected (i.e. there are areas without an assigned item), or over-representated (i.e. there are areas with a lot of assigned items) (Boone et al., 2014). These results are valuable for interpreting the findings and improving the test (e.g. adding more items with for uncovered areas of the theta-scale).

There are different IRT models that can be applied, depending on the test, data, and pursued research aim. Tasks with a clear separation between correct and incorrect answers can be analysed with a dichotomous IRT model like the classical Rasch model (Baker, 2001; Raykov & Marcoulides, 2018). Less complex models include only different difficulties of items, more complex models include furthermore a different discriminatory power of items or the chance of a participant simply guessing an answer (Geiser & Eid, 2010). Tasks that allow for partial credit (e.g. a participant in a multiple-choice test ticks of some but not all correct answers or some correct and some incorrect answers) or a graded response (e.g. a Likert Scale) can be analysed with a polytomous IRT-model that is based on Samejimas approach (Embretson & Reise, 2013; Ostini & Nering, 2006). The decision for a specific model depends on the research aim and design but also on the fit of model and data. Aikakes and Bayesian information criterion are two common indicators to

compare the fit of different models and the empirical data. The model with the smallest values for both criterions can be considered to fit the data best.

After computing a fitting model and assigning the difficulty and discrimination parameter to each item, the participants competence in relation to the samples mean competence can be estimated. These competence values draw on information about the different item difficulty and are therefore a more precise assessment of participants performance than a simple sum- or mean score of the raw data (Boone et al., 2014). Still, the participants estimated competence values can only be interpreted in relation to the sample. They have to be estimated separately for different samples (e.g., when analyzing only a sub-sample).

The role-play instrument uses observation surveys with items in a graded Likert format to assess the participants behavior. For two contributions that analyze this observation data, a separate polytomous IRT model has been computed, based on the specific sub-sample. For both analyses a graded response model (GRM) could be computed. This model is fit for ordered categorical answers like the Likert format and furthermore showed the best fit to the data according to AIC and BIC (for AIC and BIC results of these analyses see Chapter 6.3 and Chapter 6.4).

One of the two contributions focuses on the computed item parameters and applies further post-estimation to describe characteristics of the overall test as well as of single items and collection of items (see Walz & Braun, 2022). The other contribution uses the estimation of participants competence as a variable in different regression analyses. Furthermore, this contribution applies dichotomous IRT-model to estimate participants competence values for a subject specific knowledge test (see Walz et al., under review).

4. DISCUSSION OF RESEARCH GOALS AND CUMULATIVE CONTRIBUTIONS

This dissertation started by introducing the research field of higher education outcomes and in particular the area of training and assessing performance competences. In light of global changes competence-orientation in general and the 21st century performance-based skills in particular become important higher education outcomes (Allen et al., 2013; Braun et al., 2020). These performance-based competences need an appropriate method of assessment (see Chapter 2.3). Simulations are an adequate but complex assessment method for performance-based competences (Chernikova et al., 2020). Role-play methods are simulations with a strong focus on interpersonal communication (see Chapter 2.4). Interpersonal and communicative competences play an important role within the frameworks of 21st century skills (Voogt & Roblin, 2012). By examining the links between theoretical and empirical approaches to conceptualize communication competence, this dissertation aimed at gaining a better understanding of the nature of these competences in order to teach and assess them based on an evidenced role-play instrument. Furthermore, the employed assessment instrument and data sample has been thoroughly described for the first time so it could be made accessible to other researchers. This furthermore supports the transparency and replicability of the current results in the sense of open science and FAIR data management (Wilkinson et al., 2016).

The foundations to address the research goals as well as first results have been presented in the chapters before the contributions by providing a conceptual framework based on theoretical considerations and methodological applications (Chapter 2), by thoroughly presenting the KomPrü instrument and by examining its validity and supplementing it with a detailed documentation (Chapter 3). The four presented cumulative contributions aim at distinct research questions. They each present and discuss key findings relating mainly to these questions. While these questions are all connected with the two research goals of this dissertation, none of the cumulative contributions have been discussed together and in light of these higher-level research goals. The following discussion aims at relating the individual results and provide the necessary discussion to address the two research goals of this dissertation.

In a first step, both research goals are addressed separately. In a second step limitations of this dissertations approach and findings are presented, along with a brief idea of how to improve in future research. In a second step, connections between both goals as

well as open and further research questions that arise from the current results are presented. In a fourth step I reflect on the research process of this dissertation on a more general level and under special consideration of data management standards.

4.1 Discussion of the Potential of Role-Plays to Assess Performance-Based Competence in Higher Education

The first goal of this dissertation has been seeking to contribute to the thriving research field of assessing higher education learning outcomes by gaining a deeper understanding of role-plays as a method to assess performance based-competences in institutions of higher education. This goal has been addressed generally by identifying characteristics and areas of implementation of role-plays as well as in context of a particular instrument by empirically examining external characteristics of a person like gender or age during the analysis of these characteristics in context of the KomPrü instrument. Furthermore, all four cumulative contributions add to this first goal by intertwining theoretical foundations with empirical application and analysis.

I introduced seven characteristics that reveal the broad scope of role-play application in the context of higher education in Chapter 2.4. On the one hand, this certainly implies the adaptability of role-plays in a broad variety of higher education settings and makes them a powerful method to train and assess performance-based competences. On the other hand, it implies a high complexity in the development of role-play methods. Different characteristics lead to a different learning and testing environment with different outcomes. For example, a role-play where all interaction partners are peer-students will create a different environment than a role-play where students interact with trained confederates. While the first application can create a fruitful and save learning environment, the second increases the level of uncertainty and creates a powerful assessment environment. As a result, implementing a role-play for training or assessment requires careful consideration and adaption of each characteristic to match learning or assessment goals best. This has two implications for the use of role-plays in higher education. First, applications of role-plays for different fields of usage should be developed and tested in higher education research to create validated applications for higher education teachers to use. Development of a role-play should build on theoretical and didactic frameworks and differentiate between subject-oriented and generic parts. The KomPrü instrument in particular is strongly connected to a theoretical framework that allows for generic assessment of communication competence. Applying teacher-education and economics contexts

enhances the instruments approximation of reality and results in an instrument that assesses a generic competence in a subject-specific manifestation. Second, higher education teachers who want to apply role-plays for teaching and assessment need to be aware of the differences in role-play applications and their consequences. They need to be qualified to decide which application fits their requirements. Hence, training for higher education teachers providing the competences to adequately choose and apply role-play in higher education settings is required.

The complexity of role-plays and number of people involved in the assessment furthermore may lead to different biases of assessment results than traditional tests. Role-plays simulate reality and individual characteristics like age and gender that are known to subconsciously bias behavior in real-world settings, may potentially affect the assessment of a participant's performance. The KomPrü instrument included an extensive training of rater and confederates before they participated in the assessment. Furthermore, the items in the observation survey are strongly linked to the underlying theoretical frame. By examining the effects of age, gender as well as experience with the KomPrü instrument, I concluded, that for the case of the KomPü instrument, the assessment of communication competence can be regarded as unbiased by these subconscious parts of communication (see Chapter 3.4). The results indicate that a theoretical based definition of communication competence that is primarily based on conscious and learnable elements can be assessed empirically and in an authentic setting while unconscious parts of a communication like age or gender do not impact the assessment. This emphasizes the previous idea of a strong integration of theoretical frameworks in the future development of role-plays and their application in contexts of higher education.

The four contributions of this dissertation address this first research goal by identifying the broad field of higher education outcomes, by theoretically and empirically examining the usability of the KomPrü role-play test as an assessment instrument for communication competence in higher education.

In the first contribution that was published in the peer-reviewed compendium Oxford Bibliographies two colleagues and me examined the learning outcomes of higher education institutions (Braun et al., 2020). This contribution presents the impact of three main actors within this field: These are scientific actors, policy makers and economic stakeholder. These actors initiate, support and conduct research in national and international projects. We identify several international initiatives and report about ongoing research for means

of assessing outcomes in form of specialized and general skills but also in form of personal and societal returns like better health or more social and political participation (Gross et al., 2019). One important goal of higher education is, to prepare graduates for the labor market but also to shape active citizens with a strong capability to act. In this sense, the mastery of 21st century skills during higher education helps the individual to achieve a meaningful and healthy life, it supplies the labor market and economy with competent employees and leaders, and it enriches the society with individuals' capability to act (Braun et al., 2020). While research around the world highlights the importance of higher education outcome for individuals, labor market and society the article furthermore highlights the complexity of developing teaching and assessment instruments that can be used internationally and across different cultural backgrounds. Even though the 21st century society is globalized and 21st century skills can be considered as global key competences differences in education and training systems make it difficult to provide a "one fits all" solution for teaching and assessing the performance-based 21st century skills as higher education outcomes. Role-plays like the KomPrü instrument may be useful to address this assessment problem by providing instruments to train and assess these performance-based competences based on a theoretical framework.

This thought is furthermore emphasized in the second contribution, which was published in the peer-reviewed edited volume "Student Learning in German Higher Education". This contribution addresses the research goal by connecting the theoretical and holistic understanding of education and education outcomes ("Bildung") with empirical role-play assessment methods. This aims at bridging an alleged gap between complex theoretical understanding of performance skills and its empirical conceptualization and assessment (Falkenstern et al., 2020). We constituted the complexity and learnability of performance-based competences in a holistic transformative learning environment. With this in mind, we evaluated the KomPrü instrument as one example of a simulation-based approach of communication assessment. The theoretical considerations and empirical validations lead to the conclusion that simulations in general and the assessed KomPrü role-play instrument in particular, can be a useful method assess performance-based competences like communication in the context of higher education. The contribution highlights the benefits of a strong theoretical background in the development of a role-play instrument that is supposed to assess performance-based competences in higher education.

The third cumulative contribution, which was published in the peer-reviewed journal "Higher Education Forum" (Walz & Braun, 2022) addresses the research question by further strengthening the application of theoretical facets in a role-play instrument. We examined if the tasks of the simulation can be distinguished by difficulty in a way that individuals that are more competent succeed at more difficult tasks than individuals that are less competent. Interestingly these differences did not appear related to different situations of communication but related to the theoretical aspects that apply to the observation of all simulated situations. This emphasizes the advantage of a role-play instrument with a strong theoretical background: The competence level is not distinguished by the specificity of a single situation but by the general parts of the competence that are applicable to all tested situations. Furthermore, by providing a detailed analysis of the results the contribution outlines relations between different facets (e.g., by comparing their difficulty levels) and presents their interaction when describing communication. Based on these findings, we developed a competence level model that can be used to describe an individual's communication competence as assessed with the role-play instrument (Walz & Braun, 2022). This adds to the usability of the KomPrü role-play instrument as it allows for detailed feedback. Feedback has been presented as an important characteristic of role-play application in the context of higher education in Chapter 2.4. Furthermore the results of this contribution emphasize again the importance of a theoretical background for usability of role-plays in higher education for the assessment of performance-based competences.

Finally, the fourth cumulative contribution, which is currently under review for publication in a peer-reviewed journal, addresses the research question by synthesizing the previous results and applying it to the practical field of interactions between knowledge, performance skills and visible higher education results. A previous contribution (Falkenstern et al., 2020) constitutes higher education as important learning and transformation environment. Following this idea, this fourth contribution empirically examines the higher education environment in context of learning gains in communication competence (Walz et al., under review). This fourth contribution builds on established conceptualizations of teaching competence and is compatible with current research in this field. It implies invisibility of performance skills within higher education curriculums in Germany but the secondary data analysis with a small sample size allowed only for simple

operationalizations and analysis. Therefore, the contribution suggests further and more complex research to understand and verify the current results in depth.

All results discussed so far can furthermore be linked to a downside of role-play application in higher education contexts: Role-plays are a very complex and cost-intensive method. This downside is also applicable to the KomPrü instrument. Following the previous thoughts, the costs appear at three different areas.

First, it is expensive to develop and adjust role-play applications for the context of competence assessment in higher education. The development of role-play applications should therefore be assigned to dedicated research groups and oriented at requirements derived from curricula or frameworks of higher education outcomes. The KomPrü role-play instrument has been developed by a research group and addressed communication-competence, which is an important learning goal of higher education in general and for teacher and economics students in particular (Braun, Athanassiou, et al., 2018). Even though the research group did extraordinary work in developing the theoretical framework, the role-play instrument and provided first empirical validations, they did not present a detailed manual or documentation for applying the test in a higher education assessment setting. This dissertation provided some of these parts. Nevertheless, applying the KomPrü role-play method in an assessment context requires a lot of knowledge about the instrument and its theoretical background. This is also linked with the second area of complexity and costs.

Second, it is time- and money-intensive to qualify higher education teachers to adequately choose and apply a role-play method. It may be worth to examine if qualification of role-play-application in general could be integrated in the curriculum of established further education programs in higher education didactics (e.g. the "Hochschuldidaktisches Netzwerk Mittelhessen"). For the specific case of the KomPrü role-play instrument higher education teachers who want to apply the method need to make contact with members of the former research group or use the available published material. This dissertation added to this material with the publication of research articles and research data.

Third, it can be resource-intensive to apply a role-play. For example, the KomPrü role-play instrument requires the training and assignment of observer and rater. Furthermore, the assessment takes longer than traditional assessment methods.

These different costs need to be weighed against the expected gains and if the gains prevail the costs, the resources need to be provided. This weighing cannot be done by higher education teachers because they may not mandate all of the required resources. Instead, the weighing and mandating of resources needs to be anchored at the level of higher education institutions curricula and at the level of national frameworks of higher education outcomes. Following this thought I evaluate role-plays as a valuable but complex method for assessment of performance-based competences in higher education. This is true especially for role-plays with a strong theoretical background. In order to fulfil their potential, the application of role-plays need to be anchored, based on evidence of the applications impact.

4.2 Discussion of FAIR Principles and Documentation of Scientific Data of the KomPrü Instrument and Data Sample

The second goal of this dissertation has been to improve the reusability of the KomPrü role-play instrument as well as the KomPrü data under consideration of the FAIR-principles. This goal seeks to facilitate the necessity of further data analysis and instrument application, which was required separately in context of different contributions. In the sense of the FAIR-data principles and the expansion of open science, this dissertation seeks to contribute to potential future analysis and application by providing improvements and validation of the instrument as well as a comprehensible data documentation. Accordingly, the analysis of rater and confederates age, gender and experience as well as the development of the competence level model support the applicability of the role-play instrument, while the detailed documentation and publication of data provides the means necessary to further analyze this sample.

The theoretical framework of the KomPrü instrument combines several facets of communication theory to conceptualize communication. In combination with an authentic assessment method, this allows a holistic observation of communication competence (Falkenstern et al., 2020). Still, the theoretical framework cannot entail all aspects and influences that shape the process of a communication, and the framework emphasizes those aspects that are considered as conscious and learnable skills. The analysis of raters and confederates age, gender and experience address the aspect of unconscious images that can influence communication and its evaluation (Degner et al., 2009) and tests whether these images exert influence on the assessment of participants' performance. The results have been interpreted as not indicating any systematic significant influence of the tested characteristics. Thus, the instrument is considered robust against these aspects.

Accordingly, it can be assumed, that the KomPrü role-play instrument validly assesses communication competence as conceptualized in the theoretical framework. Within the discussion of the results I pointed out that the sample did not allow for elaborate operationalization or methods of analysis. In order to validate the results and to enhance the instruments quality for reuse it could be required to repeat the analysis with a better database and more elaborate methods to validate the results of potential influence of gender and experiences. Still, the current results can be interpreted as a first empirical validation of the instruments quality and robustness.

The instrument is further improved with the development of a competence level model in the second cumulative contribution (Walz & Braun, 2022). First, the competence level model localizes the area of competence assessed by the instrument. On the one hand, this allows for an accurate application of the instrument. On the other hand, by indicating which areas of competence can be assessed accurate and which cannot be assessed accurate, the finding provides a focus for possible further development of the instrument. Second, the competence level model provides descriptors to qualitatively distinguish and describe different levels of competence. This facilitates a more valuable feedback and evaluation of student's communication competence. Without the qualitative descriptors the result of the assessment would only state that one student is assessed with more or less competence than another. With the qualitative descriptors this feedback can be enriched with detailed levels of competence regarding different facets of communication like the ability to achieve content or relational goals. By using the qualitative descriptors students and teachers can identify which areas of communication competence need more training and which are already mastered. Third, the competence level model ties in with the theoretical framework but also with the design of the rating scales and strengthens the relations between theoretical framework, surveys and interpretation. These three contributions improve the instruments applicability.

Not only the KomPrü-instrument but also the empirical data collected by applying the role-play as well as further knowledge tests and self-assessments can be regarded as a valuable result of the KomPrü research group for future scientific work. Still, this sample was not processed and documented to be available for future research in a sufficient way. The FAIR principles are an internationally recognized collection of data management standards that facilitate re-usability of data and transparency of research processes

(Wilkinson et al., 2016). The documentation of the KomPrü data under consideration of the FAIR principles is another result of this dissertation.

The documentation of the KomPrü data has been enriched with meta-data like a title, authors names and affiliations, a date of publication, and a persistent identifier, making them findable and citable. The documentation of KomPrü data is accessible through the repositorium for research data of the Justus-Liebig-University Gießen, the JLUdata and licensed for reuse under CC-BY-SA conditions. In general, CC-licenses are a useful and established tool for standardized sharing of data. Simple CC-0 or CC-BY licenses facilitate the most open permission of reuse with a low threshold. A CC-BY-SA license limits this openness and reusability in the sense that the reused data has to be published under the same conditions (Creative Commons, no year). In case of the KomPrü data this limit seems reasonable as the data belongs to a new developed instrument that may benefit from further development and reuse of data under the same conditions.

The documentation includes non-proprietary and interoperable data files, a user-friendly codebook with distributions of frequencies, and original surveys that link the originally used wordings and presentation of single questions to their according variables in the data files. All three elements of the documentation follow the same structure that is explained to the user in codebook's and survey collection's introduction and makes the documentation reusable for other scientists. It is a minor limitation that not the entire KomPrü data is accessible, as the original data contains sensitive data from knowledge tests. This data is only accessible for internal use at the Professur für Hochschuldidaktik mit dem Schwerpunkt Lehrerbildung. Still, by providing the extensive documentation the core elements of the KomPrü data sample are available for further research. Hence, the presented documentation facilitates independent examination of the KomPrü data as it is requested in the standards of FAIR data and open science. By establishing and providing the documentation, the dissertation also contributes to the further recognition and establishment of these standards in the area of higher education research.

4.3 Limitations

However, there are some limitations in addressing the research goals. First, the presented literature and the derived characteristics of role-plays in higher education in Chapter 2.4 are the result of an unstructured and mostly explorative literature research. The resulting characteristics and examples cover a broad variety of applications and contexts, but the results could be supported by a more systemize literature review. Following, the

presented characteristics are not exhaustive but provide examples that give an impression of the broad range of role-play applications. The presentation of the seven characteristics as well as the presented variety of implementation could be used as a starting point for a more in-depth literature review of role-play applications in higher education. Furthermore, the examples of role-play characteristics did serve the goal to describe and distinguish the KomPrü role-play instrument and highlight the features of this particular approach.

Second, for the secondary data analysis of the cumulative contributions as well as for the analysis of age, gender and experience, all concepts had to be operationalized with the available data. Some analysis, like the testing of age, gender and experience effects of rater and confederates in Chapter 3.4 or the indicators of student's experience in the fourth contribution (Walz et al., under review) had not been an original goal of the KomPrü data collection. Therefore, the information either had to be added manually or had been operationalized rather crudely. Age and gender information could be added for most but not all rater and confederates, resulting in an even smaller sample size. Following the small and unequal number of different rater and confederates, the approach of analyzing age, gender effects relies on very simple statistical analysis. The same holds for the crude operationalization of raters, confederates (Chapter 3.4) and participants experience (Walz et al., under review), so that more complex statistical methods of analysis could not be applied. Still, the complete missing of effects in a significant pattern can be interpreted as an indicator of the instruments robustness against age, gender and experience effects. In case of the operationalization of participants experience in the fourth contribution, we carefully interpreted that institutions of higher education provide a good learning environment for traditional knowledge but not for performance skills (Walz et al., under review). Results of these analysis have to be interpreted with care and need further confirmation with larger and more accurate operationalization.

Third, while the total sample of the KomPrü data set can be considered as sufficient for most analysis, most of the contributions could only use subsamples and therefore some results lack in statistical power. Hence, some contributions address the necessity to explore the current results with methods that are more sophisticated as well as repeating analysis with larger and longitudinal samples. This holds especially for the competence level model. This model integrates theoretical facets, survey scales and empirically derived differences in competence based on IRT results. According to the IRT approach, the result of individual participants is depending on the sample and requires an understanding of the applied and

complex methods. There exist no standard values for applications of the instrument yet. Consequently, the interpretation of assessment results is enhanced by current results but can still improve its usability and user-friendliness.

Fourth, as the application of role-plays is very complex and extensive, the contributions could only draw on existing data instead of generating further samples that would fit requirements of longitudinal or larger data sets. While further application and exploration of the role-play instrument seems promising, it comes with high costs and has to be a long term project with a clear scope of research. The current results and contributions can be considered as suggestions, what this future scope might entail.

Fifth, except for the literature research regarding the characteristics of role-plays in higher education in Chapter 2.4 and the first published contribution (Braun et al., 2020) all presented results are restricted to the German case of teacher education and economics students. Because of the strong integration of the theoretical background in the development of the role-play tasks and the competence level model in Walz and Braun (2022), an application to other subjects or countries seems plausible but has not been done yet.

Sixth, while the documentation of the KomPrü data follows the principles of FAIR data management, some of these principles could be enhanced further. For example, data and documentation are only available through direct contact with the researchers at the Professur für Hochschuldidaktik mit dem Schwerpunkt Lehrerbildung or the general repositorium JLUpub, both part of the Justus-Liebig-Universität Gießen. The data would probably be better findable in a subject specific repositorium like the repositories provided by the Leibniz-Institut für Sozialwissenschaften (GESIS) or the Research Data Centre for Higher Education Research and Science Studies (FDZ DZHW). These subject specific repositories require a more standardized documentation of data and meta-data, which could not be achieved with the available time and resources for data processing and documentation. Still, as the meta-data includes authors names and a persistent identifier, the data set can be found by using general scientific search engines like GoogleScholar.

4.4 Directions for Further Research

This dissertation addresses the field of competence assessment using role-plays on various levels and aims at contributing valuable insights and results. Some of these results

reveal further gaps and open questions and thus further research, which is beyond the scope of this dissertation.

First, while current results support the use and application of role-play instruments for teaching and assessing performance competences like communication competence, further examination with larger and longitudinal samples of data seem promising and important. Hence, it would be a direction for further research to continue examination of the role-play instrument and its derivation from holistic theoretical concepts. Furthermore, it seems essential, to broaden the scope of the KomPrü role-play instrument by (a) exploring more efficient ways of application, and (b) pursuing the adaption of the KomPrü role-play instrument to different contexts of subject or nationality. In this sense, broadening the scope of the role-play instrument could increase the amount and quality of data and allow for further, general analysis. Efficiency of application could be increased by making use of the vast increase in modern technology and data processing to automatize data collection and first data preparation. Collaboration with experts from different fields or education systems could identify relatable situations to apply the role-play instruments tasks.

Second, the current state of results as well as the KomPrü role-play instrument could be part of further interdisciplinary research like collaborations between educational science, computer science or computer linguistics. The standardized observation surveys and competence level model of the role-play instrument are all based on and structured by the same theoretical framework. Together they facilitate the observation and assessment of a complex concept in a standardized way that might be implemented in computer software that could automatically assess video recordings of the communication situations.

Third, while further application and adaption of the KomPrü role-play instrument can enhance the usability of the approach and supplement the current findings with further data sets and their analysis, the available data of the KomPrü research group does still hold unexamined potentials for at least three directions of further research. (a) The research group applied several established scales to assess the non- and paraverbal communication of participants (Spitzberg & Adams, 2007), their self-assessment of competences (Braun et al., 2008) and evaluation of higher education experience (Schaeper & Weiß, 2016), or the application of established knowledge-tests (König & Blömeke, 2009; Lauterbach, 2015). Some of this data has already been analyzed (e.g. Braun, 2021; Braun, Schwabe, & Klein, 2018; Walz et al., under review), but there remain potentials to investigate the associations between these established scales and tests and the innovative approach of the theory-

derived role-play assessment. (b) In the background survey, the participants reported if and how they experienced communication training. These experiences are structured by general areas of experience or training like vocational or higher education training, volunteering or side jobs. Furthermore, they are supplemented by further explanations of these experiences open-ended information. Qualitative analysis of this information and examination of the association between participants' experiences and their assessment results could shed further light on the question which environments and learning opportunities can foster communicative competence. (c) Furthermore, the participants provided detailed information about their background, which allows for sociologically motivated research questions in combination with communication competence. With the availability of data and a user-friendly documentation of the data, these potentials of the data can be utilized by further research.

Finally, while all cumulative contributions of this dissertation presume higher education institutions as an environment that fosters the development of communication skills, this assumption is widely based on political frameworks for higher education as well as the theoretical prerequisites of a higher education institution as transformational environment. What is missing so far, is a more profound analysis on the specificities of this environment. (How) Is communication competence part of higher education curriculums? (How) Do higher education institutions provide learning opportunities to train communication competence? (How) Do higher education institutions assess progress and students' communication competence? Detailed analysis of courses or module descriptions as well as qualitative interviews with higher education practitioners and teachers could shed light on these questions. After all, a structured and evidence-based application of role-plays to assess performance-based competences in higher education can only be successful, if we have a thorough understanding of teaching and learning methods for performance-based competences like communication and other 21st century skills and their impact within the higher education environment. The examination of a specific assessment instrument and the current results of this dissertation make a start to address these questions by providing insights regarding the general usability of role-plays in higher education assessment settings, the practical application of the instrument and further the examination of the available data. However, if higher education is to keep up with the change and changing requirements for competences, further empirical as well as theoretical research is necessary, to understand and accomplish these requirements by using role-plays.

4.5 Reflection

During the work and research for this dissertation, the working conditions and the working environment changed several times. Each of these changes had an impact on my understanding and view of this dissertations topic and contributed to its results as well as to my personal and scientific development.

First, the changes referred to the difference between studying in a higher education institution to working and researching in a higher education institution. Studying sociology at the Friedrich-Schiller-Universität Jena provided me with a solid scientific training and first impressions of the German higher education system from a student's perspective. Experiences as tutor for statistics and student assistant made me aware of the different potentials and constrains of assessment methods in higher education. As a researcher at the Professur für Hochschuldidaktik mit dem Schwerpunkt Lehrerbildung I supplemented these first impressions with a sound theoretical and research-based extension, leading to my first published research results and the interest in performance-based competences and the KomPrü instrument. Experiences in teaching and assisting in different higher education courses added a practical understanding for the importance of research in teaching and assessment in higher education.

Second, the changes referred to the time researching at the Justus-Liebig-Universität Gießen. After a serious IT security problem, the university had to shut down its servers. Usual digital communication channels like e-mail could not be used neither internet nor internal networks (Justus-Liebig-Universität Gießen, 2020). Communication relied mainly on present face-to-face or postal interaction, many research projects were put on hold as relevant data, and programs were stored digitally at internal servers. This was the time for reflection and sharpening of theoretical and empirical goals for this dissertation. With a certain distance I could re-evaluate previous considerations and ideas for my research and data management. The time necessary to regain the understanding of my own research data emphasized the need for better documentation.

Third, just after most constraints due to the IT security problem have been resolved so far that most previous working flows returned, the next change affected the whole world. Caused by the Covid19 pandemic this time, the physical attendance had been shut down. Instead of returning to office, lecture room and auditorium, people stayed at home, working and studying from kitchen tables, bed- and living rooms. One could say that all these changes restricted the way in which we work, do research, learn and communicate. In some

ways it clearly did. However, when I look back, I also observe a vast amount of new competences required but also enhanced by these changes: first the ability to adapt to changing environments and the quick acquisition of new methods, tools and ways of working and communicating. Performance competences and especially communication competence became even more important characteristics. The experience of digital education can enhance a competence of digital presence (Kunz, 2020). Emergency remote teaching could transform into digital innovations and a higher use of blended-learning in university teaching (Guppy et al., 2022; Zawacki-Richter, 2020). The vast changes in learning and working environments could positively influence individual management skills as well as independence (Ahmed et al., 2021). Even though, essential parts of higher education and work environments strive for a return to local presence, they gained the opportunity and competence to make a competent decision between methods, tools and places to study and work that fit a certain situation best. I am sure, that despite all negative effects, there are also gains in competence that will stay and help to navigate successful in our modern world. For this dissertation, the changes induced by the Covid19 pandemic highlighted the need for alternative and innovative methods of teaching, assessment and communicating. The collaborative search for and development of fitting teaching formats, the sharing of knowledge and established materials within the university and across (inter-) national initiatives boosted the topics of open science and collaboration. Digital conferences and workshops set a low threshold for participation and reduced traveling expenses.

A fourth change appeared, when I left my position as a researcher at the Justus-Liebig-Universität to deepen my understanding and competences for FAIR data management and started as an associate in data management at the Service Center for Medical Data Science at the University of Bielefeld. The few months that elapsed between this change in July 2022 and handing in this dissertation have not only strengthened my belief in the benefits of open and FAIR data management but furthermore provided me with the opportunity to reflect on the management and documentation of the KomPrü data. From this new perspective, I would have advised myself to start the process of documentation at a significantly earlier point of time and with the help of a data management plan. Such a plan could have already enhanced the research and data collection of the KomPrü research group and would have saved precious time during the

end of the documentation process which could have been invested in standardizing data and meta data to fit the requirements of established and subject specific repositories.

When looking back at these changes, they all had the potential to end or reduce my scientific research but are also key moments of setting goals for this dissertation and gaining new and exciting perceptions that have influenced this dissertation. When looking forward, I hope to expect further changes and possibilities to extend my interests and competences.

After all, living in a changing world did not start with me graduating and becoming a researcher, nor with the IT security problem in Gießen nor with the Covid19 pandemic. Change and changing requirements for competences can be regarded as a key feature of modern society.

5. REFERENCES

Ahmed, D. A. K., Henari, T. F., & Maklef, A. A. (2021). *Investigating Students' Development of 21 Century and Lifelong Skills During Covid-19 Crisis Education* (Vol. 16). https://www.researchgate.net/profile/tara-henari/publication/351712322_investigating_students'_development_of_21_century_and_lifelong_skills_during_covid-19_crisis_education

Allen, J., Levels, M., van der Velden, R [Rolf@TechreportRePEc:unm:umaror:2013013], van der Velden, R [Rolf], & Humburg, M. (2013). *Skill mismatch and skill use in developed countries: Evidence from the PIAAC study: What is expected of higher education graduates in the 21st century?* ROA Research Memorandum (No. 013). Maastricht University, Research Centre for Education and the Labour Market (ROA); Maastricht University. https://EconPapers.repec.org/RePEc:unm:umaror:2013013

Allen, J., Ramaekers, G., & van der Velden, R [Rolf] (2005). Measuring competencies of higher education graduates. *New Directions for Institutional Research*, *2005*(126), 49–59. https://doi.org/10.1002/ir.147

Andersson, N., & Andersson, P. H. (2010). Teaching professional engineering skills–industry participation in realistic role play simulation. *Making Change Last: Sustaining and Globalizing Engineering Educational Reform*.

Argyle, M. (1993). Why I study… social skills. *Psychologist*, *12*(3), 142.

Atkins, M. J., Beattie, J., & Dockrell, W. B. (1993). *Assessment Issues in Higher Education*. https://eric.ed.gov/?id=ed369370

Baker, F. B. (2001). *The basics of item response theory*. ERIC.

Barrera, F., Venegas-Muggli, J. I., & Nuñez, O. (2021). The impact of role-playing simulation activities on higher education students' academic results. *Innovations in Education and Teaching International*, *58*(3), 305–315. https://doi.org/10.1080/14703297.2020.1740101

Berezovska, L. I., Kondratska, G. D., Zarytska, A. A., Volkova, K. S., & Matsevko, T. M. (2020). Introduction of New Forms of Education in Modern Higher and Vocational Education and Training. *International Journal of Higher Education*, *9*(7), 107. https://doi.org/10.5430/ijhe.v9n7p107

Blömeke, S. (2019). Lehrerbildung. In O. Köller, M. Hasselhorn, F. W. Hesse, & K. Maaz (Eds.), *Das Bildungswesen in Deutschland: Bestand und Potenziale* (1st ed., pp. 663–696). Klinkhardt Julius; UTB.

Boone, W. J., Yale, M. S., & Staver, J. R. (2014). *Rasch Analysis in the Human Sciences*. Springer Netherlands; Imprint; Springer. https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=1593304

Bosse, H. M., Schultz, J.-H., Nickel, M., Lutz, T., Möltner, A., Jünger, J., Huwendiek, S., & Nikendei, C. (2012). The effect of using standardized patients or peer role play on ratings of undergraduate communication training: A randomized controlled trial. *Patient Education and Counseling*, *87*(3), 300–306. https://doi.org/10.1016/j.pec.2011.10.007

Braun, E. (2021). Performance-based assessment of students' communication skills. *International Journal of Chinese Education*, *10*(1), 221258682110062. https://doi.org/10.1177/22125868211006202

Braun, E., Athanassiou, G., Gockel, S., & Pollerhof, K. (2016). KomPrü - Performance-based Assessment of Student's Commuication Skills. In O. Zlatkin-Troitschanskaia, H. A. Pant, C. Lautenbach, & M. Toepper (Eds.), *KoKoHs Working Papers: Vol. 11. Modeling and Measuring Competencies in Higher Education: Validation and Methodological Innovations (KoKoHs). Overview of the Research Projects* (pp. 11–14).

Braun, E., Athanassiou, G., Pollerhof, K., & Schwabe, U. (2018). Wie lassen sich kommunikative Kompetenzen messen? Konzeption einer kompetenzorientierten Prüfung kommunikativer Fähigkeiten von Studierenden. *Beiträge Zur Hochschulforschung*, *40*(3), 34–55. http://www.bzh.bayern.de/uploads/media/3_2018_Braun_Athanassiou_Pollerhof_Schwabe.pdf

Braun, E., Gusy, B., Leidner, B., & Hannover, B. (2008). Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). *Diagnostica*, *54*(1), 30–42.

Braun, E., & Mishra, S. (2016). Methods for Assessing Competences in Higher Education: A Comparative Review. In J. Huisman & M. Tight (Eds.), *Theory and Method in Higher Education Research: volume 2. Theory and Method in Higher Education Research* (2016th ed., Vol. 2, pp. 47–68). Emerald Group Publishing Limited. https://doi.org/10.1108/S2056-375220160000002003

Braun, E., Osada, J.-C., & Walz, K. (2020). Higher Education Graduate Outcomes and Destinations. In *Education.* Oxford University Press. https://doi.org/10.1093/obo/9780199756810-0233

Braun, E., Schwabe, U., & Klein, D. (2018). Performance-Based Tests: Using Role Plays to Assess Communication Skills. In S. McGrath, M. Mulder, & J. Papier (Eds.), *Handbook of Vocational Education and Training: Developments in the Changing World of Work* (pp. 1–11). https://doi.org/10.1007/978-3-319-49789-1_82-1

Brockmann, M., Clarke, L., Winch, C., Hanf, G., Méhaut, P., & Westerhuis, A. (2011). Interpretive Dictionary: Competence, Qualification, Education, Knowledge. In M. Brockmann, L. Clarke, & C. Winch (Eds.), *Knowledge, Skills and Competence in the European Labour Market : What's in a Vocational Qualification?* (pp. 149–184). Taylor & Francis Group.

Cameron, D. (2000). *Good to Talk? Living and Working in a Communication Culture*. SAGE Publications, Incorporated; NetLibrary, Incorporated [distributor]. https://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=251232

Charlton, R. C. (1993). Using role-plays to teach palliative medicine. *Medical Teacher*, *15*(2-3), 187–193. https://doi.org/10.3109/01421599309006713

Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-Based Learning in Higher Education: A Meta-Analysis. *Review of Educational Research*, *90*(4), 499–541. https://doi.org/10.3102/0034654320933544

Cook, D. A., Brydges, R., Zendejas, B., Hamstra, S. J., & Hatala, R. (2013). Technology-enhanced simulation to assess health professionals: A systematic review of validity evidence, research methods, and reporting quality. *Academic Medicine : Journal of the Association of American Medical Colleges*, *88*(6), 872–883. https://doi.org/10.1097/ACM.0b013e31828ffdcf

Creative Commons. (no year). *Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)*. https://creativecommons.org/licenses/by-nc-sa/4.0/

Crow, M. L., & Nelson, L. P. (2015). The effects of using academic role-playing in a teacher education service-learning course. *International Journal of Role Playing*, *5*, 26–34.

Curaj, A., Deca, L., & Pricopie, R. (Eds.). (2018). *European Higher Education Area: The Impact of Past and Future Policies*. Springer International Publishing. https://doi.org/10.1007/978-3-319-77407-7

Degner, J., Meiser, T., & Rothermund, K. (2009). Kognitive und sozial-kognitive Determinanten: Stereotype und Vorurteile. In A. Beelmann & K. J. Jonas (Eds.), *Diskriminierung und Toleranz: Psychologische Grundlagen und Anwendungsperspektiven* (1st ed., pp. 75–93). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91621-7_4

Dierkes, J. (2021). Planung, Beschreibung und Dokumentation von Forschungsdaten. In M. Putnings, H. Neuroth, & J. Neumann (Eds.), *De Gruyter Praxishandbuch Ser. Praxishandbuch Forschungsdatenmanagement* (pp. 303–326). Walter de Gruyter GmbH. https://doi.org/10.1515/9783110657807-018

Dietrich, J. W., & Olson, C. (2010). In Quest of Meaningful Assessment of International Learning. *The Journal of General Education*, *59*(3), 143–158. https://doi.org/10.5325/jgeneeduc.59.3.0143

Döring, N., & Bortz, J. (2015). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften: In den Sozial- und Humanwissenschaften* (5 Aufl.). *Springer-Lehrbuch*. Springer. https://doi.org/10.1007/978-3-642-41089-5

Duchatelet, D., Spooren, P., Bursens, P., Gijbels, D., & Donche, V. (2021). Explaining self-efficacy development in an authentic higher education learning context of role-play simulations. *Studies in Educational Evaluation*, *68*, 100940. https://doi.org/10.1016/j.stueduc.2020.100940

Dudek, K., Ohly, M., & Tauch, C. (Eds.). (2019). *Statistiken zur Hochschulpolitik: 2/2019. Statistische Daten zu Studienangeboten an Hochschulen in Deutschland, Wintersemester 2019/2020: Studiengänge, Studierende, Absolventinnen und Absolventen*. Hochschulrektorenkonferenz (HRK).

Ellis, R. (2009). *Communication skills: Stepladders to success for the professional* (2nd ed.). Intellect.

Embretson, S. E., & Reise, S. P. (2013). *Item Response Theory for Psychologists*. *Multivariate Applications Series*. Taylor and Francis. http://gbv.eblib.com/patron/FullRecord.aspx?p=1166563

Falkenstern, A., Schwabe, U., Walz, K., & Braun, E. (2020). The Research Group Performance-Based Assessment of Communication in KoKoHs. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student*

Learning in German Higher Education (pp. 301–314). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27886-1_15

FDZ DZHW. *Research Data Centre for Higher Education Research and Science Studies*. https://www.fdz.dzhw.eu/en

Fecke, J., & Müller, L. (2022). *Simulationen in virtuellen Lernumgebungen: Welche Vor- und Nachteile haben avatarbasierte und videokamerabasierte Formate bei der Durchführung von Rollenspielen?* https://doi.org/10.3217/zfhe-17-01/13

Felstead, A., Gallie, D., Green, F., & Henseke, G. (2019). The determinants of skills use and work pressure: A longitudinal analysis. *Economic and Industrial Democracy*, *40*(3), 730–754. https://doi.org/10.1177/0143831X16656412

Fischer, F., & Opitz, A. (Eds.). (2022). *Springer eBook Collection*. *Learning to Diagnose with Simulations: Examples from Teacher Education and Medical Education* (1st ed. 2022). Springer International Publishing. https://doi.org/10.1007/978-3-030-89147-3

Fitó-Bertran, À., Hernández-Lara, A. B., & Serradell-López, E. (2014). Comparing student competences in a face-to-face and online business game. *Computers in Human Behavior*, *30*, 452–459. https://doi.org/10.1016/j.chb.2013.06.023

Frindte, W., & Geschke, D. (2019). *Lehrbuch Kommunikationspsychologie*. Beltz. http://nbn-resolving.org/urn:nbn:de:bsz:31-epflicht-1131774

Gartmeier, M., Bauer, J., Fischer, M. R., Hoppe-Seyler, T., Karsten, G., Kiessling, C., Möller, G. E., 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. https://doi.org/10.1007/s11251-014-9341-6

Geiser, C., & Eid, M. (2010). Item-Response-Theorie. In C. Wolf & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 311–332). VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-92038-2_14

Gerecht, M., & Kminek, H. (2018). (Erziehungs-)Wissenschaftlich Publizieren: veränderte Bedingungen und neue Techniken. *Erziehungswissenschaft*, *29*(57 (2-2018)), 29–36. https://doi.org/10.3224/ezw.v29i2.04

GESIS. *Daten teilen*. https://www.gesis.org/datenservices/daten-teilen

Glaesser, J. (2019). Competence in educational theory and practice: a critical discussion. *Oxford Review of Education*, *45*(1), 70–85. https://doi.org/10.1080/03054985.2018.1493987

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*(4), 323–343. https://doi.org/10.1037/1082-989X.11.4.323

Grice, H. P. (1975). Logic and conversation. In P. Coyle & J. Morgan (Eds.), *Syntax and Semantics* (3rd ed., pp. 41–58). Academic Press.

Grose-Fifer, J. (2017). Using role-play to enhance critical thinking about ethics in psychology. In R. Obeid, A. Schwartz, Shane-Simpson Christina, & P. J. Brooks (Eds.), *How we teach now: The GSTA guide to student-centered teaching* (213-223). Society for the Teaching of Psychology.

Gross, C., Bela, A., Jungbauer-Gans, M., Jobst, A., & Schwarze, J. (2019). Educational Returns Over the Life Course. In H.-P. Blossfeld & H.-G. Roßbach (Eds.), *Education as a Lifelong Process: The German National Educational Panel Study (NEPS)* (pp. 137–153). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-23162-0_8

Grossman, P., Compton, C., Igra, D., Ronfeldt, M., Shahan, E., & Williamson, P. W. (2009). Teaching practice: A cross-professional perspective. *Teachers College Record*, *111*(9), 2055–2100.

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, *52*(3), 67.

Guppy, N., Verpoorten, D., Boud, D., Lin, L., Tai, J., & Bartolic, S. (2022). The post-COVID-19 future of digital learning in higher education: Views from educators, students, and other professionals in six countries. *British Journal of Educational Technology*, Article bjet.13212. Advance online publication. https://doi.org/10.1111/bjet.13212

Habermas, J. (1984). *The theory of communicative action*. Beacon Press.

Hargie, O. (2010). *Skilled Interpersonal Communication: Research, Theory and Practice, 5th Edition*. Taylor & Francis. https://books.google.de/books?id=jxKtAgAAQBAJ

Hargie, O. (2022). *Skilled interpersonal communication: Research, theory and practice* (Seventh edition). Routledge.

Hartig, J., & Frey, A. (2013). Sind Modelle der Item-Response-Theorie (IRT) das „Mittel der Wahl“ für die Modellierung von Kompetenzen? *Zeitschrift Für Erziehungswissenschaft*, *16*(S1), 47–51. https://doi.org/10.1007/s11618-013-0386-0

Hartig, J., & Goldhammer, F. (2010). Modelle der Item-Response-Analyse. *Enzyklopädie Erziehungswissenschaft Online*(%s), 2–37.

HEGESCO. (2022, March 16). *HEGESCO: The Project*. http://www.hegesco.org/

Hu, S., Kuh, G. D., & Li, S. (2008). The Effects of Engagement in Inquiry-Oriented Activities on Student Learning and Personal Development. *Innovative Higher Education*, *33*(2), 71–81. https://doi.org/10.1007/s10755-008-9066-z

Hyytinen, H., & Toom, A. (2019). Developing a performance assessment task in the Finnish higher education context: Conceptual and empirical insights. *The British Journal of Educational Psychology*, *89*(3), 551–563. https://doi.org/10.1111/bjep.12283

Johnston, M. P. (2017). Secondary data analysis: A method of which the time has come. *Qualitative and Quantitative Methods in Libraries*, *3*(3), 619–626.

Joyner, B., & Young, L. (2006). Teaching medical students using role play: Twelve tips for successful role plays. *Medical Teacher*, *28*(3), 225–229. https://doi.org/10.1080/01421590600711252

Justus-Liebig-Universität Gießen. (2020, February 10). *#JLUoffline: Aktuelle Informationen zu den Folgen des Cyber-Angriffs auf die JLU*. https://www.uni-giessen.de/jluoffline

Kettula, K., & Berghäll, S. (2013). Drama-based role-play: a tool to supplement work-based learning in higher education. *Journal of Workplace Learning*, *25*(8), 556–575. https://doi.org/10.1108/JWL-04-2012-0036

Kilgour, P. W., Reynaud, D., Northcote, M. T., & Shields, M. (2015). Role-playing as a tool to facilitate learning, self reflection and social awareness in teacher education. *Journal of Inovative Interdisciplinary Research*, *2*(4), 8–20.

Klieme, E., & Hartig, J. (2008). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In M. Prenzel, I. Gogolin, & H.-H. Krüger (Eds.), *Kompetenzdiagnostik* (pp. 11–29). Springer Fachmedien.

Koltay, T. (2022). Research data management. In *Research Data Management and Data Literacies* (pp. 77–108). Elsevier. https://doi.org/10.1016/B978-0-12-824475-3.00002-3

König, J., & Blömeke, S. (2009). Pädagogisches Wissen von angehenden Lehrkräften. *Zeitschrift Für Erziehungswissenschaft*, *12*(3), 499–527.

Kraska-Miller, M. (2014). *Nonparametric statistics for social and behavioral sciences* (1st edition). CRC Press. https://learning.oreilly.com/library/view/-/9781466507623/?ar

Kunz, A. M. (2020). (Online-)Präsenz als Schlüsselkompetenz. In M. Stanisavljevic & P. Tremp (Eds.), *(Digitale) Präsenz - Ein Rundumblick auf das soziale Phänomen Lehre* (pp. 61–64).

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. https://doi.org/10.1016/j.pec.2007.02.011

Lauterbach, O. (2015). *Erfassung wirtschaftswissenschaftlicher Fachkompetenzen von Studierenden in Startkohorte 5 des Nationalen Bildungspanels*. Technischer Bericht (NEPS Working Paper No. 51). Bamberg. Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel. https://www.neps-data.de/Portals/0/Working%20Papers/WP_LI.pdf

Mainz, J. G.-U. (2022, February 17). *KomPrü – KomFäh | Kompetenzen im Hochschulsektor*. https://www.kompetenzen-im-hochschulsektor.de/komprue-komfaeh/

Martin, D. A., Conlon, E., & Bowe, B. (2019). The role of role-play in student awareness of the social dimension of the engineering profession. *European Journal of Engineering Education*, *44*(6), 882–905. https://doi.org/10.1080/03043797.2019.1624691

Mayhew, M. J., Rockenbach, A. N., Bowman, N. A., Seifert, T. A. D., Wolniak, G. C., Pascarella, E. T., & Terenzini, P. T. (2016). *How College Affects Students : 21st Century Evidence That Higher Education Works*. John Wiley & Sons, Incorporated. http://ebookcentral.proquest.com/lib/unigiessen/detail.action?docID=4658582

McConville, J. R., Rauch, S., Helgegren, I., & Kain, J.-H. (2017). Using role-playing games to broaden engineering education. *International Journal of Sustainability in Higher Education*, *18*(4), 594–607. https://doi.org/10.1108/IJSHE-08-2015-0146

Medel-Anonuevo, C., Ohsako, T., & Mauch, W. (2001). *Revisiting Lifelong Learning for the 21st Century*. UNESCO Institute for Education.

Michener, W. K. (2015). Ten Simple Rules for Creating a Good Data Management Plan. *PLoS Computational Biology*, *11*(10), e1004525. https://doi.org/10.1371/journal.pcbi.1004525

Moreno-Guerrero, A.-J., Rodríguez-Jiménez, C., Gómez-García, G., & Ramos Navas-Parejo, M. (2020). Educational Innovation in Higher Education: Use of Role Playing and Educational Video in Future Teachers' Training. *Sustainability*, *12*(6), 2558. https://doi.org/10.3390/su12062558

Nemec, R., Brower, E., & Allert, J. (2021). A Guide to Implementing Role-Play in the Nursing Classroom. *Nursing Education Perspectives*, *42*(6), E163-E164. https://doi.org/10.1097/01.NEP.0000000000000678

Nestel, D., & Tierney, T. (2007). Role-play for medical students learning about communication: Guidelines for maximising benefits. *BMC Medical Education*, *7*, 3. https://doi.org/10.1186/1472-6920-7-3

Nicolaides, A. (2012). Innovative teaching and learning methodologies for higher education Institutions. *Educational Research*, *3*(8), 620–626.

Oßwald, A. (2021). Barrieren, Hemmschwellen und Gatekeeper. In M. Putnings, H. Neuroth, & J. Neumann (Eds.), *De Gruyter Praxishandbuch Ser. Praxishandbuch Forschungsdatenmanagement* (pp. 277–296). Walter de Gruyter GmbH. https://doi.org/10.1515/9783110657807-016

Ostini, R., & Nering, M. L. (2006). *Polytomous item response theory models* (Vol. 144). Sage.

Pellegrino, J. W., & Hilton, M. L. (Eds.). (2012). *Education for life and work: Developing transferable knowledge and skills in the 21st century*. National Academies Press. https://doi.org/10.17226/13398

Penprase, B. E. (2018). The Fourth Industrial Revolution and Higher Education. In N. W. Gleason (Ed.), *Higher Education in the Era of the Fourth Industrial Revolution* (pp. 207–229). Springer Singapore. https://doi.org/10.1007/978-981-13-0194-0_9

Pereira, D., Flores, M. A., & Niklasson, L. (2016). Assessment revisited: a review of research in Assessment and Evaluation in Higher Education. *Assessment & Evaluation in Higher Education*, *41*(7), 1008–1032. https://doi.org/10.1080/02602938.2015.1055233

Phillips, E. (2012). Law Games–Role Play And Simulation In Teaching Legal Application And Practical Skills: A Case Study. *Compass: The Journal of Learning and Teaching at the University of Greenwich*, *5*, 1–4.

Putnings, M., Neuroth, H., & Neumann, J. (Eds.). (2021). *De Gruyter Praxishandbuch Ser*. *Praxishandbuch Forschungsdatenmanagement*. Walter de Gruyter GmbH. https://doi.org/10.1515/9783110657807

Rao, D. (2011). Skills development using role-play in a first-year pharmacy practice course. *American Journal of Pharmaceutical Education*, *75*(5), 84. https://doi.org/10.5688/ajpe75584

Rao, D., & Stupans, I. (2012). Exploring the potential of role play in higher education: development of a typology and teacher guidelines. *Innovations in Education and Teaching International*, *49*(4), 427–436. https://doi.org/10.1080/14703297.2012.728879

Raykov, T., & Marcoulides, G. A. (2018). *A course in item response theory and modeling with Stata*. Stata Press College Station, TX.

Romijn, B. R., Slot, P. L., & Leseman, P. P. (2021). Increasing teachers' intercultural competences in teacher preparation programs and through professional development: A review. *Teaching and Teacher Education*, *98*, 103236. https://doi.org/10.1016/j.tate.2020.103236

Russell, C., & Shepherd, J. (2010). Online role-play environments for higher education. *British Journal of Educational Technology*, *41*(6), 992–1002. https://doi.org/10.1111/j.1467-8535.2009.01048.x

Sanagavarapu, P., Abraham, J., & Taylor, E. (2019). Development and validation of a scale to measure first year students' transitional challenges, wellbeing, help-seeking, and adjustments in an Australian university. *Higher Education*, *77*(4), 695–715. https://doi.org/10.1007/s10734-018-0298-2

Schaeper, H., & Weiß, T. (2016). The Conceptualization, Development, and Validation of an Instrument for Measuring the Formal Learning Environment in Higher Education. In H.-P. Blossfeld, J. von Maurice, M. Bayer, & J. Skopek (Eds.), *Methodological Issues of Longitudinal Surveys: The Example of the National Educational Panel Study* (pp. 267–290). Springer Fachmedien Wiesbaden, Imprint: Springer VS. https://doi.org/10.1007/978-3-658-11994-2_16

Schnurr, M. A., Santo, E. M. de, & Green, A. D. (2014). What do students learn from a role-play simulation of an international negotiation? *Journal of Geography in Higher Education*, *38*(3), 401–414. https://doi.org/10.1080/03098265.2014.933789

Schwabe, U., & Braun, E. (2021). From Certificates to Performance: A Paradigm Shift and its Consequences for Higher Education Research. *Journal of Supranational Policies of Education (JoSPoE)*(13), 104–124. https://doi.org/10.15366/jospoe2021.13.005

Smith, E. (2008). Pittfalls and Promises: The use of Secondary Data Analysis in Educational Research. *British Journal of Educational Studies*, *56*(3), 323–339. https://doi.org/10.1111/j.1467-8527.2008.00405.x

Spitzberg, B. H., & Adams, T. W. (2007). *CSRS, the conversational skills rating scale: an instructional assessment of interpersonal competence*. NCA, National Communication Association.

StataCorp. (2017). Ranksum: Equality tests on unmatched data. In StataCorp (Ed.), *Stata: Release 15: Statistical Software.* College Station, TX: StataCorp LLC. Reference Manual.

Statistisches Bundesamt. (2017). *Bildung und Kultur. Studierende an Hochschulen: Vorbericht. Wintersemester 2016/2017* (Fachserie 11 Reihe 4.1). Statistisches Bundesamt. https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00067055

Statistisches Bundesamt. (2019, March 12). *Studierende insgesamt und Studierende Deutsche im Studienfach Wirtschaftswissenschaften nach Geschlecht*. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Tabellen/lrbil06.html

Stewart-Spencer, S., Eubanks, R., Law, J., Wyman, D. V., Pullen, A., & James, D. (2019). Counseling Theories Role Play as a Teaching Tool. *Journal of Counseling Research and Practice*, *4*(1), 2.

Sysoieva, S., & Mospan, N. (2018). Concept of Competence in the International and National Educational Contexts. *Continuing Professional Education: Theory and Practice*(1-2), 7–15. https://doi.org/10.28925/1609-8595.2018(1-2)715

Verderber, K. S., Verderber, R. F., & Sellnow, D. D. (2014). *Communicate!* (14. ed., International ed.). Wadsworth Pub.

Vermeiren, S., Duchatelet, D., & Gijbels, D. (2022). Assessing students' self-efficacy for negotiating during a role-play simulation of political decision-making. Taking student characteristics and simulation features into account. *Studies in Educational Evaluation*, *72*, 101124. https://doi.org/10.1016/j.stueduc.2022.101124

Voogt, J., & Roblin, N. P. (2012). A comparative analysis of international frameworks for 21 st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies*, *44*(3), 299–321. https://doi.org/10.1080/00220272.2012.668938

Wagenaar, R. (2021). Evidencing competence in a challenging world. European higher education initiatives to define, measure and compare learning. *International Journal of Chinese Education*, *10*(1), 221258682110069. https://doi.org/10.1177/22125868211006928

Wallace, J. D., & Becker, D. (Eds.). (2019). *A Gower book. The handbook of communication training: A best practices framework for assessing and developing competence*. Routledge.

Walz, K., & Braun, E. (2022). A Competency Level Model For Communication Skills. *Higher Education Forum*, *19*, 45–69.

Walz, K., Schwabe, U., & Braun, E. (under review). The Role of Performance Skills in German Teacher Education: Empirical Relations between General Pedagogical Knowledge, Communication Skills and Grades.

Walz, K., Schwabe, U., Braun, E., & Justus Liebig University Giessen. (2022). *Kompetenzorientierte Prüfung kommunikativer Fähigkeiten von Studierenden (KomPrü)*. https://doi.org/10.22029/JLUPUB-1133

Watzlawick, P., Bavelas, J. B., & Jackson, D. D. (1967). *Pragmatics of human communication: A study of interactional patterns, pathologies, and paradoxes*. Norton.

Weinert, F. E. (1999). *Concepts of competence*. Munich. Max Planck Institute for Psychological Reserch.

Westrup, U., & Planander, A. (2013). Role-play as a pedagogical method to prepare students for practice: The students' voice. *Högre Utbildning*, *3*(3), 199–210.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Winterton, J., Delamare-Le Deist, F., & Stringfellow, E. (2006). *Typology of knowledge, skills and competences: Clarification of the concept and prototype* (Cedefop reference series TI-73-05-526-EN-C). Luxembourg.

Wolter, A., & Kerst, C. (2015). The 'academization' of the German qualification system: Recent developments in the relationships between vocational training and higher education in Germany. *Research in Comparative and International Education*, *10*(4), 510–524. https://doi.org/10.1177/1745499915612188

Zawacki-Richter, O. (2020). The current state and impact of Covid-19 on digital higher education in Germany. *Human Behavior and Emerging Technologies.* Advance online publication. https://doi.org/10.1002/hbe2.238

Zlatkin-Troitschanskaia, O., Pant, H. A., Toepper, M [M.], & Lautenbach, C. (2020). Modeling and Measuring Competencies in Higher Education. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student Learning in German Higher Education* (pp. 1–6). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27886-1_1

6. APPENDIX

6.1 First Paper (Published)

Authors and Contributions[3]		∑
Edith Braun	*Conceptualization, Writing (Review & Editing), Supervision, Project Administration*	33%
Julia-Carolin Osada	*Conceptualization, Investigation, Writing (Original Draft + Review & Editing)*	33%
Kristina Walz	*Conceptualization, Investigation, Writing (Original Draft + Review & Editing)*	33%

Publication Information:
This contribution has been published at the peer-reviewed compendium Oxford Bibliographies in 2020.
Furthermore, the results have been presented at 18th Biennial EARLI Conference at RWTH Aachen University, Germany (12 to 19 August 2019) as well as in the "Jahrbuch 2019" of faculty 03 at Justus-Liebig-Universität Gießen.

Bibliographical Information:
Braun, E., Osada, J.-C., & Walz, K. (2020). Higher Education Graduate Outcomes and Destinations. In E. Braun, J.-C. Osada, & K. Walz (Eds.), *Education.* Oxford University Press. https://doi.org/10.1093/obo/9780199756810-0233

[3] Contributors' roles based on the CRediT authorship contribution standard (Brand et al. 2015).

6.2 Second Paper (Published)

Authors and Contributions:		∑
Anastasia Falkenstern	*Conceptualization, Writing (Original Draft + Review & Editing)*	50%
Ulrike Schwabe	*Writing (Review & Editing), Data Curation*	10%
Kristina Walz	*Methodology, Formal Analysis, Data Curation, Writing (Original Draft + Review & Editing)*	20%
Edith Braun	*Writing (Original Draft + Review & Editing), Supervision, Project Administration, Funding Acquisition*	20%

Publication Information:
This contribution has been published at the peer-reviewed edited volume "Student Learning in German Higher Education" in 2020.

Bibliographical Information:
Falkenstern, A., Schwabe, U., Walz, K., & Braun, E. (2020). The Research Group Performance-Based Assessment of Communication in KoKoHs. In O. Zlatkin-Troitschanskaia, H. A. Pant, M. Toepper, & C. Lautenbach (Eds.), *Student Learning in German Higher Education* (pp. 301–314). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-27886-1_15

Abstract:
In recent years, educational theory and empirically oriented educational research have been viewed in contrary light. Growing interest in evidence-based expertise in education and the increase in assessment of competences reinforced this debate. The aim of this contribution is to question this contradiction, and to consider possibilities of constructive collaboration between educational theory and empirical research in assessment of competences. The aim is to stimulate new formats of assessment. We discuss education as a holistic transformation and a social process. According to this understanding, we identify communication competences as an educational aim and discuss opportunities of its training and assessment. We suggest role-plays as an adequate method to develop a competence-based instrument. Ten specific role-plays were developed, which were applied by N = 515 students. The results support role-plays as an effective and reliable assessment of communication competences. Nevertheless, further analyses are necessary to detect and explain various effects and gains in competences. The combination of empirical educational research and educational theory seems to be promising, and we hope to discover further conceptual impacts by bringing both sides together.

Keywords:
performance based test; communication competences; educational theory; social interaction; higher education; intersubjectivity; KoKoHs

6.3 Third Paper (Published)

Authors and Contributions		∑
Kristina Walz	*Conceptualization, Methodology, Formal Analysis, Data Curation, Writing (Original Draft + Review & Editing), Visualization*	80%
Edith Braun	*Conceptualization, Writing (Original Draft + Review & Editing), Supervision, Project Administration*	20%

Publication Information:
The following article has been published in the peer-reviewed journal "Higher Education Forum" in 2022.
Furthermore, a presentation of the results has been accepted to SIG1-4 Joint Conference of the European Association for Research on Learning and Instruction (EARLI) Special Interest Groups (SIG1: Assessment and Evaluation & SIG4: Higher Education). University of Cadiz, Spain (22 to 25 June, cancelled due to COVID19 pandemic). The results have been presented at the 19th Biennial EARLI Conference for Research on Learning and Instruction at University of Gothenburg, Sweden (23 August to 2 September 2021, digital).

Bibliographical Information:
Walz, K., & Braun, E. (2022). A Competency Level Model For Communication Skills. *Higher Education Forum*, *19*, 45–69. https://doi.org/10.15027/52115

Abstract:
This paper examines a communication model based on six theoretical facets. Each facet was operationalized according to two aspects of Habermas' theory of communicative action: strategic and understanding-oriented action. The aim of the empirical analyses was to ascertain whether the postulated model could be used to measure different levels of competence, employing analyses from item-response-theory. We used a sample of 515 students from 11 German higher education institutions. Our empirical study confirmed qualitatively different levels of competence in all six facets. By linking rubrics with quantitative results, we were able to describe each level of competence qualitatively and to relate the different facets to each other. The purpose of this study is to support higher education institutions in the development of concrete strategies for helping students master complex competencies that aid them in their personal and professional development.

Keywords
Communication; competence level model; Habermas; higher education; IRT

6.4 Fourth Paper (currently under review)

Authors and Contributions		∑
Kristina Walz	*Conceptualization, Methodology, Formal Analysis, Data Curation, Writing (Original Draft + Review & Editing), Visualization*	60%
Ulrike Schwabe	*Conceptualization, Data Curation, Writing (Original Draft + Review & Editing)*	25%
Edith Braun	*Conceptualization, Writing (Original Draft + Review & Editing), Supervision, Project Administration, Funding Acquisition*	15%

Publication Information:
The following article is currently under review for publication in a peer-reviewed journal. Furthermore, the results have been presented at the digiGEBF 2021 at DIPF Leibnitz-Institut für Bildungsforschung und Bildungsinformation in Germany (March to December 2021, digital), the The 19th Biennial EARLI Conference for Research on Learning and Instruction at University of Gothenburg in Schweden (23 August to 2 September 2021, digital) and the 16. Annual conference der Gesellschaft für Hochschulforschung at Justus-Liebig-Universität Gießen in Germany (16 to 17 September 2021, digital).

Bibliographical Information:
Walz, K., Schwabe, U., & Braun, E. (under review). The Role of Performance Skills in German Teacher Education: Empirical Relations between General Pedagogical Knowledge, Communication Skills and Grades.

Abstract:
This contribution examines empirical associations between general pedagogical knowledge (GPK), communication skills, and grade point average in initial teacher education.
We assessed the GPK and communication skills of 105 student teachers in Germany, using a knowledge test and a performance-based assessment.
Empirical results show a significant correlation between GPK and grade point average but not between communication skills and grade point average. GPK is strongly associated with the number of semesters; communication skills are strongly associated with age. We discuss the learning objectives of initial teacher education in higher education in the light of the found results.

Keywords
communication skills; general pedagogical knowledge; Germany; performance; student teacher; professional teaching competence; teacher education

The Role of Performance Skills in German Teacher Education: Empirical Relations Between General Pedagogical Knowledge, Communication Skills, and Grades

This contribution examines empirical associations between general pedagogical knowledge (GPK), communication skills, and grade point average (GPA) in initial teacher education. We assessed the GPK and communication skills of 105 student teachers in Germany, using a knowledge test and a performance-based assessment. Empirical results show a significant correlation between GPK and GPA but not between communication skills and GPA. GPK is strongly associated with the number of semesters; communication skills are strongly associated with age. We discuss the learning objectives of initial teacher education in higher education in the light of the found results.

Keywords: communication skills; general pedagogical knowledge; performance; professional teaching competence; teacher education

Introduction

Teachers play a central role in learning processes, knowledge transfer, and in a student's acquisition of competence (Darling-Hammond 2000; Loughran and Hamilton 2016). Studies have convincingly shown that a teacher's knowledge and competences are important determinants for the quality of classroom teaching (Hattie 2010; Caena and Vuorikari 2021). Inside and outside of the classroom, teachers face a variety of specific situations that require reflective action and complex decision-making (Griffith and Lacina 2018). Aiming at a better understanding of the prerequisites for effective and successful teaching performance, the assessment of professional teaching competence have become an important research topic within empirical educational research (Cramer, Koenig, and Rothland 2020; Ben-Peretz and Flores 2018). In this article, we contribute to a better understanding of the role of performance skills within the education phase of student teachers in Germany. In Germany, teacher education

includes two phases of initial teacher education and extra-occupational continuing education. A student's initial teacher education is characterised by a clear separation between theoretical education in higher education and its practical application afterwards at schools where student teachers gain classroom teaching experience (Terhart 2019). We refer to the first phase of a student's teacher education as the 'education phase' and to the second phase as the 'training phase'. While these two phases follow a long tradition (Blömeke 2019), they are also connected to the phenomenon of 'reality shock' caused by the lack of coherence between pedagogical theory and its practical application (Corcoran 1981; Cramer, Koenig, and Rothland 2020; Dicke et al. 2016); frustration and emotional exhaustion (Voss and Kunter 2020; Voss et al. 2017); and high dropout rates (Hong 2010; Stokking et al. 2003). Both phases of a student's initial teacher education, therefore, seek to align more closely the student's knowledge acquisition with their capacity to act (Terhart 2019; Rots, Aelterman, and Devos 2014). Looking to examine professional teaching competence, we investigated the possible connections between a student's knowledge and performance skills to their attendance at institutions of higher education that provide a specific learning environment as part of a student's initial studies, during which they may attain vocational knowledge and performance skills.

We start by introducing the core concepts of professional teaching competence, with a special focus on general pedagogical knowledge (GPK) and communication skills. We regard grade point average (GPA) as an evident and comparable indicator of a student teacher's learning success at an institution of higher education. In addition to theoretical research, we also summarise the current state of research. Our study builds on previous research that explores the relationship between knowledge and performance (e.g. König and Pflanzl 2016; Zaragoza, Seidel, and Hiebert 2021). However, we add to

such research by examining the potential effectiveness of using performance-based assessments in measuring a student teacher's communication skills.

Conceptual Framework and Previous Research

Throughout the world, institutions of higher education play an important role in structuring and organising a student's initial education as a teacher (Craig 2016). In Germany, education programs prepare student teachers for the task of teaching in a mainly theory-driven but nevertheless practice-oriented way. Education programs seek to educate future teachers about general pedagogical methodologies, such as approaches to learning or classroom instruction, and in the specific future subjects that they will teach, such as mathematics, science, foreign languages or art (Blömeke 2019; Kultusministerkonferenz 2004/2019). Within the following sections, we provide an overview of the initial education phase in teacher education programs at institutions of higher education; we draw special attention, however, to such programs as exist within Germany today and focus on professional teaching competence as a key competence for future teachers, and on its connection to classroom-performance and communication.

Professional Teaching Competence

On a conceptual level, we draw on a comprehensive definition of professional teaching competence (Weinert 1999, illustrated in Blömeke 2019 p.666). Within this multidimensional concept, a general differentiation is made between professional knowledge, on the one hand, and affective-motivational characteristics, on the other. Professional knowledge can be further broken down into subject-specific knowledge, pedagogical content knowledge, and teaching methodology, as proposed by Shulman (1987). Affective motivational characteristics can be further broken down into teachers' subject- and teaching-specific beliefs, job motivation and personality traits like self-

efficacy (based on work by Richardson 1996; Thompson 1992). Together, these factors define a construct of a teacher's professional competence that is based on domains of knowledge and characteristics that can be learned and therefore taught (Kunter et al. 2013).

Previous studies have focused on examining the relation between knowledge and affective motivational characteristics (e.g. Blömeke et al. 2014; Depaepe and König 2018). In this study, we focus on establishing a relation between a teacher's professional knowledge and learning success. As such, we neither directly assess nor analyse a teacher's affective motivational characteristics.

Subject-specific knowledge, or content knowledge, refers to a teacher's understanding of the specific subjects they will teach in the future; their knowledge of facts and terms; and to their the deeper understanding of how these facts and terms are intertwined (Kleickmann et al. 2015; Shulman 1986). During the education phase of teacher education in Germany, student teachers acquire content knowledge through basic and advanced courses as part of the subject's general higher education curriculum. Pedagogical knowledge, or pedagogical content knowledge, connects the knowledge about a particular subject with pedagogical methods of presenting and teaching this content effectively (Shulman 1986). In Germany, the notion of 'Fachdidaktik' (subject-didactics) is an important part of a student's education phase (van Driel and Berry 2012). A teacher's content knowledge and pedagogical content knowledge is strongly structured by the subject they teach. It is therefore difficult to make comparisons between student teachers in different subjects (Blömeke 2019).

Teaching methodology, or general pedagogical knowledge (GPK), can be understood as a generic teaching skill that is part of a teacher's professional knowledge. This article follows the definition formulated by König et al. (2011) that focuses on the

competence of classroom management and instruction as essential parts of GPK. In Germany, the concept of GPK is part of a student's education phase and structured by pedagogical and other core disciplines (such as sociology or psychology) that are part of every student teacher's education, regardless of the subjects in which they will specialise. We consider GPK to be a distinctive part of the higher education learning environment that applies to all student teachers – even if the specific focus slightly differes between higher education institutions (Nehls et al. 2020). Therefore, we use GPK as part of our analysis to operationalise the knowledge aspect of professional teaching competence.

From Knowledge to Performance Skills

The vocational field of teaching requires not only knowledge but also situation-specific performance skills (Kunter et al. 2013; König et al. 2020). Professional teaching competence, therefore, is linked to the ability to transform knowledge and characteristics into action to master various situations. Within the context of teaching, teachers observe and interpret situations, then make decisions about how to act that are based on their knowledge and personal characteristics. This connection between knowledge, characteristics, and situation-specific abilities develops a teacher's performance skills (Blömeke 2019). Different aspects of knowledge are generally well integrated within the education phase. However, most programs do not include the teaching of performance skills within their curriculum; students in education programs do not begin to build such skills until later during the practical, training phase that takes place at schools (Terhart 2019). This can lead to a division between theoretical knowledge and practical application of skills that excludes opportunities for reflective learning and competence development for future teachers. Studies indicate that students perceive this division as an incoherent and even shocking experience that has a negative

impact on their skills (Alles et al. 2019; Dicke et al. 2015). During the education phase students acquire knowledge and skills that impact the development of their affective motivational characteristics (Blömeke et al. 2014). If the education phase would also create opportunities for fostering a student's performance skills it would establish a stable environment in which students could already began to put into practice the knowledge and characteristics that they acquire during their studies (Toom et al. 2021; Korthagen 2017). Were education programs more strongly integrate theory with practice, it could help student teachers avoid the shock that often accompanies operating within a real classroom for the first time (Darling-Hammond, Chung, and Frelow 2002; Jenset, Klette, and Hammerness 2018).

A central goal of teacher education research is therefore to find ways to improve professional teaching competence by aligning more closely a student teachers's acquisition of knowledge with their performance skills. Researchers and political actors alike consider performance to be an important learning outcome in higher education (Kultusministerkonferenz 2004/2019; Struyven and Meyst 2010; Flores 2016).

Communication Competence as a Generic Performance Skill

One important performance skill for teachers is communication competence, as teachers are faced with multiple settings in their professional life where interaction becomes necessary (Kultusministerkonferenz 2004/2019; Struyven and Meyst 2010). Teachers interact not only with students but also with colleagues, supervisors, and parents. Each communicative situation requires more than mere knowledge. They also require performance skills (Falkenstern et al. 2020). Teachers' communication skills have received considerable attention recently: however, the majority of scholarship focuses on specific scenarios related to teachers' communication, such as parent-teacher interaction (Gartmeier, Gebhardt, and Dotger 2016; Bruïne et al. 2014); classroom

instruction (Findeisen, Deutscher, and Seifried 2021; Tomei and Gilchrist-Petty 2018); or peer-interaction (Spillane, Hopkins, and Sweet 2018; Sjoer and Meirink 2016). These distinct communication scenarios are important tasks and integral to effective teaching. However, in order to understand and improve communication competence — in the sense of a generic performance skill and as part of professional teaching competence — we need a more holistic understanding, one that captures in a more complex way, the communicative skills that teachers need to thrive professionally (Falkenstern et al. 2020). Thus, instead of focusing on a specific situation in teacher communication, we address the underlying competence that teacher's need to communicate successfully in a variety of different situations and contexts. We, therefore, use a theoretical conceptualisation of communicative competence as formulated by Braun et al. (2018).

On a theoretical level, this assessment is based on the differentiation between two types of communication: understanding-oriented and strategic communication, first established by Habermas (1984) as part of his theory of communicative action. While these types are generic and thus applicable to different situations, we go beyond previous research in teacher communication. We use the term 'understanding-oriented communication' to describe a type of communication in which the communication partners aim to reach an agreement by acting in ways that are transparent, understandable, true, and honest within the scope of socially acceptable behaviour (Grice 1975). This type of communication is connected to a solution-oriented approach to conflict management (Braun et al. 2018). We use the term 'strategic communication' to describe a type of communication in which the communication partners do not necessarily aim at achieving a common goal. They aim instead at achieving goals as defined by a particular purpose. This type of communication is connected to situations of negotiation (Braun et al. 2018). Based on these two generic types of communication,

Braun et al. (2018) developed a theoretical framework for understanding communicative skills in a holistic way and applied it to ten role-plays that require either understanding-oriented or strategic communication. This crucial distinction between the two types has been empirically confirmed, resulting in simulation-based performance test (Braun 2021).

The test consists of role-plays that simulate different communicative situations, which are very likely to take place within the context of a teacher's professional life. However, these situations are not meant to interpret situation-specific behaviour but to assess teachers' underlying communication competence. Thus, in this test, student teachers perform several role-plays, each of which takes place within a different setting for staging the interaction with a communication partner. For the analysis we used the results of all performed role-plays and combined them into one competence score for each type of communication (Walz and Braun 2022).

Following this argumentation, we understand communication competence as a generic performance skill that helps teachers navigate professional interactions. Furthermore, as such situations take place within the context of a teacher's professional life, a teacher's successful performance also depends on their professional knowledge and individual characteristics. Therefore, the role-plays fit the theoretical assumptions of professional teaching competence as well as the requirement of a holistic competence test.

GPK is mostly relevant in situations related to professional interaction. Generic performance skills, by contrast, have a broader relevance. Communication competence, in particular, can be regarded as a 21st-century skill (Voogt and Roblin 2012). Future teachers are involved with communication and communicative action that go beyond the scope of formal higher education. Furthermore, they also experience communicative

situations in a variety of contexts. This has three implications relevant for our research. First, since we understand communication as a generic skill that is widely used inside and outside of contexts related to teacher education, there might be additional, more general, learning environments for communication competence apart from formal teacher education. With this in mind, we examined not only the ways in which a higher-education environment affects the development of skills but also the ways in which the general learning environment does. We operationalised the impact of formal learning environment with the number of semesters attended as an indicator of the learning environment at the institution of higher education. We used age as an indicator for the learning environment exeeding the institution respectively. We refer to such learning as 'general life experience'. Second, we considered professional knowledge as related to the learning environment at the institution of higher education. We considered general life experience as important for general performance skills (like communication skills) but not necessarily for specific knowledge (like pedagogical knowledge). Third, while communication in this general sense may require competent use of both types of communication, education phase programs in Germany focus on communication as solution-oriented conflict mangament (Kultusministerkonferenz 2004/2019). We, therefore, argue that during the education phase of their teacher education programs, institutions of higher education in Germany offer more learning opportunities for understanding-oriented communication than for strategic communication.

Grades as an Indicator of Learning Success

Continous performance assessments and attaining higher education degrees are relevant for later professional career (Cramer, Koenig, and Rothland 2020). Before, during and after a student teacher's studies, grades are an evident and comparable quantitative indicator of general (professional) ability in the respective education phase (Magno

2011). They signal how far the defined learning objectives have been achieved. In addition to higher education certificates, grades indicate differences in individual achievement and competence at the same level of education (Lynch and Hennessy 2017), which is a core criterion for recruitment into teaching profession in Germany (Terhart 2007). Grades do not only indicate specific knowledge and skills; they also indicate different aspects of a student's motivation, attentiveness, conscientiousness time management, and applying learning strategies (Magno 2011). Thus, a student's GPA, a summative evaluation, is predictive for a variety of professional and personal competences. Grades are a meaningful indicator of the level of professional teaching competence that a student has attained during their studies at an institution of higher education (Cramer, Koenig, and Rothland 2020). While research on the empirical relation between grades as the common practice and performance-based learning outcomes is generally sparse, existing evidence shows the expected correlation between test scores in knowledge of German didactics and grades in German (König and Bremerich-Vos 2020). The low proportion of overall explained variance in grades, however, indicates that other determinants seem to be relevant. It remains an open empirical question whether performance skills — like communication skills — are reflected in grades. Our study seeks to address this gap in the research.

Hypotheses

In Germany, the development of professional teaching competence is an important part of the education phase of teacher education. Theoretical knowledge is a learnable part of professional knowledge that is deeply embedded within the education-phase curricula. Performance skills supplement theoretical knowledge and form a bridge between theory and its application. Knowledge and performance skills, we assert, are both important outcomes of the education phase. In our study, we therefore consider

these two factors but with a more refined focus on GPK and communication skills, both of which can be compared amongst student teachers with different future subjects. During their studies, student teachers earn grades that provide empirical evidence of their professional teaching competence, as related to their knowledge and performance skills. Based on these observations, we formulate our first hypothesis:

(1) GPK and communication skills are linked to a student's GPA.

We consider GPK as knowledge that student teachers learn specifically during the education phase of their studies. We consider communication skills as generic skills that are important for teaching but are also relevant outside higher education and classroom settings. Student teachers may acquire communication skills both inside and outside institutions of higher education. In this study, we used the number of semesters spent at an institution of higher education as an indicator for experience acquired in higher education and the participant's age as an indicator for general life experience, acquired inside and outside an institution of higher education. Accordingly, our second hypothesis is:

(2) Different learning and experience environments promote GPK and communication skills.
    (a) Number of semesters spent at an institution of higher education is associated with GPK and communication skills, equally.
    (b) Number of semesters spent at an institution of higher education is associated more strongly with understanding-oriented communication than with strategic communication.
    (c) Age is associated more strongly with communication skills than with GPK.

(d) Age is associated with both understanding-oriented communication and strategic communication, equally.

We examined the first hypothesis by analysing the associations between GPA, pedagogical knowledge, and communication skills. We examined the second hypotheses by analysing the associations between GPK, communication skills, and the number of semesters in attendance and age, as indicators of formal training as well as of general life experience.

Method

Sample

We used a sample of 105 student teachers from five German institutions of higher education (Braun 2021).

In Germany, teacher education programs separate student teachers according to their future educational tracks. The sample we drew on consisted of student teachers from all tracks (see Table 1). The average age of the sample was 23 years old (sd 3.65). The sample consisted of 72 female and 30 male participants. The remaining three either did not identify their gender according to a male/female binary or did not respond to the question at all. In regards to the study's breakdown according to gender and most educational tracks, our sample mirrors the general distribution for student teachers in the education phase in Germany (Statistisches Bundesamt 2017), according to official statistics. The average number of semesters spent at an institution of higher education was 6.61 (sd 4.44). The GPA was 2.02 (sd 0.55; N=73). Most of the participants who did not report any grades were in their first year of higher education (N=21; 67.7%) and had not yet have any exams.

Instruments

We used two different tests to assess the student teacher's GPK and communication skills. Furthermore, we used self-reported information about GPA, number of semesters spent at an institution of higher education and age to operationalise the remaining constructs of our hypotheses.

We assessed GPK with a test developed by König and Blömeke (2009). We used a validated shortened version of the instrument, which assesses GPK as an unidimensional construct (König and Blömeke 2010). This shortened version of the test consists of two test-booklets, each containing 15 open-and-closed questions from a pool of 18 questions. Responses were regarded as either right or wrong, according to a prepared coding scheme (König and Blömeke 2010).

We assessed communication skills with a performance-based test developed by Braun et al. (2018). The test differentiates – theoretically as well as empirically – two types of communication: strategic and understanding-oriented communication (Braun et al. 2018; Braun 2021). The test consists of five role-plays each for strategic communication and understanding-oriented communication. Each role-play includes instruction sheets for a student teacher and a trained conversation partner (confederate), as well as an observation sheet for a trained rater. The instruction sheets introduce the communicative situation to a student teacher and a confederate. Overall, to successfully perform this task, a student teacher needs to know about classroom management and to possess underlying strategic communication skills.

The observation sheets for the rating person contain items, which were derived from several established communication theories and are associated with specific observable behaviour for both types of communication (Braun 2021). After the student teacher's performance, the observer rated the student teacher's behaviour for each item

on a 4-point rating scale. In our sample, each student teacher performed four of the ten role-plays, two strategic and two understanding-oriented.

We considered the students' GPA as a comparable indicator that in sum reflects the general achievements they had attained thus far at an institution of higher education. We also considered the number of semesters that a student had attended and a student's age as general but valuable indicators of their general life experience and their experience at an institution of higher education.

Multi-Matrix Test Design

Participants filled out a questionnaire about their socio-demographic characteristics and about their studies. We assigned one test-booklet of the GPK test as well as two strategic and two understanding-oriented role-plays randomly to each participant. Using this procedure, we observed 83 percent data of the GPK assessment with 17 percent data missing completely at random and 40 percent data of the role-plays with 60 percent data missing completely at random.

Data Preparation

In a first step, we applied IRT and multiple imputation to prepare the data and estimated the competence values for each student. We used different IRT models to determine competence values for GPK and both types of communication to gain precise competence estimations for further analysis (Boone 2016; Embretson and Reise 2013). Additionally, because IRT models are sensitive to missing data (as long as the data is missing completely at random), they are useful for analysis of data from multi-matrix design (Hartig and Frey 2013; Raykov and Marcoulides 2018). The performance assessment as well as GPK have a multi-matrix design. Therefore, an IRT analysis allowed an estimation of competence for each participant.

In the case of GPK, we used a dichotomous Rasch model as suggested by König and colleagues (2011). In the case of the communication-performance assessment, we used a polytomous-graded response model (GRM), which fits ordered categorical answers (Raykov and Marcoulides 2018; Samejima 2010). Furthermore, we drew on Aikakes and Bayesian information criterion, which are commonly used to compare different IRT models. The model associated with the smallest values within a comparison is considered to fit the data best (Raykov and Marcoulides 2018). Compared with two other polytomous IRT models, the GRM was associated with the smallest values (strategic: AIC=2986.741, BIC=3329.102; understanding-oriented communication: AIC=3657.732; BIC=4082.366). In sum, we used the estimated test scores for the latent variables 'GPK', 'strategic communication' and 'understanding-oriented communication' for all 105 students.

We computed a multiple imputation to estimate missing values in GPA in our sample. Most missing values were related to students who were in their first semester and could not report a grade when the test was conducted. Apart from that, first-semester students did not show any significant characteristics that separated them from the rest of the sample. We included all variables of our analysis model (GPK and both types of communication), gender, track, self-rated success of course-completion, number of semesters, and the institution of higher education for the multiple imputation.

Analysis

For the purpose of analysing the association of communication skills and GPK with GPA, we conducted a multiple linear regression. The regression analysis uses the reported or imputed GPA as dependent, and the scores of communication skills and GPK as independent variables. We examined the strength of the associations using the beta-coefficients of the regression analysis. Furthermore, we interpreted the adjusted R^2

as an indicator of the share, hence the evidence of communication skills and GPK in the GPA.

In our analysis, we used semester and age as indicators for general life experience and learning environment at institutions of higher education. We examined whether there was an association between these two indicators and communication skills and GPK. For this purpose, we conducted three multiple linear regressions and compared the results with our four assumptions regarding the relation of associations ((a) to (d)). The regression analysis used the competence estimations of communication skills and GPK as dependent, and the reported experience indicators semester and age as independent variables. We examined the strength of the associations using the beta-coefficients of the regression analysis.

Results

GPK as well as communication skills are evident within a student's GPA earned at an institution of higher education.

We observed a significant association between GPK and grades (beta=-0.29, $p<0.01$, N=100). German grades are ranked from '1' being the highest to '4' being the lowest grade with which it is possible to pass an exam. Higher test scores for GPK were associated with a higher GPA. There was, however, no significant relation between grades and the two types of communication (strategic: beta=0.13, $p>0.1$, N=100; understanding-oriented: beta=-0.09, $p>0.1$, N=100).

Altogether, the regression model explains $R^2=0.07$ (adjusted) of the variance in grades. This overall explanatory power seems plausible, as GPK shows a significant impact but is only one among other aspects formally defined in the curricula and professional teaching competence. Table 2 summarises the results of the regression

analysis between a teacher student's GPA and their competence scores for strategic communication, understanding-oriented communication, and GPK.

The results confirmed Hypothesis 1 only in parts. We observed that GPK was evident within GPA, while communication skills were not.

Different learning and experience environments promote GPK and communication skills.

Table 3 summarises the results of three regression-analysis between strategic communication, understanding-oriented communication, and GPK respectively and the number of semesters in attendance and age. First, we present the results concerning the hypotheses:

(e) The number of semesters spent attending an institution of higher education is associated with GPK and communication skills, equally.

(f) The number of semesters spent attending an institution of higher education is associated more strongly with understanding-oriented communication than with strategic communication.

The number of semesters spent attending an institution of higher education is associated significantly with GPK (beta=0.52, $p<0.01$, $R^2=0.22$, N=101) but not with either of the communication skills (strategic: beta=0.01, $p>0.1$, $R^2=0.05$, N=103; understanding-oriented: beta=0.09, $p>0.1$, $R^2=0.08$ N=103). Therefore, we could only partly confirm (a).

Concerning (b), we could not confirm a substantially stronger association between the number of semesters spent at an institution of higher education and understanding-oriented communication than between the number of semesters and strategic communication.

Second, we present the results concerning the hypotheses:

(g) Age is associated more strongly with communication skills than with GPK.

(h) Age is associated with both understanding-oriented communication and strategic communication, equally.

A student teacher's age showed a significant and comparably strong association with both types of communication (strategic: beta=0.26, $p<0.05$, N=103; understanding-oriented: beta=0.27, $p<0.05$, N=103). This result confirms Hypothesis (d). Since this connection is weaker and non-significant for age and GPK (beta=-0.14, $p>0.1$, N=103) the results also confirm (c).

Overall, these associations do not explain much variance in communication skills (strategic: $R^2=0.05$; understanding-oriented: $R^2=0.08$) but for GPK ($R^2=0.22$). Again, these findings emphasise the strong embedding of GPK within the curricula and the relatively weak embedding of communication skills at institutions of higher education.

Summary and Discussion

Summary of the Study

In the education phase of teacher education, the acquisition of performance skills is essential as adressed by political actors (Kultusministerkonferenz 2004/2019). Existing literature suggests that higher education-based teacher education not only promotes professional knowledge but also more generic performance skills as outcomes of higher education (Cramer, Koenig, and Rothland 2020; König and Pflanzl 2016). However, studies focussing on the relation of these learning outcomes and grades are sparse. Our

study, therefore, examined their promotion within current teacher education programs in Germany. According to our hypotheses, grades should thus reflect these performance skills as well as professional knowledge. For our empirical investigation, we operationalised professional knowledge as GPK (König and Blömeke 2009), and performance skills as communication skills (Braun et al. 2018). However, empirical results did not confirm all of our assumptions: While we found a clear correlation between GPK and GPA, we did not find evidence that there is a meaningful association between communication skills and grades.

In addition, we also examined the impact of learning opportunities in higher education, operationalised with the number of semesters participants had attended an institution of higher education, versus a learning experience acquired outside of such an institution, operationalised with the participant's age, on both facets of professional teaching competence. Consequently, we assumed that students would improve in both areas during their studies in higher education. According to our results, however, an association between enhanced performance skills and the number of semesters a student had attended an institution of higher education, i.e., the amount of time they had spent in a higher education learning environment, was found for GPK, but not for communication skills. Nevertheless, there is a significant correlation between a student's age and communication competence that implies that these skills are trainable. These findings go along with the fact that knowledge is reflected in a student teacher's GPA, while communication skills are not.

Implications

Based on our findings, it might be claimed that performance skills are not acquired at institutions of higher education but in the training phase of teacher education. We argue, however, that this claim is problematic: new teachers regularly report a 'reality shock'

when starting their training phase (Voss and Kunter 2020; Dicke et al. 2015). This ‘reality shock’ leads to a significant dropout rate from teacher education programs (Dicke et al. 2016; Stokking et al. 2003) and results in high levels of frustration among this group of new teachers. To prevent such a ‘reality-shock’ university-based teacher education programs should therefore seek to integrate more thoroughly and effectively theoretical knowledge and its practical application (Blömeke 2019; Flores 2016). To do so, such programs could support relevant learning opportunities, such as creating performance-based settings. This would help close the gap between theory and practice in the sense of establishing coherence and better prepare students to become effective teachers (Cramer 2020).

This is closely related to our study’s second aim: we seek to make performance skills measurable. Role-plays, we suggest, can be used as a method for initiating and evaluating complex performance skills in a way that goes beyond testing ‘pure’ knowledge (Blömeke, Gustafsson, and Shavelson 2015).

Strengths and Limitations

Our study makes an empirical contribution to research on the assessment of a teacher’s professional competence. It shows how knowledge and performance skills are reflected in a student teacher’s grades. Specifically, we go beyond previous research by using a performance-based test for assessing a student’s general communication skills.

Despite its strengths, this study has at least three limitations. First, the test setting itself required a great deal of planning and time (Braun 2021). In addition, different learning environments inside and outside instituions of higher education were operationalised with two rough proxy variables. Learning opportunities could instead be operationalised based on the extent to which they were anchored in the curriculum or specific courses. Such questions, however, go beyond the scope of this article.

Second, our analytic sample was with N=105 cases quite small for a regression framework. It might be that some coefficients did not become significant, because they lacked statistical power. Yet, the significant correlation between GPK and grades can be trusted as being robust, because it also showed up in the small sample. More advanced statistical analyses, like structural equitation models, were not applicable because of the small sample size.

A third limitation is our general assumption that higher education programs help students acquire knowledge and skills and can therefore be systematically improved with better learning opportunities in higher education environments. Since we only had a cross-sectional data set, we could not empirically test this assumption. We would need to measure a student's knowledge and performance skills at least at two points in time: first when the students start attending a higher education program and second when they are close to the end of their studies. This would allow us to separate the effect of learning at an institution of higher education from other learning opportunities in life as well as to account for individual differences in competence levels when entering higher education.

Directions for Future Research

Additional research shows promise for helping us better understand current learning opportunities in teacher education programs and their impact on learning outcomes. As such, we have identified three directions for future empirical investigation. First, little is known about how to develop specific methods for their implementation and development: Do students who attend institutions of higher education have better communication skills if some courses offer training in such skills or if such skills are anchored in the curriculum? To answer these questions, systematic evaluations of courses addressing communication skills or other generic skills could provide valid

empirical answers.

Second, what is the potential of utilising simulations in performance-based testing? Are performance-based test settings a suitable way to initiate communicative action and to integrate communication training into teacher education? A recent meta-study shows that simulation methods, such as performance-based, role-play assessments, have become more popular within higher education programs and are effective (Chernikova et al. 2020).

Third, what other learning environments are suitable to train and thus promote performance skills? How do (which) learning opportunities outside higher education shape communication skills? To answer this question, it is worth focusing on extracurricular experiences in a more intense and systematic way.

The aim of our paper was to examine the promotion of performance skills at institutions of higher education. Based on the empirical evidence obtained thus far, our findings suggest that within current teacher education programs in Germany there is no evident, conscious promotion of communication skills. The debate about the ‘reality shock’ that graduates of teacher education programs experience after the completion of their studies is a global one. Although, our study was restricted to Germany, we believe that these empirical findings have potential application to other national contexts. To strengthen coherence within teacher education programs, in different phases of a student’s studies and their professional development afterwards, it is worth thinking more intensively about performance-based teaching within education programs to assist students in transferring their theoretical knowledge and performance skills into the classroom. We believe, therefore, that measuring and training performance skills as early as possible during a student’s studies is crucial to their future professional success (Cramer 2020; Rothland 2020). However, as a model, the performance-based test for

measuring a student's communication skills is flexible; it can be easily adapted to other communication situations and to the second phase of teacher education as well.

Acknowledgements

We would like to express our special thanks to Johannes König for providing access to the short version of the test for general pedagogical knowledge. Isabel Hofmeister also offered a great deal of support by coding open answers.

Declaration of Interest Statement

The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding Details

The data collection for this study was supported by the German research program ‘Modelling and Measuring Competencies in Higher Education – Validation and Methodological Innovations’ (KoKoHs) by the German Federal Ministry of Education and Research (BMBF), under Grant 01PK14001. Additional research was supported by the German research program, ‘The Giessen Campaign: Teacher Education 2.0 (GOL)’ as part of the German Federal Ministry of Education and Research (BMBF), under Grant 01JA1929.

Disclosure Statement

The authors report there are no competing interests to declare.

References

Alles, Martina, Jennifer Apel, Tina Seidel, and Kathleen Stürmer. 2019. “How Candidate Teachers Experience Coherence in University Education and Teacher Induction: the Influence of Perceived Professional Preparation at University and Support during Teacher Induction.” *Vocations and Learning* 12 (1): 87–112. doi:10.1007/s12186-018-9211-5.

Ben-Peretz, Miriam, and Maria A. Flores. 2018. “Tensions and paradoxes in teaching: implications for teacher education.” *European Journal of Teacher Education* 41 (2): 202–13. doi:10.1080/02619768.2018.1431216.

Blömeke, Sigrid. 2019. “Lehrerbildung.” In *Das Bildungswesen in Deutschland: Bestand und Potenziale*, edited by Olaf Köller, Marcus Hasselhorn, Friedrich W. Hesse, and Kai Maaz. 1st ed., 663–96. Bad Heilbrunn: Klinkhardt Julius; UTB.

Blömeke, Sigrid, Nils Buchholtz, Ute Suhl, and Gabriele Kaiser. 2014. “Resolving the chicken-or-egg causality dilemma: The longitudinal interplay of teacher knowledge and teacher beliefs.” *Teaching and Teacher Education* 37: 130–39. doi:10.1016/j.tate.2013.10.007.

Blömeke, Sigrid, Jan-Eric Gustafsson, and Richard J. Shavelson. 2015. “Beyond Dichotomies: Competence viewed as a Continuum.” *Zeitschrift für Psychologie* 223 (1): 3–13. doi:10.1027/2151-2604/a000194.

Boone, William J. 2016. “Rasch Analysis for Instrument Development: Why, When, and How?” *CBE life sciences education* 15 (4). doi:10.1187/cbe.16-04-0148.

Braun, Edith. 2021. “Performance-based assessment of students’ communication skills.” *International Journal of Chinese Education* 10 (1): 221258682110062. doi:10.1177/22125868211006202.

Braun, Edith, Georgios Athanassiou, Kathleen Pollerhof, and Ulrike Schwabe. 2018. “Wie lassen sich kommunikative Kompetenzen messen?–Konzeption einer kompetenz orientierten Prüfung kommunikativer Fähig keiten von Studierenden.” *Beiträge zur Hochschulforschung, 40 (3)*, 34–55.

Bruïne, Erica J. de, T. M. Willemse, Jeanne D’Haem, Peter Griswold, Lijne Vloeberghs, and Sofie van Eynde. 2014. “Preparing teacher candidates for family–school partnerships.” *European Journal of Teacher Education* 37 (4): 409–25. doi:10.1080/02619768.2014.912628.

Caena, Francesca, and Riina Vuorikari. 2021. “Teacher learning and innovative professional development through the lens of the Personal, Social and Learning to Learn European key competence.” *European Journal of Teacher Education*, 1–20. doi:10.1080/02619768.2021.1951699.

Chernikova, Olga, Nicole Heitzmann, Matthias Stadler, Doris Holzberger, Tina Seidel, and Frank Fischer. 2020. “Simulation-Based Learning in Higher Education: A Meta-Analysis.” *Review of Educational Research* 90 (4): 499–541. doi:10.3102/0034654320933544.

Corcoran, Ellen. 1981. “Transition Shock.” *Journal of Teacher Education* 32 (3): 19–23. doi:10.1177/002248718103200304.

Craig, Cheryl J. 2016. “Structure of Teacher Education.” In Loughran and Hamilton 2016, 69–135.

Cramer, Colin. 2020. “Kohärenz und Relationierung in der Lehrerinnen- und Lehrerbildung.” In Cramer, Koenig, and Rothland 2020, 269–79.

Cramer, Colin, Johannes Koenig, and Martin Rothland, eds. 2020. *Handbuch Lehrerinnen- und Lehrerbildung.* utb Professionsforschung.

Darling-Hammond, Linda. 2000. “How Teacher Education Matters.” *Journal of Teacher Education* 51 (3): 166–73. doi:10.1177/0022487100051003002.

Darling-Hammond, Linda, Ruth Chung, and Fred Frelow. 2002. “Variation in Teacher Preparation.” *Journal of Teacher Education* 53 (4): 286–302. doi:10.1177/0022487102053004002.

Depaepe, Fien, and Johannes König. 2018. “General pedagogical knowledge, self-efficacy and instructional practice: Disentangling their relationship in pre-service teacher education.” *Teaching and Teacher Education* 69: 177–90. doi:10.1016/j.tate.2017.10.003.

Dicke, Theresa, Jill Elling, Annett Schmeck, and Detlev Leutner. 2015. “Reducing reality shock: The effects of classroom management skills training on beginning teachers.” *Teaching and Teacher Education* 48: 1–12. doi:10.1016/j.tate.2015.01.013.

Dicke, Theresa, Doris Holzberger, Olga Kunina-Habenicht, Christina Linninger, and Franziska Schulze-Stocker. 2016. “„Doppelter Praxisschock“ auf dem Weg ins Lehramt? Verlauf und potenzielle Einflussfaktoren emotionaler Erschöpfung während des Vorbereitungsdienstes und nach dem Berufseintritt.” *Psychologie in Erziehung und Unterricht* 63 (4): 244. doi:10.2378/peu2016.art20d.

Embretson, Susan E., and Steven P. Reise. 2013. *Item Response Theory for Psychologists.* Multivariate Applications Series. Hoboken: Taylor and Francis. http://gbv.eblib.com/patron/FullRecord.aspx?p=1166563.

Falkenstern, Anastasia, Ulrike Schwabe, Kristina Walz, and Edith Braun. 2020. "The Research Group Performance-Based Assessment of Communication in KoKoHs." In *Student Learning in German Higher Education*, edited by Olga Zlatkin-Troitschanskaia, Hans A. Pant, Miriam Toepper, and Corinna Lautenbach, 301–14: Springer.

Findeisen, Stefanie, Viola K. Deutscher, and Jürgen Seifried. 2021. "Fostering prospective teachers' explaining skills during university education—Evaluation of a training module." *High Educ* 81 (5): 1097–1113. doi:10.1007/s10734-020-00601-7.

Flores, Maria A. 2016. "Teacher Education Curriculum." In Loughran and Hamilton 2016, 187–230.

Gartmeier, Martin, Markus Gebhardt, and Benjamin Dotger. 2016. "How do teachers evaluate their parent communication competence? Latent profiles and relationships to workplace behaviors." *Teaching and Teacher Education* 55: 207–16. doi:10.1016/j.tate.2016.01.009.

Grice, Herbert P. 1975. "Logic and conversation." In *Syntax and Semantics*, edited by P. Coyle and J. Morgan. 3rd ed., 41–58. New York: Academic Press.

Griffith, Robin, and Jan Lacina. 2018. "Teacher as Decision Maker: A Framework to Guide Teaching Decisions in Reading." *Read Teach* 71 (4): 501–7. doi:10.1002/trtr.1662.

Habermas, Jürgen. 1984. *The Theory of Communicative Action.* Boston: Beacon Press.

Hartig, Johannes, and Andreas Frey. 2013. "Sind Modelle der Item-Response-Theorie (IRT) das „Mittel der Wahl" für die Modellierung von Kompetenzen?" *Z Erziehungswiss* 16 (S1): 47–51. doi:10.1007/s11618-013-0386-0.

Hattie, John. 2010. *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement.* Reprinted. London: Routledge.

Hong, Ji Y. 2010. "Pre-service and beginning teachers' professional identity and its relation to dropping out of the profession." *Teaching and Teacher Education* 26 (8): 1530–43. doi:10.1016/j.tate.2010.06.003.

Jenset, Inga S., Kirsti Klette, and Karen Hammerness. 2018. "Grounding Teacher Education in Practice Around the World: An Examination of Teacher Education

Coursework in Teacher Education Programs in Finland, Norway, and the United States." *Journal of Teacher Education* 69 (2): 184–97. doi:10.1177/0022487117728248.

Kleickmann, Thilo, Dirk Richter, Mareike Kunter, Jürgen Elsner, Michael Besser, Stefan Krauss, Marie Cheo, and Jürgen Baumert. 2015. "Content knowledge and pedagogical content knowledge in Taiwanese and German mathematics teachers." *Teaching and Teacher Education* 46: 115–26. doi:10.1016/j.tate.2014.11.004.

König, Johannes, and Sigrid Blömeke. 2009. "Pädagogisches Wissen von angehenden Lehrkräften." *Zeitschrift für Erziehungswissenschaft* 12 (3): 499–527.

König, Johannes, and Sigrid Blömeke. 2010. *Pädagogisches Unterrichtswissen:(PUW); Dokumentation der Kurzfassung des TEDS-M Testinstruments zur Kompetenzmessung in der ersten Phase der Lehrerausbildung:* Humboldt-Univ. zu Berlin.

König, Johannes, Sigrid Blömeke, Lynn Paine, William H. Schmidt, and Feng-Jui Hsieh. 2011. "General Pedagogical Knowledge of Future Middle School Teachers: On the Complex Ecology of Teacher Education in the United States, Germany, and Taiwan." *Journal of Teacher Education* 62 (2): 188–201. doi:10.1177/0022487110388664.

König, Johannes, and Albert Bremerich-Vos. 2020. "Deutschdidaktisches Wissen angehender Sekundarstufenlehrkräfte." *Diagnostica* 66 (2): 93–109. doi:10.1026/0012-1924/a000251.

König, Johannes, Albert Bremerich-Vos, Christiane Buchholtz, Ilka Fladung, and Nina Glutsch. 2020. "Pre–service teachers' generic and subject-specific lesson-planning skills: On learning adaptive teaching during initial teacher education." *European Journal of Teacher Education* 43 (2): 131–50. doi:10.1080/02619768.2019.1679115.

König, Johannes, and Barbara Pflanzl. 2016. "Is teacher knowledge associated with performance? On the relationship between teachers' general pedagogical knowledge and instructional quality." *European Journal of Teacher Education* 39 (4): 419–36. doi:10.1080/02619768.2016.1214128.

Korthagen, Fred. 2017. "A Foundation for Effective Teacher Education: Teacher Education Pedagogy Based on Situated Learning." In *The SAGE Handbook of*

Research on Teacher Education: 2, edited by D. J. Clandinin and Jukka Husu, 528–44. London: SAGE Publications Ltd.

Kultusministerkonferenz. 2004/2019. “Standards für die Lehrerbildung: Bildungswissenschaften.” Beschluss der Kultusministerkonferenz vom 16.12.2004 i.d.F. vom 16.05.2019. Accessed October 05, 2021. https://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung-Bildungswissenschaften.pdf.

Kunter, Mareike, Uta Klusmann, Jürgen Baumert, Dirk Richter, Thamar Voss, and Axinja Hachfeld. 2013. “Professional competence of teachers: Effects on instructional quality and student development.” *Journal of Educational Psychology* 105 (3): 805–20. doi:10.1037/a0032583.

Loughran, John, and Mary L. Hamilton, eds. 2016. *International Handbook of Teacher Education.* Singapore: Springer Singapore.

Lynch, Raymond, and Jennifer Hennessy. 2017. “Learning to earn? The role of performance grades in higher education.” *Studies in Higher Education* 42 (9): 1750–63. doi:10.1080/03075079.2015.1124850.

Magno, Carlo. 2011. “A Closer Look at other Taxonomies for Learning: A Guide for Assessing Student Learning.” *The Assessment Handbook* 5: 50–58.

Nehls, Caroline, Johannes König, Gabriele Kaiser, and Sigrid Blömeke. 2020. “Profiles of teachers’ general pedagogical knowledge: nature, causes and effects on beliefs and instructional quality.” *ZDM Mathematics Education* 52 (2): 343–57. doi:10.1007/s11858-019-01102-3.

Raykov, Tenko, and George A. Marcoulides. 2018. *A course in item response theory and modeling with Stata:* Stata Press College Station, TX.

Richardson, Virginia. 1996. “The role of attitudes and beliefs in learning to teach.” *Handbook of research on teacher education* 2 (102-119): 273–90.

Rothland, Martin. 2020. “Theorie-Praxis-Verhältnis in der Lehrerinnen- und Lehrerbildung.” In Cramer, Koenig, and Rothland 2020, 133–40.

Rots, Isabel, Antonia Aelterman, and Geert Devos. 2014. “Teacher education graduates’ choice (not) to enter the teaching profession: does teacher education matter?” *European Journal of Teacher Education* 37 (3): 279–94. doi:10.1080/02619768.2013.845164.

Samejima, Fumiko. 2010. “The General Graded Response Model.” In *Handbook of Polytomous Item Response Theory Models : Developments and Applications*,

edited by Michael L. Nering and Remo Ostini, 77–107. New York, UNITED STATES: Routledge.

Shulman, Lee S. 1986. “Those who understand: Knowledge growth in teaching.” *Educational researcher* 15 (2): 4–14.

Shulman, Lee S. 1987. “Knowledge and Teaching: Foundations of the New Reform.” *Harvard Educational Review* 57 (1): 1–23. doi:10.17763/haer.57.1.j463w79r56455411.

Sjoer, Ellen, and Jacobiene Meirink. 2016. “Understanding the complexity of teacher interaction in a teacher professional learning community.” *European Journal of Teacher Education* 39 (1): 110–25. doi:10.1080/02619768.2014.994058.

Spillane, James P., Megan Hopkins, and Tracy M. Sweet. 2018. “School District Educational Infrastructure and Change at Scale: Teacher Peer Interactions and Their Beliefs About Mathematics Instruction.” *American Educational Research Journal* 55 (3): 532–71. doi:10.3102/0002831217743928.

Statistisches Bundesamt. 2017. “Bildung und Kultur. Studierende an Hochschulen: Vorbericht. Wintersemester 2016/2017.” Fachserie 11 Reihe 4.1. https://www.statistischebibliothek.de/mir/receive/DEHeft_mods_00067055.

Stokking, Karel, Frieda Leenders, Jan de Jong, and Jan van Tartwijk. 2003. “From student to teacher: reducing practice shock and early dropout in the teaching profession.” *European Journal of Teacher Education* 26 (3): 329–50. doi:10.1080/0261976032000128175.

Struyven, Katrien, and Marijke de Meyst. 2010. “Competence-based teacher education: Illusion or reality? An assessment of the implementation status in Flanders from teachers’ and students’ points of view.” *Teaching and Teacher Education* 26 (8): 1495–1510. doi:10.1016/j.tate.2010.05.006.

Terhart, Ewald. 2007. “Erfassung und Beurteilung der beruflichen Kompetenz von Lehrkräften.” In *Forschung zur Lehrerbildung: Kompetenzentwicklung und Programmevaluation*, edited by Manfred Lüders and Jochen Wissinger, 37–62. Münster: Waxmann.

Terhart, Ewald. 2019. “Teacher Education in Germany.” In *Oxford Research Encyclopedia of Education*: Oxford University Press.

Thompson, Alba G. 1992. “Teachers' beliefs and conceptions: A synthesis of the research.” In *Handbook of research on mathematics teaching and learning: A project of the national council of teachers of mathematics*, edited by Douglas

Grouws, 127–246. Reston, Virginia: National Council of Teachers of Mathematics.

Tomei, Lawrence, and Eletra Gilchrist-Petty, eds. 2018. *Deviant Communication in Teacher-Student Interactions.* Advances in Educational Technologies and Instructional Design: IGI Global.

Toom, Auli, Kirsi Pyhältö, Janne Pietarinen, and Tiina Soini. 2021. "Professional Agency for Learning as a Key for Developing Teachers' Competencies?" *Education Sciences* 11 (7): 324. doi:10.3390/educsci11070324.

van Driel, Jan H., and Amanda Berry. 2012. "Teacher professional development focusing on pedagogical content knowledge." *Educational researcher* 41 (1): 26–28.

Voogt, Joke, and Natalie P. Roblin. 2012. "A comparative analysis of international frameworks for 21 st century competences: Implications for national curriculum policies." *Journal of Curriculum Studies* 44 (3): 299–321. doi:10.1080/00220272.2012.668938.

Voss, Thamar, and Mareike Kunter. 2020. ""Reality Shock" of Beginning Teachers? Changes in Teacher Candidates' Emotional Exhaustion and Constructivist-Oriented Beliefs." *Journal of Teacher Education* 71 (3): 292–306. doi:10.1177/0022487119839700.

Voss, Thamar, Wolfgang Wagner, Uta Klusmann, Ulrich Trautwein, and Mareike Kunter. 2017. "Changes in beginning teachers' classroom management knowledge and emotional exhaustion during the induction phase." *Contemporary Educational Psychology* 51: 170–84. doi:10.1016/j.cedpsych.2017.08.002.

Walz, Kristina, and Edith Braun. 2022. "A Competency Level Model For Communication Skills." *Higher Education Forum* 19: 45–69.

Weinert, Franz E. 1999. *Konzepte der Kompetenz. Gutachten zum OECD-Projekt" Definition and Selection of Competencies: Theoretical and Conceptual Foundations (DeSeCo):* Neuchatel: Bundesamt für Statistik.

Zaragoza, Adriana, Tina Seidel, and James Hiebert. 2021. "Exploring preservice teachers' abilities to connect professional knowledge with lesson planning and observation." *European Journal of Teacher Education*, 1–20. doi:10.1080/02619768.2021.1996558.

Table 1. Descriptives of the Sample.

	N	M	sd
Age	103	23.03	3.65
Semester	104	6.5	4.48
Grade	73*	2.02	0.55
Educational Track	105		
Elementary school	6		
Elementary & Secondary school I	4		
Secondary school I	18		
Secondary school II	52		
Vocational school	5		
Other	20		
Gender	105		
Female	72		
Male	30		
Other	3		

Legend: * For the following analysis we used multiple imputations and estimated the grade for N=100; N – Number of Observations; M – Mean Value; SD – Standard Deviation.

Table 2. Associations Between GPA, Strategic Communication, Understanding-Oriented Communication (standardised regression coefficients)

	Grade
SC	0.13
UOC	-0.09
GPK	-0.29***
Adjusted R^2	0.07
N	100

Legend: SC – strategic communication; UOC – understanding-oriented communication; GPK – general pedagogical knowledge; N – number of observations; * - $p<0.1$; ** - $p<0.05$; *** - $p<0.01$

Table 3. Associations Between Strategic Communication, Understanding-Oriented Communication, GPK, Term, and Age (standardised regression coefficients)

	SC	**UOC**	**GPK**
Semester	0.01	0.09	0.52***
Age	0.26 **	0.27**	-0.14
Adjusted R^2	0.05	0.08	0.22
N	103	103	103

Legend: SC – strategic communication; UOC – understanding-oriented communication; GPK – general pedagogical knowledge; N – number of observations; Significance: * - $p<0.1$; ** - $p<0.05$; *** - $p<0.01$

EIGENSTÄNDIGKEITSERKLÄRUNG

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig, ohne unerlaubte fremde Hilfe und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht. Bei den von mir durchgeführten und in der Dissertation erwähnten Untersuchungen habe ich die Grundsätze guter wissenschaftlicher Praxis, wie sie in der ’Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis’ niedergelegt sind, eingehalten.

_____

Kristina Walz; Lemgo,