

Justus Liebig University Giessen, Germany
Macquarie University Sydney, Australia

Inclusive Education for All: Development of an Instrument to Measure the Teachers' Attitudes

by
Stephan Kielblock
M. A. (JLU)

Thesis submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

June 2018

Table of Contents

Table of Contents	ii
List of Figures.....	v
List of Tables	vi
List of Appendices	viii
Summary	ix
Statement.....	x
Acknowledgements.....	xi
Chapter 1 · Introduction	1
1.1 Introduction	1
1.2 Purpose of the Study	2
1.3 Significance of the Study	3
1.4 Structure of the Thesis	4
Chapter 2 · Literature Review	6
2.1 Introduction	6
2.2 Inclusive Education for All – Understandings and Definitions	7
2.2.1 ‘Inclusive Education for All’ as a New Global Commitment.....	7
2.2.2 ‘Inclusive Education for All’ in National and Local Contexts	13
2.3 Inclusive Education for All and Teachers’ Attitudes	24
2.3.1 Teachers and their Attitudes as a Key to Inclusive Education for All.....	24
2.3.2 Empirical Studies on Teachers’ Attitudes towards Inclusive Education	32
2.4 Specification of the Research Problem	41
2.4.1 Instruments to Measure Teachers’ Attitudes towards Inclusive Education.....	42
2.4.2 Statement of the Research Problem	44
2.4.3 Research Question	46
2.5 Chapter Summary	47
Chapter 3 · Methodology	49
3.1 Introduction	49
3.2 Purpose of the Empirical Study.....	49
3.3 Study Population	50
3.3.1 Countries in which the Present Study was Carried Out.....	50
3.3.2 Further Description of Both Contexts.....	52
3.4 Research Design	56
3.4.1 General Stance: Quantitative Research.....	56
3.4.2 Research Style: Survey Research Using a Questionnaire.....	57

3.4.3 Objective: Develop a New Attitude Questionnaire	61
3.4.4 Scope: Cross-Cultural Research.....	70
3.4.5 Establishing Quality of the Measurement Instrument	77
3.5 Procedures of Developing the Data Collection Instruments.....	79
3.5.1 Develop the English Questionnaire.....	80
3.5.2 Questionnaire Translation and Adaptation in German	86
3.6 Data Collection Procedures.....	94
3.6.1 Approval of the Empirical Study	94
3.6.2 Drawing the Sample.....	95
3.6.3 Sampling.....	96
3.6.3 Administration of the Questionnaire	98
3.7 Quantitative Methods of Analysing Data	99
3.7.1 Preliminary Analyses	99
3.7.2 Scale Analysis.....	103
3.7.3 Validation of the Scale (Hypotheses)	109
3.8 Chapter Summary.....	114
Chapter 4 • Results.....	116
4.1 Introduction.....	116
4.2 The English Attitude Questionnaire	116
4.2.1 Result of the Systematic Literature Review: Relevant Questionnaire Items.....	116
4.2.2 Item Revisions and Response Format	120
4.2.3 Outcome of the Written-Comments Pre-Test.....	123
4.2.4 Final Attitude Items in English Language	124
4.3 The German Attitude Questionnaire	124
4.3.1 Translated Versions.....	124
4.3.2 Comparisons of the Different Item Versions and Decisions.....	124
4.3.3 Final Changes to the Items due to Intensive Pre-Testing.....	128
4.3.4 Final Items in German Language.....	132
4.4 Characteristics of the Sample.....	132
4.5 Internal Structure of the Scale.....	134
4.5.1 Outcome of the Initial Data Examination.....	134
4.5.2 Exploratory Factor Analysis	136
4.5.3 Confirmatory Factor Analysis.....	142
4.5.4 Obtaining the Factor Scores.....	147
4.6 Relation of the Scales' Dimensions to Conceptually Related Aspects	148
4.6.1 Australian Teachers' Attitudes, Self-Efficacy, and Background.....	149
4.6.2 German Teachers' Attitudes, Self-Efficacy, and Background	160
4.6.3 Synopsis of the Results for the Australian and the German sample	170
4.7 Chapter Summary.....	172

Chapter 5 · Discussion.....	173
5.1 Introduction	173
5.2 Interpretation of the Obtained Results	173
5.2.1 Indicators of Inclusive Education for All	173
5.2.2 Internal Structure of the New Instrument	176
5.2.3 Conclusions on the Validation Hypotheses	182
5.3 Limitations of the Study	196
5.4 Implications of the Findings and Conclusion	197
5.4.1 Implications for Policy and Practice	197
5.4.2 Implications for Further Research	198
5.4.3 Overall Conclusion	199
References.....	201
Appendices.....	221

List of Figures

Figure 1. Visual elements that support the understanding of the seven-point scale	70
Figure 2. Depiction of the professional translation into the target language (TL _(P)) from the source language (SL), and the possible sources of discrepancies.....	91
Figure 3. Depiction of the comparisons with regard to item.....	126
Figure 4. Confirmatory factor analysis of the three-factor model	143
Figure 5. Confirmatory factor analysis of the four-factor model	146

List of Tables

Table 1. Decision framework of Weijters et al. (2010) for selecting a response scale format.	68
Table 2. Exclusion criteria for the review of the papers.....	82
Table 3. Comparison between both German versions. Item 3 as an example	89
Table 4. Coding scheme for the systematic comparison	90
Table 5. Result of the search for relevant questionnaire items using a systematic approach.	117
Table 6. Overview on the 14 Papers that Resulted out of the Literature Review.....	119
Table 7. Results of the English-English and the German-German comparisons. Absolute and relative number of items with or without issues	125
Table 8. Discrepancies, special attention needed, and decisions made.....	127
Table 9. Written and oral critique and changes to the wording	131
Table 10. Intended sample size and actual sample	133
Table 11. Exploratory factor analysis for the Australian sample.....	137
Table 12. Exploratory factor analysis for the German sample	140
Table 13. Comparative examination of the support-related items	145
Table 14. Characteristics of the four attitude dimensions	147
Table 15. Differences of the attitudes amongst Australian teachers with regards to gender .	149
Table 16. Differences of the attitudes amongst Australian teachers with regards to age.....	150
Table 17. Differences of the attitudes of Australian pre-service vs. in-service teachers	151
Table 18. Differences of the attitudes of Australian primary vs. secondary teachers.....	152
Table 19. Differences of the attitudes of Australian teachers with regards to years of teaching experience	153
Table 20. Differences of the attitudes of Australian teachers with regards to holding a postgraduate degree/diploma or not	154
Table 21. Differences of the attitudes of Australian teachers with regards to their inclusion-related knowledge	155
Table 22. Differences of the attitudes of Australian teachers with regards to their level of inclusion-related training.....	156
Table 23. Differences of the attitudes of Australian teachers with regards to their inclusive education-related experiences	157
Table 24. Differences of the attitudes of Australian teachers with regards to their self-efficacy in managing behaviour	158
Table 25. Differences of the attitudes of Australian teachers with regards to their self-efficacy in collaboration	159

Table 26. Differences of the attitudes of Australian teachers with regards to their self-efficacy in using inclusive instruction.....	159
Table 27. Differences of the attitudes amongst German teachers with regards to gender.....	160
Table 28. Differences of the attitudes amongst German teachers with regards to age.....	161
Table 29. Differences of the attitudes of German pre-service vs. in-service teachers.....	162
Table 30. Differences of the attitudes of German primary vs. secondary teachers.....	163
Table 31. Differences of the attitudes of German teachers with regards to years of teaching experience.....	163
Table 32. Differences of the attitudes of German teachers with regards to holding a postgraduate degree/diploma or not.....	164
Table 33. Differences of the attitudes of German teachers with regards to their inclusive education-related knowledge.....	165
Table 34. Differences of the attitudes of German teachers with regards to their level of inclusion-related training.....	166
Table 35. Differences of the attitudes of German teachers with regards to their inclusive education-related experiences.....	167
Table 36. Differences of the attitudes of German teachers with regards to their self-efficacy in managing behaviour.....	168
Table 37. Differences of the attitudes of German teachers with regards to their self-efficacy in collaboration.....	169
Table 38. Differences of the attitudes of German teachers with regards to their self-efficacy in using inclusive instruction.....	169

List of Appendices

Appendix A · Ethics Approval	221
Appendix B · Approved SERAP	222
Appendix C · Approval of Data Collection at the JLU	223
Appendix D · Approval of Conduct of the Study at Schools in Hesse, Germany	224
Appendix E · Analysis of the TEIP scale	226
Appendix F · Univariate Analysis of the Attitude Items	227
Appendix G · Correlative Analysis of the Attitude Items	230
Appendix H · Structure Matrices	232
Appendix I · Additional Information as they pertain to the Confirmatory Factor Analysis ..	234
Appendix J · Survey in English Language	235
Appendix K · Survey in German Language	239
Appendix L · Original Items and Revised Items	243
Appendix M · All Versions Compared	246

Summary

In recent years, the term inclusive education has played an unprecedented role in research and policies across the globe. It is relatively accepted to differentiate between a narrow and a broad understanding of inclusive education. On the one hand, the more narrow understanding focuses on the placement and the catering for specific students, such as those with identified special educational needs and/or disabilities (SEND). On the other hand, a more broad understanding of inclusive education incorporates views on the diversity of all students and supportive learning environments for all.

In order to foster inclusive education *for all*, the literature suggests that it would be of vital importance to gain empirical data about the teachers' attitudes towards inclusive education *for all*. Yet, recent review studies have uncovered that particularly empirical studies tend to utilise a view on students with SEND and that there seems to be a lack of attitude measurement instruments that operationalise a broader understanding of inclusive education *for all*.

Accordingly, the present study attempted to make a unique contribution to the field of inclusive education in that it reviewed a substantial number of studies and developed a new, sound and robust instrument to measure different facets of the teachers' attitudes towards inclusive education *for all* students. Teacher samples were drawn in Australia (n=146) and in Germany (n=238), and the data analysis revealed four dimensions of the teachers' attitudes; namely, the vision, the differentiation, the general practices, and the supports as they pertain to inclusive education *for all*. The validity of the measurement was established and the final version seemed to be ready to use in further studies that attempt to utilise inclusive education *for all*, rather than *for some*.

Statement

I hereby certify that this work has been submitted in identical form to the Justus Liebig University Giessen and the Macquarie University Sydney in fulfilment of the requirements of the Cotutelle Agreement between both institutions.

This work has not previously been submitted for a degree or diploma in any university. To the best of my knowledge and belief, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.


Stephan Kielblock

31 December 2017

Acknowledgements

My particular interest in inclusive education was instigated in July 2014, when I met Jeremy J. Monsen at a conference at Cambridge University, UK. Our discussions led to a workshop on cross-cultural perspectives on teachers' attitudes towards inclusive education that we both held together in August 2015 in Limassol, Cyprus. In this workshop, our focus was – naturally – on including students with identified special educational needs. One unexpected outcome of this workshop was written by Tinde Kovac-Cerovic and Alberto Nagle Cajes on a poster, which I have until today on my office wall: “We suggest a broader approach.” This encouraged me to study inclusive education *for all* the way it is represented in the present thesis. I would like to thank all workshop participants to share their views, and Jeremy for our fruitful research collaboration and friendship over the years.

Like this example illustrated, many substantial gains in my personal understanding of inclusive education *for all* were only possible through exchange with international colleagues. In this regard, I am more than grateful that, generally, the Justus Liebig University Giessen, Germany (JLU) and, specifically, Ludwig Stecher supported all my international travels over the years. Later, I was also affiliated to the Macquarie University Sydney, Australia (MQU), which also supported my international ambitions to travel to conferences in countries such as Finland, Korea, and the United States. I am thankful for the idealistic and monetary support.

All the ideas that I acquired over the years needed to be transformed into a concrete research project and into a written thesis. This was a fascinating journey, which was guided by my supervisors Ludwig Stecher (JLU) and Stuart Woodcock (MQU). I am very thankful that both of my supervisors were continuously encouraging me over the years to (try to) think beyond the mainstream.

There is a personal history behind reaching this point in my academic career, and I would like to thank all the unique individuals who supported me over the last three decades or so, such as my parents and siblings, my peers (and their families), and my academic teachers. A proverb says, it takes a village to raise a child; similarly, I would say, it takes a village to write a thesis. I would like to acknowledge that there were many people involved in intensively supporting me over the last few years, so that I had the capability to conduct and write the study in the present form. I would like to express my deepest appreciation to these people; namely, my beloved wife and my beloved son, our families and friends, the BiFo team and the StEG team, and the students and teachers, who participated in the present study. Without you all, nothing like this would have been written in the way I was able to write it.

Chapter 1 · Introduction

1.1 Introduction

In recent years, the term inclusive education has played an unprecedented role in research and policies across the globe, which gave rise to a variety of different understandings of this concept (Ainscow, Booth, & Dyson, 2006; Ainscow & Miles, 2008; Dyson, 2004, 2014; Göransson & Nilholm, 2014; Haug, 2017; Messiou, 2017; Miles & Singal, 2010; Nilholm & Göransson, 2017; Thomas, 2013; Waitoller & Artiles, 2013). It is relatively accepted amongst scholars (such as Arduin, 2015; Armstrong, Armstrong, & Spandagou, 2011; Miles & Singal, 2010; Opertti, Brady, & Duncombe, 2009; Shyman, 2015), to differentiate between a narrow and a broad understanding of inclusive education. On the one hand, the more narrow understanding focuses on the placement and the catering for *specific students*, such as those with identified special educational needs and/or disabilities (SEND). On the other hand, a more broad understanding of inclusive education incorporates views on the diversity of *all students*, and the changes that the schools and the school system must pass through to be able to provide a supportive learning environment for all.

In a way, both of these perspectives are represented in different global policies, mostly advocated by the United Nations Educational, Scientific and Cultural Organization (UNESCO). The perspective that *some* students need particular attention is promoted by UNESCO's 'Inclusive Education' policies; most prominent, the Salamanca Statement (UNESCO, 1994). And the perspective that education should be available *for all* students is represented in UNESCO's 'Education for All' policies; most prominent, the World Declaration on Education for All (UNESCO, 1990).

It is obvious that inclusive education (in a wider or narrower understanding) is not just a global concept, but that its content has real effects for those countries that subscribe to its ideals. One of the most visible effects can be examined on the school level and researchers such as Ainscow, Booth, and Dyson (Ainscow et al., 2006; Ainscow, Farrell, & Tweddle, 2000; Ainscow & Miles, 2008; Booth, 1995; Booth & Ainscow, 2011; Dyson, 2004, 2014) have argued that inclusion needs to be realised through school development, including the school's local community. Although schools provide the environments for inclusive teaching practices and inclusive student-teacher interactions, there are convincing arguments and there is strong

evidence that at a fundamental level the teachers and their attitudes are the key to inclusive education for all of the students.

Concerning the teachers' attitudes towards inclusive education, a large evidence base is available from a variety of previous empirical studies. However, this evidence base is challenged by recent review studies. Researchers have pointed out that the focus on some students rather than on all is much more common in studies (Messiou, 2017), and that empirical studies tend to understand inclusive education as catering for some students (e.g. with SEND) specifically, while more conceptual studies utilise inclusive education as catering for all students (Nilholm & Göransson, 2017). It is well documented that German instruments to measure attitudes towards inclusive education generally focus on students with SEND (Ruberg & Porsch, 2017).

Against the backdrop of this situation, the present study attempted to make a unique contribution to the field of inclusive education in that it reviewed a substantial number of studies and instruments and developed a new instrument to measure different facets of the teachers' attitudes towards inclusive education for all students, which is also usable in cross-cultural investigations.

1.2 Purpose of the Study

As noted before, investigations in the area of inclusive education are confronted with a variety of understandings of inclusive education. Hence, an initial purpose was to clarify the meaning of inclusive education. As it was agreed that inclusion generally needed to be understood as a normative idea which is connected to certain values (Haug, 2014, 2017), the present investigation started with examining the relevant global contexts in which the ideas of 'inclusive education' and 'education for all' evolved. As an overarching term, 'inclusive education *for all*' was coined and discussed in the present study.

After such preliminary clarifications, the main purpose of the present study was to investigate how the teachers' attitudes towards inclusive education *for all* can be measured. The attempted new measurement instrument was thought to be sound and robust. Concerning the former, the instrument should allow a valid and reliable measurement. And concerning the latter, the instrument should be ready to be used in multi-language, multicultural and multinational settings. The attitudes were assumed to comprise certain facets; accordingly, the purpose of the present study was also to establish certain dimensions of the measurement instrument.

The resulting measurement instrument of the present study was thought to provide new opportunities for researchers to study the teachers' attitudes towards inclusive education *for all*, without narrowing down possible teachers' responses to aspects as they pertain basically to mainstreaming or integration.

1.3 Significance of the Study

Teachers and their attitudes are crucial for inclusive education *for all* to take place in 'real-world' practices. Hence, the present study contributes generally to the research knowledge as it pertains to teachers and their attitudes.

As noted before, many understandings of inclusive education are apparent and considerable confusions exist in this regard. The present study introduced 'inclusive education *for all*' as a term, which is not just another understanding besides many others, but it attempts to integrate some of the existing understandings. Similarly, the study started utilising 'inclusive education *for some*' as a term that signifies students with SEND, yet, at the same time dissociates from former notions of integration and mainstreaming. The significance of these two terms is that some initial steps were made in the present study towards reducing the conceptual confusion through integrating certain understandings.

The main purpose was to find a way to measure the teachers' attitudes towards inclusive education *for all*. The whole present study represents the ambitious attempt to develop such a new instrument. Besides some limitations, all procedures of the empirical study were realised the way they were conceptualised in accordance to an in-depth discussion of the methodological literature. Hence, the new measurement instrument resulted in a sound and robust scale. This scale comprised 12 items, which formed four dimensions of the teachers' attitudes towards inclusive education *for all*.

In this way, the present study makes a unique contribution to the field of inclusive education. This pertains not only to further research, which might particularly gain new insights when utilising the new measurement instrument, but also to all relevant stakeholders in education, because they might adapt the term inclusive education *for all* (as opposed to many others who continue their sole focus on some, rather than all) and they might be informed about the new instrument (as opposed to other measurement instruments that most stakeholders in education are continuously confronted with and that continue to focus on some, rather than all).

1.4 Structure of the Thesis

The present study comprises five chapters. This Chapter one has given an overview on the fundamental issues that this study takes up, and elaborates further. The main purpose of the present study to develop a new instrument that measures the teachers' attitudes towards inclusive education *for all* is justified, before the significance of the obtained results are delineated.

Chapter two examines in the first part inclusive education from a global perspective. Particularly, UNESCO's efforts towards 'Inclusive Education' and 'Education for All' are discussed as important driving forces on the global level, which represent a narrow and a wide understanding of inclusive education, respectively. Drawing on global developments, it is noted that, conceptually, these two understandings were moving closer together over the years. Hence, 'inclusive education *for all*' was established as a term, which refers to both the 'inclusive' and the 'for all' character of education. After a detailed discussion on how and why global concepts get deflected when transposed into national, local, school, and classroom practice levels, the crucial importance of teachers and their attitudes are emphasised for implementing inclusive education *for all*. Accordingly, a variety of empirical studies were reviewed on the teachers' attitudes towards inclusive education; yet, a closer examination revealed that all these studies were not about inclusive education *for all*, but about inclusive education *for some*. As the research problem it is highlighted that in order to implement inclusive education *for all*, research needs to be carried out on the teachers' attitudes towards inclusive education *for all*, rather than *for some*.

In Chapter three, the empirical part of the study is specified with regards to the main purpose of the present study to develop a new instrument to measure the teachers' attitudes towards inclusive education *for all*, which is sound, robust and comprise different dimensions. In order to allow for developing a robust instrument, two contexts were selected in an informed way, where the study was carried out; namely Australia and Germany. Furthermore, the key parameters of the study are discussed and justified with regards to the general stance, the research style, the objective, the scope and how to establish the quality of the measurement. The procedures regarding how the questionnaire was developed in English language, and how it was translated to German, were building on all these intensive discussions with regards to the key parameters. The data collection procedures are detailed, before the procedures of analysing data are discussed and defined.

Chapter four, then, presents the results as they were obtained through conducting the study as it was determined in the previous chapter. The depiction of the results starts with the

systematic literature review, which was thought to result in a number of relevant items in English that could be utilised as indicators of the teachers' attitudes towards inclusive education *for all*. After revision and pre-testing, the final attitude items were translated and adapted in German; hence, the results of the according processes and also the German pre-test are described. The resulting samples from Australia (n=146) and Germany (n=238) are outlined, before the results of the statistical analyses are presented. This part is divided into examining the internal structure of the data, and examining the relationships to other teacher variables such as self-efficacy and experiences.

The final Chapter five presents in a first part in-depth interpretations of all of the obtained results. After describing and reflecting the obtained indicators of inclusive education *for all*, the results as they pertain to the internal structure of the new instrument are discussed. This is carried out for each of the four dimensions, and, finally, for the overall structure of the instrument. Then, conclusions are presented on all of the validation hypotheses, which were specified in order to gain insights in how the instrument and its dimensions related to other teachers' aspects. After considering limitations of the study, conclusions are reached regarding the new instrument, including four dimensions, which can be considered as sound and robust. Implications of the study's findings are discussed and an overall conclusion is given in the end of the study.

Chapter 2 · Literature Review

2.1 Introduction

It is widely acknowledged that inclusive education should consider *all* individuals, and not *some* specifically (see e.g. Thomas, 2013). Yet, the most recent reviews clearly point to the fact that studies on inclusive education are focussing to a large extent on students with identified special educational needs and/or disabilities (SEND) (Haug, 2017; Messiou, 2017; Nilholm & Göransson, 2017; Ruberg & Porsch, 2017). Furthermore, there seems to be a gap between conceptual or theoretical studies on the one hand, which advocate a more diverse learners- and ‘for all’-related perspective and empirical studies on the other hand, which utilise an understanding of inclusive education as the placement of students with SEND in the mainstream (Nilholm & Göransson, 2017).

As the first major step, an extensive discussion of conceptual understandings and definitions of inclusive education will be carried out in the following section. At international level, developments towards ‘education for all’ and ‘inclusive education’ are reviewed, and an attempt is made to think education for all and inclusive education together, which leads to the postulation of ‘inclusive education *for all*’ as guiding principle on the global level. If global policies and concepts are translated to the national, local, school, and classroom practice levels, tensions and diverse interpretations are created, which is discussed subsequently. As a second major step in this chapter, the teachers are emphasised as one of the main stakeholders in education. It will be argued that their attitudes need to be considered as crucial for implementing more inclusive practices. The empirical literature on teachers’ attitudes towards inclusive education, which is reviewed subsequently, demonstrates that the instruments to measure the teachers’ attitudes towards inclusive education were exclusively focused on the placement of particular students and/or on students with SEND. Hence, how to operationalise inclusive education *for all* for an empirical investigation of the teachers’ attitudes is established as the research problem of the present study, which is described and justified in the end of this chapter.

2.2 Inclusive Education for All – Understandings and Definitions

At the heart of inclusive education are the efforts of the United Nations, and in particular of the United Nations Educational, Scientific and Cultural Organization (UNESCO) (Burnett, 2008; Kiuppis, 2014; Mundy, 2016; Peters, 2007). Many studies give reference to UNESCO's Salamanca Statement (UNESCO, 1994), and some to UNESCO's World Declaration on Education for All (UNESCO, 1990) as the origin of the present inclusive education-related thinking.

2.2.1 'Inclusive Education for All' as a New Global Commitment

Education for all and inclusive education are commitments on a global level that are closely related to the UNESCO as one of the main driving forces towards achieving inclusive and quality education for all. Both terms have their own history and a particular meaning; yet, not much research effort was expended on how both might complement each other.

The Early Adoptions of the Idea to Provide Education for All

The idea of the education for all was laid out in the 1940s after World War II (Mundy, 2016; Roche, 2016). The notion that education should be available 'for all' was explicitly formulated in the constitution of the UNESCO, which was adopted in London on 16 November 1945. In this document, it was stated that the States' Parties are

“believing in full and equal opportunities for *education for all*, in the unrestricted pursuit of objective truth, and in the free exchange of ideas and knowledge, are agreed and determined to develop and to increase the means of communication between their peoples and to employ these means for the purposes of mutual understanding and a truer and more perfect knowledge of each other's lives” (UNESCO, 2014, p. 5; italics added).

The idea that “everyone has the right to education” was also articulated by the United Nations General Assembly in Paris on 10 December 1948 (UN, 1948) in Article 26 of the Universal Declaration of Human Rights.

Besides these early notions of the availability of education for all students, various authors (such as Anderson & Boyle, 2015; Guo, 2014; Kiuppis, 2014; Tomlinson, 2015) highlighted that UNESCO's Education for All, as a global movement, was initiated more recently after the Cold War. In 1990, the World Conference on Education for All in Jomtien, Thailand resulted in the approval of the World Declaration on Education for All. In its Preamble, the assertion of the Universal Declaration of Human Rights was repeated that “everyone has a right to

education” (UNESCO, 1990, p. 1). This notion is mirrored in the World Declaration on Education for All “that education is a fundamental right for all people, women and men, of all ages, throughout our world” (UNESCO, 1990, p. 2). The declaration states that

“to serve the basic learning needs of all requires more than a recommitment to basic education as it now exists. What is needed is an “expanded vision” that surpasses present resource levels, institutional structures, curricula, and conventional delivery systems while building on the best in current practices” (UNESCO, 1990, p. 4).

Furthermore, the declaration states that disparities must be reduced. This pertains to a large variety of underserved groups and it is emphasised particularly that

“the learning needs of the disabled demand special attention. Steps need to be taken to provide equal access to education to every category of disabled persons as an integral part of the education system” (UNESCO, 1990, p. 5).

The Jomtien World Declaration was controversially discussed with regard to its influence of developments after 1990. On the one hand, both the conference and the declaration was a unifying element that connects all countries with a common vision. Ainscow and Miles (2008) call it a ‘ground-breaking’ conference, because a number of countries worldwide committed to achieve common goals with regard to providing education for all. On the other hand, the commitments seemed not to be translated into action after the conference. Despite the available resources – such as the Framework for Action to Meet Basic Learning Needs (which was also adopted at the Jomtien conference) and over one hundred pages of background information on the ‘expanded vision’ provided by Haddad, Colletta, Fisher, Lakin, and Sutton (1990) – an implementation or initiation of relevant steps scarcely happened in the different countries. In the words of Mundy (2016, p. 7), ‘little tangible action’ was noticeable after the Jomtien conference, despite the ‘glorious pledges and commitments’. In addition, authors such as Ainscow and Miles (2008) criticised that the World Declaration on Education for All might have left too much room for interpreting the notion of ‘all’ as ‘almost all’. In their view, only the so-called Salamanca Statement ensured later that all really does mean all (Ainscow & Miles, 2008).

In 1994, the World Conference on Special Needs Education took place in Salamanca, Spain. The participating delegates of governments and international organisations approved the Salamanca Statement and its Framework for Action (UNESCO, 1994). The Salamanca Statement reaffirms the Universal Declaration of Human Rights and the Education for All agenda. The Salamanca Statement leaves no doubt that

“regular schools [...] are the most effective means of combating discriminatory attitudes, creating welcoming communities, building an inclusive society and achieving education for all” (UNESCO, 1994, p. ix).

Both conferences and related policies lay the foundations for the UNESCO pushing forward towards Education for All (as agreed in Jomtien) and Inclusive Education (as agreed in Salamanca).

The Dakar Era (2000-2014) and the Millennium Development Goals

In 2000, the World Education Forum was held in Dakar, Senegal. As Mundy (2016, p. 7) noted, this conference lead into ‘a much more productive Education for All decade’, which Mundy (2016) refers to as the ‘Dakar era’. The outcome of the conference is known as the Dakar Framework for Action, which already carries “action” in its name. In this Framework, the necessity to action is pointed out and agreed upon explicitly:

“The Dakar Framework is a collective commitment to action. Governments have an obligation to ensure that EFA goals and targets are reached and sustained. This is a responsibility that will be met most effectively through broad-based partnerships within countries, supported by co-operation with regional and international agencies and institutions.” (UNESCO, 2000, p. 8)

Since 2000, UNESCO’s Education for All and Inclusive Education were systematically fostered. The Dakar Framework specified regional frameworks for action that were thought to be achieved by the different countries within 15 years. In addition, as pointed out for example by Mundy (2016), the United Nations released eight so-called Millennium Development Goals (MDGs), which urged the countries to achieve amongst others universal primary education (goal 2). Moreover, the UNESCO established the periodically conducted UNESCO Education for All Global Monitoring Report.

Another crucial development after Dakar was the publication of the Guidelines for Inclusion in 2005. As it was emphasised by different inclusion researchers (such as Armstrong et al., 2011; Berlach & Chambers, 2011; Opertti & Brady, 2011), these guidelines relate the vision of education for all and the vision of inclusive education to each other. In the guidelines (UNESCO, 2005), it is critically acknowledged that the education for all movement had scarcely taken up issues of special needs, and vice versa that inclusive education was not recognised as an essential element of the education for all movement. In the following quote from UNESCO’s Guidelines for Inclusion, it is articulated that providing education for all

students is only possible when discrimination and exclusion is taken seriously and is explicitly tackled:

“It is important to highlight that Education for All does not automatically imply inclusion. Inclusion properly understood is precisely about reforming schools and ensuring that every child receives quality and appropriate education within these schools. To this extent, inclusion is critical to the EFA [Education for All] movement since without it, a group or groups of children are excluded from education. Thus, EFA by definition cannot be achieved if these children are excluded. Both EFA and inclusion are both about access to education, however, inclusion is about access to education in a manner that there is no discrimination or exclusion for any individual or group within or outside the school system.” (UNESCO, 2005, p. 29).

Like these guidelines, many other important developments can be recognised in the years after Dakar. For example the establishment of the Convention on the Rights of Persons with Disabilities (discussed with regard to the inclusive education movement e.g. by Armstrong et al., 2011; see also Section 2.2.2). Yet, the positive developments proceed slower than originally thought that they would. Mundy (2016) stated that “despite much good news, it is important to note that many of the promises of the Dakar era did not gain the momentum expected” (Mundy, 2016, p. 9).

The Incheon Era (Since 2015) and the Sustainable Development Goals

In 2015, the new sustainable development agenda was adopted by the General Assembly of the United Nations. The former eight MDGs were revised into seventeen Sustainable Development Goals (SDGs). The original education-related goal (MDG 2) “achieve universal primary education” was reformulated in the goal to “ensure inclusive and quality education for all and promote lifelong learning” (SDG 4).

Also in 2015, the World Education Forum 2015 took place at Incheon, Republic of Korea. The participants reaffirmed the vision of education for all and recognised “with great concern that we are far from having reached education for all” (UNESCO, 2015, p. 5). The so-called Incheon Declaration presented a “new vision for education” (UNESCO, 2015, p. 6) for the next 15 years (until 2030). The targeted direction for worldwide developments in establishing education for all students is in line with the SDG 4, as it is explicitly stated in the Incheon Declaration (UNESCO, 2015). The goals specified by the Incheon Declaration cover a wide range of topics such as access to education, inclusion and equity, gender equality, quality education, and lifelong learning opportunities. Inclusion and equity means to address exclusion and ensure that ‘no one is left behind’. The declaration states:

“Inclusion and equity in and through education is the cornerstone of a transformative education agenda, and we therefore commit to addressing all forms of exclusion and marginalization, disparities and inequalities in access, participation and learning outcomes. No education target should be considered met unless met by all. We therefore commit to making the necessary changes in education policies and focusing our efforts on the most disadvantaged, especially those with disabilities, to ensure that no one is left behind.” (UNESCO, 2015, p. 8)

As it is clear from this quote, the Incheon Declaration reflects a broad definition of inclusive education, as it is noted for example by Messiou (2017). Compared to the previous Dakar Framework, it is noticeable that the mission to tackle exclusion and marginalisation in all forms complements the importance of education for all students. Both perspectives – to provide education *for all* students, and to tackle exclusion *of some* students – are emphasised. Currently, the ‘Incheon era’ (this term is inspired by the notion of the ‘Dakar era’ used by Mundy, 2016, p. 9) has just started, and the Incheon Declaration sets the direction until 2030. It remains to be seen, what developments might be achieved by then.

An Attempt to Establish ‘Inclusive Education for All’ as a New Guiding Principle

Especially within the Dakar era, attempts were made to synchronise the efforts of Education for All, inclusive education and the broader commitments to the MDGs. Recently, UNESCO’s definitions of inclusive education and education for all have been aligned. Inclusive education attempts to ensure that “all learners have access to quality education that meets basic learning needs and enriches lives”, and education for all means providing “quality basic education for all children, youth and adults” (www.unesco.org; accessed on 20/03/2017). It was noted by scholars such as Kiuppis (2014) that these newly established directions of ‘Inclusive Education’ and ‘Education for All’ as two of UNESCO’s missions blurs the boundaries between them. According to Kiuppis (2014), both aim at achieving that education supports individuals in reaching their full potential and that discrimination comes to an end. In addition, and in line with this, the newly established UN SDG 4 works systematically towards ‘inclusive and quality education for all’, which explicitly combines the notions of inclusive education and education for all in one expression.

Researchers have tried to bring both the perspective on education *for all* students (as advocated by UNESCO’s Education for All) and the perspective on tackling exclusion *of some* students to foster education for all (as advocated by UNESCO’s Inclusive Education) together. As early as in 2004, Peters (2004) noted that Education for All and Inclusive Education needs to join forces under a new kind of thinking and planning, which she called “Education for All-Together” (Peters, 2004, p. 47). Recently, Thomas (2013) called for a new kind of inclusive

thinking, which considers inequality and equity at the same time. According to this perspective, young people need recognition, respect and identity. ‘Community’ is at the heart of this new thinking, as Thomas (2013) pointed out, and it is crucial to examine the role of schools in enabling community for students to prosper within such community. Shyman (2015) argued in a similar direction, when he emphasised a new definition of inclusive education based on social justice. In the words of Shyman (2015), this definition states that “*all* individuals, regardless of exceptionality, are entitled to the opportunity to be included in regular classroom environments *while* receiving the supports necessary to facilitate accessibility to both environment and information” (Shyman, 2015, p. 351). This perspective combines the notion of education for all and the notion that some might need additional support. Miles and Singal (2010) reiterate that UNESCO’s Education for All is likely to overlook the issue of continued exclusion of particular individuals, while UNESCO’s Inclusive Education tends to demarcate special cases as separate issues. In this way, Miles and Singal (2010) make a clear point that issues such as ‘disability’ need to be “recognised as one of many issues of difference and discrimination, rather than as an issue on its own” (p. 11).

If inclusive education and education for all are brought together in just one expression, it seems relatively obvious to use ‘education’ as a link between them and combine them into the term ‘inclusive education for all’. In the literature, this exact term was used before. For example, Miles and Singal (2010) mentioned in one sentence that “the extent to which more inclusive educational practices are promoted at country level will depend on the development of a clear understanding of the concept of ‘inclusive education for all’ in the cultural contexts in which it is developed” (p. 8). Another example is Carrington et al. (2012), who mentioned the term in their chapter title “towards an inclusive education for all” (p. 3). Similarly, the term was used as a subtitle of a book: “school without walls: inclusive education for all” (Jha, 2008). Yet, none of these references, or any other reference that mentioned this term in a way (such as Arduin, 2015; Avramidis & Kalyva, 2007; Costello & Boyle, 2013; Loreman, 2014) have defined its meaning or have used it consistently.

As noted in the previous paragraphs, recently, there is a push forward towards combining the idea of inclusive education and the idea of education for all. However, there seems to be not yet any established term to refer to education as being inclusive *for all* rather than *for some*. The present study attempts to take up the term ‘inclusive education for all’, as it was mentioned in previous works, to define its meaning, and to use it consistently. To coin this term in this way, it would add to the education for all perspective that discrimination exists and needs to be tackled and would add to the inclusive education perspective that the focus on all students should not be lost out of sight. Although the expression of ‘inclusive education for all’ might

be not new in itself, the present attempt of a definition and the present attempt to use it accordingly, might justify to refer to ‘inclusive education *for all*’ as being a new concept. This new concept gives emphasis to the fact that the principle of inclusive education is thought to be relevant for all students rather than for some particularly. However, many recent studies on inclusive education are not focussing on all students but students with SEND (see evidence in Section 2.3.2). In order to be able to refer to studies like that, the concept of ‘inclusive education *for some*’ is also introduced in this study. To differentiate between inclusive education *for all* and inclusive education *for some* allows one to describe that researchers tend to utilise and operationalise former ideas of mainstreaming and integration under the umbrella of inclusion. However, inclusive education was and is supposed to refer to all students rather than to particular groups of students, which this Section 2.2.1 clearly demonstrated (see detailed discussion in Section 2.4). In this sense, it is not suggested that ‘for all’ and ‘for some’ are both, in a normative sense, valid ways to think about and conduct research on inclusive education. The differentiation is supposed to allow to describe that the term might have changed (e.g. from integration to inclusion), but that researchers tend to remain in their former mindsets.

The term inclusive education *for all* on the global level is defined in the present study in accordance with UNESCO’s policies and with the discussed references (such as Ainscow et al., 2006; Ainscow & Miles, 2008; Miles & Singal, 2010; Peters, 2004; Shyman, 2015; Thomas, 2013): Inclusive education *for all* describes the presence, participation and achievement of *all* students in education, and the imperative to tackle exclusion and marginalisation that *some* individuals face with regards to their presence, participation and achievement in education, and to initiate all necessary steps on all necessary levels that these individuals are embraced by the notion of ‘all’. In this way, this term might be able to embrace both the education *for all* students perspective and the inclusive education (as strengthened education *for some* students) perspective; and it might therefore be able to explicate the direction for education in the next number of years, which was inscribed implicitly in the Incheon Declaration (UNESCO, 2015).

2.2.2 ‘Inclusive Education for All’ in National and Local Contexts

Global policies are relatively abstract, which means that they need to be translated into national and local contexts. These translation processes are highly complex and a variety of aspects deflect the translation in other directions than intended, such as the already existing laws on different societal levels, monetary flows, involved agencies, political directions, past events in the society, utilisation of particular words and phrases in policies, values, attitudes and so forth. The dynamics of the processes depend largely on the kind of global policy and the context in which it is supposed to be implemented, or even to be made legally binding. Hence, in the

following section, a selection of some major issues of translating global inclusive education policies into national and local contexts are highlighted as they seem to dominate current inclusive education discourses. These include diverse definitions and understandings of inclusive education in different contexts, diverse understandings as they are represented in different educational policies, the way SEND are defined and identified in certain contexts, schools as important institutions to implement more inclusive ideas, and the general pedagogic approach.

Diverse Meanings of Inclusive Education

Research that tried to clarify what inclusive education meant repeatedly have reached the conclusion that, generally, there were different understandings. Ainscow et al. (2006) developed a typology of six different understandings of inclusive education (see also Ainscow & Miles, 2008). As opposed to the previously presented reflections on the global (policy) level, Ainscow et al. (2006) were more interested in examining how inclusive education was understood within countries, government policies and schools. Being concerned with including students with SEND, or trying to reduce the exclusion of students whose behaviour is considered as being difficult, are the first two ways of thinking about inclusive education, respectively. According to Ainscow et al. (2006), the third understanding relates inclusive education to all groups that are considered as being particularly vulnerable to exclusion. Developing schools for all is the fourth, and Education for All as advocated by the UNESCO is the fifth way of thinking about inclusive education. According to Ainscow et al. (2006), there is a sixth understanding: the principled approach to education and society. This way of thinking is not in favour of any of the aforementioned understandings, but it acknowledges that the understanding of inclusive education, as adopted by particular schools, is not unalterable and solid, but it is ‘a never-ending process’ of developing and scrutinising the current view on inclusive education (Ainscow et al., 2006). This implies that the “emphasis should be less on what inclusion might look like and more on how it might be developed with schools” (Ainscow et al., 2006, p. 23).

Nearly a decade later, Göransson and Nilholm (2014) proposed four definitions of inclusive education, which were reached through a literature review of research articles. Being concerned about the placement of students with SEND is the first, and being concerned about meeting their social/academic needs is the second kind of definition. According to Göransson and Nilholm (2014), a third definition of inclusive education is to meet the social and academic needs of all students, and fourth, inclusive education as creating communities, which is based on notions of “equity, care [...], justice, honouring of subjugated knowledge and valuing

diversity” (Göransson & Nilholm, 2014, p. 270). The postulation of these four definitions has been criticised (Dyson, 2014; Haug, 2014). Haug (2014) argued that inclusive education should be understood more as an overarching normative idea which is connected to certain values. This argument is similar to the general (normative) vision of inclusive education *for all* as it has been discussed in Section 2.2.1. Dyson (2014) pointed out that inclusive education needs to be understood as a principle that “is embodied in different ways in different contexts” (Dyson, 2014, p. 282), which was described in the previous paragraph as that principled approach to education (Ainscow et al. (2006). Contrary to the attempt of Göransson & Nilholm (2014) to find a generally accepted classification, such a view would emphasise that inclusion has a substantially different meaning for each school and its community. Besides this critique, Göransson and Nilholm (2014) demonstrated at least that a great variety of conceptual understandings of inclusive education is present amongst researchers from different contexts. The issue that inclusive education seems to be a relatively unambiguous concept in international policies (e.g. in the context of the UNESCO; see Section 2.2.1), yet, seems to have diverse meanings on more local levels, needs further consideration in the following paragraphs, because it relates to the methodological approach used in the present study.

Noticing that inclusive education research across contexts is not easy to carry out because of the lack of a generally accepted definition, Dyson (2004) formulated that inclusive education is “a highly slippery concept, particularly when it is used across the boundaries of different education systems” (p. 614). Until today, efforts to clarify the concept of inclusive education across different contexts (Haug, 2017; Nilholm & Göransson, 2017; Schneider, 2015) did not result in a single embracing understanding. Hence, the conclusion might be that the ‘conceptual diversities’ (Göransson & Nilholm, 2014; Kruse & Dederling, 2017) of inclusive education cannot be resolved – at least on the national and local level. However, as argued previously in the present study, on the global level, the understanding of inclusive education *for all* seemed to have little variation across the globe (on the contrary, it was even criticised by Tota, 2014 that the policy development by the UNESCO is unanimous and therefore structurally undemocratic). Hence, the discrepancies, as they were identified in reviews (such as Ainscow et al., 2006; Nilholm & Göransson, 2017), seemed to be created when the global suggestions were translated into the national and local policies and practices.

Inclusive Education in Policies on the National and Local Level

As noted previously, there is a certain agreement on the global level of what inclusive education *for all* comprises of; yet, if understandings of the concept are compared on the national or local level, different views are apparent. The aforementioned variety of definitions is mirrored in

and/or emerges from policy documents on the national/local level that are characterised by considerable discrepancies and contradictions (Armstrong et al., 2011; Hardy & Woodcock, 2015; Slee & Allan, 2001). Three aspects seemed particularly worth mentioning with regard to the emergence of differing understandings of inclusive education. *First*, inconsistencies might arise on the national level, because of different *interpretations* of the global policies by the leading governments and responsible agencies on the national level. For example, policies as they pertain to inclusive education in the United States reflect a number of the principles of transnational policies, yet, one inclusive education policy in Canada re-iterates integrative notions of bringing those students with SEND into the mainstream (Hardy & Woodcock, 2015). Hence, some national policies are generally in line with the global policies, while others are not. *Second*, besides the actual content, another issue that might lead to inconsistencies, is the *vagueness* of some policy texts. Slee and Allan (2001) for example demonstrate through their analysis that a policy, which was supposed to support inclusive education, keeps systematically away from making any concrete suggestions or give any directives for action. The researchers conclude that the policy is “a license to do nothing” (Slee & Allan, 2001, p. 183). It seems very likely that such policies that leave much space for interpretation produce a variety of differing understandings. *Third*, inconsistencies might also arise, because the local education authorities have to concretise educational policies against the backdrop of their *specific local demands*. Only concrete policy guidance that fits the local realities, allows more inclusive arrangements, as it is emphasised by Ainscow et al. (2000). The authors point out that the local education authorities must clarify for themselves and with regard to the local situation “what is meant by inclusive education and how policies might be introduced to encourage developments in that direction” (Ainscow et al., 2000, p. 224). This can produce a variety of meanings, because a locally relevant definition of inclusive education is laid down in policy documents, which are difficult to continuously revise. Hence, a particular understanding is perpetuated to lead the understanding of inclusive education in a particular way (which was relevant to a particular time at a particular place).

The complexities of the development of education-related policies and legislations are demonstrated by Nes and Strømstad (2006), who analysed the revision process of the basic education policies in Norway starting in 2001. The Norwegian government appointed a committee in 2001 to make suggestions for improving basic education in Norway. Amongst many other suggestions, it was recommended by this committee in 2003 to change the §5.1 of the Education Act, which stated originally that students who struggled with ordinary education had a right to special education, and replace it with a legal right to adapted education for all (Nes & Strømstad, 2006). A large variety of different parties were invited to comment on these

recommendations, and with regard to the §5.1 the responses were 50:50 pro or contra the deletion of the §5.1 respectively, as Nes and Strømstad (2006) reconstructed further. The Institute of Special Education in Oslo, then, opposed that further elaborations would be needed in this regard, and they convinced the government that they need to write these recommendations themselves. In their recommendations the Institute of Special Education pointed out the value of keeping §5.1, which resulted in actually keeping the §5.1. Nes and Strømstad (2006) concluded that this was most illustrative “how special education gets most of the attention even when the broader issue of adapted education for all is on the agenda” (Nes & Strømstad, 2006, p. 375). Notably, §5.1 does still apply in Norway in this form, stating that “pupils who either do not or are unable to benefit satisfactorily from ordinary teaching have the right to special education” (Norwegian Ministry of Education and Research, 2014, §5.1).

Besides these examples, another interesting case for illustrating issues that arise from policies on national and local levels is Australia. On the one hand, Australia has a relatively advanced discourse on diversity, yet, on the other hand, Australia has a federal constitution, which produced major differences in the quality of education-relevant policy texts. In Australia, a variety of conventions and declarations guide inclusive practices, and the general trend is apparent that understandings of inclusive education were shifting from emphasising particularly students with SEND to “educational opportunities for all students, framed in terms of diversity” (Carrington et al., 2012, p. 17). Important educational guiding texts (such as the Melbourne Declarations or the Australian Curriculum) emphasise that teachers are supposed to cater to the diversity of their students (Carrington et al., 2012). Yet, the educational policies as they pertain to inclusive education in Australia are different between the States. Hardy and Woodcock (2015) demonstrate in their policy analysis that policies in New South Wales (NSW), for example, have the tendency to encourage bringing students with SEND into the ‘mainstream’ schools (see also e.g. L. J. Graham & Sweller, 2011 for statistical evidence and problematising the specific focus in NSW on students with SEND). On the other hand, in Queensland, another state in Australia, the policies are challenging homogenising tendencies and emphasise students with diverse needs and abilities instead (Hardy & Woodcock, 2015).

As opposed to Australia’s tradition in discussing diversity-related issues, there are also countries that only recently started discourses on establishing inclusive education. An example is Germany, where basically the Convention on the Rights of Persons with Disabilities (CRPD; UN, 2006) brought the German term ‘Inklusion’ (Engl. Inclusion) up for discussion in the educational field, around its ratification in 2009. At that time, Germany had established a fully separated – and highly elaborated – special education system besides the regular school system. Like Australia, Germany is divided into Federal States, and these states are responsible for all

education-related issues and policies. In 2009, none of the school laws and policies of Germany's 16 Federal States met the requirements of the CRPD (Hinz, 2015). Accordingly, all Federal States had to change their education legislation and policy considerably (Hebborn, 2014). Special educational services in Germany depended on identified SEND (Banafsche, 2013; Henry-Huthmacher, 2015). Highly complex and partially contradictory laws governed the goods and services as they pertained to identified SEND (Banafsche, 2013 exemplifies these contradictions with regard to the goods and services according to the social security codes SGB VIII and SGB XII). While recently, most school-related laws have been changed more or less in accordance with the CRPD (Hebborn, 2014; Henry-Huthmacher, 2015), the general system that distributes support to students with identified SEND has hardly changed. Hence, in many cases, the ministries strive "exclusively for new structures of 'special educational support' in the regular school" (Hinz, 2015, p. 24). The school laws keep their opportunity to assign a child to a 'more appropriate form of school', as mentioned by Kruse and Dederling (2017) with regards to the law in Lower Saxony, a German Federal State. As Kruse and Dederling (2017) formulate, many of the Federal States reserve their right that "'un-includable' children and teenagers can theoretically be transferred to special schools against their will (or that of the parental authority)" (Kruse & Dederling, 2017, p. 6).

The Medical- and Social Model of Special Educational Needs and/or Disabilities

As discussed previously, to identify SEND still plays a major role in current inclusive education discourses in different contexts. Booth (1995) noted that a new kind of thinking is only possible, if we are "prepared to jettison the language that ties us to old habits of thought" (Booth, 1995, p. 97). 'Special educational needs' and 'disability' are such terms that suggest a clear cut between the normal and the non-normal or between the general and the special (Booth, 1995; L. J. Graham & Macartney, 2012; L. J. Graham & Slee, 2008). However, the use of labels, such as certain SEND recently increased in many contexts (L. J. Graham & Jahnukainen, 2011).

Two modes of thinking about SEND are generally distinguished; namely, a medical model and a social model. The idea that SEND are inherent of particular students' minds and bodies, which limits their abilities and capacities, is referred to as the deficit model or the *medical model* of inclusive education (Dudley-Marling & Burns, 2014). In a broader sense, this means that services need to be delivered that compensate the students' inherent defects so that such exceptional students can function like *normal* students in the classroom. More recently, the medical model was criticised by different perspectives that can be summarised as the social model or the social constructivist/constructionist model of inclusive education (Dudley-Marling & Burns, 2014). In this way of thinking, SEND are no 'true defects', but emerge out of certain

discourses (Slee, 2008; Tomlinson, 2012). This is supported by research that has, for example, demonstrated that the age of a child relative to its peers has a direct effect of being identified with Attention Deficit Hyperactivity Disorder (ADHD) (Evans, Morrill, & Parente, 2010). If ADHD would refer to any inherent and unalterable trait of the child, such an effect should not exist. In other words, to identify specific SEND depend on expectations as they pertain to what is considered being normal in a particular society and context.

The particular use of categories of SEND are dependent on more general societal values as they pertain to the educational system, as Arduin (2015) demonstrated in a comparative study of four different countries. The neo-liberal values in England and Ireland converge with the utilisation of the medical model, while Norway and Finland, which are generally subscribed to more social-democratic values, utilise more of a social-interactionist model (Arduin, 2015). Many of Slee's studies (Slee, 2001a, 2001b, 2011, 2013; Slee & Allan, 2001) point in a similar direction, too, that the medical model persists and as a result the battle for (special educational) resources demands special education to identify needs and according services. In addition, a similar interaction between the values and the discourse model was also described as the 'irresistible rise of the SEN industry' (Tomlinson, 2012, 2015) or the 'manufacture of inability' (Tomlinson, 2017), in order to point out that inclusive education is intertwined with particular interests of different stakeholders and monetary considerations.

One instrument of 'market efficiency' and 'competition' in education is the standardised (achievement) testing of students, which is one of the indicators for neo-liberal values underpinning the educational sector. Inclusive education and the 'standards agenda' are in conflict with each other, because both are making different suggestions with regard to desired school developments (Ainscow et al., 2006; Ainscow et al., 2000; Glazzard, 2013). When academic achievement is assessed in standardised examination procedures, the "most vulnerable learners will continue to be singled out for specialised attention" (Glazzard, 2013, p. 186). This keeps specific learners marginalised and creates difficulties with regard to their participation and achievement (Glazzard, 2013). Similar conflicting developments were described by Engsig and Johnstone (2015), who demonstrate that the emphasis to develop more equity through inclusive education is contradicted in Denmark by the implementation of "US-inspired accountability-driven" (Engsig & Johnstone, 2015, p. 475) standardised testing.

An example of increased market efficiency and competition in the educational sector is Australia. The Australian educational system is regarded to utilise neo-liberal strategies to foster quality through competition (see e.g. Hardy & Woodcock, 2015). One instrument that is commonly mentioned in this respect is the National Assessment Program – Literacy and Numeracy (NAPLAN; see e.g. Fachinetti, 2015; Johnston, 2017). NAPLAN was introduced in

2008, and it was supposed to replace the varieties of different standardised testing that were in place before, with a common instrument for assessment throughout the country (Fachinetti, 2015). NAPLAN examines the students' literacy and numeracy achievement in school years 3, 5, 7 and 9, and the results are published online at the school level (Johnston, 2017). Johnston (2017) notes that adjustments for students with disabilities are possible. A number of students with identified SEND do not have to participate in NAPLAN (Anderson & Boyle, 2015; Elliott, Davies, & Kettler, 2012). Hence, there is a certain interest at the school level for diagnosing low-achieving students, because, as discussed before, it brings additional monetary resources and services, but it also could improve the schools' NAPLAN scores, because these particular students might get around the standardised testing, which makes such a school more attractive for families with higher socio-economic status. As mentioned previously in the present study, although all of its advancements in Australian inclusive education, there are still tendencies in the Australian school system to place students identified with SEND in special schools or to 'confine' them in special classes or units in regular schools, as the monitoring committee of the United Nations denounced (UN, 2013). Although Australia has ratified the CRPD (UN, 2006), the United Nations clearly criticised in their report the 'substandard education' and the lower school completion rates of students with SEND (UN, 2013).

A country that challenges the argument of Arduin (2015) that social-democratic values of a society suggest that a social-interactionist stance towards SEND is more likely, is Germany. Germany and its educational system is generally seen as social-democratic (see e.g. Tomlinson, 2015). Except for the general discourse on educational standards (e.g. Rödler, 2012 discusses inclusive education and educational standards as being paradoxical, while e.g. Hinz, 2015 states that they are not necessarily contrary to each other), there is no nation-wide achievement testing that is being reported on the school-level. In accordance with the aforementioned analysis by Arduin (2015), one would have suggested that Germany's discourses on SEND utilises a social model. Germany has a long tradition in special education. The special education sector is very advanced and has a vast variety of special schools for all kinds of SEND. Many of these special schools note that SEND emerges due to barriers that particular students face, and the special schools tried to mitigate these barriers for these students. This tradition was and still is present in the German inclusive education discourse. Yet, more recently, it was argued that mitigating barriers for students with particular SEND should not only take place in the special school sector, but also in the regular school sector. In the German-wide recommendations of the ministers of education and cultural affairs (KMK, 2011), 'Inklusion' (Engl. inclusion) was introduced as a comprehensive concept of 'living together', 'education for all', and that all barriers that hinder inclusive education in the regular system need to be overcome. 'Inklusion'

is about all children and youths and their active participation in common life and learning (KMK, 2011). Yet, to accomplish this, the KMK (2011) recommended to identify the SEND, by using the established categories, so that the children receive appropriate provision in the regular system according to their identified precondition and potential. As it is common in Germany's federalism, the KMK (2011) recommends that all Federal States make their own laws what kinds of goods and services are provided for each of the SEND categories (KMK, 2011). Additional goods and services are only available, if a child was identified. There is an increase of children with identified SEND; especially special needs with regards to 'emotional/social development' have recently doubled (Henry-Huthmacher, 2015). In addition, it has been criticised that the recommendations by the KMK (2011) and similar recent developments in Germany "strive exclusively for new structures of 'special educational support' in the regular school" (Hinz, 2015, p. 24). Generally, it turned out to be difficult and expensive to change the established structures of special education and develop a unified educational system with schools for all (Hinz, 2015; Klemm, 2012). The United Nations' Committee that examined the progress of implementing the CRPD was very concerned that the German educational system was still in large parts fully segregated (UN, 2015).

Schools and Inclusive Education

Inclusive education *for all* requires the actual pedagogical settings to create communities that are welcoming for all and to combat discriminatory attitudes of all (based on UNESCO, 1994). Schools are suitable institutions for inclusive education to take place in this way; namely, schools involve students and teachers, but also parents and the wider community, and they are developing so that they are able to respond adequately to demands such as becoming welcoming for all and tackle discrimination. A school that provides inclusive education *for all* in this sense needs constantly be mindful of new exclusionary dynamics and it needs to find creative ways to foster provision and prospering for all. Such "an inclusive school is one that is on the move, rather than one that has reached a perfect state" (Ainscow & Miles, 2008, p. 20). In the perspective of Slee (2008, 2011), such a task is unlikely to be carried out by special schools nor by regular schools; instead it needs a new revised kind of schooling, which he calls *irregular schooling*.

Generally, students spend a considerable amount of time in schools. Most recent data presents evidence that "Students in OECD countries and economies receive an average of 7,538 hours of compulsory instruction during their primary and lower secondary education, ranging from 5,976 hours in Latvia to almost double that in Australia (11,000 hours) and Denmark (10,960 hours)" (OECD, 2017, p. 334). About a decade ago, the OECD average of students'

time spent in schools was 6,898 hours (OECD, 2007). Yet, not only the objective time that students spend in schools has increased over the recent years, but also the subjective significance of schools from the perspectives of the students (Fraij, Maschke, & Stecher, 2015). Hence, the time students spend in schools should at best be free from discrimination and exclusion.

It was argued that improving schools in this way means developing inclusive education (Ainscow et al., 2006). School developments in this direction of ‘schools for all’ are supported in many cases around the globe by the ‘Index for Inclusion’ (Booth & Ainscow, 2011; first edition published in 2000). Another school development tool, yet not as elaborated and established as the Index for Inclusion, is the ‘school-based inclusivity framework’ (Berlach & Chambers, 2011). All in all, the school seems to be a crucial institution for the implementation of inclusive education *for all*, because of its significance for the students, and because it provides the space where conceptual ideas of inclusive education *for all* need to be transposed into ‘real world’ teaching practices and social actions.

Pedagogical Approach to Inclusive Education

At the classroom level, inclusive practices are crucial for quality teaching for all. Even in more general recommendations to ‘what makes great pedagogy?’, a review by Husbands and Pearce (2012) clearly recommended that “effective pedagogies are inclusive and take the diverse needs of a range of learners, as well as matters of student equity, into account” (p. 11). There is a great deal of research available on pedagogical strategies to teach all learners, mostly referred to as the ‘inclusive pedagogy’ for all (Black-Hawkins & Florian, 2012; Florian & Black-Hawkins, 2011; Florian & Linklater, 2010; Pantić & Florian, 2015). One aspect of inclusive teaching is that teachers should not fall for labels that particular children might carry (Daniels, 2006). In order to tackle exclusion, labels might be needed at some point to uncover specific exclusory pressures (Ainscow & Miles, 2008), and it might generally be difficult to work and think without using any labels; yet, teachers need to be very aware of the ‘trap of stereotyping’ (L. J. Graham & Macartney, 2012).

These recommendations from previous research seem to be plausible, yet, research uncovered that these ideals seemed hardly implemented in ‘real-world’ contexts. In an empirical study with chief education officers, principals, and teachers in Sweden, Nilholm, Almqvist, Göransson, and Lindqvist (2013) demonstrated that particularly practitioners at the school level highlight the need for medical diagnosis and labelling of children. A similar result was reported in the aforementioned study by Nes and Strømstad (2006) in Norway, who reported that particularly politicians seemed to be in favour of education for all, while the

majority of school authorities and teachers were in favour of retaining special education (Nes & Strømstad, 2006). This shows that, although the concepts to develop inclusive schools (Berlach & Chambers, 2011; Booth & Ainscow, 2011) and inclusive practices (Black-Hawkins & Florian, 2012; Florian & Black-Hawkins, 2011; Florian & Linklater, 2010; Pantić & Florian, 2015) are available, practitioners might still stick to the established ways of thinking about SEND.

‘Inclusive Education for All’ as a Guiding Principle on all Levels

On the global level, the directions towards education for all and inclusive education, as they were set by the UNESCO, are relatively unambiguous. The emphasis given previously to inclusive education *for all* as a new kind of thinking is on the global policy level the consequent continuance of the current ongoing narrowing of the scope of education for all on the one side and inclusive education on the other. Yet, as the previous paragraphs showed, global policies need to be translated into national and local policies, and the clash of global policies and ideas with cultural and structural realities deflects the original intents and meanings into a diverse range of understandings and thinking about inclusive education. On the one hand, in order to be effective, policies need to consider the social realities of education and schools (Armstrong et al., 2011). On the other hand, only consistent and coherent guidelines allow practices to become more inclusive (Hardy & Woodcock, 2015; Hinz, 2015; Miles & Singal, 2010), which might mean that inconsistencies between different policies (on different levels) and also vagueness in policy texts might lead to confusion amongst the relevant stakeholders.

Taken together, inclusive education *for all* can be understood as a ‘north star’ (Hinz, 2015), which explicitly encourages active engagement and intensive contention with its (abstract) content. As the previous discussion of the different aspects demonstrated, this pertains not only to policy development, but also to continuously questioning the established understandings of SEND, to develop schools to be more inclusive for all and to maintain a reflective stance towards the pedagogical approach. In this way, the notion of Ainscow and Miles (2008) that inclusive education should be understood on the school level “as a never-ending process” (p. 20), might apply to all other levels, too. Accordingly, this genuine process character of inclusive education might be re-interpreted in the context of the present study with regard to all described national, local, school, and classroom practice-related levels as a never-ending and iterative process that is oriented towards a sufficient translation of the global notion of inclusive education *for all*.

2.3 Inclusive Education for All and Teachers' Attitudes

Inclusive education *for all* was introduced in the previous section as the new concept. The discussion focused to a large extent on global policies and issues as they pertained to the translation of global commitments to the national, local, school, and classroom practice levels. It was demonstrated that the global policy messages clash with the national and local realities and become deflected into a diverse range of understandings. The previous section ended with notions as they pertained to the pedagogical approach. Yet, the pedagogical approach to inclusive education *for all* needs to be implemented in 'real world' contexts. Hence, this section considers the teachers as one of the main stakeholders in education and as being crucial for implementing inclusive education *for all*. Particular emphasis is given to teachers' attitudes.

2.3.1 Teachers and their Attitudes as a Key to Inclusive Education for All

In the Incheon Declaration (UNESCO, 2015), emphasis is given to teachers and educators being necessary for 'quality education' and for 'improving learning outcomes'. Hence, the teachers' empowerment, training, and support needs to be ensured, as it was declared by the UNESCO (2015).

Importance of Teachers for Quality Education

The importance of teachers for all of their students was highlighted in the report 'teachers matter' published by the Education Committee of the Organisation for Economic Cooperation and Development (OECD, 2005). This report emphasised the vital importance of teacher quality and quality teaching for improving student learning (OECD, 2005). Yet, because the report had a different scope (namely, to foster teacher policy) the section within the report that presented empirical evidence in this direction was relatively small, compared to other sections.

Strong evidence that teachers generally matter is derived from studies that measure students' achievement. Empirical studies repeatedly revealed that the students' achievement scores had little between-school variance, but not negligible within-school variance (Coleman et al., 1966; Goldhaber, 2016; Hattie, 2009; Mansfield, 2015). In other words, there were no specific kinds of schools or districts that did better (or worse) than others. This lead most authors to the conclusion that the variance of students' achievement needs to be explained at the classroom level; hence, it depended on the teachers. Rivkin, Hanushek, and Kain (2005) for example analysed the reading and mathematics achievement of students and the characteristics of their teachers and class size in Texas, United States. In this study, a lower class size (a

common argument to improve student achievement) was associated with slightly higher achievement, and a higher estimated teacher quality was associated with even higher students' achievement. The authors concluded that the reduction of the class size is not as effective as improving the teacher quality to raise the students' achievement (Rivkin et al., 2005). Findings of the meta synthesis of Hattie (2003) confirmed that teacher quality had the greatest effects on the learners, because it explained about 30% of the variance of students' achievement.

Until today, it remains a debate if the teachers or the teaching has the most influence. While Hattie (2003, 2009) found that it is not the teacher, but the 'excellence in teaching' which influences achievement the most, an analysis of the TIMSS 2011 data by Blömeke, Olsen, and Suhl (2016) showed that instructional quality had no direct effect on students' achievement. Instead it was the felt preparedness, experience, degree and relevant major of the teachers that had an effect on students' achievement (Blömeke et al., 2016). In a longitudinal study in Queensland, Australia, the importance of teachers' experiences were also highlighted, because the literacy and numeracy of the students of more experienced teachers had higher test score gains (Leigh, 2010). In the same study, the importance of high quality teachers was even more apparent, because the students of the 'very best' teachers (90th percentile teacher) achieve in half a year, what students of the 'worst' teachers (25th percentile teacher) achieve in a full year (Leigh, 2010). In addition, even longer-term effects could be demonstrated in longitudinal studies. Chetty, Friedman, and Rockoff (2014) analysed school district and tax records from more than one million children, and found that children, who were assigned to high-quality teachers were more likely to attend a college, to earn higher salaries and they were less likely to be teenage-parents.

In his seminal study 'Visible Learning', Hattie (2009) generated a 'model of successful teaching and learning', which highlighted the power of the teachers and their teaching, their proficiencies with regards to decision making, their engagement in establishing caring relationships with and amongst students and their critical reflection of desirable teaching outcomes. Hattie (2009) criticised that there is a mismatch between the strong empirical evidence that the working conditions (such as resources, smaller class sizes etc.) are not effective in terms of students' achievement on the one side, and the great importance of working conditions as it is represented in most policies and ongoing discourses on the other side. Quite a similar situation is present in most of the present inclusive education discourses around the world. As it was demonstrated previously in the present study, there is a strong emphasis on working condition-related issues, such as class size, aides, counselling, etc., which the diagnosing of children is hoped to improve. Although Hattie (2009) had not specifically commented on inclusive education, one might learn from his study with regard to inclusive

education *for all* that, if the learning of the child is in the centre of interest, excellence in teaching should be the main goal to be achieved. This does not mean that resources are not necessary at all (such as inclusive design of the learning environment) – but it does mean that the bemoaning of the working conditions might deflect the teachers' efforts towards more effective teaching strategies for all, which holds the most powerful effects for all students. In this sense, inclusive education *for all* demands that teachers are reflecting constantly on their teaching and their beliefs and attitudes, and find new ways of teaching all students. In the words of Hattie (2009): “the beliefs and conceptions held by teachers need to be questioned – not because they are wrong (or right) but because the essence of good teaching is that teachers' expectations and conceptions must be subjected to debate, refutation, and investigation” (pp. 239-240).

Teachers as Key to the Success of Inclusive Education for All

The importance of teachers for quality education and the learning, achievement and success of students is mirrored in global policies as they pertain to inclusive education *for all*. In the World Declaration on Education for All (UNESCO, 1990), the teachers' role for providing quality education was explicitly highlighted. In its Framework for Action, it was noted:

“The preeminent role of teachers as well as of other educational personnel in providing quality basic education needs to be recognized and developed to optimize their contribution. This must entail measures to respect teachers' trade union rights and professional freedoms, and to improve their working conditions and status, notably in respect to their recruitment, initial and in-service training, remuneration and career development possibilities, as well as to allow teachers to fulfil their aspirations, social obligations, and ethical responsibilities” (UNESCO, 1990, p. 12).

Subsequent to the World Declaration on Education for All, an ‘education for all teacher-training package’ (UNESCO & UNDP, 1995) has been published. This document embraced a large variety of teacher-related topics as they pertain to implementing UNESCO's Education for All movement. The UNESCO and UNDP (1995) emphasised the important role of teachers in implementing Education for All. In this way, teachers were considered ‘key actors’ for implementing an education that is meant for all, as it is expressed by the UNESCO and UNDP (1995). In the Framework for Action of the Salamanca Statement (UNESCO, 1994) the teacher was described as a member of a team, who collaborates with other professionals and with parents in order to foster inclusive education. The UNESCO (1994) highlighted that teachers need to be trained in their pre-service phase the necessary skills to collaborate and to adapt the

curriculum and instruction to meet all students' needs, and in their in-service phase they need to receive appropriate further training at school level respectively. Throughout the Dakar Framework for Action (UNESCO, 2000), it is emphasised that basic education of quality for all needs well-trained teachers. The Dakar Framework included in its extended commentary even an own sub-section on enhancing the status, morale and professionalism of teachers, where the teachers are described as advocates for change and being catalysts of change. Furthermore, it is pointed out that teachers must be able to understand diversity in the students' learning and their development so that the teaching and learning environment can be adapted to the diverse needs (UNESCO, 2000). Recently, the Incheon declaration (UNESCO, 2015) reassured the commitments with regards to teachers, as they were made in the previous global policies.

Similar notions that the teachers are 'agents of change' towards inclusive education are represented in recent studies (such as Pantić & Florian, 2015) and research summaries (such as Ballard, 2012). Ballard (2012) highlighted the connection between inclusive education and democracy. As Ballard (2012) pointed out, both require the active participation of all individuals. In order to fully embrace inclusive education and democracy, teachers need to understand that exclusion and oppression exist, and how to tackle them (Ballard, 2012). Because teachers might unintentionally be part of a system that excludes some students, they need to be critical of their own position, as Ballard (2012) emphasised. Similarly, Pantić and Florian (2015) argued that teachers need to work towards social justice through not falling for exclusion of some individuals and making a serious attempt to gain positive outcomes for all. Yet, the authors claimed that it remains unclear for teachers how to carry out such a complex task. As a solution, Pantić and Florian (2015) suggest to bring together the inclusive pedagogy (Florian & Black-Hawkins, 2011; Florian & Linklater, 2010) and the teacher agency for social justice model (Pantić, 2015). The resulting heuristic guides the teachers' thinking along the four aspects of teacher agency: purpose, competence, autonomy, and reflexivity, and for each of these aspects more concrete issues of inclusive pedagogy are provided to guide the teachers' thinking and reflexion process (Pantić & Florian, 2015). All these encouragements and guidance to support teachers working towards democracy and social justice start from the premise that the teachers are the key to the success of inclusive education *for all*.

Teachers' Attitudes with regard to Inclusive Education for All

As demonstrated in the previous paragraphs, the global policies from the 1990s emphasised teachers as being of great importance. Yet, in these documents, teachers' competencies and their work conditions were put to the foreground, while their attitudes were not mentioned at all. Only later on, the focus on teachers was widened to embrace teachers' attitudes. The

Guidelines for Inclusion (UNESCO, 2005) explicitly stated that “teacher attitudes and tolerance are the vehicles for the construction of an inclusive and participatory society” (p. 17). The UNESCO (2005) acknowledged that research had indicated that negative attitudes were a major barrier to inclusion; hence, teachers’ attitudes needed to be reflected when trying to improve inclusive education.

Positive teacher attitudes can play an important role for inclusive practices and positive educational outcomes, as it is proposed in the Framework for Inclusive Education (Peters, 2004). Peters (2004) postulated a variety of *contextual factors* such as politics, standards etc. that influence the inputs and outcomes of an inclusive educational system. The *input* perspective includes school-, student- and family-/community-related characteristics, while the *processes* comprise the school climate and the teaching and learning. As *outcomes*, Peters (2004) proposed students’ achievement (e.g. literacy, numeracy, good citizenship etc.), students’ attainment (e.g. preparation for adult life, etc.) and standards (such as impact on family and community or on governmental policy etc.). In this framework, positive teacher attitudes are part of the school climate, which is interrelated with the teaching and learning practices (e.g. active teaching methods and also e.g. active student participation), and both – the climate and the teaching and learning – lead to the students’ outcomes. In other words, the model described a relationship of teachers’ attitudes towards inclusion and the inclusiveness of the teaching and learning, and it also pointed out that the positive teacher attitudes are only one school climate factor out of nine, according to Peters (2004).

In a variety of studies, it was highlighted that attitudes are crucial for the success of inclusive education, such as Mahat (2008) or Ainscow and Miles (2008). More recently, Opertti and Brady (2011) also gave emphasis to the attitudes, roles and competencies of teachers for implementing inclusive education successfully. On the contrary, negative attitudes can also be seen as a barrier to inclusive education, as Anderson and Boyle (2015) pointed out. Besides other barriers (such as confusions with regard to the definition of inclusive education, or lack of resources), negative teacher attitudes towards inclusive education are one important aspect that can hinder its successful implementation (Anderson & Boyle, 2015).

Taken together, teachers and their attitudes can be understood as a key to inclusive education *for all*. The presented evidence underlines the importance of teachers for students’ learning in general, and policies and research highlight teachers to be of particular importance to implement inclusive education *for all*. Through their excellence in teaching, teachers allow for the presence, participation and achievement of all students in their classrooms, and through a ‘reflective process’ (as it is emphasised by Carrington et al., 2012) effective teachers are able to uncover exclusion and marginalisation and to find strategies to tackle them at least at the

classroom level. If teachers hold positive attitudes towards inclusive education *for all*, they are likely to understand that inclusive education *for all* is a never-ending and iterative process; not only on the national, local policy level, and school level, but also for them as teachers who are engaged in ongoing cycles of reflection, critical thinking, and problem solving.

Attitudes: Concept and Definition

As previously discussed, the attitude a teacher holds towards inclusive education *for all*, makes it more or less likely that s/he will be concerned with continuously finding ways to carry out more inclusive practices or not. A vast number of theoretical reflections on the concept of attitudes are available (such as Eagly & Chaiken, 1993; Katz, 1960; Rosenberg & Hovland, 1960; Triandis, 1971). Most of them are rooted in social psychology, and only a few studies discuss attitudes as a concept from disciplines such as sociology or education (such as Bergman, 1998; P. S. Cohen, 1966; Hitlin & Pinkston, 2013; Voas, 2014). In the following review of the literature, some of the cornerstones of the attitude concept are highlighted, as they are relevant for the present investigation.

The point of departure is to understand attitudes in the most basic sense as a tendency of an individual to evaluate a particular symbol or object in a favourable or unfavourable way (Eagly & Chaiken, 1993; Katz, 1960). Many attitude theorists would subscribe to this minimal definition and many empirical researchers commonly operationalise the degree of favour or disfavour towards an attitude object. Three aspects of attitudes are described in more detail. First, that attitudes can be understood as having different components, second, that attitudes and behaviour are somehow related, and third, that attitudes are interrelated with the social environment of an individual.

In his pioneering study on attitudes, Thurstone (1928) noted that the evaluation of a specified topic, which he termed attitude, comprises a range of very different individual processes, such as “feelings, prejudice or bias, preconceived notions, ideas, fears, threats, and convictions” (Thurstone, 1928, p. 531). Over the years, this conception was developed further and certain agreement was established amongst social psychologists that attitudes should be understood as having three distinct dimensions or components: an affective, a behavioural and a cognitive dimension (Eagly & Chaiken, 1993; Rosenberg & Hovland, 1960; Triandis, 1971). Until today, this three-fold model of attitudes was popularised under the label ‘ABC model of attitudes’ (‘A’ for affective, ‘B’ for behavioural, and ‘C’ for cognitive). The idea of this model is that the attitude of an individual comprises, what somebody knows about the object in question, what s/he feels about it, and what kind of behavioural options s/he associates with the attitude object as an appropriate response to it. This three-fold understanding of attitudes is the

foundation of the definition given by Triandis (1971), which comprises each of the three components explicitly: “an attitude is an idea charged with emotion which predisposes a class of actions to a particular class of social situations” (p. 2). According to this, researchers (such as Rosenberg & Hovland, 1960; and, Triandis, 1971) propose that, if an attitude is activated, the individual should demonstrate three distinguishable responses that should be more or less empirically measurable: an affective, a behavioural and a cognitive, respectively. Researchers in the field of inclusive education recently emphasised that attitude scales should operationalise these three components (De Boer, Pijl, & Minnaert, 2011; Vignes, Coley, Grandjean, Godeau, & Arnaud, 2008) and some attitude towards inclusive education scales actually tried to implement this notion (De Boer, Timmerman, Pijl, & Minnaert, 2012; Mahat, 2008).

The conception of attitudes as having these exact three dimensions is not unambiguous. For the cognitive component it seems plausible to assume that it can be explicated by the respondents in an attitude questionnaire. To complete a questionnaire demands cognitive activity (Tourangeau, Rips, & Rasinski, 2000), hence, the cognitive component of attitudes (if explicit) seems likely to be measurable by using survey methodology. Yet, for the affective and the behavioural component, a survey researcher needs to assume that affect and behavioural predispositions can actually be studied in this way. Amongst empirical studies, there is disagreement concerning this. For example Mahat (2008) used questionnaire items that began with “I believe...” to operationalise the cognitive component, “I get frustrated...” or “I get upset...” etc. to measure the affective component, and “I am willing...” for items that are supposed to measure the behavioural component. On the one hand, this seems to be generally in line with social psychology; as previously stated, social psychologists are in agreement concerning the three dimensional structure of attitudes. Yet, from a pragmatic point of view, it seems not justified to narrow down the attitude concept to these three components of attitudes. Many of the rich and insightful dimensions of established instruments that illuminated attitudes with regards to physical, academic, behavioural, and social SEND (Wilczenski, 1992), attitudes concerning core perspectives, expected outcomes, and classroom practices (Stoiber, Gettinger, & Goetz, 1998), or general philosophy, classroom behaviour of SEND children, perceived ability to teach SEND children, classroom management with SEND children, and academic and social growth of SEND children (Larrivee, 1982; Larrivee & Cook, 1979); just to name a few. It seems to make a difference, if the research purpose is to study (the psychology of) attitudes, or if the research purpose is to study a persons’ attitude towards a certain attitude object. In the latter case, it seems more promising to conceptualise attitude dimensions with respect to the attitudes’ content and not with respect to the three presumed psychological processes underlying the attitudes.

As previously discussed, attitudes can be understood as having different components. In addition, it can be assumed that attitudes and behaviour are somehow related. In his seminal summary of early research on attitudes, Allport (1935) pointed out that attitudes play a role for a particular response to objects to which the individual holds a particular attitude. In his words, “an attitude is a mental and neural state of readiness, organised through experience, exerting a directive or dynamic influence upon the individual’s response to all objects and situations with which it is related” (Allport, 1935, p. 810). Fishbein and Ajzen (1975) noted that it would have been of great practical significance to understand the relation of attitudes and the actual behaviour of an individual. Yet, most attitude researchers, such as Fishbein and Ajzen (1975) or Eagly and Chaiken (1993) are very aware of the fact that there is strong evidence that the attitude-behaviour relationship is considerably weak or not existent. A recent review (Chaiklin, 2011) and meta-analysis (Glasman & Albarracín, 2006) underlined the relative absence of a strong attitude-behaviour relation. Two basic issues concerning the attitudes as predictors of (subsequent) behaviour has been discussed. On the one hand, the causality of attitudes and behaviour is unclear. According to Triandis (1971), the relation ‘(stimulus-)attitude-response’ would be plausible in the reverse order of causality, too. Indeed, the comprehensive Handbook of Attitudes (Albarracín, Johnson, & Zanna, 2005) comprises chapters on both, the influence of attitudes on behaviour and the influence of behaviour on attitudes. On the other hand, the attitude-behaviour relation seems to be influenced by a range of further variables (see Eagly & Chaiken, 1993 for a comprehensive overview). One of the most respected theories on the attitude-behaviour relation is the Theory of Planned Behaviour (Ajzen, 1988, 1991), which includes the original attitude-behaviour relation, yet, it postulates other key variables. According to this theory, the behaviour is guided by behavioural intents which are themselves influenced by the attitude toward the behaviour, the subjective norm as it is found to be relevant for the behaviour, and the perceived behavioural control (Ajzen, 1988, 1991). Only the perceived behavioural control influences besides the behavioural intent also directly the behaviour, while the others solely influence the intention (Ajzen, 1988, 1991). Many empirical researchers in the field of attitudes towards inclusive education build on the Theory of Planned Behaviour, such as Batsiou, Bebetos, Panteli, and Antoniou (2008), Kuyini and Desai (2007), MacFarlane and Woolfson (2013), and Mahat (2008), just to name a few.

Besides these first two aspects, that a persons’ attitudes have certain components, and that attitudes and behaviour are somehow related, the third aspect is that attitudes are intertwined with the social environment of an individual. As described, the Theory of Planned Behaviour postulates the perceived social pressures have an effect on the attitude-behaviour relation. Hence, in this model, the perception of the social environment plays a role in shaping the

behavioural intents; in other words, it is not directly related to the attitudes. It is common in most attitude conceptualisations to see attitudes as something residing permanently within the individual, or as something emerging spontaneously within the individual (see e.g. the continuum between ‘stored in memory’ vs. ‘constructed on the spot’ to systematise attitude definitions, suggested by Bohner & Dickel, 2011). Yet, it is not so common to see attitudes intertwined with the social environment of an individual. Besides early attempts to illuminate attitudes through disciplines beyond social psychology (such as Bergman, 1998; P. S. Cohen, 1966), recently, Hitlin and Pinkston (2013) compared the terms values, attitudes, and ideologies. They concluded that values and ideologies are idealised versions of the world, while attitudes are closely related to the actual action of an individual. The authors state that sociology plays a crucial role in understanding attitudes, and that values, attitudes and ideologies are “fundamentally social in development, enactment, and consequences” (Hitlin & Pinkston, 2013, p. 332). In another recent study, Voas (2014) remarks that although sociology used attitudes as a term since long ago, it is basically defined by psychology as an internal mental state. From a sociological point of view, attitudes should not be seen as subjective expressions of favour or disfavour, but more as normative statements about social order (Voas, 2014). The interrelatedness of the individual and the society is strongly emphasised by Voas (2014), and he points out that attitudes always refer to both, the personal view and the social world that surrounds the person.

A comprehensive discussion of the concept of attitudes is beyond the scope of the present study. Yet, the three aspects discussed previously lead to the understanding of attitudes as it is relevant for the present study. The attitude of an individual towards an attitude object is a tendency of the individual to evaluate each of the attitude objects’ facets in terms of favourableness; which is both, a mirror of previous social experiences in particular social environments, and a precursor of subsequent social action. In other words, the relatively stable tendency of a teacher to think about inclusive education *for all* in a more (un)favourable way, is related in a way to certain abstract or concrete experiences and to a certain set of (more inclusive or more exclusive) practices that the teacher would find most appropriate in certain classroom situations.

2.3.2 Empirical Studies on Teachers’ Attitudes towards Inclusive Education

As discussed previously, teachers are considered one of the key factors to the success of inclusive education *for all*, and one important prerequisite of implementing more inclusive teaching practices is the teachers’ attitude towards inclusive education *for all*. As inclusive education has led to worldwide transitions of educational systems and practices in the last few

decades, a vast amount of empirical studies – especially with a focus on teachers and their attitudes – is available. Many of these studies focus on specific teachers' characteristics that are interrelated with or predictive for the teachers' attitudes. Avramidis and Norwich (2002) conducted an extensive review of the literature published between 1984 to 2000 on teachers' attitudes towards integration and teachers' attitudes towards inclusion. They found that many studies on integration and later on inclusion attempted to match a variety of student-related (e.g. the kind of disability), teacher-related (e.g. training) and environment-related (e.g. support) variables to the teachers' attitudes.

Recent empirical studies that explicitly *attempted* to measure the teachers' attitudes towards a contemporary understanding of inclusive education were selected for the following review. Hence, the following literature review considers the most recent findings in terms of teachers' attitudes and takes notice of the different views on inclusive education, as they were represented in each of the different studies. In other words, not only each study's findings will be reported, but also the understanding of inclusive education, as it is represented in each of the studies.

Teacher Background

The personal background of teachers, as it is present in inclusive education-related attitude research, comprises gender and age as two relevant aspects. Concerning the *gender* of the teachers, the research evidence is not consistent. Some studies operationalised the teachers' attitudes with regards to the placement of particular students in the regular classroom. One of these studies found for in-service teachers in Greece that female teachers had more positive attitudes (Tsakiridou & Polyzopoulou, 2014). A similar approach was used by Sharma, Shaukat, and Furlonger (2015) and Bhatnagar and Das (2014), who found for pre-service teachers in Pakistan, and for in-service teachers in India, respectively, that male teachers were more positive. Other studies used an empirical approach that focussed on the teachers' attitudes with regards to SEND-related aspects. Such studies in Pakistan (Ahmad, 2012) and in Bangladesh (Ahmmed, Sharma, & Deppeler, 2012) found that male in-service teachers tend to have more positive attitudes. Alghazo and Gaad (2004) reported male in-service teachers in the United Arab Emirates having significantly less positive attitudes towards inclusive education compared to female teachers. Avramidis, Bayliss, and Burden (2000a) found that English female pre-service teachers had more positive attitudes towards inclusive education. This was also reported by, Boyle, Topping, and Jindal-Snape (2013) with regard to Scottish in-service teachers. In their study on in-service teachers' attitudes towards inclusive education in Germany and Finland, Saloviita and Schaffus (2016) found a similar difference between males and

females; yet, the difference was only significant for the Finnish sample. With regards to the teachers' attitudes towards inclusive education *for all*, there does not seem to be any research evidence available within the searched literature. Taken together, studies do not seem to be consistent with regards to female or male teachers having more or less positive attitudes towards inclusive education. There seemed to be a tendency with regard to the region; namely, it was generally found that male teachers were more positive in Pakistan, India and Bangladesh in the southeast of Asia, while female teachers were more positive than males in some European countries and in the United Arab Emirates.

Another aspect of the teachers' background is their *age (in years)*, and the research evidence suggests that the younger teachers tend to have more positive attitudes. When in-service teachers in India were asked about their attitudes towards the placement of particular students into the regular school, the '40 years or younger' group answered in more positive way than the older teachers (Bhatnagar & Das, 2014). Studies with a more specific focus solely on aspects of students with special educational needs, found similar results. One of these studies found for in-service teachers in England that teachers with the most positive attitudes are on average ten years younger than those holding more negative attitudes (Monsen, Ewing, & Kwoka, 2014). Similarly, Saloviita and Schaffus (2016) report a negative correlation between age and attitudes, with regards to German in-service teachers, which means that the younger teachers tend to hold more positive attitudes than their older counterparts. No study was found that related the teachers' age to their attitudes towards inclusive education *for all*. Taken together, the evidence suggests that younger teachers tend to have generally more positive attitudes towards the placement of exceptional students in regular schools and towards students with SEND. Yet, age and attitudes is a difficult relation, which needs to be handled and interpreted with certain care. None of the cited studies analysed longitudinal data that followed individual teachers over a period of time. Hence, the association of age and attitudes might have different reasons or causes. In general, the majority of teachers around the world have their initial teacher training in their early twenties. Hence, teachers are confronted early in their careers with the current thinking as it pertains to educational concepts. If some decades later educational concepts might have changed, teachers with a certain distance to their initial training (and with a lack of appropriate further training) might not adopt the new educational concepts, and might, then, appear to have negative attitudes towards the new concepts. For example, the teachers in the aforementioned study by Bhatnagar and Das (2014) tended to be more positive, if they were younger than 40 years old.

An interpretation of these results could be that the 40 and above year old teachers in the study of Bhatnagar and Das (2014) might have had their initial training 20 years prior to the

study, which would have meant that they were trained before policies in India were aligned with the Sarva Siksha Abhiyan (Education for All Movement) in 2001 and the Action Plan for Inclusive Education of Children and Youth with Disabilities in 2005 (Bhatnagar & Das, 2014). Hence, that the younger generation of teachers might have already trained under these new policies, which might explain in part why they were more positive towards the placement of particular students in the regular schools. In this light, an alternative explanation of the age-attitude relation might be that it can be understood as an indicator of issues of the teachers to keep pace with such radical developments in educational thinking.

Another interpretation of the association between attitudes and the age of the teachers is that age correlates with experiences as a teacher, which affects the attitudes. Indeed, the teachers' *professional teaching experience (in years)* was found to be associated to their attitudes towards inclusive education in the way that less experienced teachers tend to have more positive attitudes. Findings like this were found for the teachers' attitudes towards the placement of particular students into regular schools, such as the study of Bhatnagar and Das (2014), who report that Indian in-service teachers with less than 10 years of experience were more positive with regards to their attitudes compared to teachers with more years of experience. Bhatnagar and Das (2014) found significant results for the attitude-age and for the attitude-experiences relation; yet, the authors did not report the relation of attitudes, age and experience statistically in more depth. For in-service teachers in South Africa and Finland, Savolainen, Engelbrecht, Nel, and Malinen (2012) found in a regression analysis that years of teaching experience is a negative predictor for teachers' attitudes towards inclusive education. In other words, teachers with less years of experience tend to have more favourable attitudes and vice versa. In this study attitudes were understood with regards to placement issues. A similar result was found for in-service teachers in Japan by Yada and Savolainen (2017), who found teaching experiences to be a negative predictor of the attitudes. Other studies operationalised the teachers' attitudes with regard to SEND, such as Boyle et al. (2013) who differentiated in their study in Scotland newly qualified teachers (less than one year of experience; e.g. probationary year) from those with 1-5 years, 6-10 years and so forth. They reported that especially the newly qualified teachers held more positive attitudes than the teachers with experiences in teaching practices. Alghazo and Gaad (2004) found for in-service teachers in the United Arab Emirates that teachers with more than twelve years teaching experience tended to have less positive attitudes than the less experienced. However, the one to five years teaching experience group had similar unfavourable attitudes, and only the teachers with six to eleven years of experience held attitudes that were more positive (Alghazo & Gaad, 2004). Besides these placement-related and SEND-related attitude studies, no study was found

that presented results concerning the duration of teaching experience and the teachers' attitudes towards inclusive education *for all*. Overall, the results mirror those already found with regard to the age of the teachers that younger teachers tend to have more positive attitudes. Yet, the results of Alghazo and Gaad (2004) seemed to suggest that teaching experiences and attitudes might not be a linear relationship. To consider this would be of particular relevance, especially when conducting a *linear* regression analysis, as it is carried out by a range of studies (such as Savolainen et al., 2012; Yada & Savolainen, 2017).

Not many studies are available that compare attitudes towards inclusive education between *primary versus secondary school teachers*. Yet, McGhie-Richmond, Irvine, Loreman, Cizman, and Lupart (2013) found for Canadian primary in-service teachers hold more positive attitudes towards inclusive education, compared to teachers in secondary schools. This investigation used an attitude towards inclusive education measure that was concentrated on notions of special educational needs. Studies on differences of primary and secondary school teachers concerning their attitudes towards inclusive education *for all* was not found in the literature. Primary schools are commonly ascribed as being more of a small and caring environment with generalist teachers, while secondary schools were larger, academically driven with subject teachers; in this way, Coldron, Crawford, Jones, and Simkins (2015) refer to primary and secondary schools in England as often being described as 'different worlds'. From such a systems' point of view, primary schools tend to be more inclusive per se compared to the secondary school sector; which might be reflected in the aforementioned finding of McGhie-Richmond et al. (2013).

At least some evidence suggests that the teachers' attitudes is related to the teachers holding a *higher degree qualification* or not. In a study that used a SEND focus with regards to measuring teachers' attitudes, Ahmmed et al. (2012) reported, in-service teachers in Bangladesh with lower qualification tended to hold more positive attitudes towards inclusion; the higher the degree, the more negative was the attitude in this study. Concerning the teachers' attitudes towards inclusive education *for all*, no research evidence seem to be available, so far. This evidence-base is not very solid. One further consideration in this respect might be that the kind of training affects the attitudes of teachers, as Woodcock and Hardy (2017) demonstrated for special education professional development in schools. In their study, Woodcock and Hardy (2017) reported detrimental effects of special education-related professional development on teachers' thinking and understandings of inclusive education. Other results were presented in the study of Van Reusen, Shoho, and Barker (2001), who found that the subjective 'expertise in special education' played a significant positive role for how positive the teachers perceived inclusion. In other words, if the teachers were trained (e.g. professional development or felt

expertise) in special education, they were more positive towards inclusive education *for some*, and more negative towards inclusive education *for all*.

Inclusive Education-Related Professional Background

Besides the general professional background, the teachers' inclusive education-related background comprises their knowledge, specific training and experiences as they pertain to inclusive education. Concerning the *knowledge with regard to inclusive education*, the literature suggests that more knowledge is associated with more positive attitudes. From research that focussed the placement of particular students in regular schools it is known that pre-service teachers' self-perceived knowledge about inclusive education is positively related to their attitudes (Loreman, Forlin, & Sharma, 2007). This result, which was drawn from samples of pre-service teachers in Australia, Canada, Hong Kong, and Singapore, was also found in a similar study in Mexico (Forlin, García Cedillo, Romero-Contreras, Fletcher, & Rodríguez Hernández, 2010). The evidence from these two studies seems to be consistent with regard to the positive knowledge-attitude relation. Notably, in these studies the knowledge is self-reported. Hence, it is a broad and relatively unspecific measure. There are attempts to create a knowledge test for inclusive education using vignettes (Sucuoğlu, Bakkaloğlu, İçsen Karasu, Demir, & Akalin, 2013, 2014); yet, this research did not find a significant relationship between knowledge and attitudes and relied on a disabilities-oriented understanding of inclusive education.

Evidence was found in the literature that more *training in inclusive education* is associated with more positive attitudes. More positive attitudes towards the placement of particular students in the regular school were found to be related in this way in a sample of Greek in-service teachers (Tsakiridou & Polyzopoulou, 2014), and similar findings were reported for in-service teachers from India (Bhatnagar & Das, 2014). Besides these results for in-service teachers, Forlin et al. (2010) reported a similar result for pre-service teachers in a study in Mexico; previous training in inclusion was associated with positive views on inclusion. All these studies point in a similar direction, yet, there is one study that challenges this evidence with regard to a placement-oriented understanding of inclusive education. For Pakistani pre-service teachers, Sharma et al. (2015) reported that those pre-service teachers with 'nil' inclusion related training reported the most positive attitudes. Maybe, it could be important to note that the pre-service teachers were asked to indicate their 'level of training in special education' in this study. The detrimental effect of some of the special education-related professional development on teachers' attitudes towards inclusive education (Woodcock & Hardy, 2017), which was previously mentioned, could be an explanation here, too. If a high

‘level of training in special education’ might mean that the pre-service teachers become convinced of special affordances for special children, and if they are then asked within the attitude measure to indicate, if these special children should attend regular classes, then the pre-service teachers might express an unfavourable attitude towards such a placement. From research with a stronger focus on special educational needs and/or disabilities, the positive relationship between the in-service teachers’ amount of training in inclusive education and attitudes was found in a variety of studies in different countries. Examples include Canada (Sokal & Sharma, 2013), England (Avramidis, Bayliss, & Burden, 2000b), Greece (Avramidis & Kalyva, 2007), and Scotland (Boyle et al., 2013). Although there is a strong evidence-base that training in inclusive education and attitudes with regard to the placement of particular students and with regard to SEND are related to each other, from a perspective of inclusive education *for all*, such evidence is lacking. Yet, generally, this training-attitude association seems to be stable across different countries.

The teachers’ *experiences with inclusive educational settings* was found in the literature to be associated with more positive attitudes towards inclusive education. Several studies that focussed on attitudes concerning the placement of particular students in regular classrooms found this relationship. In this way, a study on Chinese in-service teachers found positive attitudes to be positively related to experiences on teaching students with disabilities (Malinen, Savolainen, & Xu, 2012). A similar finding was revealed from a study in Mexico on pre-service teachers, who were at the end of their studies and about to start teaching (Forlin et al., 2010). Teachers’ contact with a person with disability is associated with more positive attitudes, which is known from the Indian context (Bhatnagar & Das, 2014). More positive attitudes for those teachers who have taught students with special educational needs were reported for Greek in-service teachers (Tsakiridou & Polyzopoulou, 2014). These findings are all relatively similar, yet, like in the previous paragraph, the study by Sharma et al. (2015) contradicts these findings. Those pre-service teachers from Pakistan who had no level of experience in teaching students with a disability reported more positive attitudes (Sharma et al., 2015). From studies that used an attitude measure, which were more focussed on SEND, a positive relation between inclusive education-related experiences and attitudes was found. Having experiences on teaching students with disabilities was positively related to in-service teachers’ attitudes in Germany (Hellmich & Görel, 2014). In a study conducted by Ahmmed et al. (2012) on in-service teachers in Bangladesh, the contact with a student with a disability and the success in teaching a student with a disability were examined separately. Both were positively related to the teachers’ attitudes towards inclusion (Ahmmed et al., 2012). For teachers from England, Avramidis et al. (2000b) reported more positive attitudes for those who have ‘active experience of inclusion’,

and for Greece more positive attitudes were reported for those teachers who were ‘working in schools with integration units’ (Avramidis & Kalyva, 2007). Attitudes that are more positive were also reported for Greek and Cypriot teachers who have taught students with SEND (Batsiou et al., 2008). With regards to the teachers’ attitudes, understood as the attitudes towards inclusive education *for all*, there is a lack of evidence. Comparable to the findings reported previously concerning the inclusive education-specific training, there is strong evidence that those teachers with inclusive education-related experiences tend to have more positive attitudes towards inclusion.

The Effects of Teachers’ Attitudes on Teaching Practices

Studies in the field of teachers’ attitudes towards inclusive education start on the plausible presupposition that new policy, as they pertain for example to inclusive education, needs to be implemented in real-world practices; hence, the practitioners are critical in bringing these new policies into action. Furthermore, it seems convincing that a teacher who holds certain unfavourable attitudes towards a particular policy might obstruct the realisation of it. It follows that positive attitudes towards inclusive education are crucial for inclusive teaching practices to take place. Some scholars articulated this attitude-practice relation as the ‘basic assumption’ (Avramidis & Norwich, 2002, p. 130), or the ‘common belief’ (Sharma & Sokal, 2016, p. 21) of attitude researchers in the area of the inclusive education. Yet, some empirical studies are available that demonstrate the positive effects of favourable attitudes towards inclusive education with regards to inclusive teaching strategies, the creation of a positive classroom learning environment and student outcomes.

Empirical evidence suggests that those teachers who hold more positive attitudes towards inclusive education utilise more inclusive teaching strategies. In a study conducted by Kuyini and Desai (2007) in Ghana, in-service teachers and principals were asked about their attitudes towards inclusive education concerning the placement of particular students, and the teachers were observed in three teaching sessions, using the Effective Teaching Practices Checklist (ETPC; Kuyini & Desai, 2007). Using correlations and a regression model, Kuyini and Desai (2007) demonstrated that positive inclusive attitudes of the teachers and effective inclusive teaching practices were associated with each other. In a study that used an attitude measure that was focused more on general issues as they pertain to students with SEND, Sharma and Sokal (2016) asked in-service teachers from a former study in Canada (Sokal & Sharma, 2013), if they would allow to be observed in their classroom. Observations were carried out using the Inclusive Practices Classroom Observation Scale (IPCOS; Sharma & Sokal, 2016). A small number of teachers consented and the results show that positive attitudes and inclusive

classroom practices are interrelated (Sharma & Sokal, 2016). Results in this direction that focussed the teachers' attitudes towards inclusive education *for all* does not seem to be available. As suggested by the aforementioned conceptual Framework for Inclusive Education (Peters, 2004), a variety of variables can have an influence on the practices, such as the teachers' knowledge of inclusive education or the school norm. Yet, the evidence suggests that the teachers' attitude towards inclusive education plays – at least in part – a role for explaining the teachers' inclusive classroom practices.

Furthermore, research evidence suggests that the classroom learning environment, as perceived by the students, actually benefits if the teacher has a more favourable attitude towards inclusion. Studies were carried out with respect to a SEND-related understanding of attitudes; yet, no study seems to be available with an inclusive education *for all* perspective. In a study that was conducted in New Zealand, Monsen and Frederickson (2004) asked a sample of in-service teachers to indicate their attitudes towards inclusive education. In addition, all of the teachers' students were asked to answer the My Class Inventory (MCI; Fraser, Anderson, & Walberg, 1982), which was originally developed in order to assess the classroom learning environment with regards to satisfaction, cohesiveness, friction, competitiveness, and difficulty from the perspective of the students. Monsen and Frederickson (2004) compared the students' ratings of the teachers with positive, middling, and negative attitudes and found that the classrooms of those teachers, who saw inclusion positively, were perceived by the students as more satisfactory. In addition, although not statistically significant, Monsen and Frederickson (2004) reported a tendency of a relationship between more positive attitudes and less friction in the classrooms. More recently, a similar study was conducted in England with a greater sample of in-service teachers and all of their students (Monsen et al., 2014). Similar research procedures were used and Monsen et al. (2014) found that the classrooms of teachers with more positive attitudes towards inclusive education were perceived by the students more positively with regard to satisfaction (higher score), cohesiveness (higher score), friction (lower score), competitiveness (lower score), and difficulty (lower score) (Monsen et al., 2014). In a further study, Monsen, Ewing, and Boyle (2015) elaborated more on this data from England. The authors were able to demonstrate that the satisfaction and the cohesiveness of the classroom environments as perceived by the students largely depended on just one rather than on all four dimensions of the teacher attitudes towards inclusive education measure. The quality of the learning environment depends on how teachers scored on the dimension 'social benefits for all of inclusion of SEN pupils in mainstream classes' (5 items such as 'The inclusion of SEN students can be beneficial for non-SEN students'; Monsen et al., 2015). Taken together, positive

attitudes seem to be interrelated not only with the practices, as reported in the previous paragraph, but also with how the students perceive the classroom.

Research evidence suggests an interrelation of the teachers' attitude towards inclusive education and inclusion-related student outcomes. Although there is a lack of research on attitudes towards inclusive education *for all*, a study was found in this direction that focused more on attitudes towards the placement of particular students in regular schools. Surveying a large number primary school students with special educational needs and their teachers in the Netherlands, Van der Veen, Smeets, and Derriks (2010) aimed at analysing the likelihood of these children being referred to special education and their literacy and numeracy attainment levels. Concerning the role of the teachers' attitudes for the students' outcomes, Van der Veen et al. (2010) reported that students with SEND are less likely to be referred to a special school if their teacher holds a positive attitude towards inclusive education. With regard to the students' literacy, Van der Veen et al. (2010) found no influences of the teachers' attitudes. Yet, the authors report that the more positive the teachers' attitude was, the more progress was noticeable in the numeracy attainment levels of the students from the 6th to the 8th school year (Van der Veen et al., 2010). There are certain limitations of such large-scale studies (see e.g. Van der Veen et al., 2010, pp. 38-41 for a critical discussion), yet, the evidence at least suggests that there is a covariation of teachers' attitudes and the outcomes of these teachers' students.

2.4 Specification of the Research Problem

As previously discussed, the global commitment to all of the students' thriving within the education system needs to be translated to the national and local contexts, manifested and continuously revised in relevant policies, and be adopted and implemented by schools and teachers. The crucial importance of teachers for all of their students to prosper was highlighted. For the same reason, the significance of the teachers' attitudes towards inclusive education *for all* became apparent in some of the global policies and in conceptual research. The literature review presented previously confirms that the vast array of relevant attitude research from 2000 to 2017 tends to be focussed on the placement of particular students in regular classes and on students with SEND, as opposed to include education *for all*.

In their systematic literature review, Nilholm and Göransson (2017) recently provided evidence that conceptual position papers were generally arguing towards a more contemporary understanding of inclusive education, while empirical research papers tended to use inclusive education in an outdated way; partly even synonymous with mainstreaming and integration.

This conceptual divide (Nilholm & Göransson, 2017) appeared clearly in the literature review of the present study, too. This is an important issue, because if inclusive education concentrates on the placement of particular students, the development of the pedagogical quality is neglected as an important task for inclusive education, as was recently argued by Haug (2017).

The way teachers' attitudes were measured within the reported empirical studies is elaborated further in the next paragraphs, which will subsequently lead to the research problem of the present study and the research question.

2.4.1 Instruments to Measure Teachers' Attitudes towards Inclusive Education

A closer examination of the empirical studies presented previously revealed that these studies used different instruments to measure the teachers' attitudes towards inclusive education. Two perspectives of the measurement instruments can be distinguished; one focused the placement of particular students and the other focused students with SEND. Both perspectives reiterate a clear cut between regular and special education and between regular and special students.

An understanding of teachers' attitudes towards inclusive education that appeared amongst a variety of empirical studies was the *placement of particular students in the regular classroom*. Several studies in this respect (Bhatnagar & Das, 2014; Kuyini & Desai, 2007; Loreman, Forlin, et al., 2007; Sharma et al., 2015; Tsakiridou & Polyzopoulou, 2014) used the Attitudes Toward Inclusive Education Scale (ATIES), which was developed by Wilczenski (1992, 1995). This scale attempts to measure, whether the respondents think that the integration of children with physical, academic, behavioural, and social difficulties into the regular classroom is feasible. Notably, the ATIES tried to avoid the use of special educational needs and/or disability-related rhetoric. Items are worded like "students who cannot move without the help from others should be in regular classes" ('physical' dimension; Wilczenski, 1992, p. 311) or "students whose academic achievement is 1 year below the other students in the grade should be in regular classes" ('academic' dimension; Wilczenski, 1992, p. 311). The respondents are asked to rate their dis/agreement to these statements. Although the wording was developed in a neutral way, the fundamental idea behind the scale was to measure the feasibility of placement in particular cases into the mainstream, as it was discussed in the 1980s and the early 1990s, when Wilczenski (1992, 1995) developed the ATIES. The idea that there could be a special placement option for particular children, contradicts contemporary thinking about inclusive education and is in no way in accordance with recent thinking towards inclusive education *for all*.

A number of empirical studies (Forlin et al., 2010; Hecht, Niedermair, & Feyerer, 2016; Malinen et al., 2012; Montgomery & Mirenda, 2014; Savolainen et al., 2012; Yada & Savolainen, 2017) used the so-called Sentiments, Attitudes and Concerns about Inclusive

Education (SACIE) scale, developed by Loreman, Earle, Sharma, and Forlin (2007; SACIE) and revised by Forlin, Earle, Loreman, and Sharma (2011; SACIE-R). The SACIE(-R) comprises three factors that measure the sentiments, the attitudes and the concerns, respectively. Although the SACIE and its revised version were developed when inclusive education was already discussed in more contemporary terms, the attitude dimension is in fact a selection of five items from the ATIES scale, such as “students who have difficulty expressing their thoughts verbally should be in regular classes” (‘attitude’ dimension; Forlin et al., 2011, p. 58). Although being an elaborated instrument in times when mainstreaming and integration were on the agenda, from a current perspective on inclusive education *for all*, it might be worth considering jettisoning attitude instruments with the sheer focus on the students’ placement.

An in-depth examination of the studies, which utilised the ATIES or the SACIE, demonstrated that most of them were focused on the teachers’ attitudes towards inclusive education *for some*, with a focus on students with disabilities. The operationalisation and the particular understanding of inclusive education seemed to converge in a way. This comes not as a surprise, because the ATIES is constructed along certain labels of disabilities (such as physical, academic, behavioural, and social). As mentioned before, the ATIES tried to avoid using explicit labels; yet, until today, the implicit labels that the ATIES utilised tied this instrument to a disabilities-related understanding of inclusive education *for some*.

Besides the placement, other empirical studies had focussed the operationalisation of the teachers’ attitudes on a variety of issues with regards to *students with SEND*. A measurement instrument that has a long tradition in this direction is the Opinions Relative to Mainstreaming (ORM) scale, which was developed in the late 1970s by Larrivee and Cook (1979) and its factor structure was described by Larrivee (1982). This scale was used amongst the previously discussed empirical studies (Avramidis et al., 2000a, 2000b; Monsen et al., 2015; Monsen et al., 2014; Monsen & Frederickson, 2004) in slightly different variations. Most prominent, the original mainstreaming-oriented ORM (Larrivee & Cook, 1979) was updated in the 1990s with a more contemporary integration-oriented wording by Antonak and Larrivee (1995; called the ORI), and more recently with an even more contemporary inclusion-oriented wording by Monsen et al. (2015; called the TAIS). The developments of the items’ wording over recent decades can be observed. For example the item “mainstreaming the special-needs child will promote his/her social independence” (Larrivee & Cook, 1979, p. 322), in the 1990s became “integration of the student with a disability will not promote his or her social independence” (Antonak & Larrivee, 1995, p. 145). This item was updated more recently as being “including the SEN child in the regular classroom promotes his or her social independence” (Monsen et al., 2015, p. 68). This example illustrates that the wording might have changed slightly over

recent decades due to updated terminology, yet, the general logic of the items maintained nearly unaltered compared to the original from the 1970s.

Some of the previously described empirical attitude studies (Ahmmed et al., 2012; Sharma & Sokal, 2016; Sokal & Sharma, 2013) used a scale that was originally developed to measure school principals' attitudes toward the inclusion of students with disabilities in regular schools (PATIE; Bailey, 2004). It was argued in some of the empirical studies that these items were formulated in a general way so that teachers can respond to them, as well. The scale used statements like "regular teachers are not trained adequately to cope with the students with disabilities" (Bailey, 2004, p. 85) or "regular students will be disadvantaged by having special needs children in their classroom" (Bailey, 2004, p. 86). Using statements for attitude measurement to refer specifically to students with SEND is common in a variety of the aforementioned empirical studies. Yet, from a current point of view, an adoption of an inclusive education *for all* agenda would require one to think more in the direction of *for all*, and discard the thinking that presumes some individuals to be fundamentally and inherently different to all of the others. Contemporary attitude measurement instruments should incorporate this new *for all* perspective, rather than re-iterating previous perspectives.

A closer examination of these studies (which utilised more or less updated versions of the ORM, or similar instruments, such as the PATIE), demonstrated that most of them were focused on the teachers' attitudes towards inclusive education *for some*, with a specific focus on students with special educational needs. Similar to the ATIES, which was discussed before, the operationalisation and the particular understanding of inclusive education seemed to converge in a way. Indicators that address particular students by using labels need to be updated from time to time, in order to keep pace with the developments as they pertain to (politically) correct language use. However, updated wording does not necessarily update the underlying construct that is measured. In this way, it does not matter which exact labels are used to signify that some students are different, compared to all the regular students. This logic is inscribed in the ORM and related instruments in a way that ties such studies to an understanding of inclusive education *for some* students with SEND.

2.4.2 Statement of the Research Problem

The evidence provided previously suggests that all of the instruments that measure teachers' attitudes towards inclusive education utilised an understanding, which is characterised by a clear cut between a special and a regular educational system and between special and regular students. The different empirical measurement instruments operationalised inclusive education on the one hand with regards to the placement of particular students in the regular system, and

on the other hand with regards to different issues as they pertain to students with SEND. If inclusive education *for some* would be in the focus of these attitude measurement instruments, they should touch upon a variety of aspects of inclusive education, as tackling inequalities and exclusion of those groups of children, *whoever* can be considered as being vulnerable to exclusionary pressures (such as Roma children, street children, child workers, indigenous people, rural people, etc.). The constricted view on SEND, as it is represented in contemporary empirical research on the teachers' attitudes towards inclusive education, narrows down the scope of inclusive education considerably to only one aspect of a whole variety of important issues, such as inclusive education for students from disadvantaged families, for students with Learning Disabilities (LD), or for students with Attention Deficit Hyperactivity Disorder (ADHD). Moreover, if the divide between "children with average abilities" (as used in an item in the study of Avramidis & Kalyva, 2007) and those whose identified predispositions were determined as being below (or above) what is normal is reiterated in surveys for teachers, this might have detrimental effects on the teachers' thinking. In other words, if teachers are repeatedly (over decades) confronted in teacher surveys with ideas of mainstreaming and integration, this might make it difficult for teachers to think beyond these older concepts and develop an inclusive education *for all*-related thinking. Especially in times, when teachers are supposed to keep themselves up-to-date with the most recent empirical studies and to base their practices on empirical evidence, the re-adjustment of their views on outdated concepts of mainstreaming and integration, as they are represented in empirical research on the teachers' attitudes towards inclusive education, might have counter-productive effects. As previously discussed, the teachers and their attitudes play a central role in implementing new policies. Hence, to adjust the empirical research on teachers' attitudes towards a more contemporary understanding of inclusive education *for all*, seems to be imperative. It might be time to raise the question, "if inclusion is about all, why do we still mostly focus on some?", as Messiou (2017, p. 152) recently asked, and to rigorously apply this question to all of the studies in the field of inclusive education *for all*, including those investigations with a focus on the teachers' attitudes.

Most recent literature reviews (Nilholm & Göransson, 2017; Ruberg & Porsch, 2017) supported the view that current empirical studies were acknowledging contemporary thinking about a wider understanding of inclusive education; yet, these empirical studies were not able to operationalise this understanding appropriately and, hence, the empirical parts of these studies were basically focussed on ideas of mainstreaming and integration. Ruberg and Porsch (2017) found in their systematic review a similar disparity between conceptual and empirical understandings of inclusive education, as it was reported by Nilholm and Göransson (2017).

Yet, while Nilholm and Göransson (2017) concluded that future research needs to be clearer with defining inclusive education in order to mitigate some of the existing confusion about this term, Ruberg and Porsch (2017) concluded that future research needs to consider that most empirical researchers are in principle aware of a wider and more contemporary understanding of inclusive education *for all*, but that empirically the placement-/SEND-related understanding of inclusive education *for some* is still present in current empirical investigations. They clearly state that this demands further developments in the operationalisation of the concept:

While all studies theoretically refer to a wider understanding of inclusion that is focussed on the societal participation of all humans, nearly all of the conducted surveys define inclusion in the sense of classes for both students with and without special educational support needs. In order to strengthen the significance and comparability of empirical studies in the field of ‘attitudes towards inclusion’, it seems necessary to clarify how measurement instruments should be designed that are not concentrated on disability, but which pick up on and mirror a wider understanding of inclusion. (Ruberg & Porsch, 2017, p. 409; original quote in German, which was translated by the present author)

Although conceptual clarification, as Nilholm and Göransson (2017) emphasised, seemed to be generally important, the issue that Ruberg and Porsch (2017) uncovered seemed to be more urgent. If no empirical research instruments are available to study inclusive education *for all*, then the gap between conceptual and empirical understandings of the concept cannot be neither narrowed nor resolved. Hence, each single empirical study contributes in a way to enhancing the confusion as it pertains to the meaning of inclusive education, not only amongst researchers who conduct and consume research, but also amongst all relevant stakeholders in education (such as teachers, principals, parents, policy makers, etc.), who consume research, but who are also participants of such research studies.

2.4.3 Research Question

According to the previously stated research problem, there seems to be a need to identify ways of measuring the teachers’ attitudes towards inclusive education *for all*, as there seems to be no adequate attitude measurement instrument available. The overarching research question in this study is: How can the teachers’ attitudes towards inclusive education *for all* be measured? As inclusive education *for all* can be considered a global commitment, and because the different countries that develop towards reaching more inclusive education *for all* need to learn from each other through international (research) exchange, the present study attempted to find a way to measure the teachers’ attitudes towards inclusive education *for all* in different countries, with

different cultural backgrounds, and with different languages. Hence, the sound and robust measurement instrument that the present study attempted to develop, needed to be created in and for cross-cultural contexts. In order to demonstrate the quality of the new measurement instrument, teacher samples in Australia and Germany will be drawn and the teachers' responses will be analysed to find relevant dimensions of the teachers' attitudes towards inclusive education *for all*.

2.5 Chapter Summary

Considerable evidence emphasised the importance of teachers for student learning. Yet, from the teachers' perspective to make sense out of contradicting policies, concepts, and ideas seems particularly difficult; especially when quality teaching is supposed to be achieved. Particularly when it comes to inclusive education, the present chapter demonstrated that from the individual teachers' point of view, a large diversity of understandings and policies are apparent. Hence, within this conceptual confusion, it might be no wonder that questions, as they pertain to where particular students with SEND need to be placed, are still in the forefront of teachers' thinking (and of empirical research) on issues as they pertain to inclusive education.

One argument, as it was laid out in the previous sections, was that such great confusion does not exist on the global (policy) level, and that the diverse understandings that are apparent from the teachers' or the researchers' point of view emerge due to the translation of these global ideas and policies into national, local, school, and classroom practice levels. This insight was relevant for several reasons (e.g. it encourages all stakeholders to keep or start questioning the established views on inclusive education), but most important it explained to some extent why researchers found conceptual diversities (Göransson & Nilholm, 2014), sometimes even *within* research papers (Nilholm & Göransson, 2017; Ruberg & Porsch, 2017). According to this, the hypothesis was developed that, conceptually, inclusive education started to embrace more and more the notion of *all*, rather than *some*; but that the development of empirical research instruments lag behind.

Concerning the concrete implementation of inclusive education *for all* into 'real world' practices, the schools were emphasised as important institutions for inclusive education to take place. Schools can work towards the presence of all, and they can provide a sufficient culture that fosters more inclusive practices; yet, ultimately, the participation and achievement of all students and to tackle discrimination in daily (school) practices, demands more than school level developments. It demands best teachers and the best teaching. This chapter presented

evidence in this way, that the teachers and particularly their attitudes are crucial for implementing more inclusive education *for all*. However, within the large body of empirical research in this area on teachers' attitudes, which was discussed in quite some detail in this chapter, the focus on the placement of particular students and the views on students with SEND was prevalent.

If the teachers' attitudes towards inclusive education *for all* are crucial for proceeding some steps in education, and if there is a lack of research instruments, which operationalise up-to-date understandings of the concept, then, the development of such an instrument seems to be an urgent matter. Accordingly, the research problem and research question are both formulated in this way. The next chapter will provide the methodology that was utilised in the present study to fill this research gap and to make a substantial contribution to research in the area of inclusive education.

Chapter 3 · Methodology

3.1 Introduction

The literature review suggested that the teachers' attitudes are an important factor for inclusive education to take place in schools. A variety of studies demonstrated the importance of teachers and their attitudes for the implementation of most recent policies and concepts as they pertain to inclusive education. None of the reviewed empirical studies on teachers' attitudes seemed to touch upon inclusive education *for all*. These studies utilised the term inclusive education not as referring to all students, but as the catering for particular groups of students with special educational needs and/or disabilities (SEND). This study aimed to construct a sound and robust measurement instrument, which operationalises teachers' attitudes towards a wide/r understanding of inclusive education (in the sense of catering for *all* instead of catering for *some specifically*), which would complement the existing knowledge about teachers' attitudes towards inclusive education.

In the following section, the methods used to construct and substantiate such a new measurement instrument are described. First, the purposes of the empirical study will be specified, before describing the contexts in which the present study was situated. Then, the research design, as it will be utilised to work towards the study's purposes, is presented. This step comprises all of the methodological decisions with regard to the key parameters of the study. After the procedures of developing the data collection instruments are described, the data collection procedures and the quantitative methods of analysing the data are outlined.

3.2 Purpose of the Empirical Study

While the statement of the research problem and the research question were previously formulated in a more general way, more concrete purposes of the study needed to be specified, in order to be able to align the research design and the other key parameters of the present study with achieving these purposes. First and foremost, the primary purpose was to find a way to measure teachers' attitudes towards inclusive education *for all*. In this way, each of the indicators, which were utilised for measuring teachers' attitudes needed to reflect inclusive

education *for all*, rather than inclusive education *for some*. Moreover, according to the discussion of this new concept and the particular understanding of (teachers') attitudes in the previous chapter, teachers' attitudes towards inclusive education *for all* were assumed to touch upon different aspects. Hence, the sound and robust measurement instrument, which the present study attempted to develop, was considered to comprise multiple dimensions.

Developing a new questionnaire is a difficult undertaking, especially when attempting to accomplish for use in cross-cultural settings, like it was attempted in the present study. Harkness, Edwards, Hansen, Miller, and Villar (2010) cautioned that research instruments, which were developed in one context, should not be assumed to be valid in another context. The authors pointed out that research instruments for cross-cultural studies should be particularly designed for multinational, multicultural and multilingual use (Harkness et al., 2010). For the development of the instrument in the present study, this meant that it needed to be developed in (at least) two languages (for multilingual use). In addition, these items in two languages needed to use terminology and phrases that could be understood in different contexts (for multicultural use), and the content of these items needed to be correct and valid in different countries (for multinational use).

3.3 Study Population

In order to develop an inclusive education *for all* survey instrument for cross-cultural use, the countries, in which the questionnaire was thought to be developed, was selected in a purposeful way. The outcome of this informed decision was to develop the questionnaire in Australia and Germany. As will be discussed in the following section, the study population comprises teachers in different stages of their careers; namely, pre- and in-service teachers are included in the sample. In this way, the development of the attitude instrument was thought to be valid for all teachers; no matter if they were pre-service or in-service teachers. The concrete sampling strategy is discussed according to the research question to initially develop a new attitude towards inclusive education *for all* instrument.

3.3.1 Countries in which the Present Study was Carried Out

According to the study's research question and the specified purposes, a new questionnaire for cross-cultural use was thought to be developed, which was able to measure the teachers' attitudes towards inclusive education *for all*. This brought up the question, in which countries

and, accordingly, in which languages this new measurement instrument was attempted to be developed.

In the literature review (see especially Section 2.2.2), a variety of issues as they pertained to developments towards inclusive education for all were reported for a number of countries, such as the United States, Canada, or Norway. Amongst these examples was *Australia*, which seemed particularly interesting in terms of inclusive education *for all*. As discussed earlier in the present study, due to its federal constitution, major differences in relevant policies were apparent in Australia's States, while, generally, Australia was thought to have already undergone major steps towards inclusive education *for all* (Carrington et al., 2012). As previously discussed, Australia's school system is built on neo-liberal values, which means that standardised testing is established and schools are competing in terms of test scores on the school level (Fachinetti, 2015; Hardy & Woodcock, 2015; Johnston, 2017). Low-achieving students who would lower the school's test score could be excluded from standardised testing, which, again, pushed diagnosing and labelling practices forward, as discussed previously (Anderson & Boyle, 2015; Elliott et al., 2012).

Another example, which turned out in the literature to be of particular interest with regards to inclusive education *for all* was *Germany*. Like Australia, Germany had a federal constitution with regards to the educational sector, and policies vary considerably across the Federal States (Hebborn, 2014; Hinz, 2015), which was discussed previously in the present study. In this regard, both Australia and Germany seemed to be comparable; namely, both countries leave the authority of education-related questions basically to the Federal States. Yet, as opposed to Australia, Germany's school system has a long tradition in fully separated special schools for students who were identified with SEND (Banafsche, 2013; Henry-Huthmacher, 2015). Another difference between Australia and Germany is that Germany's school system is mostly considered as social-democratic as opposed to neo-liberal (Tomlinson, 2015), which meant among others that no nation-wide students' testing is in place.

Taking together these arguments, it seemed feasible to select Australia and Germany as relevant contexts for the present study. As described, the both countries have a comparable constitution (federalism), which might make it less difficult to carry out an empirical study in both countries in a meaningful way. Yet, both countries are considerably different to each other (standardised testing vs. none; less special schools but still diagnosing vs. many special schools; etc.). These differences might allow one to assume that the attitude instrument that is thought to be developed in these contexts might be robust across relatively different contexts.

Within Australia and Germany, specific areas were chosen in a convenient way (see convenience sampling in L. Cohen, Manion, & Morrison, 2007). Namely, the teacher

population in Sydney (New South Wales, Australia) and Giessen (Hesse, Germany) were chosen as contexts of the study. The study population was narrowed down in this way, because the present study's purpose was not to draw a potentially *representative* sample from the whole country's population. The purpose was to draw a sample comprising *relevant* individuals with a considerable variety of characteristics, so that the developed instrument can be considered sound and robust across many individuals (see 'development sample' in DeVellis, 2011), which is elaborated further upon later in this chapter. Hence, the wider study population were teachers in Australia and Germany, yet, the study population that the present study's results can be applied to are teachers in Sydney and Giessen. The actual sampling and the potential generalisation of the empirical results are reflected later in this chapter.

3.3.2 Further Description of Both Contexts

Although a direct comparison of both contexts was *not* intended to be carried out in the present study, some additional information as they pertain to Australia and Germany are provided in the following section. Most obviously, both countries are multinational, multicultural and multilingual, which is a good point of departure for developing a comparative questionnaire, as Harkness et al. (2010) noted. In order to gain more insights of both study contexts, the brief overview in the following section adapts the 'locational levels' dimension as it was suggested by Bray and colleagues (Bray, Adamson, & Mason, 2007; Bray & Thomas, 1995). The description of a variety of aspects were thought to illustrate some of the commonalities and differences of both contexts in the present study. Many further insights, what Australia and Germany can learn from each other in terms of education and particularly with regards to inclusive education, can be drawn for example from the study of Harrington, Kastirke, and Holtbrink (2016).

A first aspect is in what *world regions* both contexts are localised. Geographically, Germany is on the northern hemisphere, while Australia is on the southern hemisphere. A great deal of inclusive education research have explicitly discussed differences between the global North and the global South, and the need for more collaboration between these countries in the future to allow for more inclusive practices to take place (Ainscow & Miles, 2008; Miles, Lene, & Merumeru, 2014; Miles & Singal, 2010; Moberg, 2003; Savolainen et al., 2012; Tota, 2014). Yet, most of these studies refer specifically to developing countries on the southern hemisphere and to developed countries on the northern hemisphere. In this way, Australia and Germany are not good examples of the development-related North-South divide, because both countries are considered the most developed countries worldwide. According to the recent report of the United Nations Development Programme (UNDP, 2016), Australia has together with

Switzerland the second highest Human Development Index in the world, while Germany has the fourth highest score worldwide. For the present study, this meant that the study contexts were considerably different to each other, because they were localised on different sides of the planet. On the other hand, both contexts were comparable to each other, because both needed to be considered as being most developed countries. Hence, for developing the new instrument to measure the teachers' attitudes towards inclusive education *for all*, these two contexts seemed appropriate.

A second aspect is, which *countries* are considered within the wider world regions. While Australia is not only a country, but at the same time a continent, Germany is a country within Europe. According to the World Bank statistics (<https://data.worldbank.org>; available data from 2016; accessed on 17/12/2017), Australia has a land area of 7,682,300 square kilometres and a population of about 24,127,000. Germany would fit into Australia about 22 times with a land area of 348,900 square kilometre, while Germany's population is 82,668,000 inhabitants (three times bigger compared to Australia's population). In other words, both countries are quite different to each other (in terms of land area and population, but also with regards to many other facets; see e.g. each country's website on www.oecd.org). Some similarities, but also some major differences as they pertain generally to Australia and German were previously discussed in the Literature Review Chapter. On the one hand, both countries are similar in that both have signed and committed to several documents of the United Nations, such as the Universal Declaration of Human Rights (UN, 1948), or the Convention on the Rights of Persons with Disabilities (UN, 2006). On the other hand, as described earlier in the present study, these global policies were confronted with two contexts with very different values (Australia's values could be considered as more neo-liberal while Germany's values could be considered as more social-democratic) and very different histories (Australia seems relatively developed in terms of inclusive education, because it has closed down most special schools, while Germany has established a comprehensive and nationwide special education system, which was not easy to be integrated into the regular system). Hence, both countries are different in terms of their general practices with regards to inclusive education (UN, 2013, 2015). For the present study, this meant that both contexts have some fundamental similarities, yet, they also differ considerably. It seemed appropriate to have such different contexts, which would make an instrument relatively robust, if it would be developed successfully in these diverse contexts.

Third, the *states* of the countries are examined. In Australia, there are six States and two Territories, while in Germany, there are sixteen Federal States. As mentioned before, both Australia and Germany have a federal constitution in questions as they pertain to education, and as previously discussed, these States differ considerably in terms of education. In a recent

study by Hardy and Woodcock (2015), it was demonstrated (amongst other results) that in Australia the State New South Wales had more exclusive policy texts compared to the more inclusive policy texts, which were present in Queensland (see L. J. Graham & Sweller, 2011 for discussion of further issues as they pertain to NSW's school system). In Germany, a recent study conducted by Berkemeyer, Bos, Hermstein, and Abendroth (2017) highlighted (amongst other results) that there were considerable differences between the Federal States in terms of inclusion and exclusion. Amongst a variety of recommendations, Berkemeyer et al. (2017) emphasised for Hesse that this Federal State might need some more developments in the direction of more inclusive education, in order to provide education for all students. For the present study, it seems appropriate to consider New South Wales (NSW; Australia) and Hesse (Germany) as two relevant States, because both seem to have not yet tapped their full potential in terms of inclusiveness of their educational systems.

A fourth consideration pertains to the *districts*, which were examined in these two States. NSW and Hesse are divided into a variety of different districts. Politically, NSW is divided into 93 State electoral districts (www.elections.nsw.gov.au; accessed on 17/12/2017), and Hesse is divided into 21 districts and five urban districts. Sydney is the largest city in NSW (and also Australia), comprising a variety of districts and a population of 5,029,768 (www.abs.gov.au; accessed on 17/12/2017). Giessen, on the contrary, refers to the city of Giessen and also the district of Giessen, which comprises not only urban but also some suburban areas. Giessen district comprises a population of 249,040 (<https://statistik.hessen.de>; accessed on 17/12/2017). For the present study, both contexts seemed to be most feasible, because Sydney is a very large metropolis with a range of different schools and universities, while Giessen district had an urban core with a university, yet, also some more rural areas. In this way, to have such a variety of contexts seemed most appropriate.

As a fifth aspect, the *institutions* are examined, which were relevant for the present study. In Australia and Germany, pre-service teacher training and the in-service teaching in schools depend to a great extent on the context (e.g. State, university, etc.). The following notions as they pertain to the institutions are only supposed to give a very brief overview. Teacher training in both contexts take place in higher education institutions (mostly universities). In Australia, teachers, generally, complete a 4-year undergraduate program to become a teacher. Alternatively, those, with a professional qualification (3-year undergraduate program) can pass a 1-2-year postgraduate program to become a teacher (e.g. Master of Teaching). The pre-service training is, generally, differentiated in the primary and the secondary sector (and there is a path that leads to being a special education teacher). In Germany, teachers, generally, pass a 4-5-year program to become a teacher. A variety of options for lateral entry into different phases of

the teacher training exist, depending on the Federal State and the university. Generally, the training is differentiated in a similar way to Australia in a primary and a secondary (and a special) track; yet, in many cases to become a secondary school teacher is further differentiated into two separate tracks, where one track (so-called ‘Gymnasium’) is only for higher-achieving students and the other track (so-called ‘Haupt-/Realschule’) is for other students. This differentiation into a primary sector and a secondary sector (and a special sector) in teacher education is mirrored in the existence of primary schools and secondary (and special) schools. Notably, while the primary sector in Australia, generally, comprises of grades K-6 (Kindergarten [Ger. Vorschule] to Grade 6), the primary sector in Germany, generally, comprises Grades One to Four. With regards to the institutions, the comparability of both educational systems was found to be sufficient for the present study, and the differences as they existed between Australia and Germany were found to contribute to the robustness of the instrument that this study attempted to develop.

A final consideration pertains to the *individuals*; namely the teachers themselves. In research contexts, it is common to clearly differentiate teachers according to their availability in different institutions. Student teachers are commonly approached at universities and mostly, these teachers are in their pre-service phase. Other research approaches teachers at schools, which are mostly, in-service teachers. Accordingly, research on pre-service teachers (sampled in universities) and on in-service teachers (sampled in schools) are embedded in different research discourses. Yet, this picture seems not to be comprehensive. In-service teachers can appear at universities to attend further training courses, to study for a higher degree qualification, and so forth. In the pre-service teacher study of Hecht et al. (2016), for example, the sample comprised seven percent of participants, who had already completed their higher education teacher training and were at the entry into their professional life. On the contrary, pre-service teachers can appear at schools in their practicum phase, or in their first probationary year/s after initial training (in Germany, this probationary phase before reaching the “second Staatsexamen” is part of the initial training, yet, although it is on-the-job; hence the teachers in this phase are not considered full in-service teachers), and so forth. In the study of Boyle et al. (2013), for example, the in-service teacher sample comprised seven percent probationary/newly qualified teachers (‘under one year and/or probationary year’). Accordingly, the present study attempted to examine the attitudes of teachers in general, no matter which phase of their professional careers (initial training, newly qualified teacher, in-service, further training, etc.) the teachers were in. Hence, the teacher population was specified as being both, pre-service and in-service teachers at universities and schools. From a more conceptual perspective, there seems to be no plausible argument, why e.g. a teacher at the end of her/his study program might have

a completely different attitude structure compared to an in-service counterpart in her/his first years at school. There might be gradual differences of the favourableness of teachers with regards to inclusive education *for all* in different steps in their careers; yet, it was not assumed that there were actually fundamental differences in this regard.

All these arguments clearly demonstrate that both countries have many similarities such as both being one of the most developed countries in the world, both having ratified several UNESCO policies (e.g. the Convention on the Rights of Persons with Disabilities), both having a federal constitution and both being comparable with regards to the school system (e.g. a primary vs. a secondary sector) and the teacher training (basically at universities). Hence, a study on the teachers' attitudes can be carried out in these two contexts in a meaningful way. On the other hand, evidence was presented, that both contexts can be considered differently in many aspects, too. An example is that Australia is large with a small population, while Germany is small with a large population. Australia seems to be some steps ahead in implementing change as it is necessary in order to comply with e.g. the Convention on the Rights of Persons with Disabilities, compared to Germany. The Australian primary sector is mostly K-6 while in Germany the primary sector usually comprises only the first four school years. Hence, it can be summarised that besides the comparability, Australia and Germany are different to each other. A scale development in these particular contexts is likely produce a robust new instrument.

3.4 Research Design

The key parameters of the research design need to reflect the attempt to find a way, how teachers' attitudes towards inclusive education *for all* can be measured in cross-cultural settings. The empirical part of the present investigation was generally based on established text books (such as L. Cohen et al., 2007; Döring & Bortz, 2016) and recent recommendations for test and scale development (Bühner, 2011; DeVellis, 2011; Lane, Raymond, Haladyna, & Downing, 2015). All research design-related decisions as they pertain to the general stance, the research style, the objective, the scope and the quality (validity, reliability and fairness) are justified in the following sections.

3.4.1 General Stance: Quantitative Research

The research question of the present study was to find a way, how teachers' attitudes towards inclusive education *for all* can be measured. This determined some of the present research study's key parameters. To assume that the *measurement* (see Michell, 1999) of the teachers'

views is actually possible, the study took a *quantitative* stance (also: a normative paradigm, as opposed to an interpretative paradigm; see L. Cohen et al., 2007). Quantitative research attempts to study a large number of individuals. In this way, it is also possible to study many individuals' responses across different contexts. On the contrary, utilising a qualitative methodology would also have been an appropriate way to study attitudes. A qualitative stance would suggest to value each individuals' interpretation of the world and to start with in-depth considerations of individual viewpoints (see L. Cohen et al., 2007; and see Savin-Baden & Major, 2013 for a more detailed discussion). As opposed to a quantitative approach to attitudes, which basically differentiates degrees of favourableness of particular aspects of the attitudes, the qualitative view would be able to gain rich insights into the individual content and meaning of the individuals' attitudes.

The qualitative and the quantitative research stance both have advantages and disadvantages (see also the detailed discussion in Döring & Bortz, 2016). As L. Cohen et al. (2007) noted, quantitative research over-emphasises the individuals' common and standardisable views, and neglects the actual individual perspectives, while qualitative research over-emphasises the individual views and neglects in part the structural forces that shape the individuals' views and actions. As discussed previously in the literature review, the phenomenon under investigation – namely, inclusive education *for all* – starts as a global commitment, which, then, becomes relevant in many countries across the globe. In this way, the particular interest of the present study is to identify and investigate common aspects of inclusive education *for all*, from the viewpoint of the teachers. In other words, the present study takes a macro- or system-level perspective on the teachers' attitudes, while the individual idiosyncrasies of the teachers' views are not of particular interest in the present study. This more abstract focus suggests utilising a quantitative stance, and, in this way, suggests that *measuring* attitudes is feasible for obtaining information from a large variety of different teachers (as opposed to gaining rich insights from only a few teachers). In the next few paragraphs, it will be argued that more qualitative research styles (such as case study or ethnography) are not as adequate for the present study compared to more quantitative research styles (such as survey research or tests). Hence, in the following sections, more arguments are given to support the choice for a quantitative stance.

3.4.2 Research Style: Survey Research Using a Questionnaire

The research study needs not only to be specified in terms of a quantitative or a qualitative methodology. To the contrary, L. Cohen et al. (2007) cautioned to put this differentiation too much to the forefront. It seems more effective, as L. Cohen et al. (2007) emphasised, to

differentiate different types of research styles, which are possible to be carried out: case study, ethnography, action research, experiment, survey research, and assessment. All these methodologies could be carried out in the present study, and they all have different advantages and disadvantages with regards to answering the research question of the present study. In the following section, arguments are presented as to why survey methodology seemed to have the greatest potential for the present study, and why this style of carrying out the present study was selected.

Choosing a Research Style

Case study and *ethnography* research are used, if the uniqueness of particular individuals/situations are of particular interest of a study (L. Cohen et al., 2007). Although the present study did not aim at portraying a unique case nor at portraying a specific situation in the subjects' terminology, such approaches would certainly benefit to understand the perspective of the teachers in greater depth. In order to develop a new measurement instrument that is supposed to have a high quality (e.g. validity across different contexts), a range of such case studies or ethnographies would be needed to be carried out in a range of contexts. Such an approach would start with understanding the teachers' viewpoints, and would allow to potentially construct the questionnaire in a way that is most relevant to the actual teachers. This strategy to develop a new questionnaire seems promising, because it values the perspectives of the teachers and starts with real-world experiences of the teachers (similar arguments are for example used by researchers, who emphasise cognitive techniques for pretesting new questionnaires, such as Collins, 2003; Lenzer, Neuert, & Otto, 2015; Willis, 2005). However, there are three issues with the sole use of such an approach. First, case studies and ethnographies are time-consuming and need a great amount of resources (Savin-Baden & Major, 2013). If several such studies would have to be carried out in order to grasp a variety of individual cases (e.g. from different institutions, or countries), such expenses seemed to go beyond what seemed feasible for the present doctoral study. Second, it seemed unlikely that a sufficient variety of contexts could be studied in this way to develop a measurement instrument that can be assumed to be valid for cross-cultural use. In this way, Savin-Baden and Major (2013) emphasised that ethnographic research, but also case studies tend to be not easily generalizable to other contexts. Third, another issue was that the new agenda of inclusive education *for all* might not be found in the practitioners' views, yet. In the literature review, it turned out that particularly practitioners at the school level highlighted the need for medical diagnosis and labelling of children (Nes & Strømstad, 2006; Nilholm et al., 2013). Hence, if an investigation to develop a new measurement instrument would start with the practitioners' views, the medical model

might be in the foreground of responses, which would not be in line with the purposes of the present study. To construct a measurement instrument that represents the new kind of thinking towards inclusive education *for all* might need to start from a more conceptual viewpoint, and not from established views of the teachers.

Action research is utilised, if an intervention is planned, implemented, reviewed and evaluated, and an *experiment* is used, if controlled conditions are needed, for example in order to make generalisations about the effectiveness of a treatment (L. Cohen et al., 2007). Although the present study did not specifically include an intervention or a treatment, specific knowledge, on how teachers would for example respond to new policies (such as inclusive education *for all*) and how they transpose their perception of such policies into action, would benefit the aim of developing a new measurement instrument. This would allow one to focus on specific aspects of the attitudes, which turn out to be relevant in terms of teaching and related teacher action. Yet, like the previous approaches, to carry out action research and experiments is time-consuming and is strongly dependent on particular contexts (L. Cohen et al., 2007). This would make it unnecessarily laborious to carry out such research in different contexts to reach a conceptual saturation and generalisability. In addition, the interventions and treatments that action research and experiments would require (L. Cohen et al., 2007), would focus not so much the attitudes and perceptions of teachers, but more the functioning and the consequences of the teachers' attitudes and perceptions. This would go one step beyond the present study's scope. In other words, the study of an intervention or treatment would already presume an understanding of the underlying attitudes of the participating individuals. Hence, to carry out action research or an experiment seems informative for developing a new measurement instrument, yet, it seems that these methodologies are not the first choice, because it is not the intervention or the treatment that is in the focus of the study.

Survey research is utilised if the goal is gathering large-scale statistical data about opinions of respondents in an economical and efficient way, while *assessments (or tests)* are used to measure achievement, abilities and performance (L. Cohen et al., 2007). The latter style of conducting research seems to be of general interest for the present study, because certain emphasis is given to the measurement. Hence, literature as it pertained to (cross-cultural) assessments and tests are explicitly included, when the present study was developed, carried out and interpreted. Yet, the assessment goes beyond what is actually intended in the present study. It is not intended to assess e.g. the knowledge concerning inclusive education *for all* or the performance levels in carrying out inclusive practices *for all*. To conduct research in this way has the advantage to be able to consider many responses of individuals and to collect the data in a more standardised way. Yet, both advantages are at the same time disadvantages,

because if many individuals are asked for giving responses to a number of standardised items, there are less options for an in-depth study of particular cases and for giving the respondents the opportunity to express their thought apart from what the instruments want them to do. The aim of the study was to gain insights into how a larger variety of respondents viewed inclusive education *for all*; if they see it in a more favourable or unfavourable way, which seems to fit well with the strength of survey research (L. Cohen et al., 2007). Hence, survey research seemed to be the most appropriate style of conducting the present study. In the following section, this methodological decision to utilise survey research is concretised.

Choosing a Mode of Conducting Survey Research

L. Cohen et al. (2007) differentiates several modes of conducting survey research. One option is so-called interview questionnaires, which are administered by an interviewer. The presence of an interviewer is supposed to improve response rates and to allow the respondent to ask questions. Yet, the interviewer's presence also makes the process of completing the questionnaire less casual, which is, on the one hand, positive in terms of controlling the environment in which the questionnaire is completed. On the other hand, the interviewer's presence might evoke social pressures to complete the survey, and might lead to social desirability bias of the responses (Davis, Couper, Janz, Caldwell, & Resnicow, 2010). An approach to mitigate the interviewer effects are telephone survey (L. Cohen et al., 2007). Yet, telephone surveys have other issues, such as that the response format must be easy enough so that the respondents can understand the answer options, which are read out on the phone by the interviewer. Accordingly, it has been reported that telephone surveys produce more missing data compared with face-to-face surveys (L. Cohen et al., 2007). Both modes involve interviewers that are supposed to improve response rates and foster responses to all of the items, while accepting that the participation and responding to items is not fully voluntary. Other options, as described by L. Cohen et al. (2007), are to use self-administration, administration via post, or administration via the internet. In these modes, no interviewer guides the interview. All these modes have different biasing effects, as discussed by Bowling (2005).

For the present study, it was important to allow the respondents to choose to participate in a voluntary way. As described later in this chapter, the sampling utilised in the present study was non-probability and convenience sampling; hence, there was no need to try to convince (or even force) potential respondents to partake. In addition, the non-response to particular items was found to be informative. If an item had a particularly high number of missing answers, it might indicate that this item was not well developed (e.g. bad wording, offensive, not understandable). To have an interviewer, who forces explicitly or implicitly responses to each

of the items, would be counter-productive in this sense. Hence, a mode of conducting survey research was chosen, without an interviewer. In the present study, to gather all of the respondents' e-mail or postal addresses seemed to make the sampling procedure unnecessarily difficult, and it was assumed that the cognitive burden, which might be imposed on the respondents by utilising a self-administered mode (as discussed by Bowling, 2005) is acceptable. As discussed later in this chapter, the questionnaire was administered at university classes and schools; in a way that the respondents completed the survey at their own convenience and wherever they liked, and, then, submitted their responses anonymously.

3.4.3 Objective: Develop a New Attitude Questionnaire

While the previous paragraphs only proceeded on the assumption that a new research instrument was thought to be developed, it is clear from the literature review that this instrument was meant to measure *attitudes* specifically. There is a large research body on attitude measurement, and some core characteristics are discussed in the following sections, as they are generally relevant for the present study.

General Recommendations for Scale Development

In the methodological literature (such as Bühner, 2011; Jonkisz, Moosbrugger, & Brandt, 2012), different procedures for developing a new questionnaire are differentiated. If a theory is available, which comprises all of the constructs of interest, the search for relevant items should be focussed on operationalising exactly these constructs. Bühner (2011) and also, for example, Jonkisz et al. (2012), call this a *rational construction strategy*. Yet, as the literature review in the present study demonstrated, the notion of inclusive education *for all* represents a new kind of thinking. Hence, although extensive research was available with regards to former and related concepts, inclusive education *for all* as a newly established term lacked a solid theoretical basis, which was why a rational strategy to develop the questionnaire seemed not to be feasible for the present research study.

A range of strategies were suggested, if the newly developed instrument was not based on an underlying theory. According to Jonkisz et al. (2012), an *intuitive construction strategy* can be recommended in cases, where no theory is available. This strategy would mean to completely rely on the researcher's intuition, and to build methodological decision-making and item construction on plausible arguments. Such an approach seems appealing because of its pragmatism; yet, because of its dependency on what a particular researcher would find being appropriate, it seems difficult to be justified as a solid research method. Another method would be the *external construction strategy* (Bühner, 2011; Jonkisz et al., 2012), which is a pragmatic

approach to developing a questionnaire that is thought to identify particular individuals belonging to particular groups. Items are collected that were thought to gain differing responses by the different groups of potential study participants. Through empirical testing, those items which were most discriminating between the groups of interest are considered most appropriate. This strategy seemed not adequate to be used in the present study, because the external construction is focused more on the discriminating functioning of the items and not so much about actually measuring to construct that was supposed to be measured.

Another method that was suggested by Bühner (2011) and Jonkisz et al. (2012) was an *inductive or internal construction strategy*. This approach of developing a new instrument comprised creating a large number of items, which are all related to a specified research topic. These items are, then, administered to respondents and the collected data is analysed with exploratory statistical methods. Bühner (2011) pointed out that through the empirical data analysis, a conceptual idea of specific dimensions is developed. And Jonkisz et al. (2012) added that the interpretation of the dimensions, which were reached through statistical exploration, can go beyond the actual items.

The inductive strategy seemed particularly feasible for the present study, because the item construction is done in an exploratory – but still systematic – way. The cited literature in the previous paragraphs cover particularly the German perspective. In the international methodological literature, especially the inductive strategy was recommended generally for developing a new scale. DeVellis (2011) stated to start with determining clearly what was attempted to be measured and, then, to generate an item pool. After defining the format for measurement, DeVellis (2011) recommended to let the item pool be reviewed by experts and to consider other items/scales to be included, in order to be able to validate the newly developed scale. Then, according to DeVellis (2011), after administering the questionnaire to a development sample, the items are evaluated using statistical methods and the scales' length is optimised.

Overall, an inductive strategy, as described before, was utilised in the present study. As a next step, it was discussed, if this general strategy to develop a new instrument was also feasible to develop a new instrument to measure *attitudes*.

Measurement of Attitudes

The common assumption that attitudes can be measured (Thurstone, 1928) and the search for a technique for the measurement of attitudes (Likert, 1932) goes back at least to the late 1920s and early 1930s. Until today, a variety of attitude measurement techniques have been proposed and used. It is common to distinguish between direct methods and indirect methods (Antonak

& Livneh, 1988; Hitlin & Pinkston, 2013; Schwarz, 2015). The latter methods are applied to measure implicit attitudes that the individuals are not aware of. In such cases, the measurement would involve for example to measure the response time to a given question or to measure how participants categorize pictures or terms rapidly, without thinking too much about them (Hitlin & Pinkston, 2013; Schwarz, 2015). The direct methods measure directly the explicit attitudes of individuals. If it can be assumed that the individuals can articulate their attitude towards a particular topic, then it seems to be feasible to rely on the answers that respondents give on questions regarding their attitudes (Schwarz, 2015). In inclusive education research, most available studies on attitudes assumed that individuals could express their thoughts about inclusive education, because this topic was widely discussed, not only in professional contexts, but also in public media, so that the teachers were familiar at least with some aspects of it, have a particular standpoint towards it, and can express this standpoint. Accordingly, most widely used in inclusive education research are direct methods of measuring attitudes. There are not many indirect methods, such as Lüke and Grosche (2017), who recently developed an attitude towards inclusive education test, based on an implicit method, in order to tackle social desirability in participants' responses. Although the argument of social desirability response bias is a valid argument, indirect measurement of attitudes are difficult to be carried out and are considered to be more of a supplement (Antonak & Livneh, 1995) to traditional direct attitudes testing and not so much a better method on its own. Hence, the present study attempted to measure attitudes in a direct way.

Four direct measures are generally recommended to conduct research on attitudes (Eagly & Chaiken, 1993; Fishbein & Ajzen, 1975; Maio & Haddock, 2009; Procter, 2008; Schwarz, 2015). A first method is to utilise scales on the basis of *equal-appearing intervals* (Thurstone, 1928; Thurstone & Chave, 1929). For establishing equal-appearing intervals a number of judges need to be involved, who sort a variety of statements into different (mostly from 1 to 11) increments of general favourableness. Similarly, the *scalogram analysis* (Guttman, 1944) involves finding statements that represent different sections on the continuum that is thought to be measured. Yet, the aim is to find statements with regard to a chosen topic that can be sorted into a clearly hierarchical order. In other words, the agreement with a particular statement involves the agreement with all other hierarchically lower items. Another method to measure attitudes is to use *semantic differentials* (Osgood, 1952; Osgood, Suci, & Tannenbaum, 1957), which involves using a variety of pairs of bipolar adjectives (such as bad vs. good, or dirty vs. clean). The respondents are asked to rate the particular topic with regard to each pair of adjectives (e.g. closer to 'good' as opposed to 'bad'). The fourth method is to use *summated ratings* (Likert, 1932), which involve asking the participants to rate a variety of topic-related

statements with regards to how strong they would disagree or agree with each of them. Likert (1932) argued that the sum (or the mean) of these ratings can be used to determine the individual's attitude.

Of these four direct methods for measuring attitudes, Likert's method is most common and most feasible. In order to determine which scaling method has generally the most practical relevance, Döring and Bortz (2016) conducted a database search using PsycINFO. The outcome of this search was that papers that were tagged with 'Likert scale' (1,815 hits) were by far most common amongst the studies in the database. Other methods were less common, such as the 'Guttman scale' (201 hits) or the 'Thurstone scale' (61 hits) (Döring & Bortz, 2016). The fact that Likert scales are most common in attitude research was emphasised by other researchers, too (such as Antonak & Larrivee, 1995; Bohner & Wänke, 2002; DeVellis, 2011; Schnell, Hill, & Esser, 2011). As Schwarz (2015) noted, Likert scales need considerable time for a careful development of appropriate statements and to pre-test, if the respondents understand the statements as they were intended. Yet, Thurstone and Guttman scales are even more time-consuming in their development. Semantic differential scales are less costly to be developed, yet the format of the scale makes it difficult to include many different aspects of the research topic, because the adjective pairs would become repetitive. If more facets might be used in larger studies (with a range of items), Likert's method seems most feasible. The direct measurement using Likert scales are most feasible and effective in larger survey studies, which include a range of constructs (Schwarz, 2015). In the present study, an attempt was made to use a Likert-type direct method for measuring teachers' attitudes.

Development of a Likert Scale to Measure Attitudes

In the previous paragraphs, it was clear that a direct Likert-type questionnaire was to be developed in order to measure teachers' attitudes towards inclusive education *for all*. Likert (1932) recommended to review other questionnaires in the relevant field and adopt questionnaire items from these established questionnaires. In the words of Likert (1932): "whenever it was possible to use questionnaire material which had previously been extensively tried out [...], we preferred to use the questions exactly as they stood" (Likert, 1932, p. 12). In this way (and under additional consultation of other relevant materials that might inspire the researcher to develop new items), Likert (1932) created a large set of potentially relevant questionnaire items. These items were then administered to respondents, who were asked to indicate their degree of favourableness or unfavourableness on a five-point response scale. Likert (1932) used the numeric data that was gathered in this way to select the most appropriate

items (e.g. those with most of the responses in the neutral middle of the response scale) and to create a scale value for each individual by calculating the sum of the responses.

The recommendations on how to construct a Likert scale have not changed much, compared to the original ideas of Likert (Döring & Bortz, 2016; Schnell et al., 2011). According to Döring and Bortz (2016), the Likert scale construction starts with creating a pool of many items (e.g. about 100), and the refinement of these items (e.g. to ask other researchers to give feedback concerning the quality of the items). Second, the pool of items needs to be administered to a sample of the population, which the scale is supposed to be applicable to. The collected data is analysed in a statistical way with regards to each item's univariate quality, but also with regards to the presence of sufficient bivariate associations of the items. Third, as Döring and Bortz (2016) described, items with non-sufficient quality need to be discarded from further analysis, and the dimensional structure needs to be demonstrated using a factor analytic statistical approach. Finally, the outcome of the scale development needs to be described in detail so that other researchers can judge the scale's quality before using it in their studies (Döring & Bortz, 2016). As an overarching plan for developing a scale, these recommendations were adapted in the present study, because these procedures are described in detail and are widely accepted for reaching high-quality scales.

In principle, these procedures seemed feasible, yet, issues arise if these procedures are carried out in practice. As noted before, textbooks recommend to write a large number of new items (e.g. 100 items), test them statistically, and select only a few items that are the best. Such an approach was utilised recently for example by Saloviita (2015), who developed a new scale to measure the teachers' attitudes towards inclusive education *for some* (children with special educational needs). Saloviita (2015) originally wrote 65 statements, and narrowed down the number of items to 10, using a sample of 168 final-year pre-service subject teachers. In subsequent studies, Saloviita (2015), then, applied this 10-item-version to other samples of pre-service and in-service teachers. In other words, the scale was developed for pre-service teachers, and then, in further steps, applied to pre-service and in-service teachers. Notably, it takes a great amount of time for teachers to complete a questionnaire with a large number of attitude items (which are newly written; hence they vary in terms of quality of wording), plus other items, such as demographics and items for validation purposes. Hence, given the time constraints of in-service (and pre-service) teachers, it seemed not feasible for the present study, which clearly attempted to develop the instrument for pre-service and in-service teachers (and not adapt from one sample to the other), to include such a great number of items.

As cited before, Likert (1932) himself adapted items from established questionnaires (and also other materials such as newspapers, books) that inspired newly formulated statements for

his new questionnaire. This would only be justified, if one assumes that some of the established items, which measured former constructs, can be considered to be sufficient indicators for the new construct in question. In this way, Rost (2004) utilised the notion of the ‘item universe’, to refer to a universe of potential indicators (see for discussion also Shoemaker, 1975). Accordingly, a particular questionnaire, which comprises a set of indicators can be understood as a sample from this item universe (Rost, 2004). In the present study, a new kind of thinking about education was introduced under the umbrella term inclusive education *for all*. As discussed earlier, this term was drawn from the continuation of developments towards education for all and inclusive education. In this way, the concept of an item universe would suggest that inclusive education *for all* might share some appropriate indicators with inclusive education *for some*. In other words, the ‘*for some*’-related item universe might overlap in parts with the ‘*for all*’-related item universe. Although the concept of an item universe (Rost, 2004) seems to be relatively abstract, it gives some justification for trying to find these items, which can be considered as being part of both item universes, and adapt them for the present study, instead of trying to write completely new items. This might mediate both perspectives of the described trade-off: On the one hand, writing completely new items has the potential to align the wording fully with the construct in question without being distracted by former operationalisations, yet, the quality of the wording is to a great extent questionable, because it is considerably difficult to write effective new items. On the other hand, if former operationalisations are used or adapted, the items are to a large extent bound to the former constructs, although the quality of the wordings might be adequate. The mediating position of the present study was to try to find only sufficient indicators from established questionnaires, which might allow to construct an appropriate measure of the new concept inclusive education *for all*, while, at the same time, adapt the wording of the established items, which might allow these items to effectively communicate with the respondents.

Item Construction: Wording

In the methodological literature, high quality survey questions are described as being “clear, concise, and straight-forward” (Likert, 1932, p. 45), as well as “short, sweet, and to the point” (Alwin, 2007, p. 181; summing up the position of Sir Francis Galton). Yet, questionnaire items are also indicators of latent constructs, which necessitates – in addition to being just well-worded – a thorough reflection of their content. The general importance of the item wording for the adequate measurement of a certain construct, was repeatedly emphasised by scholars (Bühner, 2011; L. Cohen et al., 2007). In a recent empirical study, Blasberg, Hewitt, Flett, Sherry, and Chen (2016) showed that slight modifications to the items’ wording of a

perfectionism scale changed its association to other measures considerably. Hence, the wording needs to have the best possible quality.

For investigations across languages, a variety of recommendations is available. In their study on adapting tests for cross-cultural use, Bracken and Barona (1991) point to the costs of questionnaire translation when they recommend that it is imperative to ascertain that the source language version has reached the best possible quality even before the translation and adaptation have begun. If a survey instrument is written in translatable English, a quality translation and adaptation of this instrument into another language is more likely, as several researchers pointed out (Ercikan & Lyons-Thomas, 2013; Geisinger, 2003; Van de Vijver & Leung, 1997). Translatable English, as it was detailed by O. Werner and Campbell (1970) and by Brislin, Lonner, and Thorndike (1973), comprises recommendations with regard to the *grammar*, such as to repeat nouns instead of pronouns, to employ the active and avoid the passive voice, and leave out hypothetical phrasings of the subjunctive tense. In addition, it was recommended that items should be *specific* in the sense of avoiding for example metaphors, colloquialisms and words that indicate vagueness, and that simple sentences should be used to avoid using two different statements within one item (Brislin, 1980; Brislin et al., 1973; O. Werner & Campbell, 1970).

According to these recommendations, it seemed most feasible to try to adapt all items for the present study from other established studies, in order to not start from the beginning, but to start with well-developed and established items that could be developed even further with regards to their wording. Hence, in the present study, items were adapted from other studies (as recommended e.g. by Likert, 1932; Schnell et al., 2011), and their wording was revised according to general item wording recommendations (Bühner, 2011; L. Cohen et al., 2007) and according to recommendations to write translatable items (Brislin, 1980; Brislin et al., 1973; Ercikan & Lyons-Thomas, 2013; Geisinger, 2003; Van de Vijver & Leung, 1997; O. Werner & Campbell, 1970). In addition, the wording was repeatedly examined, as to whether each items' formulation was actually able to be an appropriate indicator of inclusive education *for all*.

Item Construction: Response Formats

The included items were constructed for use with different rating formats. Hence, a common answer format for all items needs to be chosen and the items need to be revisited, and if necessary revised, so that their wording is in line with this common rating format. According to their decision framework for selecting a response scale format, Weijters, Cabooter, and Schillewaert (2010) suggest considering the study's objective as well as the study's population

in order to find an appropriate response scale format (see Table 1). As this present study attempted to develop a new scale and was concerned with pre-service and in-service teachers, who can be considered as an academic population, the framework of Weijters et al. (2010) suggested to use a 5- or a 7-point scale. All five or seven response options should carry a label according to this framework.

It is noticeable that in Table 1 all recommended scale formats have an odd number of response categories, which means that the scale has a middle point that the respondents can choose a neutral point. In their study, Weijters et al. (2010) tested different odd- and even-numbered scales and found out that the mis-response to reversed items was greater when participants responded to scales without a middle category. This result lead the authors to the conclusion that scales should have a midpoint. This finding is supported for example by a literature review conducted by Lietz (2010), who found in several studies that odd-numbered scales with a middle category seem to increase the validity and reliability of scales slightly. Yet, in another review of previous research on this issue, Alwin (2007) sees only marginal support of the statement that odd-numbered scales are better in terms of reliability.

Table 1. Decision framework of Weijters et al. (2010) for selecting a response scale format

Study objective	Study population	
	Academic	General
Scale development	5 or 7 points; fully labelled	5 points; fully labelled
Opinion measurement	5 or 7 points; fully labelled	5 points; fully labelled
Estimation of relations	5 or 7 points; endpoints labelled	5 points; endpoints labelled

Note: This table was drawn from the 'Preliminary decision framework for selecting a response scale format' as it was proposed by Weijters et al. (2010, p. 246). Compared to the original table, this depiction is simplified and it represents only the part of the original table that is important for this study.

Some empirical researchers have criticised scales with a middle category for different reasons. In their revision of the Opinions Relative to Mainstreaming scale (Larrivee & Cook, 1979), Antonak and Larrivee (1995) criticised the original 5-point scale with regards to the middle category being non-informative, hence, they introduced a 6-point scale. Six answer categories were also used by Mahat (2008), who argued that a neutral response option would be used by the respondents for three different reasons such as not comprehending the question, not wishing to participate in the survey or wanting to express agreement and disagreement at the same time. These kinds of respondents' reactions seem to be realistic. Yet, if the midpoint is a melting pot for all three of these scenarios, then the question appears how these individuals would respond without such a midpoint. The middle category may comprise individuals who ticked the midpoint for different reasons, which makes this category difficult to be interpreted. Yet, if the

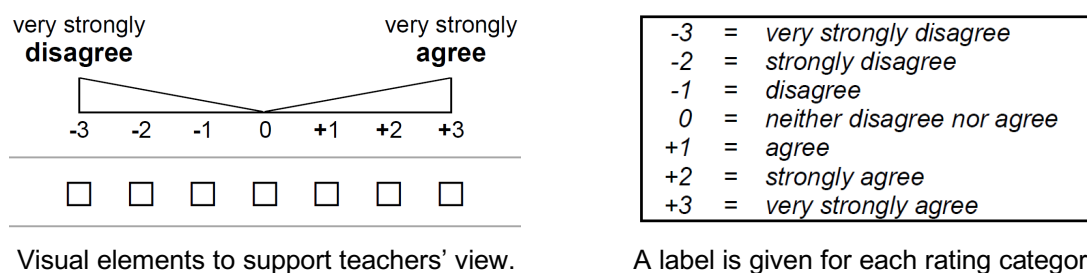
individuals are forced into either the agreement or the disagreement side of the scale, this would mix up the serious responses with the forced choices for different reasons. Hence, the critique of the middle category is justified, but not to include a middle category is not a comprehensive solution to the critique.

With regard to *how many* response categories should be offered to the respondents, there is a range of different evidence-based suggestions. In a simulation study conducted by Lee and Paek (2014), the authors found that a scale should have at least four response options or more. This finding echoes the results of Lozano, García-Cueto, and Muñiz (2008), who found in their simulation that four to seven categories would be ideal. Lozano et al. (2008) emphasised that the discriminative capacity of the respondents needs to be taken into account when making a decision concerning the number of categories. In other words, the respondents need to understand the differences between the response categories. Using data from the European Social Survey, Revilla, Saris, and Krosnick (2014) compared different scales with each other and found that 5-point scales are most valid and reliable. Yet, the European Social Survey is conducted as a standardised telephone interview (see European Social Survey, 2016). Hence, too many response categories that are read out loud by the interviewer on the phone might overwhelm the respondent on the phone. No substantial differences between respondents who answered to different scales were found by Dawes (2008) or by W. P. Jones and Loe (2013). And also Alwin (2007) was sceptic about claims in the direction of positive effects of increased numbers of answer categories. Alwin (2007) pointed out that a perfect question did not exist and that researchers need to try to communicate, what s/he wanted the study participants to respond to, in a way that the respondents' comprehension is maximised. For the Spanish version of the Eysenck Personality Questionnaire, Muñiz, García-Cueto, and Lozano (2005) used a variety of scale formats and found out that in this particular case the 7-category version of the response format was the most favourable. Similarly, Lietz (2010) suggests scale lengths of five to eight being sufficient. While most other studies used measures of the distribution, reliability, validity, etc. as indicators of differences between scale formats, Preston and Colman (2000) gave the study participants different rating scales and asked the respondents, if the scale was easy to use, quick to use, and whether the format allowed them to express their own feelings adequately. According to the judgement of the respondents, Preston and Colman (2000) concluded that seven, nine and ten response categories should be considered most adequate.

The presented literature search with regard to the middle point and the number of response categories did not provide a clear and definitive picture. Some of the arguments that supported the inclusion of a mid-point seemed to be relatively convincing, such as mitigating the mis-response to reversed items (Weijters et al., 2010) or giving an opportunity to express any kind

of discomfort with indicating an agreement or disagreement. Although there were also counterarguments, there seemed to be certain evidence that supported seven response categories, such as the simulation study results of Lozano et al. (2008), and the positive valuation of the respondents themselves, as it was reported by Preston and Colman (2000).

Figure 1. Visual elements that support the understanding of the seven-point scale



Note: The full design of the questionnaire (how these two elements are arranged on the questionnaire pages) can be found in Appendix J and Appendix K.

For the present study, this evidence seemed to justify using a seven-point scale for the newly developed scale (see Figure 1). Seven response categories have also been used in other inclusive education-related studies (such as T. Bennett, Deluca, & Bruns, 1997). As recommended in the framework of Weijters et al. (2010), for scale development, all answer options should be labelled. Yet, seven labels from ‘very strongly disagree’ to ‘very strongly agree’ would be difficult to be included, because not enough space was available on the A4 paper sheet. Hence, in order to support the respondents to interpret each of the answer options correctly, labelled end-points were given, but each box also had a number attached to it, supporting the interpretation of unfavourableness with negative numbers (-3, -2, and -1), positive numbers for favourableness (+1, +2, and +3), and zero for the neutral neither disagree nor agree (as recommended e.g. by Lietz, 2010). In addition, a box with further explanations (fully labelled) is included on each page of the newly developed questionnaire (see Figure 1).

3.4.4 Scope: Cross-Cultural Research

As this study attempted to develop a new instrument for use in cross-cultural studies, the notion of ‘cross-cultural’ needed to be clarified. As already specified, the study’s population was (pre- and in-service) teachers in Australia and Germany; to be exact, in Sydney and in Giessen, respectively. These contexts were thought to be comparable in a way, but also considerably different. Australia and Germany are different countries with different cultural-historical backgrounds and different languages. In this way, both populations are different with regards to what Harkness et al. (2010) coined “3M”, referring to the fact that research on these

populations needs to be considered as being multinational, multicultural and multilingual. Research in contexts like these are generally referred to as cross-cultural research. In the following examination of this issue, some core aspects are reflected as they pertain to cross-cultural research. Previously, it was already specified that in the present study survey methodology was utilised. Hence, to carry out a survey in cross-cultural research is reflected. It will be noted that a questionnaire was needed in two languages, which should be understandable in two cultural settings. As the present study attempted to develop a new questionnaire, it is reflected how a questionnaire for cross-cultural use can be developed. Questionnaire translation will turn out as one important aspect.

Aspects of Conducting Research in Cross-Cultural Settings

According to Geisinger (2003), cross-cultural research refers to investigating individuals across cultural groups *and* across different countries. Van de Vijver (2015a, 2015b) puts the meaning of cross-cultural research slightly more general, as conducting research with persons from different countries *and/or* ethnic groups. A great deal of confusion, what cross-cultural research actually refers to, exists due to differing understandings of the term culture. As Matsumoto (2005) pointed out, cross-cultural researchers commonly studied individuals across different contexts and attributed differences in responses to cultural differences. One major issue with this research practice was that culture was never operationalised (properly) and, hence, was never actually controlled for in such studies; which is why it did not seem justified to assume culture being the cause for certain individual differences in cross-cultural studies (Matsumoto, 2005). Matsumoto (2005) emphasised that differences/similarities in culture and differences/similarities in nationality are not necessarily the same; in addition, differences/similarities in culture and differences/similarities in language are not necessarily the same. Accordingly, cross-cultural research needs to not only be reflected in terms of *culture*, but also in terms of *nationality* and *language*. These three aspects are combined in the previously mentioned notion of “3M” (multinational, multicultural and multilingual; Harkness et al., 2010), which seems to be an adequate embracing conception of cross-cultural research as being multinational, multicultural and multilingual research.

In studies that are conducted across national, cultural, and lingual boundaries the question arises, if all utilised constructs are actually comparable (e.g. understood in the same way) across countries, cultures, and languages. Geisinger (2003) noted that the terms ‘etic’ and ‘emic’ have been established to explicate the scope of research. The *etic perspective* assumes that certain constructs are universally applicable to research studies across different contexts, as Geisinger (2003) sums up this position. This does not only mean that it is assumed that the data collection

can be done in a meaningful and comparable way in different contexts, but also that generalisations are valid across such contexts. The *emic perspective*, on the contrary, emphasises the idiosyncrasies of local contexts, and questions the comparability of different contexts (Geisinger, 2003).

The Etic and Emic Perspective in Research on Inclusive Education for All

The level of the general perspective (global vs. national/local), the conceptualisation of inclusive education *for all* (unified meaning vs. diversified meaning) and the research perspective (etic vs. emic) seem to converge. In the present study, it was reflected in the literature review that inclusive education *for all* is on a global level a truly etic concept. There is clear guidance and little ambiguity how to interpret the most recent policies on the global level (UNESCO, 2005, 2015). Yet, it was also clearly stated in the literature review that idiosyncrasies emerge when global policy is transposed into national and local policies. Hence, these differences of inclusive education *for all* on the national and local levels can only be highlighted as an emic concept that has a particular meaning solely in a particular context.

To make the relation of the understanding and the scope of research explicit in this way, has certain implications for carrying out the present study. If the focus would be on the *national/local understandings* of inclusive education *for all*, an emic stance would guide the empirical research, such as suggesting to ask teachers how they perceive certain (idiosyncratic) aspects of inclusive education (for all), which might be personally and locally relevant for them and their practices. Such an investigation would concentrate on in-depth single contexts. As presented previously, Germany, for example, struggles with merging the special school system with the general school system (UN, 2015). Hence, the teachers' perspective on inclusive education in Germany would very likely be focussed on special students, special schools, and their relation to the general school system (as it is clearly represented in German studies; see for a most recent systematic review Ruberg & Porsch, 2017). Cross-cultural research in this emic sense would be particularly sceptic about the comparability of research across contexts. This does not mean that cross-cultural exchange of research results is considered to be impossible; but this does mean that cross-cultural comparison always starts at the assumption of difference, as it was articulated most vigorously by a range of researchers (Ainscow, 1991; Ainscow et al., 2006; Ainscow et al., 2000; Ainscow & Miles, 2008; Booth, 1995; Booth & Ainscow, 2011; Dyson, 2004, 2014).

On the contrary, if the focus would be on the *global understanding* of inclusive education *for all*, an etic stance would guide the empirical research. In such a case, the assumption that teachers around the world are more or less familiar with a comparable concept of inclusive

education *for all* would guide the empirical research. For research with teachers, this would mean, for example, to ask teachers how they perceive general aspects of inclusive education *for all*, as it is conceptualised globally. Such an investigation could investigate multiple contexts, if two assumptions are made. First, it has to be assumed that such a common concept of inclusive education *for all* actually exists on a global level. In the present study, this was demonstrated in the literature review. Second, it has to be assumed, that the respondents can actually gauge the concepts on the global level. There is evidence, that the global *for all* perspective of inclusive education has been advocated especially in conceptual articles (Nilholm & Göransson, 2017). Most recent textbooks include the *for all* perspective (Carrington & Macarthur, 2012). And there is a range of researchers famously advocating the *for all* perspective; for example Roger Slee in Australia and Andreas Hinz in Germany. Hence, it is likely that teachers were potentially confronted with recent global thinking towards inclusive education *for all* (irrespective of the local demands, and irrespective of the fact that empirical research on teachers' views did not yet pick up on this *for all* perspective). Furthermore, to ask them how they feel about this kind of thinking seems feasible from this point of view. Cross-cultural research in this etic sense would be convinced that similar concepts exist across contexts and that such research starts at the notion of commensurability.

Surveys and Questionnaires in Cross-Cultural Research

Recommendations how to carry out cross-cultural research commonly emphasise that research instruments need to be translated. In most cases such research instruments are questionnaires (for surveying or testing). As Harkness et al. (2010) pointed out, multinational, multicultural and multilingual research would demand so-called 'comparative instruments', which are "deliberately designed for use with multiple populations" (Harkness et al., 2010, p. 34). As it was noted before, to develop such a sound comparative instrument for Australia and Germany is the aim of the present study.

Generally, if a new questionnaire is thought to be developed, which is supposed to be used in more than one context, two strategies can be applied: first, one questionnaire can be developed for one context and can in a second step be adapted for another context; or, second, both questionnaires can be attempted to be developed together for both contexts (Harkness et al., 2010). Although some researchers argued for the latter approach (Erkut, 2010; Erkut, Alarcón, García Coll, Tropp, & Vázquez García, 1999), it needs considerable resources and, hence, Harkness et al. (2010) cautioned that such a parallel or simultaneous strategy would not be feasible in most studies. The former approach seemed more promising and is by far the most common way to develop an instrument for cross-cultural use. Yet, such a strategy would involve

the development of a new instrument for use in one context (as already discussed in the previous section) and the translation and adaptation of the instrument in another language, which is going to be discussed in the following section.

Instrument Translation and Adaptation for Cross-Cultural Research

Against the intuition that the translation of questionnaires for cross-cultural research is easy to be carried out (e.g. by a person who is relatively fluent in both languages), a number of methodological studies emphasise to not only rely on a single non-professional translation. Yet, there is no agreed standard guideline (Maneesriwongul & Dixon, 2004) or ‘gold standard’ (Epstein, Santob, & Guillemina, 2015) on how to carry out the translation and adaptation of a questionnaire. After in-depth examination of the relevant questionnaire translation literature, two translation strategies seemed to be generally recommended. The first strategy was the so-called back-translation, comprising a translation into the target language and another translation back into the source language. The second strategy was to conduct two translations from one source language version into two target language versions. In addition, some of the methodological literature recommended to put considerable effort into comparing the different translated versions and to reconcile them into one high-quality version in the target language. These three aspects are discussed in the following paragraphs.

The translation from the source in the target language, followed by a translation from the target language version back into the source language (so-called back-translation) was emphasised in many studies as an effective procedure to gain evidence for the quality of the translation. Although this procedure was used even in earlier studies, one of the first methodological discussions of this procedure within cross-cultural research was presented by O. Werner and Campbell (1970), who discussed repeated back-translations, systematic comparisons of the different versions by judges and iterative adjustments of both the source and the target language versions (which is also known as ‘decentering’). O. Werner and Campbell (1970) noted that this kind of iterative adjustment procedure has certain power in the sense of making both language versions of a questionnaire more comparable to each other – yet, it has the limitation that both versions tend to become more banal, which might affect the reliability and validity of the items. Hence, in the present study, the decentering approach was not utilised. In a study on translation quality and equivalence, Brislin (1970) recommended back-translation and serious scrutinising of the quality of the translations to ensure equivalence between the different versions. Bracken and Barona (1991) recommend for psychological testing in multiple languages a similar approach, yet, the equivalence of the different versions should be scrutinised in the end by a committee comprising bilingual individuals from multiple national

and/or regional backgrounds. Similar guidelines were advocated by the World Health Organisation (WHO, n.y.), yet, they suggested to convene the committee already after the initial translation to review the translated version even before the translation back into the source language. In another study in the health sector, Da Mota Falcão, Ciconelli, and Ferraz (2003) recommend that one back-translation cycle (translation to target language and then translation back into source language) is enough, if the translator, who has done the translation, and the translator, who has done the back-translation, meet afterwards to clarify divergences and agree upon one target language version. This one reconciled version is finally discussed by a panel of experts, who only need to be competent in the target language, as Da Mota Falcão et al. (2003) point out. After experiences with translating materials for their study, Weeks, Swerissen, and Belfrage (2007) recommended that after the translation and back-translation a pre-test with participants from the target-population should be carried out, before a committee is convened to finalise the instruments. In all these studies, the translation and back-translation was emphasised as an important element of preparing the instruments for multi-language use. The studies also illustrated that the translation and back-translation allowed various modes of comparing and scrutinising the emerging differing versions.

Many studies successfully utilised a different approach, where two different translations into the target language were carried out from one source language version (so-called split-forward translation). As early as the late 1950s, Phillips (1959) initially used back-translation to translate items from English to Thai and back to check for discrepancies. Yet, Phillips (1959) was upset because of the longwinded procedures and disappointing translation results, so he decided to change the strategy. Two independent translators translated the items from English to Thai and then they both met to discuss their results. The finalised and fully agreed target language items were then discussed with Phillips (1959) himself, together with the two translators. In the 1990s, Guillemin, Bombardier, and Beaton (1993) proposed for use in health-related research that at least two separate translations should be carried out, and all these translations should be back-translated separately. A committee then compares all the produced versions. A comparable approach is proposed by the influential International Society for Pharmacoeconomics and Outcomes Research (ISPOR) guidelines (Wild et al., 2005). Yet, Wild et al. (2005) recommend to convene a committee directly after the two separate translations into the target language are available, in order to reconcile the two different versions into one high-quality version. This reconciled target language version is then subjected to back-translation and the different versions are reviewed by another committee, which makes the final adjustments. Recommendations with regard to translation procedures that are even more complex than the ISPOR guidelines exist, such as Sousa and Rojjanasrirat (2011) or P. S. Jones,

Lee, Phillips, Zhang, and Jaceldo (2001). These guidelines share the extensive and repeated use of multiple forward translations from one source language version into two (or more) target language versions, which is then followed by reviewing and reconciling the different versions. All these examples, where different translations added value to the translation quality, suggested not only to rely on one single professional translation into the target language, but also to create a *different* translation into the target language in order to gain creative tensions, which might encourage critical discussions of the adequacy of particular conceptual decisions of the professional translator.

Steps of questionnaire translations are sometimes referred to as techniques, without making mention of how to *review the emerging versions* and how to use the produced versions effectively in a *revision process* that ultimately aims at producing a high quality target language version (see Maneesriwongul & Dixon, 2004). Tyupa (2011) emphasised the reviewing procedures as most critical stages of the translation and adaptation process. Guidance how to perform this comparison was drawn from the literature on cross-cultural research in contexts with more than one language. Weeks et al. (2007) recommended that the translation of instruments should try to maximise the equivalence of meaning and to minimise data contamination (also referred to as bias). With regard to *equivalence*, Herdman, Fox-Rushby, and Badia (1997) found in a literature search 19 different types of equivalence. The so-called ‘conceptual equivalence’ seems by far to be the most common type amongst a number of studies, as reported by Herdman et al. (1997), yet, there seems to be no consensus what conceptual equivalence comprises. In the most general understanding, *equivalence* refers to the comparability of the source language version and the target language version (Tyupa, 2011). Other studies draw attention to the fact that full equivalence is impossible to reach; hence, it should be attempted to *minimize bias* instead, as suggested for example by Eremenco, Cella, and Arnold (2005). Besides general reflections on sources of bias (Eremenco et al., 2005; Hambleton, 2005; Hambleton & Patsula, 1998; Van de Vijver & Leung, 1997), the theory of test translation error (Solano-Flores, Backhoff, & Contreras-Niño, 2009; Solano-Flores, Contreras-Niño, & Backhoff, 2013), allows in-depth insights into why translation errors occur in three dimensions: item design, language, and content. Solano-Flores et al. (2013) demonstrated with PISA-2006 data that test item translations that are flawed with regard to the latter two dimensions (language and content) affect the student performance on these particular items. Solano-Flores et al. (2009) pointed out that translations should never be assumed to be correct, but to be within a range of acceptability. The ‘checklist of possible translation differences or errors’ (Ercikan & Lyons-Thomas, 2013) was also of particular help for carrying out the comparisons in the present study.

The reviewed literature on instrument translation and adaptation suggested both back-translation and split-forward translation as being effective procedures. In addition, in the literature it was recommended to take serious steps to review the different versions and reconcile them, in order to reach for equivalence and avoid bias. Some of the reviewed translation and adaptation procedures involved many different steps. Maneesriwongul and Dixon (2004) concluded that there is not a single perfect translation technique; hence, multiple steps should be taken during translation and adaptation of questionnaires. However, according to a functionalist approach (Colina, Marrone, Ingram, & Sánchez, 2016; Fourie & Feinauer, 2005), there is not one single correct translation of questionnaires. Hence, there seems to be a certain benefit in including more than one translation step (at least in order to be able to use the differences of the translations for discussing the appropriateness of the translation), yet, to use a great number of steps seems to produce a complex array of many options for translation, which might become more difficult to reconcile the more steps are involved. Hence, it was decided, in accordance with the reviewed literature, that a back-translation and a split-forward translation would produce a sufficient range of different versions that can be reconciled through in-depth discussion.

3.4.5 Establishing Quality of the Measurement Instrument

The Standards for Educational and Psychological Testing (in most studies abbreviated as ‘the Standards’; AERA, APA, & NCME, 2014) comprise guidelines for testing and assessment. The Standards were established in the 1950s, and have been revised repeatedly until today. They comprise a full range of general recommendations concerning nearly all aspects of the conduct of quantitative testing. Although survey methodology, as it is carried out in the present study, and test methodology are not identical, the general issues and recommendations as they pertain to both methodologies overlap to some extent (L. Cohen et al., 2007). For example, it is important for surveys and for tests that the measurement is valid, reliable and fair. These aspects are reflected in quite some detail in the Standards; hence, it seems recommended to consider these guidelines. According to the Standards, the quality of the measurement needs to be established with regards to validity, reliability and fairness. In short, *validity* “refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (AERA et al., 2014, p. 11), while *reliability*, in general, refers “to the consistency of scores across replications of a testing procedure” (AERA et al., 2014, p. 33). The issue of *fairness* comprises reflections how to establish a fair measurement for all “subgroups of test takers” (AERA et al., 2014, p. 49). Although there might be some differences between the methodologies as they pertain to surveys and tests, the general recommendations of the

Standards to establish validity, reliability and fairness are valid for surveys, too, as the following discussion will demonstrate.

How to Establish Validity

L. Cohen et al. (2007) emphasised that validity should be established as best as possible, while acknowledging that perfect validity is not possible to be reached. A large range of different kinds of validity, and different kinds of systematisations of validity types have been described in the literature, as summed up by L. Cohen et al. (2007). In line with the Standards, two major steps were considered for establishing validity in the present study. The first step was to provide *evidence regarding the internal structure* of the scale (AERA et al., 2014). More specific, evidence on the dimensionality of the measurement, on the distinctiveness, reliability and interrelationships, and on the degree of confidence in order to discourage over-interpretation was provided. This step involved the univariate and correlative examination of the data, exploratory factor analysis and confirmatory factor analysis (which will be discussed in Section 3.7.2). The second major step of analysis was to provide *content-oriented evidence* for establishing the validity of the scale and *evidence regarding relationships with conceptually related constructs* (AERA et al., 2014). The former was carried out as an interpretation of each of the factors that were established through the factor analytic approaches. The latter involved examining statistical associations of the attitude factors and other inclusive education-related aspects of the teachers (which will be discussed in Section 3.7.3).

How to Establish Reliability

L. Cohen et al. (2007) and other researchers (Bühner, 2011; L. Cohen et al., 2007; Döring & Bortz, 2016; Schnell et al., 2011) noted that the reliability of a measurement can be understood as stability, as equivalence, and as internal consistency. *Stability* would be established, if the instrument is applied to similar respondents over time, and similar data is yielded. There are different ways of carrying out research that is supposed to demonstrate stability of an instrument. Generally, it would be required to repeat the measurement using the same instrument (so-called retest), and through comparing the resulting data of both measurements, it could be demonstrated if the instrument is stable over time. Another method to establish reliability is through *equivalence*, as pointed out by L. Cohen et al. (2007). Reliability as equivalence would mean, to use an instrument and another similar instrument for data collection and compare the gathered data of both. If the instrument in question is reliable, it would be expected to find similar results for the instrument, compared to another equivalent instrument. Finally, the *internal consistency* of an instrument can be understood as an indicator of the

reliability, too, as L. Cohen et al. (2007) noted. In multi-item instruments, the internal consistency refers to how strong each item is correlated with all the other items of an instrument. Measures of the internal consistency can be calculated without repeated measurement (which would be needed for establishing stability) and without including additional instruments (which would be needed for establishing equivalence). In the present study, it did not seem feasible to repeat the measurement; in addition, other instruments that were able to measure the teachers' attitudes towards inclusive education *for all* did not seem to be available (see discussion in Section 2.4). Hence, the internal consistency was examined using Cronbach's Alpha (see Taber, 2017 for a most recent discussion on Alpha as a measure of the internal consistency).

How to Establish Fairness

In the most recent version, the Standards (AERA et al., 2014) recommend with regards to instrument-use in a variety of languages that the methods for translation and adaptation need to be described, especially concerning how *fairness* between the different versions was established with regard to all facets of the testing. The Guidelines for Translating and Adapting Tests of the International Test Commission (ITC, 2010) give 22 recommendations with regard to the context, the translation and adaptation procedure, the administration of the questionnaire and the documentation and interpretation of the results. The ten statements that describe the translation and adaptation procedures within the ITC (2010) guidelines comprise the demand to provide evidence for different aspects of equivalence between both versions and to ascertain that the potential participants are familiar with all aspects of the questionnaire content and format. Several researchers (such as Ercikan & Lyons-Thomas, 2013; Geisinger, 1994; Hambleton & Patsula, 1998) emphasised the Standards and the ITC guidelines as being important general rules for translations and adaptations of questionnaires. Although it was not attempted in the present study to compare any of the obtained attitude scores, the recommendations as they pertained to the quality aspect of fairness were followed throughout all procedures that involved constructing the questionnaire for use in different contexts and throughout all procedures to approaching the participants in these contexts.

3.5 Procedures of Developing the Data Collection Instruments

Previously, the study population and the research design have been discussed. In the following sections, concrete steps to develop the data collection instruments are described. The procedures comprise developing the English questionnaire, comprising to select related constructs/aspects

and their operationalisation. As a second step, the procedures of translating and adapting the questionnaire is presented (see Appendix J and Appendix K for the final versions of the questionnaire in English and in German language, respectively).

3.5.1 Develop the English Questionnaire

Although the literature review demonstrated that there seems to be hardly any teachers' attitude measures that capture inclusive education *for all*, a large body of research seemed to be available in the English-speaking literature on former understandings of teachers' attitudes towards the inclusion of students with SEND. In order to gain more understanding of the available research and in order to learn from the available research experiences, a systematic literature search was carried out. This search resulted in a range of relevant items, which were carefully revised and pre-tested.

Systematic Literature Search

A systematic literature search was carried out in order to find relevant questionnaire items. The literature search was conducted in January 2015. It was supported by the reference management software Citavi (see www.citavi.de/en for further details). The systematic approach started with a search for relevant empirical studies that utilised a quantitative survey methodology for illuminating the teachers' attitudes towards inclusive education. If these studies reported questionnaire items that were used for data collection, these were extracted and critically examined. This process resulted in a selection of questionnaire items that seemed to be formulated in a way so that they could be potentially used for measuring the teachers' attitudes towards inclusive education *for all*.

There seemed to be no databases or compendia available that comprised scales and items to measure teachers' attitudes towards inclusive education. Antonak and Livneh (1988) gathered a variety of hands-on information about the measurement of attitudes toward people with disabilities. More recently, some measurement instruments have been bundled on websites (e.g. the Database for Quality of School in Germany (Germany); or the Measurement Instrument Database for the Social Sciences (Ireland); etc.). Yet, these compendia and websites were area and discipline specific and only provided a small part of information about the existent survey instruments.

Under the assumption that newly developed measurement instruments are usually published in the form of research articles or other text documents (report, thesis, etc.), the search for relevant items was carried out by doing a systematic search for such research texts. Hence, a number of research literature databases were searched for relevant studies that might have

included information about the concrete measurement of teachers' attitudes towards inclusive education.

Six different databases were used for the literature search, namely the Education Resources Information Center (ERIC), Education Research Complete (ERC), Fachinformationsdienst Bildung (FIS Bildung), PsycINFO, PSYINDEX, and Wiley Online Library. This selection covered a variety of different disciplines. The first three databases have their specific focus on education, the fourth and fifth more on psychology, and the sixth database provides more sociological records. Although journal papers might be in the centre of interest of these databases, they also index books, dissertations and reports. There are governmentally funded databases (e.g. ERIC) and databases from for-profit companies (e.g. ERC, which is held by EBSCO Information Services).

In order to find appropriate search parameters, a series of preliminary search trials were carried out. It turned out that to include the search term "measurement" (or variations) excluded relevant records that contained explicit information about the operationalisation of the used constructs. "Integration" and "inclusion" seemed to produce relatively similar results in the preliminary database queries; hence, only "inclusion" was chosen. Yet, some relevant records were tagged with 'mainstreaming', which is an outdated terminology, but was used as a search term, too, in order to not exclude potentially relevant papers. To receive information about scales from a variety of practitioners (e.g. pre-service teachers, early childhood practitioners), the term "teacher" was not used for the search. The final search terms that were used were "attitudes toward/s inclusion/mainstreaming" and the German pendant "Einstellung/en zu Inklusion".

Reducing the Number of Papers to those that Include Potentially Relevant Items

Four iterations were carried out in January, February and March 2015 to narrow down the number of papers found to those that were most relevant for the present study (see depiction of the steps in Table 2). As Table 2 demonstrates, the titles and abstracts of the collected references were examined in a first cycle. Papers with an obvious irrelevant topic were deleted. If the title and/or the abstract indicated that solely qualitative methods or no empirical methods at all were used, the paper was also excluded. Although there were remarkably insightful papers using qualitative approaches, those papers were deleted, because the ultimate goal of this literature search was to find relevant items, which were quantitative studies. In a second iteration, the abstracts and methods sections were examined if detailed information about the methodological procedure was presented. If such information were missing or only presented very briefly, the paper was excluded from further analysis. Third, the methods sections were analysed in more

depth, if the teachers' attitudes towards inclusive education as a construct was operationalised. The different variations of this (e.g. "concerns" instead of "attitudes") were carefully considered with regard to the usefulness of a certain operationalisation within the context of this study. The final iteration was to check if the wording of the used questionnaire items was included in the methods, the results, or the appendix. Studies that failed to present how they measured the attitudes were not examined any further.

Table 2. Exclusion criteria for the review of the papers

Exclusion criteria	Description	Section to examine
1. Not relevant	A paper was deleted, if it had an obviously other topic, if it used qualitative methods or no empirical methods at all.	Title, abstract
2. Methods not described	A paper was deleted, if the methods were not described in detail.	Abstract, methods
3. Other focus	A paper was deleted, if teachers' attitudes towards inclusion was not one of the operationalised constructs.	Methods
4. No wording presented	A paper was deleted, if it did not present the actual questionnaire items.	Methods, results, appendix

Analysing the Items' Content and Select only the Relevant Questionnaire Items

Papers that remained after applying the criteria presented in Table 2, were considered being relevant, having a proper methods section, having operationalised teachers' attitudes towards inclusive education and having presented the wording of the items used for data collection. These papers were examined, if there were at least some items that did not solely refer to the students' SEND. If all items clearly and exclusively utilised a SEND rhetoric, these papers were not examined any further. Those papers that included items without such notions were analysed in more depth. From these papers, all attitude items were extracted. These items were, in the next step, carefully examined for eligibility. For the present study, those items were of specific interest that might serve as indicators for the teachers' attitudes towards inclusive education *for all*. Items were excluded a) if they used explicit special educational needs and/or disability related terminology, b) if they utilised principles that were related to special educational needs and/or disabilities, and c) if they were so broad in their scope that they might not be able to indicate the attitudes of a person at all.

Revision of the Attitude Items

After extraction of relevant items, each items' wording was examined in-depth. As previously discussed, the wording was revised due to recommendations from the methodological literature

and a common rating format was used. Some items were reverse-phased, in order to have both positively and negatively phased statements, which was recommended in the literature for developing a Likert scale (Döring & Bortz, 2016; Likert, 1932; Schnell et al., 2011). This led to the preliminary version of the attitude part of the questionnaire.

Select Further Variables to be Included in the Questionnaire

The attitude part of the questionnaire was developed through the systematic literature review, a thorough selection process, and a careful revision of each of the items. This pool of attitude items corresponded relatively well with the present study's purpose to develop a set of relevant items. This set of items could be used for collecting data and examine the internal structure of the items statistically. Yet, one of the purposes of the present study was to examine the relation of the attitudes to other relevant variables, which was also introduced in the previous section as an aspect of the validity of a new measurement instrument. Accordingly, further items were selected to be included in the present questionnaire. In order to select a set of relevant items in an informed way, the literature review was considered. Notably, previous studies, as they were discussed in the literature review, were focussed on inclusive education *for some*, while the present study attempted to operationalise how teachers viewed inclusive education *for all*. As to differences between the *for some* and the *for all* perspective, it cannot be assumed that the significant associations that were reported previously for teacher aspects (such as gender, age, experiences) and their attitudes towards inclusive education *for some* are necessarily comparable to similar associations with the teachers' attitudes towards inclusive education *for all*. Nevertheless, all these different aspects as they were introduced in the literature review might still be important to consider in the present study. Accordingly, for validation purposes, it would be important to gain insights into the relationship of the new attitude instrument to other teachers' aspects. Hence, the literature review was utilised for an informed selection of potentially other relevant variables that were included in the questionnaire. Section 3.7.3 comprises more elaborations on hypothetical relationships of attitudes and further teachers' aspects.

In the following paragraphs, the additional aspects as they pertained to the teachers are presented in the order as they appeared in the questionnaire (see Appendix J and Appendix K); hence, the self-efficacy items are presented first, followed by the other items concerning the teachers' background. The description is focussed on selecting an appropriate operationalisation for each of the different aspects.

Recently, Sharma and George (2016) pointed out that, if self-efficacy and attitudes are both used in a survey, the self-efficacy should not be too general or too specific. With regards to

studies on inclusive education, Sharma and George (2016) recommended to use the Teacher Efficacy for Inclusive Practices (TEIP; Sharma, Loreman, & Forlin, 2012) scale. The present study investigated a more broad understanding of inclusive education *for all*, for which the TEIP might be especially suitable: Sharma et al. (2012) emphasised that the TEIP tried to avoid notions of ‘specific disabilities’, and that the items tried to grasp the teachers’ ability to include *all learners*, which seemed suitable to the present study. However, five of the original eighteen items explicitly addressed students with disabilities. This was changed for the present study into ‘all students’. Three dimensions were assumed for the TEIP, as it was proposed originally by Sharma et al. (2012), which was also confirmed in many studies (Forlin et al., 2010; Hecht et al., 2016; Malinen et al., 2012; Montgomery & Mirenda, 2014; Sharma et al., 2015; Sharma & Sokal, 2016). The three sub-scales of the TEIP were Efficacy in Managing Behaviour (e.g. “I am confident in my ability to prevent disruptive behaviour in the classroom before it occurs.”), Efficacy in Collaboration (e.g. “I am able to work jointly with other professionals and staff (e.g., aides, other teachers) to teach all students in the classroom”), and Efficacy to use Inclusive Instructions (e.g. “I am confident in designing learning tasks so that the individual needs of all students are accommodated”). In other studies such as in the study by Forlin and Sin (2010) the TEIP was successfully utilised. Forlin and Sin (2010) reported high reliabilities for the total scale score (TSS) (Pre $\alpha=.90$; Post $\alpha=.92$), and also for the subscales Efficacy to use inclusive instructions (Pre $\alpha=.73$; Post $\alpha=.81$), Efficacy in managing behaviour (Pre $\alpha=.90$; Post $\alpha=.90$), and Efficacy in collaboration (Pre $\alpha=.82$; Post $\alpha=.85$). In the German context, the TEIP was utilised for example by Hecht et al. (2016) and they reported some minor issues with regard to the “collaboration” factor, but generally sufficient reliability values: Inclusive instruction (Alpha=.75), managing behaviour (Alpha=.82), and collaboration (Alpha=.76). The TEIP was therefore included in the questionnaire for the present study.

As the literature review revealed, the personal background of the teachers, their professional background and their inclusion-related professional background were found to be associated with their attitudes. These aspects were operationalised according to how these studies measured these aspects. As demographic variables, gender (female; male) and age (up to 30 years; 31-40; 41-50; 51-60; above 60 years) were utilised in the questionnaire. As there were not so many teachers found in both samples to be in the above 60 of age group, this category was merged with the 51-60 years category to above 50 years.

The professional background comprised of information as to whether the teachers were in their pre-service or their in-service phase. This variable was created due to the institutions, where the teachers were approached. Although exceptions might be possible, it was at least very likely that nearly all students in the regular degree programs were actually pre-service

teachers and that nearly all school teachers were actually in-service teachers. Survey questions related to the professional background were, for which level the respondents were trained (primary/elementary; secondary/high school; other). In addition, the teachers were asked how many years of experience they had had, and if they had held a postgraduate degree/diploma (no; yes). Each respondent was writing a number in order to indicate the years of teaching experience. For data analysis, these numbers were categorised into ‘up to five years’, ‘6-11 years’, and ‘12 and more years’. This differentiation was made due to the results that suggested that more experienced teachers held more negative attitudes and less experienced teachers held more positive attitudes (Bhatnagar & Das, 2014; Boyle et al., 2013; Savolainen et al., 2012; Yada & Savolainen, 2017). Yet, the analysis should be also open for findings that suggested that those with some years of teaching experience were the most positive (in the study of Alghazo & Gaad, 2004 those with 6-11 years teaching experience were most positive).

The inclusive education-related background was measured by asking how well their own knowledge of the local legislation and/or policy as it pertains to inclusive education *for all* was. A five-point scale from ‘very good’ to ‘none’ was offered to the respondents to rate their knowledge. The teachers were also asked to rate their amount of inclusive education *for all*-related training on a three-point scale using ‘high (at least 40 hours)’, ‘some’, or ‘none’. Finally, the respondents were asked if they have experienced inclusive classroom settings before or not. If they responded with ‘yes’, they were asked to indicate the quality of past experiences with an inclusive setting using ‘positive’, ‘neutral’ or ‘negative’. For analysis, both variables were combined into ‘no’, ‘yes, positive’, ‘yes, neutral’, and ‘yes, negative’.

Design of the Questionnaire

The previously discussed parts were brought together in a questionnaire. Part 1 incorporated the newly developed attitude items. Part 2 included the self-efficacy items, and Part 3 of the questionnaire was where the personal and professional background items were placed. All items in Part 1 and Part 2 were presented in random order. These and other aspects of the questionnaire design were developed in accordance with the recommendations given by Jonkisz et al. (2012).

Pre-Test the Whole English Questionnaire: Written Comments

Before a newly constructed questionnaire can be used for data collection, it is usually imperative that all items are pre-tested with individuals from the target population. All items were adapted from former empirical studies. Hence, it might be assumed that each single item underwent a serious construction, refinement and actual use for data collection within the

different empirical studies. The items were refined with regard to their wording and to a common rating format. Despite this thorough revision, the items still originate from different contexts (the meaning might differ) and from different years (appropriate wording might have changed). Therefore, the constructed and revised version of the questionnaire was given to different individuals from the Australian target population to receive their written feedback.

Those four teachers who consented to partake in the pre-test received a copy of the questionnaire. The first task was to complete the questionnaire, and, while working through the different statements, to notice if they understood the content of a statement immediately or not. If a statement was difficult to understand, the teachers were asked to indicate how often they had to re-read a particular statement until they finally understood what it meant. Second, the pre-test participants were asked if the grammar and the spelling was correct throughout the questionnaire and if they would suggest any changes to the wording of the statements. In addition, they were told to feel free to make any other recommendations, too, such as with regard to the introductory text, the design, the rating format, etc. All these comments, corrections and recommendations were to be made in written form, so that the teachers had enough time to carefully and intensively work through the whole document and give their feedback.

For analysing the written feedback from all four pre-test participants (pre- and in-service teachers), a first step was to include all hand-written corrections and comments into one single Word file. The number of re-reads were attached to each single statement, the corrections were implemented using track changes, and the different recommendations were included by using the comment function. Then, the statements were examined that had to be re-read by the teachers. If comments were available for these statements, they were used to improve the readability. If no comments were available for the statements that needed to be re-read, the wording was revised to improve clarity. In a last step, all other statements were examined and decisions were made to implement the suggested corrections and improvements. The result of this revision was the finished questionnaire in English language, which was used for data collection within this study (see Section 4.2 for further details as they pertain to the English questionnaire).

3.5.2 Questionnaire Translation and Adaptation in German

The questionnaire translation and adaptation involved the actual translation and adaptation procedures, but also the reconciliation of the emerging versions of the questionnaire. In order to make the questionnaire ready to be used with a broader sample, it was pre-tested, and the procedures of the pre-test are described. All procedures, as they are presented in the following

section were carried out for all items in the questionnaire (namely, items concerning attitudes, self-efficacy and demographics).

Translation and Adaptation

As previously stated, there is no agreed way to carry out the translation and adaptation of a questionnaire (Epstein, Santob, et al., 2015; Maneesriwongul & Dixon, 2004). Hence, the steps that were used and the decisions that were made in the present study were systematically derived from the most relevant (e.g. some of the classical recommendations from the 1970s seemed to be still valid) and most recent methodological literature on questionnaire translation and adaptation, as described previously as part of the research design.

In order to allow the research to be conducted in Australia and Germany, a translation and adaptation of the questionnaire into German language was carried out, in line with the most recent methodological suggestions from the research literature. The translations comprised two different German versions; one was carried out by a professional translator and one by the present author. The professional translation was professionally translated back into English. Several iterations followed, where these versions were systematically compared and scrutinised, until one final German version was reached. The pre-final version was then subjected to the critical written feedback of pre-service and in-service teachers, and in order to collect evidence that the newly developed German statements communicate the original meaning of the English statements effectively, two think-aloud interviews with a pre-service and an in-service teacher were conducted.

The developed English questionnaire was translated into German language and then this version was translated back into English language (this procedure is referred to as ‘back-translation’). As it was argued in the study by Colina et al. (2016), such translations need to be carried out by a professional translator, because only qualified professionals are able to reach a high quality of the translation. Hence, a professional for-profit translation agency in Germany was employed to translate the English version of the questionnaire into German language. This agency was instructed that the German version needed to be ready for usage in a study in Germany. Within the agency the translation was performed by a professional translator who was a German native speaker. A final check for spelling, grammar and adequacy of language was performed within the agency by a professional lector, who was a German native speaker, too. The product of this process was a questionnaire in German language (see ‘professional translation’ column in Appendix M). After the translated questionnaire was returned from the professional translation agency the manuscript was checked for completeness and adequacy. The draft German questionnaire was given to the same agency for translation into English

language. Comparable to the initial translation process, two different professionals were involved. As agreed with the agency, both were not aware that an original English version existed and that their translation from German into English was actually a back-translation. The first person was a native English speaker and specialist for German-to-English translations. The second person was a native speaker in English language, too, specializing in proofreading and copyediting. The product of this process was the back-translated questionnaire in English language (see ‘back-translation’ column in Appendix M).

Parallel to the professional translation and the professional back-translation, another translation from the original English source version into German language was carried out by the present researcher, whose mother tongue is German and who is relatively fluent in English. Colina et al. (2016) pointed out that according to the translation studies literature, professional translators use a conceptual rather than a literal approach, while non-professional bilinguals are generally not able to go beyond a word-by-word translation approach. Although Colina et al. (2016) used this argument to support the necessity of expert involvement in the translation process, it might also be pointed out that the comparison of a professional German translation (see previous step) and a more literal German translation yield a certain potential to understand specific conceptual adaptations that the professional translator decided to carry out. In his translation, the present researcher tried to achieve the best possible quality for each item and to ensure that the meaning of each German item was related to the underlying construct. Comments and memos were written in cases where an appropriate translation was hard to find. The product of this process was another questionnaire in German language (see ‘literal translation’ column in Appendix M).

The need for more than three translations was recommended by researchers such as Sousa and Rojjanasrirat (2011; recommended two split-forward translations, which would involve four translations) or P. S. Jones et al. (2001; recommended two split-forward translations and two translations back into the source-language, which would involve six translations). However, it needs to be considered that each of these translation steps is very expensive (as noted e.g. by Epstein, Osborne, Elsworth, Beaton, & Guillemin, 2015), and, as argued previously, a great amount of additional information is produced through each step. The more information that is produced in this way, the more difficult it will be to reconcile all these versions into one final version of the questionnaire. Hence, for the present study a split-forward translation (one by a professional translator and one by the present researchers) and a back-translation (by a professional translator) were found to create enough creative tensions in order to reach the best possible quality of the translated and adapted version.

Reconciliation of the Translated and Adapted Versions

The four questionnaires comprise the original English version, the professional translation into German, the professional back-translation into English, and the more literal translation of the original version into German, which was carried out parallel to the professional translations (see Appendix M for an overview on the different versions). In other words, there are two versions in German language and two versions in English language. If translations would be rather unambiguous than contingent (which they are certainly not), one would presumably find both German versions and both English versions to be similar to each other. Yet, a certain meaning in one language can be expressed in various ways in another language, and often dilemmas arise from the fact that there are different possibilities to embrace aspects of the intended meaning. There is not one single way to grasp the full original meaning in one expression, as exemplified for example by Lonner (1968) or as for example described by O. Werner and Campbell (1970). In what way both German versions and both English versions are different to each other might elucidate these dilemmas to some extent.

Both German versions were compared and both English versions were compared with each other. This was done in two steps. The first step was a pure word-by-word comparison. Comments were made with regards to each difference, to try to explain why the wording might differ. Within the second step, similar meanings, synonym words and comparable phrases were thought to be found. Those differences that could not be identified as comparable in this regard were of particular interest for further analysis, because they might indicate an improper translation. These comparisons of both English versions and of both German versions were carried out in some detail (see Table 3 for an example).

Table 3. Comparison between both German versions. Item 3 as an example

(A) · Prof. translation	(B) · Literal translation	Comment
Ich <u>fühle</u> mich überfordert,	Es überfordert mich,	(A) expresses a feeling, (B) not.
wenn ich im Unterricht <u>differenzieren</u> <u>muss</u> ,	<u>differenziert</u> zu <u>unterrichten</u> ,	(A) is formulated as being forced to: "if I have to", (B) not. (A)'s verb is 'differentiate' and (B)'s verb is 'teach'.
um den Bedürfnissen <u>all</u> meiner Schüler	um auf <u>alle</u> Bedürfnisse der Schüler/innen	In (A) "all" refers to the students. In (B) "all" refers to the needs.
	<u>in meiner Klasse</u>	(B) adds a "in my class".
<u>gerecht</u> zu werden.	<u>einzuweisen</u> .	(A) seems stronger, (B) more 'respond to'.

In Table 3, this comparison is demonstrated for Item 3 "I get overwhelmed when I have to differentiate to cater for all of the students' needs in my classroom." Both German versions have many similarities and some differences. As can be seen in Table 3, "im Unterricht

„differenzieren“ and „differenziert unterrichten“ have the basic meaning to “differentiate in the classroom”. Yet, both versions are obviously not identical. In the former, the verb is “differenzieren” (Engl. “to differentiate”) and in the latter the verb is “unterrichten” (Engl. “to teach”). Furthermore, the “wenn ich im Unterricht differenzieren muss” (Engl. “when I have to differentiate in the classroom”) has a “muss” (Engl. “have to”) in the professionally translated version, which is not apparent in the other version. For this reason, the former version gave slightly more emphasis to the fact that to differentiate is nothing that the teacher would normally do, but s/he might be forced to do. Another aspect is the position of the “all” or “alle” (Engl. “all”) in both of the German sentences in Table 3. In the professional translation the “all” refers to “my students”, but in the own translation it refers to the “needs”. Furthermore, in the literal translation there is the explicit notion of “in meiner Klasse” (Engl. “in my class”), which is not there in the professional translation. A closer look reveals that the translation agency has added a “meiner” (Engl. “my”) to the students (Ger. “Schüler”). In the literal version this reference to the “own class” is presented separately using “in meiner Klasse” (Engl. “in my class”) which basically has the same meaning but stresses more the class and not the students. A different wording can be found at the end of the sentence in Table 3. “Den Bedürfnissen...gerecht werden” and “auf alle Bedürfnisse...eingehen” are very close to each other (Engl. “cater for...needs”). Yet, the “gerecht werden” implies more or less to be successful with meeting the needs of somebody, while “eingehen” is more or less focussed on the process and the intention of meeting the needs of somebody. One last difference can be found in the beginning of item 3 (see Table 3). The professionally translated version starts with “Ich fühle mich” (Engl. “I feel”) and the other version obviously not. This means that the professional translation frames the item with the feeling to be overwhelmed. In contrast, the literal translation starts the item the notion of being overwhelmed.

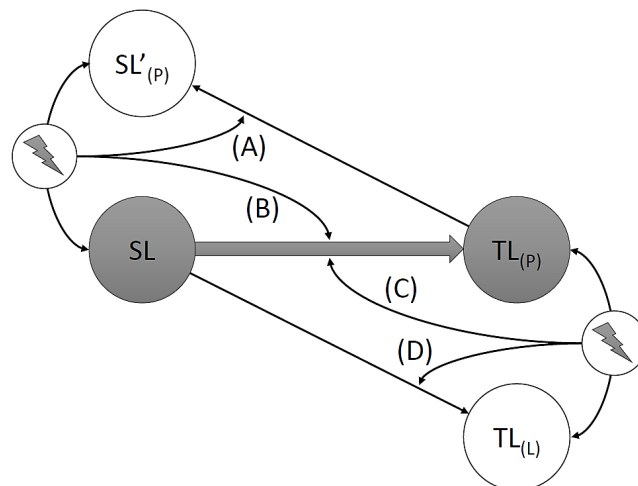
Table 4. Coding scheme for the systematic comparison

Coding Category	Description	Example
identical	This category is used for items or parts of items that have an identical wording.	(<i>identical wording</i>)
synonymous	This category is used for items or parts of items that have a differing wording but a related meaning.	“are able to” vs. “can” or “personnel from outside school” vs. “external staff”
meaning slightly different or other minor issues	This category is used for items or parts of items that have differing wording and slightly different meaning.	“differentiated adjustments can be carried out” vs. “that differentiation is possible”
different meaning or other major issues	This category is used for major differences between the items or parts of items. Also e.g. missing or new words.	“labelling students” vs “grouping pupils”

After the first cycle of English-English and of German-German comparisons, which were carried out in the demonstrated way for each of the items, a coding scheme was developed (see Table 4). The similarities and differences were represented in this scheme from identical, over synonymous and minor differences to major differences. For applying these codes, all statements were divided into units of meaning (as demonstrated in Table 3), and each part of the sentences were analysed if the meaning was comparable or differed to some extent.

Through the comparisons of the English versions and the German versions, some insights into the appropriateness of the translations were gained, yet, the causes of any kinds of discrepancies cannot be detected. For one thing, differences between the professional and the literal translation can be caused by one of the translations being effective and the other not, or even by both translations being error-prone. For another thing, differences between the original and the professionally back-translated version can be caused by a flawed translation into the target language, by a problematic translation back into the source language or even both. Hence, inadequate conclusions might have been drawn, if one had relied solely on the comparison of the English versions or on the comparison of the German versions.

Figure 2. Depiction of the professional translation into the target language ($TL_{(P)}$) from the source language (SL), and the possible sources of discrepancies



Note: SL=source language (English) version; $TL_{(P)}$ =professional translation into the target language (German); $SL'_{(P)}$ =professional back-translation into the source language; $TL_{(L)}$ =literal translation from the original in the target language. Flash symbol=comparison indicates discrepancies.

Both comparisons (professional translation vs. literal translation; original vs. professional back-translation of the professional translation) include the professionally translated German version. Figure 2 depicts this relation; the professional translation (SL to $TL_{(P)}$) is in the centre (see Colina et al., 2016 for arguments, why a professional translation is relatively trustworthy). This version is professionally back-translated ($TL_{(P)}$ to $SL'_{(P)}$), and the comparison of both English

versions (SL vs. SL' _(P)) allows to scrutinise the adequacy of the professional translation (see B in Figure 2), yet, also the adequacy of the professional back-translation (see A in Figure 2) at the same time. On the other hand, the additional literal translation (SL to TL _(L)) allows a comparison of both German translations (TL _(P) vs. TL _(L)). This also allows to scrutinise the professional translation (see C in Figure 2), but only under the premise that the discrepancy could also emanate from an ineffective literal translation (see D in Figure 2). Hence, if for one statement both comparisons indicated minor or major discrepancies (see both flash symbols in Figure 2), then it seems relatively likely that the professional translation might be peculiar and might need additional attention and careful consideration. In this sense, the number of three translations maximises the potential of scrutinising the professional translation (SL to TL _(P)).

For all items, it was noted, if the comparison of the English versions had minor or major discrepancies and if the comparison of the German versions turned out to have minor or major differences. For items with no indication of problematic translations, the professional translation was thought to be adequate. Those items, for which one or both of the comparisons indicated minor or major issues, were subjected to more or less intensive scrutinising. Through an iterative and discursive process all suspected items were carefully revised if necessary.

Pre-Test: Written Comments and Think-Aloud Interviews

The translation and adaptation of the original English questionnaire and the revision cycles led to a German questionnaire version that was thought to have the best possible quality. Yet, all these iterations with regard to the German version did not yet involve any individuals from the German-speaking target population. The literature suggests, no matter how careful a translation was carried out, a pre-test of the translated version with individuals from the target population is imperative (Brislin et al., 1973; Hambleton, 2005). The pre-tests have the aim to scrutinise the questionnaire from the perspective of the individuals from the target population.

Comparable to the English pre-test, German pre-service and in-service teachers were asked to give written feedback on the questionnaire. Yet, while the majority (yet, not all) of the English items were already constructed and tested within English-speaking contexts, there was no evidence available that the German translations communicate the different aspects of inclusive education *for all* effectively to the target population. Hence, additional to the written pre-test, so-called cognitive methods for pre-testing were used, as it was for example emphasised by Collins (2003).

Written feedback was drawn from pre-service and in-service teachers in the German context in a similar way as described before in regards to the Australian context. In addition to the written feedback, an interview approach was chosen to complement the German pre-testing.

This approach is often referred to as think-aloud interviews (also think-aloud protocols, or TAP, or cognitive debriefing, cognitive interviewing, or cognitive pre-testing). The pre-service teacher Robert (pseudonym) and the in-service teacher Mr. Giesser (pseudonym) confirmed their participation. In order to mitigate coercion for the participants, the pre-test interviews were not conducted by the present researcher, but by a student assistant of the research team in Giessen, who had specific training and experiences conducting interviews (especially in the area of the problem-centred methodology; Kielblock & Lange, 2013; Witzel & Reiter, 2012). Both were separately invited to come to the research office at the Justus Liebig University Giessen at an agreed time. After arrival, they were welcomed and were informed about the think-aloud protocol procedure, that the interview will be audio recorded, that all analyses of the audio file will be done anonymously and that the participation is voluntary. Think-aloud interviews can be relatively time-consuming, hence, both participants were solely confronted with the attitude items, not with the whole questionnaire, like the other pre-test participants. The interview with Robert took 49:37 minutes and the interview with Mr. Giesser was 21:52 minutes long.

The first part of the think-aloud interview consisted of the interviewees completing the questionnaire. They were asked to explicate (speak aloud) their thoughts while filling out each question. The interviewer scarcely intervened; only if necessary the interviewer encouraged the interviewee. In addition, the interviewer made written notes whether the interviewee a) had to read an item several times, b) had problems to understand an item and c) presented an unusual interpretation concerning the content of an item. These notes guided the second part of the interview, where the interviewer asked the interviewee to retrospectively elaborate more on the items that seemed to be difficult or to cause misinterpretations.

According to the research literature (e.g. Willis, 2005), the first part is commonly referred to as concurrent (the interviewee speaks about her/his thinking concurrently while working on the items) and the second part as retrospective (the interviewee speaks about her/his thinking retrospectively after completion of the questionnaire). Generally, the pre-test participants were encouraged to verbalize their thoughts and these verbalisations were analysed as being indicators of the ongoing cognitive processes while working through the questionnaire items. Ercikan et al. (2010) pointed out that such verbalisations of thought processes can provide an understanding of the constructs under investigation and can provide evidence of the validity of the (translated) items. The recommendations of the practice of think-aloud protocols (Collins, 2003; Ercikan et al., 2010; Ercikan & Lyons-Thomas, 2013; Lenzer et al., 2015; Willis, 2005) were followed when the think-aloud interviews for the present study were planned and conducted.

The written comments were analysed in a comparable way to the English written comments. All comments were pooled in one file. Through a systematic comparison of these comments, ideas for the improvement of the items were generated. The interview data (audio that was transcribed verbatim) was analysed in a process that involved three stages. First, content analysis was utilised to gain understanding of the thought processes of both interviewees regarding each item. Then each passage of both interviews was coded according to what kinds of wording issues was expressed. Finally, these codes were analysed in-depth with the focus on drawing a picture of which items the interviewee found problematic and what the interviewee would suggest to change. The results from the analysis of the written comments and the results from the analysis of the think-aloud interviews were then used to find reciprocal support (or non-support) for the different critiques. All critique was carefully implemented and the suggested improvements were made. The result of this in-depth analysis of the pre-tests was the final questionnaire in German language, which was used for data collection within this study (see Section 4.3 for further details as they pertain to the German questionnaire).

3.6 Data Collection Procedures

In the following section, the actual sampling procedures are described in more detail. This comprises the procedures for selecting the participants on the one hand and the procedures for data collection on the other hand. The teachers in Sydney were approached in certain units (in other contexts units might be called subjects) at the Macquarie University Sydney and in schools in Sydney, and the teachers in Giessen at the Justus Liebig University Giessen, and at schools in Giessen. Those pre-service teachers were eligible to partake in the study who were enrolled at the time of the study at one of these universities in one of the different programs that lead to an award that allows them to teach at a regular public school as a teacher. With regard to the in-service teachers, those individuals were eligible to partake in the study who were working as an in-service teacher at a regular public primary/elementary or public secondary/high school in Sydney or in the area of Giessen at the time of the study's data collection. These in-service teachers were approached through the school's principal.

3.6.1 Approval of the Empirical Study

All procedures with regard to the data collection and the management and analysis of this empirical data were conformed to the ethical conduct of research involving humans. The procedures of this study were in line with the Australian statement on ethical conduct in human

research (National Health and Medical Research Council, 2015). Ethical issues were intensively examined and approved before conducting this study (see Appendix A). For data collection in schools, the state education research applications process (NSW Department of Education, 2015b) was passed (see Appendix B). The procedures of this study were also in line with the “Satzung der Justus-Liebig-Universität Gießen zur Sicherung guter wissenschaftlicher Praxis” and with the “Promotionsordnung des Fachbereichs Sozial- und Kulturwissenschaften der Justus-Liebig-Universität Gießen” which were the relevant guidelines for the conduct of higher degree research at the Justus Liebig University Giessen. The commissioner for data protection of the Justus Liebig University Giessen examined and approved the procedures (see Appendix C). For collection of data in German schools the commissioner for data protection at the Hesse Ministry of Education and Cultural Affairs (HKM) examined and approved the procedures (see Appendix D).

3.6.2 Drawing the Sample

Due to institutional differences, the sampling procedure was slightly different for pre-service teachers and in-service teachers. Accordingly, in the following section, the sampling is described for Australia and Germany and for the universities and schools.

With regards to selecting Australian pre-service teachers, there were four relevant degrees at the Macquarie University that trained students to become teachers in primary education or in secondary education. These four degree programs were compared in order to find units (subjects) that these four degrees shared. Overall, amongst the eligible units, one unit was randomly picked comprising first-year, one with second-year and one with third-year students. Picking units randomly was carried out with the primary goal to avoid selecting specific unit conveners. *Second*, the procedures for selecting the Australian in-service teachers involved a search for public schools in urban New South Wales (Sydney). The outcome of this search was a number of 211 schools. Over two thirds of these schools were primary schools and nearly one quarter were secondary schools. There is a small number of so-called infant schools (lower primary), environmental education centre and central or community schools, as well. Schools were chosen randomly. *Third*, the German pre-service teachers at the Justus Liebig University Giessen comprised three relevant degrees in general education. One of these prepared the pre-service teachers to work in primary schools. The other two degrees prepared for the secondary sector. In all of these degrees, students had to attend four “Fundamental Science” units (Ger. Grundwissenschaften), such as research in education, politics, psychology and sociology. Generally, pre-service teachers were supposed to attend these units in their first two years of studies; some might also attend them in their third year. One of these units was selected to

approach the students. *Fourth*, the procedures for selecting the German in-service teachers started with collecting a list of all schools in the area of Giessen. The list comprised 70 schools. About three quarters were primary schools and one quarter were secondary schools. A random sample of schools was drawn.

3.6.3 Sampling

According to L. Cohen et al. (2007) sampling can be considered a key feature of survey research. In the present study, the sampling strategy and the sample size needed to be aligned with the primary purpose of the study to develop a robust new measurement instrument.

Sampling Strategy

The aim of the empirical study was not to present a representative picture of the teachers' attitudes, but to find a way to measure the teachers' attitudes towards inclusive education *for all*. This general aim of the present study has an effect on the sampling strategy, which was utilised in the present study. Generally, probability sampling strategies and non-probability sampling strategies are differentiated, as described by L. Cohen et al. (2007). The former type of sampling tries to draw the sample in a way so that it can be interpreted as an adequate representation of the population. In such samples, as Stecher (2005) has argued, it would be allowed to interpret the absolute scores of the measurement (e.g. the obtained attitude scores). The latter type of sampling draws the sample in a way so that it comprises a variety of 'relevant' individuals, yet, probably in a different composition compared to the whole population. In such samples, the absolute scores (such as the obtained attitude scores) are likely to be biased, but the associations of variables (e.g. the relationship of attitudes and self-efficacy) can still be interpreted (Stecher, 2005). A similar argumentation can be found in terms of the (scale) development sample, which was described by DeVellis (2011). He emphasised that it would not necessarily be important that the sample, which is used for scale development, could draw a representative picture of the population, but it would be of crucial importance that the assumption is justified that the relationships among items and dimensions would be similar in the sample and the population.

In the present study, the general strategy was to carry out a relevance-oriented (Stecher, 2005) convenience sampling without any claims of representativeness of the sample. However, random sampling was carried out on the level of the units and the schools, to avoid choosing particular units (e.g. unit of supervisor) or particular schools (e.g. schools that were already collaborating with the university). Yet, the core focus was still to draw a relevant sample, which

was most adequate for developing the new instrument in a way that it can be considered sound and robust across many different individuals.

Sample Size

In order to carry out certain statistical methods, an adequate sample size was needed. Three different considerations needed to be taken into account with regard to the intended analyses. First of all, the sample should allow one to *judge the single item quality*. According to Ellis and Mead (2004), an item analysis (within the so-called classical test theory) can be performed with any kind of sample sizes, yet, a number of 200 cases can, in general, be considered as a sufficient sample. Penfield (2013) also recommends to collect 200 cases, if the goal is to analyse the item quality (e.g. item mean, item-total correlation etc.). Yet, as Penfield (2013) adds, a minimum sample size of $n=100$ might still be acceptable. Second, *exploratory factor analysis* should be possible to be carried out. As Field (2013) sums up recommendations from the literature, a sample of 300 or more cases is generally very likely to be sufficient to conduct an exploratory factor analysis. Yet, the sample size is dependent on the number of variables included in the analysis, their actual factor loadings, and their communalities, as Field (2013) points out. According to De Winter, Dodou, and Wieringa (2009), the exploratory factor analysis demands an absolute minimum of 50 cases, and their study shows that under certain circumstances, an even smaller sample might still be possible to utilise. Third, besides the item analysis, the sample should also allow *structural equation modelling*. As one of the leading scholars in structural equation models, Kline (2011) states that studies using these kinds of methods usually have sample sizes of about 200 cases. Yet, Kline (2011) notes that the adequacy of a certain sample size might to some extent depend on the model complexity (less complex=smaller sample needed), the distribution of the variables (more normal-distributed=smaller sample needed), and the associations between the variables (more linear=smaller sample needed).

Other, more conceptual considerations as they pertain to the sample size were discussed by L. Cohen et al. (2007). On the one hand, L. Cohen et al. (2007) emphasised that in actual research projects, the number of cases is not only dependent on statistical considerations, but “sample size might also be constrained by cost – in terms of time, money, stress, administrative support, the number of researchers, and resources” (p. 102). On the other hand, the sample size depends on the population that the sample is thought to represent and the precision that the study attempts to achieve through the sample. Concerning the former aspect, L. Cohen et al. (2007) noted that there is a contra-intuitive relation of sample size and population size; namely that larger population sizes demand lower proportions of sampled individuals for ideal

representation and vice versa. In this regard, L. Cohen et al. (2007) gives reference to Krejcie and Morgan (1970), who noted that no matter how large the population is, a sample size of about 380 cases might be sufficient to represent the population. In addition, L. Cohen et al. (2007) cautioned that the population-sample-relation needs to be considered for each group that is attempted to be analysed separately (such as Australia and Germany in the present study). Concerning the latter aspect about the sample precision, L. Cohen et al. (2007) emphasised that the confidence levels (such as 95%) and confidence intervals (such as 3%) need to be reflected, too, when determining an adequate sample size. If for example 200 cases were envisaged for each country – as Kline (2011) noted as a common sample size for studies utilising an analytic approach that is similar to the present study's approach –, and if a confidence level of 95% and confidence intervals of three percent were assumed, then, this number of cases would be able to represent 250 individuals of the population (see the table of population-sample-relations in L. Cohen et al., 2007, p. 104). In other words, the opportunities to interpret the obtained attitude scores beyond the boundaries of the present study are relatively low. However, as noted before, the present study attempted not to realise a representative sample, but a relevant selection of individuals (DeVellis, 2011; Stecher, 2005).

A vast range of very different recommendations were reviewed, with regards to an adequate sample size. For the present study, it would be crucial to be able to carry out an exploratory factor analysis; hence, the number of cases per sample should not be less than 50 (De Winter et al., 2009), and in the best case considerably larger (Field, 2013). In addition, more complex analyses (such as structural equation modelling) should be possible to be carried out for Australia and Germany, demanding about 200 cases; depending on the complexity of the model (Kline, 2011). Accordingly, a sample of about 100-200 cases for Australia and 100-200 cases for Germany was thought to be obtained.

3.6.3 Administration of the Questionnaire

The data collection procedures were relatively similar in the Australian and the German cases. *First*, unit convenors and principals were contacted in order to ask for participation of their students or staff. The pre-service teachers were approached through the unit convenors of the selected core units, and, if approved, all students within the selected units were invited to partake in the study. The in-service teachers were approached through the school principals. If a principal consented, all teachers within the school were asked for participation in the study. These responsible persons (unit convenors and principals) were informed about the study's content and the aims of the data collection. The *second* step was to inform the potential study participants concerning the study's content and aims as well as hand out the questionnaire to

them. Once approved by the responsible persons, in most cases, the researcher attended at an agreed time lectures at both universities (concerning the pre-service teachers) and staff meetings at schools (concerning the in-service teachers). The aim was to give relevant information to the potential participants and to inform them about the study and their rights. Then the questionnaire was handed out to the potential study participants. The questionnaire was four A4-sized pages printed on one A3-sized paper which was folded one time in the middle so that the questionnaire looked like a booklet. In the booklet the potential study participants found an information sheet on an extra page. It was printed on an extra page so that the questionnaire could be submitted while the relevant information about the study could be kept. The *third* step, was the submission of the questionnaire of those individuals who decided to partake in the study. Generally, all pre-service and in-service teachers had the option to complete the questionnaire or not. Teachers, who did not want to participate, had several options to remain anonymous. They could simply not submit a questionnaire. Or they could submit a blank questionnaire. They could also submit a questionnaire without valid answers. Questionnaires without any valid answers (in the sense of appropriately using the standardized format) were not considered in any further analysis. In most cases, a locked box was left in the lecture theatre/classes or in at the school's reception for a certain time so that the participants were able to submit their questionnaire whenever they were ready to.

3.7 Quantitative Methods of Analysing Data

After administering the questionnaire to the described sample, and after data coding, data cleansing and recoding the reverse phrased attitude items (as was recommended by Penfield, 2013), the statistical analysis was carried out, supported by the Statistical Package for the Social Sciences (IBM SPSS Statistics 24). Only for confirmatory factor analysis and obtaining the factor scores, R's lavaan package (Rosseel, 2012) was utilised. The methods for data analysis followed to a large extent the recommendations given by Field (2013).

3.7.1 Preliminary Analyses

Before the scale was analysed in more depth, preliminary analyses were carried out. After examining the missing values of each of the items, the adequacy of each item's distribution was analysed with regards to outliers, the central tendency, and the normality of the data. In anticipation that a factor analytic approach was to be used, it was examined to see if some of the items were associated with each other (intercorrelations), and that not too many strong

associations existed (multicollinearity). In addition, the Kaiser-Meyer-Olkin measures and Bartlett's test were used to further scrutinise the eligibility of the items for factor analytic procedures.

Missing Data

Missing values can be crucial problems in data sets (D. A. Bennett, 2001; J. W. Graham, 2012). The individual items were analysed with regards to missing values, and it was examined how great the loss of cases would be, if listwise deletion procedures would be used (e.g. in exploratory factor analysis in SPSS). According to D. A. Bennett (2001), if over 10% of a variable are missing, results are likely to be biased. Hence, missing responses were only further discussed as a particular issue, if they resulted in this amount. In the following section, for all statistical procedures, the number of individuals with valid information is explicitly given, because procedures in SPSS involve listwise deletion. In lavaan (Rosseel, 2012), full information maximum likelihood (FIML) estimation was used.

Outliers

Outliers are a threat to many statistical procedures. Field (2013) noted that "outliers can bias a parameter estimate, but has an even greater influence on the error associated with that estimate" (p. 166). Outliers were spotted using the z-transformed distributions of the items. Extreme outliers ($|z| > 3.29$) were identified and individually examined. Extreme outlying values were set as missing, and it was examined how this affected the central tendency, the skewness and the kurtosis of each of the variables (see the detailed second and third table of Appendix F).

Central Tendencies

Arithmetic means are commonly used to indicate the central tendencies of items, which is referred to in the methodological literature as the item difficulty. Penfield (2013) stated, that the item difficulty for each variable is an indicator for which area on the scales' continuum a variable is particularly informative. Hence, an appropriate set of variables within an instrument comprises variables with central tendencies at different locations across the scales' continuum; yet, extreme item difficulties should be flagged for review, as Penfield (2013) suggested. In other words, a variety of item difficulties are desirable; yet, extreme item difficulties are problematic. Accordingly, the *central tendency* of an item was used as an indicator of the items' quality. The central tendencies of the items were determined by calculating the arithmetic mean. The attitude items were rated by the respondents on a seven-point scale, coded from one to

seven. Hence, any arithmetic mean that was below 2.0 and above 6.0 was considered as being unusually low or high, respectively.

Normality

The statistical procedures applied in this study demand that the assumption of normality is fulfilled. According to Field (2013), normality cannot be tested directly, but within a pragmatic stance, this demand can be translated into the question, if the actual data is normally distributed, which can be tested. To test items in this regard, z-transformed skewness and kurtosis measures need to be calculated, as suggested by Field (2013). Concerning violations of normality, Ferguson and Cox (1993) noted in their recommendations for conducting exploratory factor analysis the trade-off that researchers might want to “retain the maximum breadth of sampled variables while minimizing the possibility of spurious results” (p. 87). Hence, Ferguson and Cox (1993) recommended that some items with critical skewness and kurtosis might be allowed in the set of variables. Using this argument, the recommendation of Field (2013) to use a stricter cut-off value ($|z| > 1.96$; equals $p < .05$) for deciding if skewness and kurtosis is present, was lowered to $|z| > 3.29$ (equals $p < .001$).

Besides the central tendency, the *dispersion* of an item was examined, too. The z-transformed skewness and kurtosis values for each item were examined if it was $|z| > 3.29$.

Discard Items from Further Analysis

Informed decisions were made with regards to these three indicators of unusual distribution: extreme central tendency (arithmetic mean below 2 or above 6), significant skewness and/or significant kurtosis. If at least two indicators in one sample and at the same time at least one indicator in the other sample demonstrated the distribution of this item being peculiar, it was generally discarded from further analysis. All other items were retained (see the detailed fourth table in Appendix F).

Intercorrelations

As stated before, all reverse phrased items were already reverse coded. Hence, all attitude items should be positively related to each other, and at least some of the positive bivariate correlations per item should be substantial. In other words, each of the attitude items should have at least some substantial associations with some of the other attitude items, because it would be difficult to justify that an item that has no associations with any other attitude items would be a proper indicator of attitudes. Besides this content-related consideration, this will also be relevant for statistical reasons concerning the exploratory factor analysis. As it is noted by Field (2013), the

exploratory factor analysis tries to identify clusters of variables that are related to each other more than to other variables. Yet, these clusters of associated variables can only be formed statistically, if each of the items is correlated with at least some of the other items. With regards to this, Field (2013) recommends to visually scan the correlation matrix, search for correlations below .30, and consider excluding items which mostly have such low correlations. In factor analysis, factors should include an absolute minimum of three items with high loadings on a factor (Hogarty, Hines, Kromrey, Ferron, & Mumford, 2005; MacCallum, Widaman, Zhang, & Hong, 1999). If a factor comprises a minimum of three variables, that are all substantially associated with each other, then each of these variables has at least two substantial associations; one to each of the two other variables. Hence, the absolute minimum number of substantial correlations was set to two.

In order to test if enough substantial *intercorrelations* were present in the set of items, all bivariate correlations (Pearson correlation coefficients; listwise deletion) were calculated for the Australian and the German sample separately, and for each attitude variable, the number of correlations below .30 was counted. As discussed earlier, the absolute minimum number of substantial correlations was two, and if an item had a lower number of substantial correlations, it was discarded from further analysis. After deletion of one or more items, the absolute number of substantial correlations per variable may have changed. Accordingly, the calculations of the number of correlations below .30 for each variable and the deletion of variables with a number of only none or one substantial correlations were done iteratively until a correlation matrix was reached that necessitated no further deletions. After carrying out these procedures for the two samples separately, the outcomes were compared. Decisions were made separately for Australia and Germany at this stage of analysis, because it was anticipated that the factorial structure might not be identical in the two samples.

Multicollinearity

While the former step of analysis involved examining correlations that were too low, the contrary – correlations that were too high and correlations with too many other variables – need to be considered as being problematic for exploratory factor analysis, too. According to Field (2013), this so-called *multicollinearity* can cause problems in carrying out a factor analysis, because within a set of highly interrelated items, it is difficult to determine the unique contribution of an item to one particular factor and the unique contribution of another item to another factor. Field (2013) recommended using the so-called determinant as an indicator of the presence of multicollinearity in a correlation matrix. The determinant should be greater than 0.00001, otherwise multicollinearity is present. If multicollinearity was an issue within one or

both samples, according to Field (2013), the so-called variance inflation factor (VIF) can be used for identifying items that potentially make a major contribution to the overall multicollinearity. Higher collinearity of an item is indicated by a higher VIF value. Field (2013) recommended that the VIF should be below 10.

Multicollinearity was examined within the set of items of both samples, using the determinant. If multicollinearity was present, for all variables (thirty-eight, minus those that were discarded from further analysis within the previous steps of analysis) all VIFs were calculated, separately for both samples. The item with the highest VIF was discarded and the determinant was re-calculated and examined if it was above 0.00001. If not, all VIFs were calculated again, and the procedure was repeated until the determinant was sufficient.

Kaiser-Meyer-Olkin Measures and Bartlett's Test

Another correlative measure that is important to be examined before carrying out an exploratory factor analysis, relates the correlations and the partial correlations. In this way, the *Kaiser-Meyer-Olkin (KMO)* measure examines, if patterns are present within the set of items, which can potentially be found by an exploratory factor analysis. The KMO measure for the whole set of items, and the KMO measures for each single item should all be above .50 (Field, 2013). *Bartlett's test* was performed to confirm that the correlation matrix and the identity matrix are significantly different from each other, which is a standard procedure to be carried out before conducting exploratory factor analyses. Bartlett's test should be significant, as Field (2013) emphasised.

3.7.2 Scale Analysis

After carrying out the preliminary analyses, the items were ready for a scale analysis. Analysing the scale started with an exploratory factor analysis for the Australian data and for the German data separately. Then, the results were compared and the common structure was analysed using multiple group confirmatory factor analysis. As a final step of the scale analysis, the final confirmatory factor analytic model was used to obtain the factor scores. These steps were carried out in order to examine the internal structure of the instrument, which was previously discussed in the present study as one aspect of the validity of the measurement.

Exploratory Factor Analysis

The first step was to determine the number of factors that should be extracted. Generally, three methods seemed to be available, as described by Field (2013). The first one is *Cattel's method*, which involved plotting a line graph with the Eigenvalues on the y-axis and the number of

factors on the x-axis (commonly referred to as scree plot). The number of factors to retain can be found by visually spotting the point of inflexion of the graph. As demonstrated by Field (2013), this method is ambiguous, because usually the visual inspection of the graph allows for different interpretations where the point of inflexion is. The second method is *parallel analysis*. This method also uses a scree plot; yet, another line is introduced, representing the 95th percentile of a number of random data sets (O'Connor, 2000). The line of the empirical data and the line of the random data are compared, and the highest number of factors are retained, where the empirical data has still a higher Eigenvalue compared to the random data. Although parallel analysis is considered to lead to most stable factor solutions (Eid, Gollwitzer, & Schmitt, 2011; Field, 2013), this method was most recently criticised in situations where factors were assumed to be correlated, because a too low number of factors was suggested by this method (Braeken & van Assen, 2017). The third method is the so-called *Kaiser's criterion*, which is to assume all factors with Eigenvalues above 1.00 to be substantial. Compared to Cattell's method and parallel analysis, Kaiser's criterion tends to overestimate the number of factors, as emphasised by Field (2013). However, to obtain a more nuanced picture of the factor structure (as opposed to the large factors with many items, as obtained for example by parallel analysis) was found in this particular exploratory study to be an advantage of the Kaiser's criterion. Hence, Kaiser's criterion was used to determine the number of factors.

For the actual extraction of factors, descriptive methods and inferential methods are available (Field, 2013). The latter assume that respondents were selected randomly from the population, and, if this assumption is fulfilled, the extracted factors could be generalised to a larger population. Such extraction methods comprise for example the maximum-likelihood method, as discussed in Field (2013). Amongst the descriptive methods, Field (2013) noted that principal component analysis and principal axis factoring are most commonly utilised. Although both methods obtain similar results in many situations, only principal axis factoring can be considered a factor analytic approach, as discussed in detail by Field (2013). In the present study, the sampling was not drawn randomly from the population, hence, inferential methods for factor extraction were not considered. According to the critique as it pertained to principal component analysis (Field, 2013), *principal axis factoring* was utilised in the present study for factor extraction.

Besides the extraction, a rotation method needed to be chosen. According to Field (2013), two types of rotation can be differentiated; namely, an orthogonal rotation and an oblique rotation. Generally, the main difference between these two kinds of rotation techniques is that the former technique assumes all factors as being unrelated, while the latter technique allows the factors to be correlated (Field, 2013). Within these two groups of rotation techniques,

different methods can be distinguished. Field (2013) clearly recommends to utilise the varimax method, if the factors are assumed being uncorrelated and to utilise the direct oblimin method, if the factors are assumed to correlate with each other. If particular factors would be obtained in the present study, they would all be indicators of aspects of the teachers' attitudes towards inclusive education *for all*; hence, they would supposedly be correlated. Accordingly, an *oblique rotation* was calculated. The direct oblimin procedure was chosen, and delta, which defined how strong correlations between factors were allowed, was set to zero, as suggested by Field (2013).

Ideally, the exploratory factor analysis is carried out and leads directly to an interpretable result. In this way, most of the available information can be used for interpretation. Yet, if the initial result of the exploratory factor analysis was difficult to be interpreted, and if items needed to be discarded in order to reach an interpretable result, the iterations and decisions were indicated in the results section.

The fit of the model was determined by the number of residuals above .05. If this number exceeded 50% of the (non-redundant) residuals, the fit was explicitly indicated as being non-sufficient (Field, 2013).

Generally, the pattern matrix was interpreted and presented. The structure matrix was also examined in order to clarify the relationships of the items. Within the structure matrix, the shared variance is not excluded, hence, it is commonly found to be “a useful double-check” (Field, 2013, p. 702). While the pattern matrices will be included as tables in the text body, the structure matrices can be found in Appendix H. Loadings of the items on the factors above .30 were considered being substantial (Eid et al., 2011).

Cronbach's Alpha can be interpreted as a score for the internal consistency for each of the factors; hence, Field (2013) recommends Cronbach's Alpha values around .70 or .80; yet, he also cautions not to overemphasise specific cut-off values. In a most recent review, Taber (2017) demonstrated and discussed for empirical research practice that Cronbach's Alpha values even below .70 are considered as still sufficient in specific situations. Accordingly, in the present study, a Cronbach's Alpha above .70 is going to be accepted as indicating sufficient internal consistency. Yet, values below .70 are not per se non-sufficient, but they will need further explanation and explicit discussion in the text.

Comparison of the Results, which were Obtained for both Samples Separately

After the exploratory factor analysis was carried out separately for the Australian and the German sample, the results were systematically compared. It was examined, if there were factors which had similar meanings, and which comprised a similar set of items. If such

common factors and items could be found, this would allow postulating a hypothetical global structure of the teachers' attitudes towards inclusive education *for all*. In other words, the aim was to reconcile both empirical results for the Australian and the German sample to develop a data-driven hypothesis about the dimensional organisation of inclusive education *for all*-related attitudes. Common factors and items were examined and discussed.

Confirmatory Factor Analysis

According to the result of their simulation study, Gerbing and Hamilton (1996) argued that exploratory factor analysis can prove to be useful to develop a model that is further analysed using confirmatory factor analysis. The authors add, that surely a model that is developed by using exploratory and confirmatory factor analysis needs to be scrutinised in further studies with new data (Gerbing & Hamilton, 1996). Yet, as Van Prooijen and Van der Kloot (2001) pointed out, a confirmatory study with new data is not likely to confirm the structure obtained from the initial exploratory study, if within the exploratory study a confirmatory factor analysis was not able to confirm the results in the first place. Hence, within exploratory studies, the confirmatory factor analysis might be a useful analytic strategy not to confirm the results from the exploratory factor analysis (which certainly would not be conceivable on the same data), but to further scrutinise, explore and extend them.

There are some major similarities between exploratory and confirmatory factor analysis, yet, there are also major differences. A detailed discussion of the similarities and differences can be found in Brown (2015). For the present investigation, one difference is of specific interest; namely, how cross-loadings are handled by both analytical approaches. While the exploratory factor analysis allows all items to load on all factors, the confirmatory factor analysis fixes cross-loadings to zero by specifying that each of the items loads only on one factor (Brown, 2015). This allows further specifications of, for example, competing (so-called nested) models. In addition, the confirmatory factor analysis provides more specific fit statistics and advanced options for specifying groups within one model; in the present study for example the Australian and the German sample as two groups for which estimates are calculated separately within *one* model.

Appropriate *fit indices* within confirmatory factor analysis is still debated. Early notions of adequate fit in covariance structure models were formulated as rules-of-thumb. For example Bentler and Bonett (1980) noted that from their experience, fit indices of less than .90 indicated that the model might need revisions. More recently, Hu and Bentler (1999) examined in a simulation study a range of fit indices, and they recommended, in accordance to their results, specific cut-off criteria for the different fit statistics (and especially for different combinations

of fit statistics). To sum up some of their core findings, a good model has the following fit values: comparative fit index (CFI), or Tucker-Lewis index (TLI) $>.95$, root mean square error of approximation (RMSEA) $<.06$, standardised root mean square residual (SRMR) $<.08$ (Hu & Bentler, 1999). In his book on scale construction, Furr (2011) cites slightly different benchmark criteria: CFI $>.93$, RMSEA $<.08$, SRMR, $<.08$. Brown (2015) reported that some authors have criticised Hu and Bentler (1999) as “far too conservative for many types of models” (Brown, 2015, p. 75). One aspect of this critique is that to propose static cut-off criteria is not adequate for model fit that is dependent on a variety of aspects, such as the number of factors or the size of the factor loadings.

According to a review on how a variety of studies reported on confirmatory factor analyses, Schreiber, Nora, Stage, Barlow, and King (2006) criticised that researchers tend to just mention numerical fit statistics without discussing how good fit was established. In a similar review study, Jackson, Gillaspay, and Purc-Stephenson (2009) found that only about half of the 194 examined studies stated criteria for the fit indices they reported and only one-third elaborated why they used particular fit measures. According to the data provided by Jackson et al. (2009), many studies seemed to not reach the cut-off scores proposed by Hu and Bentler (1999). For example, for the CFI, the recommendation of Hu and Bentler (1999) was a value higher than $.95$, yet, the mean CFI, which was reported in 78% ($n=152$) of the studies was $.93$ (with a standard deviation of $.06$) (Jackson et al., 2009). Similarly, the mean of the TLI's, which were reported in the number of studies that Jackson et al. (2009) analysed, was below $.95$, too. These findings seem to be in line with the aforementioned criticism of Brown (2015) that the Hu and Bentler (1999) criteria were too strict for applied research. However, cut-off values seemed to have practical relevance, as they supported applied empirical researchers, to try to reach the best models. Yet, they seem to not be absolute in the sense of inevitability to establish the quality of a model. In this way, in their detailed discussion on setting cut-off values for fit indices, Marsh, Hau, and Wen (2004) emphasised “not to overgeneralize” the results by Hu and Bentler (1999). Similarly, with regards to the RMSEA test statistic, Chen, Curran, Bollen, Kirby, and Paxton (2008) argued that “it is not optimal to strive for single-test accept/reject decisions” (Chen et al., 2008, p. 490), and that attention needs to be given to the sources of model misfit and that statistical models need to be combined with human judgement. Van de Vijver and Leung (1997) recommended to use a NFI $>.90$, GPI $>.90$, AGFI $>.90$, RMSEA $<.05$, and a TLI $>.90$. Weiber and Mülhhaus (2010) recommended stricter cut-off values, but also cut-off values, which were less strict: CFI, TLI $>.90$, RMSEA $<.08$, and SRMR $<.10$ (Weiber & Mülhhaus, 2010).

According to the arguments given before, it seemed difficult to make an informed decision. Yet, due to the exploratory character of the present study, less strict cut-off values that were suggested by Weiber and Mülhhaus (2010) were adopted for the present study, to scrutinise the fit of the models.

According to Brown (2015), the classic goodness-of-fit measure is chi-square. Weiber and Mülhhaus (2010) recommended to divide chi-square by the degrees of freedom, and the resulting value should be below three. Hence, the chi-square test is also included as a measure of fit for each of the calculated models and the χ^2/df values were included in the discussion of the models.

For nested model comparisons, chi-square is used to decide which of two models fits better with the empirical data (Brown, 2015). If specified factors were tested against a one-factor model, the differences between the chi-square values and between the degrees of freedom were calculated and the resulting measures were used as indicators if one of the models fits better than the other. Regarding this, the procedures as suggested by C. Werner and Schermelleh-Engel (2010) were followed.

As noted before, lavaan (Rosseel, 2012) package in R was used to calculate the model, using the syntax “standardized=t” to obtain standardised parameter estimates (“Std.all”). The model was specified according to the results obtained from both exploratory analyses for the Australian and the German data and the comparison and reconciliation of both. Using the ‘group=’ syntax in lavaan, the confirmatory factor analysis was estimated for multiple groups; namely for the Australian and the German sample at the same time. Fit statistics were obtained for all cases, and parameter estimates were obtained for the Australian and the German sample separately.

Obtain the Factor Scores

There are three procedures to obtain scale scores for a set of items that can be assumed to be indicative for a latent dimension. The first way is to sum the values of the individual items by using the “sum()” function in SPSS. A disadvantage of this approach is that all cases with missing values have to be deleted, in order to allow for calculating a fair score. An opportunity to retain more information is to use the mean of the items’ values, and to specify that, for example, at least three of four variables should have a valid value (in SPSS, this could be achieved by using the syntax “mean.3()”). Both approaches share the basic assumption that all items contribute the same proportion of information to the scale score. At least empirically, this assumption is in most cases not valid, because factor analyses demonstrate that the loadings of items on one factor are not identical. Most sufficient is the third approach that *draws the scale*

scores directly from the estimated model. In order to obtain the factor scores from the final model, the function “lavPredict()” was used, which printed the estimated values for each of the latent variables of the confirmatory factor analysis. The values on these new variables represented the different attitude dimensions. These variables could be used in the next step as manifest variables to explore the attitude dimensions further.

3.7.3 Validation of the Scale (Hypotheses)

The scale analysis resulted in obtaining factor scores for each of the proposed dimensions of the attitude model. The next steps involved using these factor scores for the attitude dimensions and relate them to other relevant aspects. These analyses were carried out in order to examine how the dimensions of the new instrument were related to other constructs, which was previously discussed in the present study as one aspect of the validity of the measurement. More specific validation hypotheses are specified, before the procedures are described how the empirical data was examined.

Validation Hypotheses

According to the literature review (see Section 2.3.2), certain hypotheses could be formulated in regards to how the teachers’ attitudes might be related to other teacher variables, and, as noted previously, the literature review informed which other variables to include into the questionnaire (see DeVellis, 2011 for discussion of how to select particular validation items). As discussed before, the literature review comprised only studies, which had operationalised the teachers’ attitudes towards inclusive education *for some*. Hence, particular hypotheses as they pertained to the teachers’ attitudes towards inclusive education *for all* could not be derived directly from studies, which solely operationalised a variety of *for some*-related understandings of inclusive education (such as inclusive education for students with SEND). In this way, a validation in the stricter sense of construct validity (L. Cohen et al., 2007; Schnell et al., 2011) was difficult to be established, because the present study’s construction of inclusive education *for all* could not be rooted in the empirical literature on inclusive education *for some*. Nevertheless, the literature review provided at least some heuristic guidance, which teacher aspects might be relevant for explaining the teachers’ attitudes. Moreover, particular validation hypotheses, in what way these aspects were related to the teachers’ attitudes, were developed to resonate with the empirical findings of the previous research. Further arguments were provided, why certain relationships are plausible to be assumed in terms of inclusive education *for all*.

- VH1: Gender is not assumed to have a substantial effect on the attitudes towards inclusive education *for all*.

The literature review demonstrated contradicting evidence concerning the gender of the teachers. While some studies found that female teachers had more positive attitudes, other studies found males being more positive. Generally, there seemed to be no specific reasons why gender might play a role, if inclusive education is understood as being *for all*.

- VH2: Younger teachers hold more positive attitudes towards inclusive education *for all* compared to older teachers.

Concerning the age of the teachers, the evidence presented in the literature was clearer. Although there was not one definition what young(er) meant, and although alternative explanations for the correlative results were discussed (such as experience), the literature suggested relatively unambiguously that younger teachers might hold more positive attitudes compared to their older counterparts. The evidence from the literature was considerably strong, which suggested to assume that this relation might also be found in the present study.

- VH3: Pre-service teachers hold more positive attitudes towards inclusive education *for all* compared to in-service teachers.

Another validation hypothesis was drawn not directly from the literature, but from plausible reflections. Although no direct evidence was found in former research studies, it was assumed that there might have been differences between pre-service and in-service teachers. Because it was also assumed that younger teachers tended to have more positive attitudes (see VH2), the hypothesis was formulated to find pre-service teachers to be more positive compared to their in-service counterparts.

- VH4: Teachers from the primary school sector hold more positive attitudes towards inclusive education *for all* compared to teachers from the secondary school sector.

Comparing the primary and secondary school sector, the evidence, which was presented in the literature review supported the hypothesis that primary teachers were more positive compared to their secondary school counterparts. As to primary schools cater in many cases for a larger variety of students compared to the secondary school sector, it seemed valid to assume to find this relationship in the data of the present study.

- VH5: Teaching experiences (in years) is not assumed to have a substantial effect on the attitudes towards inclusive education *for all*.

The general teaching experience (in years) was found in the literature to be related to the attitudes, too. In some studies, it was found that less experienced teachers tended to have more positive attitudes. Yet, the categories that the studies used varied considerably. In addition, a

study reported that the most positive were not the least experienced, but the middle category of those with six to eleven years of teaching experience (Alghazo & Gaad, 2004). The evidence was not clear enough to form an appropriate validation hypothesis.

- VH6: A higher degree qualification is not assumed to have a substantial effect on the attitudes towards inclusive education *for all*.

Only one study was found which reported a negative effect of attitudes on whether teachers were holding a higher degree qualification. No particular argument was found to support any strong validation hypothesis as it pertained to holding a higher degree qualification and the attitudes towards inclusive education *for all*.

- VH7: Teachers with a higher knowledge of the inclusive education *for all*-related legislation and/or policy hold more positive attitudes compared to teachers with less knowledge.

With regards to the teachers' knowledge of the inclusive education-related legislation and/or policy, the literature review suggested that more knowledge is associated with more positive attitudes.

- VH8: Teachers with more training in inclusive education *for all* hold more positive attitudes towards inclusive education *for all* compared to teachers with less training.

According to findings from previous studies it was assumed that more training in inclusive education was associated with more positive attitudes. Only recently, Woodcock and Hardy (2017) found that receiving training in inclusive education *for some* has detrimental effects on the teachers' attitudes to inclusive education *for all*. Hence, the direction of the association would depend to some extent on how the respondents understood the item in the questionnaire. Yet, the item in the questionnaire was explicitly focused on "training on inclusive education of all students", hence, it was assumed to find a positive relationship between training in inclusive education *for all* and attitudes towards inclusive education *for all*.

- VH9: Teachers having positive experiences with inclusive classroom settings hold more positive attitudes towards inclusive education *for all* compared to those with less positive or no experiences.

A particularly strong evidence base was found for the hypothesis that the teachers having experiences with inclusive settings do have more positive attitudes towards inclusive education *for some*. A similar hypothesis seemed plausible for attitudes towards inclusive education *for all*.

- VH10: Teachers with a higher self-efficacy to carry out inclusive practices hold more positive attitudes towards inclusive education *for all* compared to teachers with weaker self-efficacy.

Finally, the teachers' self-efficacy appeared in the literature review as relevant aspect for their attitudes. A range of empirical evidence suggested that the teachers' self-efficacy to carry out inclusive practices has a positive relationship to the teachers' attitudes, in the sense that stronger self-efficacy beliefs go together with more positive attitudes. A corresponding hypothesis was found to be valid for a *for all*-related understanding of attitudes and self-efficacy.

Preparing the Self-Efficacy Variables for Validation

Before the items of the Teacher Efficacy for Inclusive Practice Scale (TEIP; Sharma et al., 2012) could be used for analyses, their internal structure needed to be statistically examined and the variables needed to be calculated according to the obtained dimensions. For the present study, a multi-group confirmatory factor analysis revealed an insufficient fit (see Appendix E), suggesting that the three dimensions of the original instrument were not well represented by the data of the present study. An informed selection of only the three items (as opposed to six items, as originally proposed by Sharma et al., 2012) with the strongest loadings in both samples revealed in a second multi-group confirmatory factor analysis a sufficient fit (see also Appendix E). Factor values were obtained from this analysis using the "lavPredict()" function in R's lavaan package (Rosseel, 2012).

If three-items per dimension were used (instead of six), a sufficient internal consistency was found for nearly all three-item TEIP dimensions of both samples. For the managing behaviour dimension a Cronbach's alpha of .84 and .83, for the collaboration dimension a Cronbach's alpha of .79 and .73 and for the using inclusive instruction dimension a Cronbach's alpha of .75 and .59 was found, for the Australian and the German sample, respectively. All of these alpha values are above .70, which was generally found to be sufficient; yet, using the inclusive instruction items in the German sample were found not to be closely related to each other. Indeed, in many studies, a Cronbach's alpha of .59 has been called acceptable or sufficient (see literature review and discussion of Taber, 2017); and notably, the Cronbach's alpha of .59 was reached with a number of only three items. Yet, it should be made clear at this point that the internal consistency of the using inclusive instruction items in the German sample was not given. Further studies might clarify possible causes. A first direction to look at would be to investigate if especially the using inclusive instruction items were actually adequate for teachers very early in their careers (e.g. in their first years of initial training). It is a subjective measure, but in the field phase, some pre-service teachers bemoaned that some of the self-efficacy items did not apply to them as pre-service teachers.

Calculations as they Pertained to the Validation

A challenge for the statistical analysis was to estimate the influence of each of the categorical independent variables on the dependent attitude variables. The outcome of the analysis should provide evidence about the influence of the independent variable on each of the attitude dimensions, but also an estimate with regards to the overall influence on all attitude dimensions. It is common to specify so-called path models, if several dependent variables need to be estimated at the same time. Yet, in a path model, the independent categorical variables with more than two values would need to be specified either, as being ordered and being treated as a numeric covariate, or they would need to be specified as a set of dummy variables. Concerning the former, as noted previously, for example the results of Alghazo and Gaad (2004) suggest that the effect of teaching experience in years might not be ordered in a linear way; those teachers in the middle category 5-11 years were more positive compared to the 4 and under years and the more than 11 years group. Hence, concerning the latter, the analysis would need to accommodate each groups' difference, which would mean to specify each value of an ordinal variable with a set of separate dummy coded variables (e.g. one 0/1-variable coding 4 and under years yes/no, one 0/1-variable coding 5-11 years yes/no, etc.). Another critical aspect is that a path model only provides evidence for each of the specified regression paths and the overall fit of the specified model. Yet, it does not directly estimate the overall effect of one independent variable on all dependent variables (although there might be ways to get around this).

The attitude dimensions were tested with regards to their association with conceptually related constructs. A statistical approach that allows multiple continuous dependent variables and a categorical independent variable, that estimates group differences (without assuming the values to be in a specific order), and that estimates effects both for each dependent variable and for all dependent variables together, is the *multivariate analysis of variance* (MANOVA). Hence, MANOVAs were calculated, with the attitude dimensions specified as dependent variables. For each relevant independent variable (e.g. gender, age, pre-/in-service teacher, etc.) a MANOVA was calculated for the Australian sample and for the German sample separately.

Box's test was utilised to examine, if the *covariance matrices are equal*. If Box's test is significant, it is explicitly mentioned, and Field (2013) suggests in such a case to interpret the findings with caution. In MANOVA, it is common to use *Levene's test* to examine the equality of error variances. Yet, Field (2013) noted that this test is only exact in conditions (equal group sizes and large samples) where it is actually not relevant – in other circumstances this test does not work well. Because applied researchers tend to use Levene's test, it was also included for each MANOVA in the present study, and it was reported if significant. Comparable to Box's

test, if Levene's test was significant, the results were interpreted with caution. For both tests, a 0.1 percent significance level was assumed.

With regards to the *multivariate test statistic*, Field (2013) recommended to use Pillai's trace (V) or Roy's largest root (Θ) for different reasons. Pillai's trace was thought to be especially robust for equal sample sizes, while Roy's largest root is powerful in a variety of situations, as Field (2013) noted. Yet, in order to make the results section not too confusing by using many different statistics, Pillai's trace was indicated for all MANOVAs, and if Roy's largest root differed from Pillai's trace, it was explicitly mentioned. Significance levels were five and one percent. For categorical variables with more than two values, *post hoc procedures* were utilised to obtain information about which sub-groups differ significantly from each other. Gabriel's procedure (for slightly different sample sizes) and Hochberg's GT2 procedure (for very different sample sizes) were used, as recommended by Field (2013). In order not to overemphasise significance testing, the *95% confidence intervals* were calculated and interpreted, too.

All of these analyses were carried out in SPSS using the general linear model (GLM) procedures. Specifying the relevant independent variables as fixed factors, makes SPSS to return a MANOVA. Similarly, for the three self-efficacy items, which were continuous as described before, the GLM was utilised, too. Each of the three self-efficacy dimensions were specified as a covariate; hence, SPSS returns a multivariate (multiple) linear regression analysis. The multivariate test statistic is similarly interpreted, as described before. Yet, instead of means and confidence intervals of means, the regression slopes and intercepts are presented, with 95% confidence intervals, respectively.

3.8 Chapter Summary

This chapter discussed the utilised methodology of the present study. In addition to the general reflections, informed decisions were presented as to the concrete procedures that were carried out in performing the research steps of the study. Concrete purposes of the study were identified, and the study population and the research design were specified accordingly. The purpose to find a way to measure the teachers' attitudes towards inclusive education *for all* was pursued through a systematic literature review of relevant indicators. Using the concept of an item universe (Rost, 2004), it was argued that this strategy to find relevant indicators in previous studies to construct a new instrument was feasible, although not any sufficient instrument that measured a *for all*-related understanding of inclusive education was found in the literature. The

purpose that this instrument was supposed to be sound and robust was established through carefully selecting very different (but still comparable) contexts for the study, and through planning rigorous procedures and through research-informed methodological decisions. Many of the procedures (such as the item revision, the translation, the pre-testing, etc.) were informed by designing the instrument particularly for multinational, multicultural, and multilingual use (Harkness et al., 2010). The purpose to establish certain dimensions was represented by the factor analytic approaches, as previously described. The outcomes as they pertain to the described procedures are presented in the next chapter.

Chapter 4 · Results

4.1 Introduction

The previously described procedures led to the Attitude Measure of Inclusive Education for All. The results, as they pertain to the different steps of the scale construction are reported in this chapter; namely the outcome of the systematic review, the outcome of the translation and adaptation, the characteristics of the sample, the outcome of analysing the internal structure of the scale and the outcome of the validation. Researchers (such as Antonak & Larrivee, 1995; Mahat, 2008; Triandis, 1971) have cautioned that many attitude scales had not been carefully constructed and/or have not reported sufficiently the construction process and the psychometric properties. Hence, all the results of the scale construction process are described in detail, to document the scales' development and to make the procedures and decisions as transparent as possible (as it is claimed e.g. by Erten & Savage, 2012, in order to move inclusive education research forward).

4.2 The English Attitude Questionnaire

As described in the Methodology Chapter (see Section 3.5.1), the English version of the attitude questionnaire was developed in three major steps. First, a systematic literature review resulted in a number of relevant questionnaire statements. Second, these items were revised and, third, they were pre-tested. Accordingly, all outcomes as they pertain to these steps are presented in the following three sections.

4.2.1 Result of the Systematic Literature Review: Relevant Questionnaire Items

The steps of the systematic literature review were carried out within three months from January to March 2015. Hence, the latest papers that had a chance to be included in the body of literature for the item and scale development were papers published before 2015. As described previously in the methodology section, six different databases and six variations of search terms were used, which resulted in 2,679 records.

Studies that Comprised Relevant Items

In order to narrow down the number of records to those that contain relevant items, five steps of analysis were carried out, as depicted in Table 5. Within the first two steps the records were examined, which allowed the researcher to narrow down the number considerably. In the *first* step (see A in Table 5), 2,209 records were excluded. The careful examination of all records uncovered some doublets that emerged due to discrepancies between the different databases with regard to spelling (e.g. ‘&’ instead of ‘and’) and with regard to the number of the specified authors (e.g. only first author specified vs. all authors included). In addition, the examination of the titles and abstracts revealed irrelevant reports (such as conference invitations, conference proceedings, etc.) and papers with other topics (e.g. ‘*attitudes* of biomedical and clinician scientists *toward* the *inclusion* of social scientists into the field of health research’). As stated in Section 3.5.1, those papers that utilised no quantitative research methods were excluded from further examination, too. After this initial step, 470 records retained. The range of years covered records from the late 1960s to 2014. The *second* step (see B in Table 5) resulted in the exclusion of 203 records. These records had no sufficient information included in regards to how they selected or constructed the utilised items or scales. The records retained for further examination resulted in 267.

Table 5. Result of the search for relevant questionnaire items using a systematic approach

Steps of screening the 2,679 records for eligibility that were found through the systematic review	Excluded records	Retained records
(A) Is the record relevant?	2,209	470
(B) Are the methods described?	203	267
(C) Is the focus/content relevant?	191	76
(D) Is the items' wording presented?	39	37
Examining of the items' wording within the 37 papers for eligibility of the papers	Excluded papers	Retained papers
(E) Are items free from special needs/disability logic?	23	14
Steps of screening the 273 items for eligibility that were found within the 14 papers	Excluded items	Retained items
(F) Direct use of special needs, disability, etc. notions?	194	79
(G) Indirect use of special needs logic?	11	68
(H) Too broad to capture attitudes at all?	22	46

A number of 191 records were excluded within the *third* step (see C in Table 5). The methods sections of the papers were examined, to see if the teachers' attitudes towards inclusive education (in the broadest understanding) were operationalised. Yet, some of the scales did not focus on pedagogic practitioners in general, but for example parents (e.g. the PATCH scale; Rosenbaum, Armstrong, & King, 1987) or school-age children (e.g. the CATCH scale; Rosenbaum, Armstrong, & King, 1986). Some of the scales did not operationalise attitudes, but

awareness (e.g. the M-GUDS scale; Miville et al., 1999), barriers (e.g. the BSECI scale; Buysse, Wesley, & Keyes, 1998), comforts (e.g. the TCC scale; Huang & Diamond, 2009), concerns (e.g. the CIE scale; Sharma & Desai, 2002), expectations (e.g. the PES scale; Kuyini & Desai, 2007), knowledge (e.g. the IKT scale; Sucuoğlu et al., 2014), prejudices (e.g. the PCSN scale; Unianua, 2012), and supports (e.g. the PSSIE scale; Ahmmed et al., 2012). In addition, there were scales that explicitly did not measure views on inclusive education, but the views on children with disabilities (e.g. the ATCD scale; Marom, Cohen, & Naon, 2007), on disability/special needs (e.g. the ATDS scale; Cowen, Rockway, Bobrove, & Stevenson, 1967), or for example on ‘disabled persons’ (e.g. the IDR scale; Favazza & Odom, 1997). After deletion of these records, 76 records retained.

The next two steps allowed the researcher to narrow down the number of records again significantly from 76 to 14, which was the final number of relevant studies for further in-depth analysis. Within the *fourth* step (see D in Table 5), 39 records were excluded due to the fact that no item wording was presented. The retaining 37 papers, were then, in a *fifth* step (see E in Table 5) examined with regard to the items’ wording. If an item contained explicit notions of special educational needs, disabilities, or if it pointed solely to specific groups of individuals using labels such as special educational needs, disabilities, but also to gender, ethnicity etc., these items were not considered for use within the present study. Papers, which obviously and explicitly made sole use of those labels, were excluded, because they would only be indicators of the attitudes towards inclusive education *for some*, such as students with special education needs and/or disabilities (SEND). Yet, as elaborated in the first part of this study, items were searched for, that could be used as indicators of the teachers’ attitudes towards inclusive education *for all*. Fourteen papers retained after this step. These papers contained at least some items, which did not explicitly and solely addressed individuals with SEND, or use similar labels (see overview in Table 6).

The 14 relevant studies represent a variety of contexts, such as Asia (Ahmad, 2012; Al Zyoudi, Al Sartwai, & Dodin, 2011), Australia (Mahat, 2008), Europe (Beacham & Rouse, 2012; Bosse & Spörer, 2014; Moberg, 1997; Vanderfaeillie, De Fever, & Lombaerts, 2003), and North America (Andrews & Clementson, 1997; Barnett & Monda-Amaya, 1998; T. Bennett et al., 1997; Horne, Timmons, & Adamowycz, 2008; Hsieh & Hsieh, 2012; Stoiber et al., 1998; Taylor & Ringlaben, 2012). Some of the included items were also tested for example in an African sample (Moberg, 2003). This cultural diversity of the original study contexts was desired, because the set of items was supposed to be eligible to be used in cross-cultural settings. These 14 papers cover a period from 1997 to 2014. Generally, they represent both perspectives of pre-service teachers and in-service teachers, and some of the studies have sampled students

of education in general (Vanderfaeillie et al., 2003), early childhood practitioners (Stoiber et al., 1998), and principals (Barnett & Monda-Amaya, 1998). The sample sizes used in the studies range from 20 (Horne et al., 2008) up to 518 (Ahmad, 2012). The number of items range from four (T. Bennett et al., 1997) to 37 (Vanderfaeillie et al., 2003). Nine of the 14 papers present scales and five present batteries of single statements concerning the attitudes. Notably, the papers vary greatly with regard to how they report statistical analyses, statistical measures, and the psychometric properties.

Table 6. Overview on the 14 Papers that Resulted out of the Literature Review

First author	Year	Sample	Sample size	No. of items	Response format	Final cut
Ahmad	2012	In-service teachers et al.	518 all ^a	33	5-point	✓
Al Zyoudi et al.	2011	Pre-service teachers	300	20	5-point	✓
Andrews et al.	1997	Pre-service teachers	67	9	5-point	✓
Barnett et al.	1998	Principals	65	6	4-point	✓
Beacham et al.	2012	Pre-service teachers	216	15	5-point	✓
Bennett et al.	1997	In-service teachers et al.	84	4	7-point	✓
Bosse et al.	2014	Pre-service teachers	241	12	4-point	✓
Horne et al.	2008	In-service teachers	20	22	4-point	✓
Hsieh et al.	2012	In-service teachers	130	19	4-point	✓
Mahat	2008	In-service teachers	115	18	6-point	✓
Moberg	1997	Pre-service teachers	125	20	6-point	✓
Stoiber et al.	1998	Early child. teachers et al.	128	28	5-point	x
Taylor et al.	2012	Pre-service teachers	295/190 ^b	30	5-point	✓
Vanderfaeillie et al.	2003	Students of Edu.	150	37	5-point	x

Note: 'et al.' in the sample column refers to other groups that were also examined in these studies. The last column on the right side show, if indicators from each of the 14 extracted studies were used in the questionnaire.

^a 432 in-service teachers, 108 administrators and 50 policy makers (Ahmed, 2012, p. 67).

^b In the study by Taylor and Ringlaben (2012), 295 pre-test responses were followed by 190 post-test responses of the same course. "It is assumed this was due to students withdrawing from the course after the pre-test is given or students were absent the day the post-test was administered" (Taylor & Ringlaben, 2012, p. 19).

Extracting Relevant Items

All attitude items from the 14 papers were extracted, which resulted in 273 attitude items. An in-depth analysis of these items was carried out, in order to extract only those items from the papers that could be utilised in the present study to measure teachers' attitudes towards inclusive education *for all*.

The *first* step of the analysis of the items (see F in Table 5), 194 items were excluded, because they used explicit notions of 'special educational needs' and 'disability'. These items comprised notions of integration (e.g. "I feel children with special needs have the right to receive their education in the same classroom as typically developing children"; Hsieh & Hsieh, 2012), problems that might arise because of integrating a child with special educational needs

and/or disabilities (e.g. “Inclusion of most special needs students can cause more problems than it can solve”; Horne et al., 2008), and benefits of integration (e.g. “Inclusion benefits the children in the class who do not have disabilities”; T. Bennett et al., 1997). Seventy-nine items retained for further analysis. *Second* (see G in Table 5), 11 items were excluded due to their wording, which pointed implicitly towards a more integration-oriented understanding of inclusion. This pertains for example to items that suggested that there is a separate responsibility of special teachers for ‘their special students’ (e.g. “Special education staff in this school want their students to be fully included in the regular classroom”; Horne et al., 2008), or that it could be legitimate to think about separate placement options other than the regular one (e.g. “Some children are better educated outside mainstream schools”; Beacham & Rouse, 2012). After exclusion, 68 items remained. A further 22 items were excluded within the *third* step (see H in Table 5). These items were so broadly formulated that they might not be able to indicate the inclusion-related attitudes of a person at all. This pertains to items, such as “People like to be with others with whom they share common characteristics and concerns” (Moberg, 1997), or for example “There is a gap between theory and practice” (Al Zyoudi et al., 2011). After exclusion of these items, the final number of relevant items was 46.

Taken together, the 46 items represent a considerable variability of possible *themes* with regard to inclusive education *for all*. The range of topics start at the very basic notion of inclusion and equality being a right of humankind in general. Other items also touch upon why inclusive education *for all* can be considered desirable; namely, that inclusive education leads to valuable experiences, permits academic and social progression for all, and leads to social inclusion and an inclusive society, as some items state. Besides some critical views, why inclusion was fostered in recent years (such as inclusion might be a way to lower the expenses for the educational sector), the included items also touch on how straight-forward (e.g. without asking too many questions) more inclusive practices should be implemented. Adaptations and adjustments (e.g. with regard to the assessment) are needed for being more flexible and learner-centred, according to some items. And finally, some items refer to supports that are necessary for carrying out more inclusive practices, such as additional personnel. This variety of topics, as it is represented by the items, seemed to cover a large range of relevant aspects of inclusive education *for all* (in the terminology of Rost, 2004 the ‘item sample’ seemed to be a sufficient portrayal of the ‘item universe’).

4.2.2 Item Revisions and Response Format

The items were revised through several iterations. Each revision attempted to improve the wording. Yet, for the revision, the final response format, also played a role. Through this

iterative revision, the number of items was reduced to 38 items (see Appendix L for a comparison of the original items and the revised items).

Within several item revision iterations, general and specific wording issues were solved, and an attempt was made to revise the items to be as *clear* and *readable* as possible. For example “I think it is impossible to try and accommodate too many differences in one classroom” (orig. item from Al Zyoudi et al., 2011) was slightly revised to read “It is too difficult to accommodate all students’ differences in an inclusive classroom”. In addition, some items were made *more specific*, such as for example “Teachers should be responsible for the learning of all children in the classes they teach” (orig. item from Beacham & Rouse, 2012). Here, the statement “should be responsible for the learning...” was changed into “are able to meet the needs...” to give more emphasis to the challenge of differentiation within classes. Also “teachers” was specified as being “effective teachers”, which resulted in “Effective teachers are able to meet the needs of all children in the classes they teach”. Some of the items were made *broadier* in their scope. For example, the “Department of Education” and the “School Board” were mixed into “Education Department/Board” in order to make the items compatible with the broadest variety of different educational contexts. Some items were revised due to the use of ‘*inclusive*’ instead of ‘regular’ and ‘general’ education. For example “All students will receive appropriate education [...] in regular education” (orig. item from Moberg, 1997) was revised into “[...] in inclusive education”. There were items, that were, on the one hand, thought to be more precise *without* a notion of “*I believe...*” or “*I feel...*”, such as “I believe that inclusion facilitates socially appropriate behaviour [...]” (orig. item from Mahat, 2008). On the other hand, some items were felt to require a notion of “*I feel*”, or “*I feel from my experience*”; for example, if the item included relatively abstract stakeholders, like the Education Department. The reference to the own experience was thought to give the respondent space to think of a variety of different experiences (e.g. from the own school time as a student etc.). Hence, such a notion of “I feel from my experience” was added to items such as “There is support for inclusion from the Department of Education” (orig. item from Horne et al., 2008).

Another result was that some items comprised *multiple statements within an item*, which were disentangled into two or several items with single statements each. If a statement is actually a combination of several statements, it would make it unnecessarily complicated to read and answer it. In addition, the answers might be difficult to interpret. An example is the statement “Inclusion offers mixed group interaction that will foster understanding and acceptance of differences” (Taylor & Ringlaben, 2012), which was drawn from the Opinions Relative to Integration scale (ORI; Antonak & Larrivee, 1995). The statement contains two outcomes of inclusion: understanding of differences on the one hand, and the acceptance of

differences on the other. Both are quite different, and so this item was split into two statements. The statements resulted in “Inclusion will foster acceptance of differences among students”, and “Inclusion will foster understanding of differences among students”. The same was done with “separating and labelling” (in an orig. item from Andrews & Clementson, 1997) or for example “resources and personnel” (in an orig. item from Moberg, 1997); these notions were each split into different statements.

Some items were intentionally reverse-phased, in order to have both *positively and negatively phrased statements*. Because most of the statements that were extracted from the literature were positively phased, some negative statements were constructed by the present researcher, as it was recommended when attempting to develop a Likert scale. Those negative statements cover a range of topics, for example a negative view on differentiation (such as “I get overwhelmed when I have to differentiate to cater for all of the students’ needs in my classroom”), on support (such as “I feel that external support services are a waste of time”), on the practicability of inclusion (such as “The philosophy of inclusion cannot be implemented in ‘real world’ practices”), or on parents (such as “Parents hinder the successful implementation of inclusive education”). As can be seen in Appendix L, these negatively phrased items were newly developed; yet, their content was based on other items.

There were also items that needed no corrections, such as “Education is a right that should be available to all children” (orig. item from Beacham & Rouse, 2012), or “I am willing to adapt the assessment of individual students in order for inclusive education to take place” (orig. item from Mahat, 2008). These items were used for the pre-test in the formulation, as they were drawn from the original studies. A number of items were *deleted* after serious consideration. A reason for deleting an item was that it was not clear enough, if an item actually represented an attitude (e.g. “I am familiar with inclusion”; Andrews & Clementson, 1997). Another reason for discarding an item was that it might refer to situations that might not be relevant for all respondents (e.g. “I prefer an inclusive school for my own child”; Vanderfaeillie et al., 2003). Moreover, items were discarded, if two or more items had a relatively similar meaning. For example the item “I feel inclusion is a good idea” (Hsieh & Hsieh, 2012) and “I feel that inclusive education is a practical idea in my country”, which was adapted from Ahmad (2012), have a similar structure and a comparable meaning. Although good and practical might not be identical, both items ask (without pointing to any concrete aspects) the respondent, if s/he generally thinks about inclusion in a more favourable or unfavourable way. Items that were similar to each other in this way, were merged, so that the respondents did not have the feeling that the particular items appear repeatedly in the questionnaire.

4.2.3 Outcome of the Written-Comments Pre-Test

The target population for the English questionnaire in this study comprised pre-service teachers and in-service teachers in Australia as discussed in Section 3.3 and 3.6. Accordingly, two pre-service teachers and two in-service teachers from this population participated in a pre-test. The first participant was a female pre-service teacher who was 30 and under years old. She stated in the questionnaire that her knowledge with regard to inclusion was good and that she received some training in this direction but had no experiences with inclusive settings. The second participant was female and 30 and under years, too. Her inclusive education knowledge was poor although she had received some training and had some experiences with inclusive settings. The third pre-test participant was a female, 30 and under years old in-service teacher with good knowledge concerning inclusion and some training in this area, and she had experiences with inclusive settings. All these participants were trained to teach in the secondary sector. The fourth teacher was trained to teach in the primary sector. He was a male in-service teacher and 30 and under years old. His knowledge as it pertains to inclusion was average although he reported having received some inclusion-related training and having experiences with inclusive settings. The convenience sampling strategy included those teachers who were willing to partake. They were not systematically chosen or picked. For example, there is not a range of age groups represented; both pre-service and both in-service teachers were relatively young (30 and under years). Yet, on the other hand, the pre-test participants represented, for example, a variety of different knowledge levels as it pertained to inclusive education.

The written feedback from pre-service and in-service teachers resulted in some revisions. Generally, the teachers did not indicate many issues, which might suggest that the quality of the questions and the design of the questionnaire already seemed relatively acceptable. There were two items that more than one teacher found difficult to answer. One item was “All children are capable to learn in inclusive settings”, which had to be read twice by a pre-service. An in-service teacher made the suggestion to change the “to learn” to “of learning”. As this might make the item more reader-friendly, it was changed in this way. The other item was “I feel all differentiated adjustments in an inclusive classroom can be done”, which also had several comments. One pre-service teacher and one in-service teacher had to read it twice. An in-service teacher indicated a problem with the term “all”. The other in-service teacher suggested to add “planned” after the “all” and to change the “can be done”. After serious consideration, the item was reworded into “I feel differentiated adjustments can be carried out in an inclusive classroom”. An item with only one person commenting on it was “Inclusion facilitates socially appropriate behaviour amongst all students” that was read twice by an in-service teacher, and who commented that it was not clear if “amongst all students” actually referred to “all cases, or

in general”. Hence, the wording was changed into “for all students”. Two items were corrected due to spelling.

4.2.4 Final Attitude Items in English Language

The aim of the systematic literature search and the selection, revision and pre-testing of specific items was to find a set of items which could be used to operationalise the teachers’ attitudes towards inclusive education *for all*. Thirty-eight statements were extracted through these procedures. From 14 relevant studies, 12 studies contributed ideas to the final set of items (see Table 6). No items in the final form were adapted from Stoiber et al. (1998), and Vanderfaeillie et al. (2003), although both studies were generally found to be of great importance. Which attitude items were drawn from which studies can be found in Appendix L. The product of all these steps was the final questionnaire in English language, which can be found in Appendix J.

4.3 The German Attitude Questionnaire

As discussed in the Methodology Chapter, besides the English questionnaire version, a German version was developed. The English questionnaire was translated using different translation procedures and comparing the translated versions, in order to develop one reconciled German version. This version was pre-tested in order to validate the newly translated items for the use with German teachers.

4.3.1 Translated Versions

As discussed in Section 3.5.2, a professional back-translation was carried out, which resulted in a professionally translated German version and a professional translation of this version back into English. Parallel to this procedure, another – more literal – translation from the English original into German was carried out by the present researcher, which included annotations concerning translation difficulties. These translations resulted in four versions: two German versions (the professional and the literal translation), and two English versions (which were the original and the professionally back-translated version). Appendix M gives a summary of these different versions.

4.3.2 Comparisons of the Different Item Versions and Decisions

The four versions of the questionnaire were used for a systematic multiple comparison of each item. As described in Section 3.5.2, a coding scheme was applied to differentiate major

discrepancies, minor discrepancies and similar/identical passages. These discrepancies between both English and between both German versions were then compared. This resulted in a variety of changes to the initial professionally translated German version, in order to reach the best possible quality of the German questionnaire. As described previously, the English version was not changed at this stage. It is evident from Appendix M that all of the steps as they pertain to the translation, comparison and reconciliation were not only carried out for the attitude items, but for all items of the questionnaire (including the self-efficacy items and the demographics items). Furthermore, because the attitude items were in the centre of interest of the present study, the following depiction of the results focussed only the 38 attitude items.

Outcome of the Comparisons

Concerning the 38 items, about half of the item comparisons in English and half of the German-German comparisons indicated being identical or synonymous (Eng-Eng 50% and Ger-Ger 55%; see Table 7). It can also be seen in Table 7 that minor issues were slightly more common in the German-German comparison (Eng-Eng 21% vs. Ger-Ger 26%) and that major issues seemed to be more common in the English-English comparison (Eng-Eng 29% vs. Ger-Ger 18%).

Table 7. Results of the English-English and the German-German comparisons. Absolute and relative number of items with or without issues

	English vs. English		German vs. German	
	absolute	relative	absolute	relative
Identical or synonymous	19	50.0%	21	55.3%
Meaning slightly different or other minor issues	8	21.1%	10	26.3%
Different meaning or other major issues	11	28.9%	7	18.4%
Number of items	38	100%	38	100%

The outcome of the comparison of the English-English comparison and the German-German comparison, is shown in Figure 3. The comparison of the English original and the back-translated English version is depicted in each of the upper rows in Figure 3 (labelled ‘E-E’). The grey boxes and the black boxes suggest which items needed further of much further attention respectively. The comparison of the professional German translation and the translation by the present researcher is depicted in each of the lower rows in Figure 3 (labelled ‘G-G’). Further information is provided in Table 8 which indicates that 14 items (37%) of the 38 items had no critique at all, whereas 63% were questioned due to major or minor differences between both English and/or both German versions.

Figure 3. Depiction of the comparisons with regard to item

E-E		x	x		x	x	x		x		x						x		x
No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
G-G		x	x	x	x	x	x		x	x	x					x		x	x

E-E		x	x	x	x	x			x	x		x			x			x	
No.	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38
G-G		x	x							x					x	x			

Note: “E-E” refers to the comparison of the English original and the English back-translation. “G-G” refers to the comparison of the professional German translation and the German translation by the present researcher. ‘No.’ are item numbers as they appeared in the questionnaire (see Appendix J and Appendix K). A grey box is used for suggested minor discrepancies and a black box for major discrepancies.

Figure 3 shows that for five items (2, 6, 11, 29, & 34) both comparisons indicated major discrepancies. Four items (3, 5, 9, & 19) have a black and a grey box, which also might indicate critical translations. These nine items required much further attention (see the resulting action column in Table 8). Four items (16, 24, 28, & 31) resulted in a major critique only from one of the comparisons, and three items (7, 21, & 22) resulted in both sides considered to have minor issues, as can be seen in Figure 3. These seven items required close attention (see Table 8). There were also eight items (4, 10, 17, 18, 23, 25, 35, & 37) that faced minor questioning from one side. Fourteen items (1, 8, 12, 13, 14, 15, 20, 26, 27, 30, 32, 33, 36, & 38) were not addressed at all by both of the comparisons. These 22 items seemed to be appropriately translated. Yet, the translations – especially those with one minor discrepancy – were looked through briefly.

Outcome with Regards to Changing the Professional German Translation

While 14 of the 38 items seemed to be adequately translated, there were 24 items that required further attention (see Table 8; and see Appendix M for all versions in one table). After intensive examination of these items, it was decided that for eight items, *no changes were needed* with regards to the professional German translation (see two columns on the right of Table 8). These were items 2, 4, 10, 17, 25, 28, 31, and 35. Items 4, 10, 17, 25, and 35 only had one minor discrepancy, and after examination they were retained in the wording as the professional translator suggested.

As can be seen in Table 8, for one of these items both comparisons indicated major problems with the E-E and G-G translations, and despite this, the wording of the professional translator was retained. The explanation is that *Item 2* was especially criticised with regard to the term “gute Lehrer” (good teachers) which was in the original item “effective teachers”. The translation “effektive Lehrer” appears in the own translation and might be a translation that is obviously closer to the English original “effective”. On the other hand, “effektiv” is not a

common term amongst practitioners, especially not when referring to persons such as teachers. Even the educational effectiveness research in Germany is normally translated into “*Wirkungsforschung*” and not into “*Effektivitätsforschung*”. It might be assumed that the study participants consider ‘good’ teachers as being ‘effective’ and vice versa. Hence, there might be good reasons why the professional translator chose the less obvious “gute” instead of the more obvious “effektive” for a high quality German version of this item. Moreover, for two items with one major discrepancy, the wording was not changed. First, *item 28* was back-translated from “I am willing...” to “I am prepared”, while the German versions were identical (“Ich bin bereit”). The term ‘bereit’ has a variety of meanings, hence, it might still be considered an appropriate translation. And second, *item 31* was originally “so that they can teach all students” and the back-translated version was “and thus addresses the needs of all of the pupils”. The difference between “teach” and “address the needs” can be explained in part by the professional translation which is “und so allen Schülern ihrer Klasse/n gerecht werden”. The back-translation of “gerecht werden” into meet or “address the needs” is appropriate, because “gerecht werden” has a variety of meanings, e.g. “to cope with” and “to satisfy” or “to fulfil something”.

Table 8. Discrepancies, special attention needed, and decisions made

Discrepancies	Number of items		Resulting Action	Decisions made	
	Absolute	Relative	Attention needed	Retained	Changed
Two major	5	13.2%	Very close	1	4
One major, one minor	4	10.5%	Very close	0	4
Two minor	3	7.9%	Close	0	3
One major	4	10.5%	Close	2	2
One minor	8	21.1%	Rough	5	3
None	14	36.8%	Only skim	14	0
Overall	38	100.0%		22	16

Note: This table combines the outcome of the comparison of the E-E and G-G comparisons (see also Figure 3), the resulting action in the sense of how close attention needs to be paid to the items, and the outcomes of the whole procedure in the sense of retaining the professional German translation or making some changes to the professionally translated wording.

Through the intensive examination of the twenty-four questionable items, it was found that sixteen of them *needed changes* (see Table 8). Hence, of the thirty-eight professionally translated items, 42% were actually changed (to some degree) (see the column on the right in Table 8). For *items 6, 11, 29, and 34*, major discrepancies were found in both the English-English and the German-German comparison, which made corrections necessary. *Item 6*, for example, was criticised because the professional translation of “inclusive education” into “inklusive Bildungssystem” introduced the “-system” that was not apparent in the English original. The more literal German translation contained only “inklusive Bildung”. Accordingly,

the “-system” component was found in the back-translated version: “inclusive education system”. It was decided to delete the notion of “system” in the German version. With regards to *items 3, 5, 9, and 19* one major and one minor discrepancy was found, which made it necessary to revise the wording. To give an example, in *item 3* the original “I get overwhelmed” is back-translated into “I find it difficult to cope”. The professional translation is “Ich fühle mich überfordert” which would be literally “I feel overwhelmed”, while the more literal translation puts it shorter and more direct: “Es überfordert mich”. Hence, it was decided to use this literal translation, which is shorter, easier to understand and more comparable to the English version. *Items 7, 21, and 22* were changed because both comparisons revealed minor discrepancies. One example is *Item 21*, which has the notion of “in ‘real world’ practices”. The professional translator translated “real world” into “echte Welt” and the present researcher translated it into “wahre Welt”. Both German versions are relatively similar. It was considered, if maybe a more adequate translation would be the German word “Praxis”. In German language, giving explicit reference to the “real world” is relatively unusual. It is important for the item that the statement explicitly divides “theory” (here: the *concept* of inclusion) and “practice” (here: the *implementation* into ‘real world’ practices). The German “Praxis” (literally Engl. “practice”) is usually being used in exactly this sense – as a counterpart of theories and concepts. Hence, the “echte Welt” was changed into “Praxis”. For *items 16 and 24*, one major discrepancy was found, and for *items 18, 23, and 37* one minor discrepancy was found. All these items were changed due to these discrepancies. One example is that the back-translation of *item 37* resulted in a differing wording. The original wording was “Working collaboratively with parents...” whereas the back-translation was “The involvement of parents...”. The German versions seemed to be more comparable, yet also not similar. The professional translator’s version was “Die Einbindung der Eltern” (Engl. to involve the parents) and the more literal translation was “Die Zusammenarbeit mit Eltern” (Engl. to work together with the parents). In order to adjust the wording more in line with the English version, the German version was changed in “Die Zusammenarbeit mit den Eltern”. A comprehensive overview on all versions for each of the items is included in Appendix M.

4.3.3 Final Changes to the Items due to Intensive Pre-Testing

In order to receive feedback from individuals from the target population, teachers were asked to give written feedback to the questionnaire. In addition, oral critique was gained from think-aloud pre-testing. The combination of both kinds of critique offers insights into improving the wording so that the questionnaire communicates most effectively with the German pre-service and in-service teachers.

The pre-test included the feedback from six German individuals. Pre-service and in-service teachers were selected in a way to represent a variety of different individuals from the target population. The variety of individuals covered male and female teachers and a range of different ages from the 30 and under to the 51-60 years category. Teachers were teaching or in training to teach in the primary and also in the secondary sector. The pre-service and in-service teachers in this pre-test reported different knowledge and training levels and all of them reported to have experiences with inclusive settings. Two pre-service teachers and two in-service teachers commented on the questionnaire in written form. And one pre-service and one in-service teacher took part in so-called think-aloud interviews.

At the time of the initial construction of this German version of the questionnaire, empirical studies on inclusive education in Germany were relatively scarce. Hence, it was subject to negotiation, how the different German individuals would react with regards to the different items. While most of the items in the questionnaire stem from the English-speaking world, and while inclusive attitude surveys for teachers already have a certain tradition in English-speaking countries, for the German population (both pre-service and in-service) this topic is relatively new. Hence, the pre-test for the German sample was extended compared to the pre-testing in Australia.

Outcome of the Written Critique

The outcome of the analysis of the written critique was all written comments given by the teachers collected in one document. The teachers were asked to give feedback with regards to three kinds of issues: necessary corrections, suggestions to improve the wording and re-reads, which were thought to indicate inconvenient wording. Although no substantial corrections were suggested by the pre-test participants, there were several items for which *suggestions* were formulated in the written critique. Most important seemed to be those suggestions, where two independent teachers found it necessary to make suggestions for improving the wording of a particular item. This applies to items 7 and 16. Item 7 starts with “Meiner Erfahrung nach denke ich, dass...” (Engl. “I feel from my experience”), and two different teachers made suggestions to improve the wording of this item. One teacher suggested to change the “denke ich” (Engl. “I think”) into “weiß ich” (Engl. “I know”), and the other teacher suggested to leave it out completely and to change the beginning of the sentence to “Meine Erfahrung ist” (Engl. “My experience is”). Item 16 was also questioned by two teachers. Here, the formulation “wenn die richtige Hilfestellung geleistet wird” (Engl. “with the right supports in place”) was found to be not adequate. One teacher suggested to use a different German word for supports and to add that “praktische Unterstützung” (Engl. “practical supports”) were meant, and the other teacher

suggested to use, “wenn die Voraussetzungen stimmen” (Engl. “with the right preconditions in place”). Other suggestions were for example to *switch verb forms* (e.g. to change in item 5 the “das mir hilft” into “das mir helfen könnte”; Engl. “to support me” into “might support me”), or to *make the sentence shorter* (e.g. to change in item 5 the “Meinem Gefühl nach gibt es” into “Es gibt”; Engl. “I feel there are” into “There are”). Moreover, besides the suggestions to improve the items, the *re-reads* might indicate unclear wording. Items that needed to be re-read were items 10 and 18. There were also items with unclear meaning such as items 19, 23 and 24, which had a “(?)” beside each sentence. It was not clear from the comments what exactly was not understood in each case.

Outcome of the Oral Critique

Compared to the outcome of the written critique, which comprised relatively rich suggestions concerning the wording, the oral critique especially highlighted issues with the understanding of words or whole statements. No corrections, nor many direct suggestions to improve the wording and several explicit notifications of unclear wording could be discovered from the two think-aloud interviews. All the other rich insights into the pre-test participants’ thinking while completing the questionnaire were interpreted as a support for the wording being adequate.

There were several words and formulations unclear to the think-aloud interview participants. The outcome of the analysis revealed that there were major issues on the one hand and minor issues on the other hand. One of the most obvious issues with regard to item wording concerned item 28. Item 28 is about the willingness “to adapt the assessment of the individual students”. Mr. Giesser had to re-read the item and then he said: “I do not understand the question. I completely do not understand the question”. The interviewer, then, tried to encourage Mr. Giesser to give more details. He explained that, first, he was thinking about “Noten” (Engl. “marks”), then about judging the performance of the students. He gave an example, and then concluded that marks have to be adapted in inclusive settings. He suggested that the item might have meant exactly this, but he was not sure about that. As he was not able to understand the question, he ticked “0”, as he pointed out. There was a long passage about the same item and the word “Bewertung” (Engl. “assessment”) in Robert’s interview, too. Just like Mr. Giesser, Robert read the statement and asked: “What is the meaning of “Bewertung”?”. He gave several examples and tried to understand the statement. Later he said that he would agree with the statement, if “Notengebung” (Engl. “marking”) was meant here. Another wording issue concerned item 32. It was questioned by Mr. Giesser with regard to the term “Vielfalt” (Engl. “diversity”, but also variety, richness, pluralism etc.; a broad and quite neutral concept with no specific connotation). He asked if this was supposed to mean that a variety of different

materials is available in the classroom. He then assumed that the diversity of human individuals might be meant, and he added that this actually enriches the classroom. With regard to this item, Robert had a different interpretation. He was talking about the plurality of influences and about different didactical settings and methods. One might draw the impression that Robert associated “Vielfalt” (Engl. “diversity”) with “Methodenvielfalt” (Engl. variety of teaching methods) which is a common term in teacher training.

Other issues that were expressed in the think-aloud interviews pertained to the meaning of specific terms. An example was the actual meaning of additional “personnel” in item 5 and 9. Concerning the staff, Mr. Giesser had to re-read the item and he said that he was not sure what ‘external staff’ was supposed to mean. He explained that nowadays, there is a vast array of different staff at schools, and because it was not clear to him, he ticked “0” for undecided. The pre-service teacher, Robert, also ticked both questions with “neither disagree nor agree”. He explained his decision with the fact that he, as a student, had not enough practical experience in schools to answer questions about additional staff at school sites.

Changing the Items’ Wording due to the Critique Formulated in the Pre-Tests

The written and the oral critique were compared. As can be seen in Table 9, five items (13%) were questioned in both the written and the oral pre-test. Another eight items (21%) were only questioned in written form and three (8%) in the think-aloud interviews. Twenty-two (58%) of the 38 items were not questioned at all. After in-depth consideration, the wording of five of the items were changed (see Table 9).

Table 9. Written and oral critique and changes to the wording

	Pre-tests		Decisions made		Items that were changed due to critique
	Absolute	Relative	Retained	Changed	
Critique from both	5	13.2%	4	1	Item 7
Only written critique	8	21.1%	7	1	Item 6
Only oral critique	3	7.9%	0	3	Items 22, 28, 32
No critique	22	57.9%	22	0	/
Overall	38	100.1%*	33	5	

Note: This table combines the outcome of the written and the oral pre-testing with regards to the German version of the attitude items. *=Overall is exact 100%, the .1 is due to rounding.

With regards to *item 6*, the written comments suggested that the notion of “related services” were not clear enough. There was no notion in the interviews with regard to the services not being clear. Both interviewees seemed to have no problem with understanding the wording. On the contrary, Mr. Giesser had a direct idea for another wording. He said that the “receive” should be a “will receive”, and if this was meant here, then he would be easily able to answer the

question. Hence, the “erhält” (Engl. “receive”) was changed into “wird erhalten” (Engl. “will receive”). For *items 7 and 22* it was suggested that the terminology “Bildungsministerium/Bildungsausschuss” (Engl. “Ministry of Education/Education Committee”) was not clear, because the right word in Hesse, Germany would be “Kultusministerium”. Hence, in *item 7 and 22* the “Bildungsministerium/der Bildungsausschuss” were changed to “Bildungs-/Kultusministerium” (Engl. “Ministry of Education and Cultural Affairs”), which might be more related to the teachers’ knowledge. Concerning *item 28* it was indicated that the term “assessment” might be misinterpreted. Hence, the “Bewertung” (which is in English a very broad range from assessment, evaluation, up to judgement) was changed into “Notengebung” (which would be close to “marking” in English). For *item 32* it was noted that the term “diversity” lead to wrong interpretations. In order to support the understanding to which aspect the notion of diversity refers, the relatively unusual notion of “im Klassenzimmer” (Engl. “in the classroom”) was changed to “in der Klasse” (Engl. “in the class”).

4.3.4 Final Items in German Language

After the development of the English questionnaire, the aim was to translate and adapt the English version into German. Several steps were taken, in order to ensure a high quality translation. In addition, the 38 German items underwent in-depth pre-testing, because most of the items were completely new for German teachers and it needed to be tested, if the teachers could understand each item’s intention clearly. The product of all these steps was the final questionnaire in German language, which can be found in Appendix K.

4.4 Characteristics of the Sample

After thorough development of the English questionnaire (comprising part 1 about attitudes, part 2 about self-efficacy, and part 3 about personal background variables), and after translating and adapting this questionnaire into German, samples were drawn in Australia and Germany, in order to empirically examine the internal structure of the attitude items, and in order to empirically examine the relationship of the attitude dimensions with other variables.

As depicted in Table 10 the intended sample size of about 100 to 200 cases was realised for the Australian sample ($n=146$) and the German sample ($n=238$). With 80%, female teachers were more represented in the German sample, compared to the Australian sample with about 70% female teachers. In the recent official statistics from 2015, the proportion of female

teachers in Germany was 71% (Statistische Ämter des Bundes und der Länder, 2017), and in Australia this proportion was 74% (NSW Department of Education, 2015a). Although the proportions in the sample and in the official statistics are not identical, they are relatively similar.

Table 10. Intended sample size and actual sample

	Australia		Germany	
	n	% ^a	n	% ^a
Intended sample	100-200		100-200	
Actual sample	146		238	
Personal background				
Gender				
female	102	70.3	189	80.1
male	43	29.7	47	19.9
Age				
30 and under years	95	65.5	172	72.9
31-40 years	23	15.9	27	11.4
41-50 years	17	11.7	18	7.6
Above 50 years	10	6.9	19	8.1
Professional Background				
Pre-/in-service				
Pre-service	89	61.0	169	71.0
In-service	57	39.0	69	29.0
Primary/secondary				
Primary	96	69.6	131	57.7
Secondary	42	30.4	96	42.3
Teaching experience (in y.)				
5 and under years	22	36.1	35	40.2
6-11 years	17	27.9	16	18.4
12 and more years	22	36.1	36	41.4
Higher degree qualification				
Higher degree	28	19.3	14	6.1
No higher degree	117	80.7	215	93.9
Incl. ed.-related background				
Incl. ed. knowledge				
Very good	17	11.7	3	1.3
Good	25	17.2	44	19.0
Average	66	45.5	71	30.6
Poor	22	15.2	95	40.9
None	15	10.3	19	8.2
Training in incl. ed.				
High (at least 40hrs)	20	13.9	12	5.3
Some	75	52.1	75	33.0
None	49	34.0	140	61.7
Experience in incl. ed.				
Yes, positive	56	40.3	50	21.7
Yes, neutral	29	20.9	47	20.4
Yes, negative	7	5.0	28	12.2
No	47	33.8	105	45.7

Note: Overall intended sample n=200-400; overall actual sample n=384. Missings are not separately depicted.

^a Only valid percentages.

The German sample comprised younger teachers compared to the Australian sample. Teachers in the primary school sector (including pre-service teachers who were in training to teach in the primary sector) were more represented in the Australian sample compared to the German sample. Notably, no secondary schools in Sydney participated in the study. With regards to the years of teaching experiences particularly the 6-11 years category was ticked more by the Australian teachers compared to the German teachers. The Australian sample comprised 19% of teachers who answered having a higher degree qualification; while only six percent of German teachers noted this. The samples differ with regards to the self-perceived knowledge of the legislation and/or policy as it pertains to inclusive education *for all*. The Australian teachers indicated to have relatively good knowledge in this regard, compared to German teachers' responses. The difference between 41% German teachers saying their knowledge is 'poor', compared to only 15% of the Australian teachers saying so, was striking. While one third of Australian teachers had no training in this area, almost two thirds of German teachers indicated that they have had no inclusive education *for all*-related training. Another difference is that 40% of the Australian teachers reported having had positive experiences with inclusive classroom settings, while only 22% of the German teachers reported positive experiences in inclusive classrooms.

4.5 Internal Structure of the Scale

The English and the German version of the questionnaire for teachers were used for data collection in Australia and in Germany. The obtained data was analysed with regard to exploring the internal structure of all of the attitude items. After the initial data cleansing, an exploratory factor analysis for the English version and the German version was carried out. The obtained factor structure was then further examined using a multi-group confirmatory factor analysis.

4.5.1 Outcome of the Initial Data Examination

The initial examination of the data comprised to examine missing values and the univariate distribution of each variable. In addition, using correlative analysis, the associations between the variables were examined. All the methodological decisions were discussed previously in Section 3.7.1.

Many of the variables have no missing value, as the first table in Appendix F demonstrates. Within the Australian sample, the item "The differentiated practices that inclusive education

would require cannot be achieved” (item 12) had by far the highest number of missing values with six participants (4%) not responding to this question. For the German sample, most missing values were found for “Diversity within the classroom enriches the learning environment” (item 32). Five individuals (2%) did not respond to this question. None of the variables in the present data set were even close to the cut-off of 10% missing per variable (D. A. Bennett, 2001).

The examination of the items’ distributions resulted in discarding eight items from further analysis (see second to fourth table in Appendix F). All of these items had unusually high arithmetic means and/or were significantly non-normally distributed. An example is “Education is a right that should be available to all children” (item 4), which was approved by nearly all respondents in Australia and Germany by ticking ‘very strongly agree’. Although the result that this item was approved by as many seems to be desirable, from a methodological point of view, this item needed to be discarded due to a lack of variance. The discarded items comprised different topics, such as general judgements with regards to inclusive education (items 4, 16, & 17), outcomes of inclusive education (items 10, & 32), supports as they pertain to inclusive education (items 20, & 26) and the role of parents for the success of inclusive education (item 37).

After univariate analysis, bivariate associations between the attitude items were examined. As discussed in the methodology chapter, the attitude items need to have substantial correlations with at least two other attitude items; otherwise they are not likely to form any interpretable factor. As can be found in Appendix G, within the Australian sample, nearly all of the attitude items had substantial correlations with several other items. Only item 11 (“Labelling students [...] is necessary to provide a quality education to them”) was discarded from further analysis, because this item had only one substantial correlation with item three (“I get overwhelmed when I have to differentiate to cater for all of the students’ needs in my classroom”). Within the German sample, a variety of items (items 3, 5, 7, 23, 29, & 34) had to be deleted in a first cycle, and then, items 11 and 22 needed to be discarded in a second cycle, in order to obtain a set of variables with each variable having a sufficient number of bivariate associations (see Appendix G). Hence, for the German sample, a variety of supports-related items (items 5, 7, 22, 29; and with regards to parents also item 34) were discarded at this stage; and items three and 23, which were excluded from further analysis, too, touched upon differentiation and adaptations in classrooms.

The next step of the correlative examination was to examine multicollinearity, using the determinant as an indicator. This analysis lead to discarding three further items for the Australian sample (see the second table in Appendix G). All three items (items 13, 30, & 36) had a relatively similar focus on a general judgement of inclusive education, in the sense that

children should be in the inclusive classroom (item 13), inclusive education as the best way (item 30) and a valuable experience (item 36). The set of items for the German sample had a sufficient determinant (see Appendix G). Within the Australian and the German sample, the Kaiser-Meyer-Olkin (KMO) statistic both for the overall set of items and for each individual variable was above .50; and Bartlett's test was significant. The exact values of the KMO test and Bartlett's test for each of the exploratory factor analyses are given in the present study (see the notes underneath Table 11 and Table 12 in the following section).

4.5.2 Exploratory Factor Analysis

The exploratory factor analysis was carried out for the Australian and for the German sample separately. Both samples reached different factorial solutions. After describing the results of each exploratory factor analysis, a hypothesis for a common internal structure of the attitudes is drawn from a systematic comparison of the obtained results for both samples.

Seven Factors Found within the Australian Sample

For the Australian sample, the exploratory factor analysis of the set of items resulted in a number of factors, which are depicted in Table 11. The size of the factors ranges from factors with six items to factors with only two items. Most loadings on factors are substantial and nearly all items load clearly on one of the factors.

A first factor comprised items that were generally concerned with outcomes of inclusive education. The strongest loading on this factor had the item "inclusive education ultimately leads to social inclusion" (item 38). Other items on this factor emphasised that "inclusion will foster understanding of differences among students" (item 35) and that "inclusion facilitates socially appropriate behaviour for all students" (item 1). The recoded item 15 (original wording: "Inclusion represents a negative change in our education system") also loads substantially on this factor. The four items together have an internal consistency of .77. Notably, item 15 loads not only on factor one, but also on the fifth factor. Hence, the first factor is represented best and unambiguously by items one, 35, and 38.

A second factor brings together a range of support-related items. This factor indicates a covariation of the feeling of the respondents that there is support from the Education Department/Board (items 7 and 22), and that they are supported by personnel from inside (item 9), and outside (item 5) school. Another item, which loads substantially on this factor is about "adequate resources" to support the teacher (item 33). It seemed that in item six, the respondents focussed particularly on the "related services", because this item also loads (yet, not

substantially) on this support factor. The internal consistency of these six items is relatively strong with .83.

Table 11. Exploratory factor analysis for the Australian sample

	1	2	3	4	5	6	7
38 Inclusive education ultimately leads to social inclusion.	.66	-.11	-.02	.10	.02	.06	-.07
35 Inclusion will foster understanding of differences among students.	.53	.08	.01	.13	-.13	.05	-.18
1 Inclusion facilitates socially appropriate behaviour for all students.	.47	-.13	.09	.03	-.01	-.02	-.08
15 (rec.) Inclusion represents a negative change in our education system.	.45	-.17	-.04	.08	-.32	.08	.01
7 I feel from my experience that there is support for inclusion from the Education Department/Board.	.12	-.73	.02	.12	-.04	-.08	.10
9 I feel there are adequate personnel within school to support me to address the [...] needs of all students.	.03	-.69	.04	-.07	.05	.05	-.11
5 [there are] personnel from outside school to support me to address the [...] needs of all students.	.15	-.68	.05	-.09	.05	.01	.07
33 I feel there are adequate resources to support me to address the unique educational needs of all [...].	-.20	-.67	.06	-.09	.05	.15	-.16
22 [...] the Education Department/Board supports efforts at including all students into the classroom.	.01	-.60	-.06	.23	-.08	-.07	-.05
6 All students will receive appropriate education and related services in inclusive education.	.19	-.38	-.19	.11	-.22	.19	-.02
23 (rec.) I get frustrated when I have to adapt the curriculum to meet the individual needs of all students.	.17	.09	.72	.15	-.07	.04	.02
3 (rec.) I get overwhelmed when I have to differentiate to cater for all of the students' needs in my classroom.	-.10	-.18	.56	-.02	-.14	.01	.03
28 I am willing to adapt the assessment of individual students in order for inclusive education to take place.	.05	.13	-.08	.78	-.07	.03	.10
24 I am willing to adapt the curriculum to meet the individual needs of all students [...].	-.03	-.10	.11	.65	.08	.05	-.11
25 I feel differentiated adjustments can be carried out in an inclusive classroom.	.20	-.07	.26	.51	.09	.05	-.08
21 (rec.) The philosophy of inclusion cannot be implemented in 'real world' practices.	.20	.06	.08	.00	-.66	.05	-.07
12 (rec.) The differentiated practices that inclusive education would require cannot be achieved.	-.01	.03	.18	.01	-.62	.09	.05
27 (rec.) It is too difficult to accommodate all students' differences in an inclusive classroom.	.04	.03	.43	-.02	-.49	.09	-.15
34 (rec.) Parents hinder the successful implementation of inclusive education.	.02	.06	-.04	-.05	-.15	.74	.10
29 From my experience, I feel that the community is supportive of the implementation of inclusion.	-.04	-.14	.09	.15	.17	.47	-.08
19 I believe that any student can learn in an inclusive school if the curriculum is adapted [...].	.29	.02	.02	-.11	.12	.02	-.75
8 It is possible to organise classes in a way that is suitable for all children.	-.05	-.13	.03	.16	-.28	-.12	-.55
14 All children are capable of learning in inclusive settings.	.31	-.10	-.27	.08	-.10	.07	-.41
2 Effective teachers are able to meet the needs of all children in the classes they teach.	-.13	-.20	-.17	.14	-.31	.03	-.35
18 Separating students is not necessary to provide a quality education to them.	.15	.12	.07	.12	.04	.22	-.33
31 Good teachers can differentiate their practices so that they can teach all students in their class/es.	.02	-.19	.03	.18	-.12	.08	-.33
Initial Eigenvalues	7.30	2.50	1.93	1.42	1.23	1.18	1.12
% of variance	28.10	9.62	7.42	5.48	4.73	4.52	4.29
Cronbach's Alpha	.77	.83	.64	.76	.76	.51	.78
No. of items for Cronbach's Alpha	4	6	2	3	3	2	6

Note: Pattern matrix of the exploratory factor analysis (principal axis factoring; Kaiser's criterion; direct oblimin rotation). N=121 Australian pre- and in-service teachers; determinant 0.00001069; KMO: .81; Bartlett's test: $\chi^2(325, n=121)=1264.825, p<.01$. Factor loadings over .30 are bold face. Only highest loading per variable is black font, others are grey font.

^a Part of the item was omitted due to limited space on this page (see for full wording Appendix J).

The third factor only had two items, which had their strongest loading on this particular factor. For further analyses, a two-item factor is not feasible, as discussed before (in terms of intercorrelations in Section 3.7.1); and, in addition, the internal consistency is not sufficient. Both items were focussed on inclusive practices; namely to “differentiate” (item 3) and to “adapt the curriculum” (item 23). Notably, both items three and 23 were reverse phrased. Hence, it might be possible that their unique variation as opposed to the other three items is in part the result of the reverse format of the item (see for discussion of the so-called method effect Maul, 2013).

The fourth factor was also about teaching practices with regards to “adapting the curriculum” (item 24), carrying out “differentiated adjustments” (item 25), and “adapt the assessment” (item 28). This fourth factor had an internal consistency of .76.

Similar to the third factor, all items on factor five were reverse phrased, too. Their content is only slightly different on “differentiated practices” (item 12), to “implement” inclusion (item 21), and to “accommodate all students’ differences” (item 27). The latter item also has a substantial loading on factor three. Factor five has an internal consistency of .76. This is a sufficient internal consistency as discussed in the Methodology Chapter; yet, all items are negatively worded, which makes it likely that their covariation is actually a method bias, and item 27 has a very strong loading on another factor. Hence, this factors’ quality is questionable.

While the sixth factor is again a two-item factor, which comprised notions with regards to the “community” (item 29) and the “parents” (item 34), factor seven seems to be more substantial with six items and an internal consistency of .78. The strongest loading on this factor has the notion that “any student can learn in an inclusive school if the curriculum is adapted to meet their individual needs” (item 19). Other notions, such as that “it is possible to organise classes in a way that is suitable for all children” (item 8), or that “all children are capable of learning in inclusive settings” (item 14) have a substantial loading on this factor, too. Similar to factor four, adaptations and classroom organisation are central; yet, these items seem to be more focussed on the students/children. A lower loading have items that focus teachers and their ability to teach all students (item 31) and that “separating students is not necessary to provide a quality education to them” (item 18). Only item two seemed problematic in the sense that it had a substantial loading not only on factor seven, but also on factor five.

Three Factors Found within the German Sample

The initial exploratory factor analysis with the German data was not interpretable, because there were many variables that had strong loadings across many other variables. No clear picture of the factorial structure could be obtained. In order to identify the items that might cause most of

the issues, the variance inflation factors (VIFs) were examined. Item thirty-six (“It is a valuable experience for all children to be educated in inclusive classrooms”), which had the strongest VIF within the present set of items with a value of 4.94, was discarded from further analysis. In addition, item two (“Effective teachers are able to meet the needs of all children in the classes they teach”) had a loading on one factor of more than .90, and ‘pulled’ a large number of very different items with low loadings with it in this factor. Hence, it was excluded, too. Item fourteen (“All children are capable of learning in inclusive settings”) had substantial loadings on nearly all factors and was discarded accordingly. During these steps of analysis, items nine and 33, which were both about the supports, and also item twelve, which was that “differentiated practices [...] cannot be achieved”, completely changed the factors that they belonged to; hence, because they were found not to be stable indicators of specific dimensions, they were discarded.

After these initial analytic trials, the exploratory factor analysis, as it is depicted in Table 12, resulted in three factors.

The *first factor* is relatively large with nine items and an internal consistency of .89. The highest loading on this factor is that “inclusion is the best way to meet the needs of all students” (item 30). Furthermore, there are items focussing on the “implementation” of inclusion (item 21), that “all students will receive appropriate education and related services” (item 6) and that classes can be organised “in a way that is suitable for all children” (item 8). Furthermore, items load on this factor that emphasise that “all children should be educated in the inclusive classroom” (item 13) and that “good teachers can differentiate” in order to be able to teach all (item 31). Difficulties with accommodating “all students’ differences” (item 27), and that “any student can learn in an inclusive school” (item 19) is also part of this factor, just like the reverse-phrased notion that “inclusion represents a negative change in our education system” (item 15).

As opposed to the former factor that comprised items that touched upon some general ideas of inclusive teaching practices, the *second factor* included items that were more concrete about the willingness to carrying out differentiation in the classroom. In items 24 and 28, the respondents indicated their willingness to adapt the curriculum (item 24) or the assessment (item 28). The third item of this factor was about the feeling of the respondent that “differentiated adjustments can be carried out” (item 25). This factor had an internal consistency of .73.

With an internal consistency of .76, the *third factor* comprised notions of desirable outcomes of inclusive education *for all* (items 1, 35, & 38). With a considerably low loading on all factors, the notion that “separating students is not necessary to provide a quality education to them” (item 18) seemed to be not relevant for any of the three factors.

Table 12. Exploratory factor analysis for the German sample

	Factor		
	1	2	3
30 Inclusion is the best way to meet the needs of all students.	.75	-.03	.08
21 (rec.) The philosophy of inclusion cannot be implemented in 'real world' practices.	.64	.13	.07
6 All students will receive appropriate education and related services in inclusive education.	.64	-.17	.06
8 It is possible to organise classes in a way that is suitable for all children.	.57	.28	-.14
13 All children should be educated in the inclusive classroom.	.56	-.15	.30
31 Good teachers can differentiate their practices so that they can teach all students in their class/es.	.56	.14	.03
15 (rec.) Inclusion represents a negative change in our education system.	.51	.07	.29
27 (rec.) It is too difficult to accommodate all students' differences in an inclusive classroom.	.51	.24	-.08
19 I believe that any student can learn in an inclusive school if the curriculum is adapted to meet their individual needs.	.39	.15	.26
24 I am willing to adapt the curriculum to meet the individual needs of all students within inclusive classrooms.	.12	.63	.13
25 I feel differentiated adjustments can be carried out in an inclusive classroom.	.10	.54	.28
28 I am willing to adapt the assessment of individual students in order for inclusive education to take place.	.08	.30	.26
35 Inclusion will foster understanding of differences among students.	-.07	.11	.75
38 Inclusive education ultimately leads to social inclusion.	.14	.07	.60
1 Inclusion facilitates socially appropriate behaviour for all students.	.11	.00	.60
18 Separating students is not necessary to provide a quality education to them.	.21	-.01	.29
Initial Eigenvalues	6.96	1.16	1.02
% of variance	43.49	7.24	6.36
Cronbach's Alpha	.89	.73	.76
Number of items for Cronbach's Alpha	9	3	3

Note: Pattern matrix of the exploratory factor analysis (principal axis factoring; Kaiser's criterion; direct oblimin rotation). N=219 German pre- and in-service teachers; determinant: .001; KMO: .92; Bartlett's Test: $\chi^2(120, n=219)=1514.771$, $p<.01$. Factor loadings over .30 are bold face (item 13 loads on factor 3 with less than .30 and appears only as '.30' due to rounding). Only highest loading per variable is black font, others are grey font.

Systematic Comparison and Hypothesis for Common Dimensions

The data-driven exploration of factors within the Australian and the German sample revealed a number of different factors. In order to develop a hypothesis about a possible *common* structure of attitudes towards inclusive education *for all* that is valid for both contexts, the factorial structures, which were obtained for the Australian sample and the German sample were compared with each other.

The most obvious similarity between both solutions of the exploratory factor analyses was that the *differentiation-related items* formed an own factor. This factor comprised in both samples item 24 ("I am willing to adapt the curriculum [...]"), item 25 ("I feel differentiated

adjustments can be carried out [...]”) and item 28 (“I am willing to adapt the assessment of individual students [...]). Within the Australian sample the internal consistency of these three items was .76, and for the German sample it was .73.

A similarity between the Australian and the German factor structure is the *outcomes-related factor*, indicated by item one (“facilitates socially appropriate behaviour”), item 35 (“will foster understanding of differences among students”) and item 38 (“ultimately leads to social inclusion”). Another substantial item on this factor within the Australian data was item 15, which asked in a reverse-phrased way if “inclusion represents a negative change”. Yet, this item loaded substantially on another factor, too, and it was in the German sample an indicator for a practices-related factor. Hence, it might not be understood in a similar way in both countries. Thus, it was assumed that this kind of outcomes-related factor (comprising items 1, 35, & 38) could be postulated for the Australian and the German context.

A comparison of both *general practices-related factors* of the Australian sample (see factor 7 in Table 11) and the German sample (see factor 1 in Table 12) revealed that there are three similar indicators; namely, item eight, item 19 and item 31. For the Australian case, item 19 (“I believe that any student can learn in an inclusive school if the curriculum is adapted to meet their individual needs”) has the strongest loading on the factor (.75), while in the German case, it has the weakest loading (.39). Item 31 (“good teachers can differentiate their practices so that they can teach all students in their class/es”) is the weakest for the Australian sample (loading: .33) and has a considerable loading of .56 amongst German teachers. Item eight (“it is possible to organise classes in a way that is suitable for all children”) is relatively strong in both samples; in the Australian sample with .55 and in the German sample with .57. Using the cut-off value of .30 (see Methodology Chapter; see also Eid et al., 2011), all these loadings can be interpreted as being sufficient. Hence, it seemed justified to postulate a general practices-related factor (comprising items 8, 19, & 31) for the Australian and the German context.

All three factors that could be extracted from the set of items within the German sample (see Table 12) have been described so far. *Only in the Australian sample, further factors* had a substantial internal consistency; namely, factor 3 and factor 5 (see Table 11). The latter factor 5 comprised only reverse-phrased items, which were all with regard to their content relatively similar to the general practices- and the differentiation-factors. As discussed previously, their loading on a separate factor might be due to their reverse wording. Factor 3 (see Table 11) had a strong internal consistency of .83, and, hence, might deserve further attention. Five items had substantial loadings between .73 and .60. The items comprised notions of supportive internal personnel (item 9) and external personnel (item 5), support from the Education Department/Board for inclusion (item 7) and for efforts at including all students (item 22), and

supportive resources (item 33). Item six (“all students will receive appropriate education and related services in inclusive education”) also loads on this factor with .38. The strong supports-theme that is obvious in the other five items is not so strong in this item. In the German case (see exploratory factor analysis with the German data in Table 12), item six clearly represented the general practices factor. Against this background, it became apparent that the Australian participants focussed more on the “will receive [...] related services” part of the statement when completing the English version (hence, it indicated the support dimension), and the German participants focused on the “will receive appropriate education” of the statement when completing the German version (hence, it indicated the practices dimension). Accordingly, it was decided to discard this particular item from further analysis. Yet, in the Australian data, five support-related items (items 5, 7, 9, 22, & 33) had a clear structure (see Table 11, but also Appendix H); hence, they were not discarded.

Notably, the supports-related factor is only present in the Australian sample. The corresponding items in the German sample were discarded previously due to intercorrelations that were too low (items 5, 7, & 22; see Appendix G), and due to inconsistent loadings on different factors (items 9, & 33). Due to the strength of this supports-factor within the Australian data, and because from the teachers’ point of view inclusive education *for all* demands strong supports, it might be worth considering postulating such a factor for the German sample, too. That the teachers’ empowerment, training, and *support* needs to be ensured, was for example emphasised by the UNESCO (2015). Hence, it seemed promising to elaborate further on possibilities to establish a support-related factor of the teachers’ attitudes in both, the Australian and the German data. This is examined empirically in the present study.

The remaining factors in the Australian exploratory factor analysis had an insufficient internal consistency and/or an insufficient number of items. Hence, no further common factors were postulated.

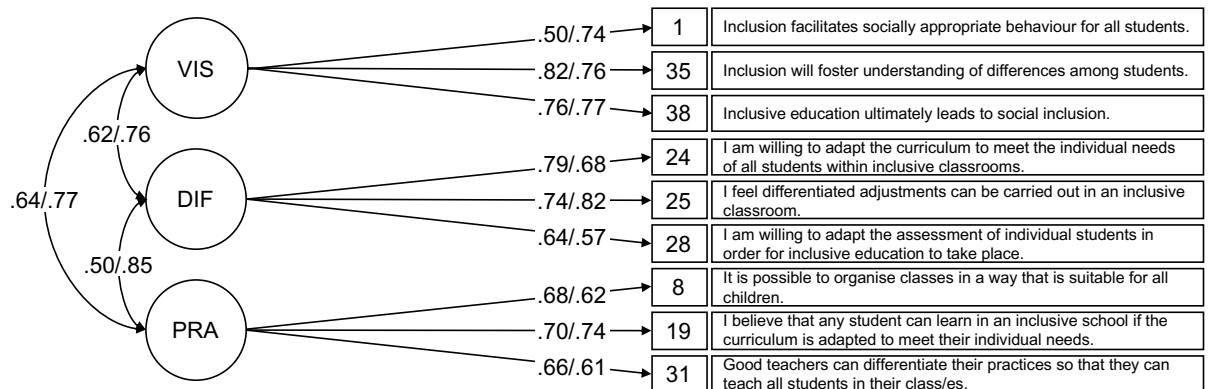
4.5.3 Confirmatory Factor Analysis

In order to elaborate the initial hypothesis of three attitude dimensions (vision of particular outcomes, differentiation, and general practices), a multi-group confirmatory factor analysis was carried out. After establishing these three factors, the supports dimension, as it was derived from the Australian data, was examined for the Australian and the German sample in a multi-group confirmatory factor analyses, too. Due to empirical and conceptual considerations, it seemed feasible to postulate this additional factor; accordingly, the full model comprising four dimensions as they pertain to the teachers’ attitudes towards inclusive education *for all* was estimated in a subsequent step.

Three-Dimensional Structure as it was Obtained through the Systematic Comparison

The outcome of the systematic comparison of both exploratory factor analyses suggested to assume three factors that were present in the Australian and the German data. As it was described previously, for each of the three factors, three common indicators were found.

Figure 4. Confirmatory factor analysis of the three-factor model



Note: Groups are Australian teachers (n=146) vs. German teachers (n=238); parameter estimates are indicated for both groups divided by a slash, respectively. Multiple-group confirmatory factor analysis. $\chi^2(48, n=384)=97.548$, CFI=.96, TLI=.93, RMSEA=.07, SRMR=.04. Full information maximum likelihood (FIML) estimation. Parameter estimates were obtained using ML, are standardised (std.all), and all are significant on the $p<.001$ level. Numbers in rectangles are the item numbers in the questionnaire (see Appendix J).

Figure 4 depicts the result of the multi-group confirmatory factor analysis of the three factor-model. Abbreviations were used to refer to the vision of particular inclusive education-related outcomes (vision=VIS), to the differentiation carried out for inclusive education to take place (differentiation=DIF) and for general practices as inclusive education would require them (pactices=PRA). The absolute fit of the model is sufficient, given that the SRMR of .04 is below .10 and chi-square divided by df is 2.03, which is below three. The RMSEA as an index of fit, which includes a model parsimony correction, is with .07 below the less strict cut-off of .08. Hence, it indicates still sufficient fit; although researchers who use the stricter .06 cut-off might suspect a threat to parsimony of the model. The comparative fit indices are both above .90 (CFI=.96; TLI=.93), indicating relatively good fit, while the TLI falls below the stricter .95-cut-off. The TLI compensates for the complexity of the model, which might be interpreted in a similar direction as the lower RMSEA value that the model is less parsimony and more complex than needed. In addition, the correlations between the factors for the German sample (see Figure 4) seemed particularly high, which lead to the hypothesis that it might be the case that all three factors could be better explained by assuming only one dimension. The model was re-calculated with only one latent variable (correlations were constrained to a correlation of 1; nested model).

Yet, the model fit decreased drastically. The absolute fit decreased from SRMR=.04 to .07, and the ratio of chi-square and the degrees of freedom increased from 2.03 to 4.20. While the SRMR of .07 can still be considered as being sufficient, the chi-square ratio is not sufficient. The RMSEA increased from .07 to .13, which is high above even the less strict cut-off of .08. The CFI and the TLI were with .85 and .80 respectively, too low. The three-factor model is significantly ($\chi^2_{\text{diff}}(8, n=384)=19.13, p<.05$) better compared to the one-factor model. This result seemed to suggest that the three-factor solution was built on a relatively solid data-driven evidence base.

Proposal of a Support Factor for both Samples: Full Four-Dimensional Model

The exploratory factor analysis showed for the Australian sample a strong factor comprising a variety of supports-related items. This factor had a considerable internal consistency of .83 (if item six was discarded, as discussed previously, the internal consistency was .84). However, these items were not functioning in the same way in the German sample.

An exploratory attempt was made to understand how the support items would work for Australia and for Germany. A multiple-group confirmatory factor analysis was carried out, including one latent support dimension with five indicators (items 5, 7, 9, 22, & 33; see model on the left in Table 13). The presence of such a factor for the Australian sample is suggested by the considerably high loadings of each item on the factor (.63 up to .78). For the German sample, items 7 and 22 resulted in high loadings (.73 and .88), while the other three loadings were considerably low (.25 up to .36).

The wording of item seven (“support for inclusion from the Education Department/Board”) and item 22 (“Education Department/Board supports efforts at including all students”) were very similar. Both items were differing in the nuance that item seven was broader on ‘inclusion’, while item twenty-two was focussing the respondents’ attention to ‘efforts at including all students’. After additional examination of the intercorrelations between the five items, a two-factor model was specified. Both ‘Education Department/Board’ items were specified to load on one factor and the other three items on support from external personnel (item 5), support from internal personnel (item 9), and supportive resources (item 33) were specified to load on the second factor. The resulting model (see model on the right in Table 13) has a considerable better fit compared to the one-dimensional model. The SRMR improved from .09 to .03, and the chi-square to degrees of freedom ratio decreased from eleven to three. The RMSEA and the CFI and TLI are an acceptable range within the two-factor solution. The two-dimensional model has a significantly ($\chi^2_{\text{diff}}(2, n=384)=89.755, p<.001$) better fit compared to the one-dimensional model. The covariance between both factors suggested that the supports from the Education

Department/Board is strongly related to other kinds of support in Australia (covariance within the Australian sample: .75); while this seemed to be less the case within the German sample (covariance: .38). In other words, the Education Department/Board items do not fit well with the other supports-related aspects within the German sample. As previously discussed, Education Department/Board was difficult to be translated. The initial “Bildungsministerium/Bildungsausschuss” and “Kultusministerium” were decided to be used as “Bildungs-/Kultusministerium” in the survey. Yet, in terms of supporting the teachers to implement inclusive education *for all*, the “Bildungs-/Kultusministerium” is a relatively abstract stakeholder in Germany; especially compared to other supports facets. Hence, future investigations might need to clarify the relevance of these stakeholders from the perspective of the teachers in different contexts.

Table 13. Comparative examination of the support-related items

	One-factor model		Two-factor model	
	Australia	Germany	Australia	Germany
Factor 1			Factor 1	
Item 5	.63	.25	Item 5	.71
Item 7	.78	.73	Item 9	.78
Item 9	.74	.26	Item 33	.73
Item 22	.71	.88	Factor 2	
Item 33	.68	.36	Item 7	.89
			Item 22	.76
			Covariance	.75
Chi-square	108.888			19.133
df	10			8
χ^2/df	10.89			2.39
n	384			384
CFI	.81			.98
TLI	.62			.95
RMSEA	.23			.09
90% CI	.19, .27			.04, .14
SRMR	.09			.03

Note: Multiple-group confirmatory factor analyses. Compared to the one-factor model, the two-factor model fits significantly better (chi-square difference tests were used to decide between the competing models as suggested by C. Werner & Schermelleh-Engel, 2010): $\chi^2_{diff}(2, n=384)=89.755$.

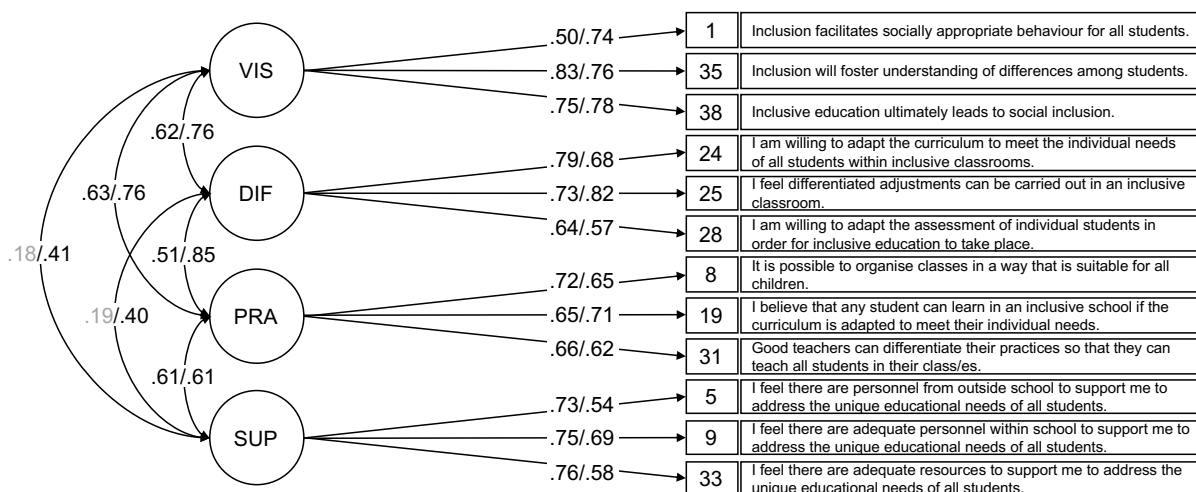
Although, the multiple-group confirmatory factor analysis demonstrated that the factor comprising items five, nine and 33 was more substantial for the Australian sample, compared to the German sample (see model on the right in Table 13), it was proposed that such a dimension might exist in both samples. According to the strong emphasis recently given in Germany to different kinds of support systems in order to allow for more inclusive practices (Kielblock, Gaiser, & Stecher, 2017; UNESCO, 2015), it seemed valid to formulate the hypothesis that in the next few years, German teachers might place more and more emphasis on the importance of supports for carrying out inclusive education *for all*. In other words, to

assume a support dimension of teachers' attitudes might be future-oriented with regards to the German situation.

Full Model

According to the previous considerations, the original model with three factors (see Figure 4) was extended to a model with an additional supports factor (supports=SUP), as it is depicted in Figure 5. As can be drawn from Figure 5, the model has an acceptable fit. The SRMR of .06 is below .10 (and even below the stricter threshold of .08). The chi-square to degrees of freedom ratio is 1.89, hence, it is very good. The RMSEA is with the value of .07 acceptable (although it is slightly above the stricter criteria of .06). Both CFI (.94) and TLI (.92) are above the cut-off value of .90; yet, the stricter criterion of above .95 was not reached. Generally, the model fits to the underlying data.

Figure 5. Confirmatory factor analysis of the four-factor model



Note: Groups are Australian teachers (n=146) vs. German teachers (n=238); parameter estimates are indicated for both groups divided by a slash, respectively. Multiple-group confirmatory factor analysis. $\chi^2(96, n=384)=181.493$, CFI=.94, TLI=.92, RMSEA=.07, SRMR=.06. Full information maximum likelihood (FIML) estimation. Parameter estimates were obtained using ML, are standardised (std.all), and nearly all are significant on the $p<.01$ level – non-significant estimates are in grey font colour. Numbers in rectangles are the item numbers in the questionnaire (see Appendix J).

As demonstrated in Figure 5, nearly all of the loadings were above .60. For the Australian sample, only the path from the vision-factor to item one had a lower loading. For the German sample, this was true for the path from the differentiation-factor to item 28. As anticipated, the loadings for the German sample from the supports-factor to items five and 33 were also just below .60. Surprisingly within the Australian sample, the support-factor is not significantly correlated with the vision- and the differentiation-factor. Yet, a high correlation was found between the supports and the general practices. A similar picture that the correlations between

vision and supports and between differentiation and supports were slightly lower compared to the others, can be found within the German group, too.

4.5.4 Obtaining the Factor Scores

As the exploratory factor analyses showed, a factorial structure for the Australian and for the German sample could be found. The systematic comparison revealed a common structure of three factors, which could be extended to a four-factor model, as the confirmatory factor analyses indicated. Each of the four factors comprised three items. Table 14 gives an overview on the mean scores (M) and standard deviations (SD). In addition, the internal consistencies were depicted in Table 14, using Cronbach's alpha as an indicator. All alpha values were at least .70 or higher, except for the German support dimension, which had an alpha of .63. The low Cronbach's alpha was anticipated, given the previous discussion of the support factor. In applied empirical research it would be not unusual to call a Cronbach's alpha of .63 still acceptable or at least moderate (see Taber, 2017). The corrected item total correlation (ITC; Ger. Trennschärfe) in Table 14 indicates how informative an item is with regards to the dimension (Penfield, 2013). Most items can be considered to provide a large amount of information (ITC greater than .50; Penfield, 2013), and only some items in Table 14 had an ITC between .30 and .50, which Penfield (2013) considered to provide a moderate amount of information.

Table 14. Characteristics of the four attitude dimensions

	Australia				Germany			
	Alpha	M	SD	ITC	Alpha	M	SD	ITC
VIS	.72				.76			
Item 1		5.93	0.95	0.42		5.33	1.33	0.60
Item 35		6.00	0.95	0.67		5.57	1.19	0.64
Item 38		5.81	1.25	0.57		5.09	1.54	0.57
DIF	.76				.73			
Item 24		5.99	0.99	0.65		5.09	1.33	0.59
Item 25		5.74	0.99	0.58		5.25	1.40	0.57
Item 28		5.74	1.12	0.54		5.10	1.49	0.51
PRA	.71				.70			
Item 8		5.13	1.52	0.54		4.68	1.66	0.51
Item 19		5.46	1.44	0.54		5.00	1.51	0.50
Item 31		5.92	1.18	0.53		4.50	1.66	0.55
SUP	.79				.63			
Item 5		4.30	1.57	0.63		4.57	1.61	0.45
Item 9		4.45	1.47	0.63		4.24	1.53	0.50
Item 33		4.06	1.63	0.63		4.18	1.49	0.38

Note: Cronbach's alpha, arithmetic mean, standard deviation and corrected item-total correlation.

As a next step, the factor scores were obtained for each of the four dimensions. The “lavPredict()” function of lavaan (Rosseel, 2012) was utilised. Four variables were created out of the model, which was presented in Figure 5. Each individual score on each of the variables was estimated according to the available information, and according to the loadings estimated within the model. The more positive (or negative) an individual score was the more positive (or negative) were the attitudes relative to the other scores (see properties of the obtained scores in Appendix I).

Taken together, the evidence concerning the internal structure of the teachers’ attitudes towards inclusive education *for all* revealed a common structure that proved to be stable across cultural contexts. Four components were proposed: the vision of particular desirable outcomes of inclusive education *for all* (VIS), the differentiation that teachers were willing to carry out (DIF), the general practices as they would be possible to be carried out under certain circumstances (PRA) and the support that the teacher receives as it pertains to inclusive education *for all* (SUP). As will be elaborated more in the discussion section, this result might be interpreted as a first step towards (a heuristic model of) the measurement of the teachers’ attitudes towards inclusive education *for all*. Yet, although the internal structure of the teachers’ attitudes were elaborated in quite detail, it remained unclear how the proposed dimensions of the attitudes were associated with conceptually related constructs.

4.6 Relation of the Scales’ Dimensions to Conceptually Related Aspects

The analysis of the literature review resulted in an English questionnaire, which was translated into German. Both language versions were used for data collection in Australia and Germany. The attitude items were tested with regards to the internal structure. Four attitude dimensions were proposed after in-depth analysis using different analytic strategies. However, it remained an open question, how these dimensions, which were thought to represent different aspects of the teachers’ attitudes towards inclusive education *for all*, were associated to conceptually related constructs. As specified in Section 3.7.3, ten validation hypotheses (VHs) were derived from the literature review. These VHs guided not only the selection of other items in the questionnaire, but also the analyses, which are presented in the following sections. As described before (see Section 3.7.3), for analysis a Generalised Linear Model (GLM) was utilised. To be more precise, a Multivariate Analysis of Variance (MANOVA) was carried out for most of the variables, and a multivariate linear regression analysis was carried out for the self-efficacy dimensions. The overall effects (all four dimensions together) and effects for the four

dimensions are reported. First, analyses are presented as they pertain to the Australian teachers, and then, the analyses for the German teachers are presented. The depiction in the following sections is concentrated solely on the obtained results; the discussion of the results against the backdrop of the initial validation hypotheses will be presented in the Discussion Chapter (see Section 5.2.3).

4.6.1 Australian Teachers' Attitudes, Self-Efficacy, and Background

The analysis of the Australian teachers' data is organised according to the validation hypotheses (VHs) as they were presented in Section 3.7.3.

Gender

The analysis of gender (see VH1), as a relevant variable that were thought to explain differences in attitudes, showed that there was no overall effect of gender on the four attitude dimensions ($V=0.06$, $F(4, 140)=2.16$, $p>.05$). However, as Table 15 depicts, a significant difference between female and male teachers was found with regards to concrete adjustments and curriculum/assessment adaptations, which were represented by the differentiation dimension. Female teachers tended to be more in favour in this regard. Yet, the 95% confidence intervals overlap (see 95% CI in Table 15), which means that there is a certain chance that the central tendencies of female and male teachers are not differing from each other. Taken together, there seems to be no evidence that gender plays a role for the present sample of Australian teachers' attitudes towards inclusive education *for all*.

Table 15. Differences of the attitudes amongst Australian teachers with regards to gender

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	2.13	1	.15			
Female				0.03	-0.05	0.11
Male				-0.08	-0.21	0.05
DIF	5.68*	1	.02			
Female				0.08	-0.05	0.22
Male				-0.22	-0.42	-0.01
PRA	0.05	1	.82			
Female				0.02	-0.18	0.21
Male				-0.02	-0.32	0.27
SUP	0.40	1	.53			
Female				-0.04	-0.24	0.16
Male				0.08	-0.23	0.39

Note: $V=0.06$, $F(4, 140)=2.16$, $p>.05$; $n=145$ (female $n=102$; male $n=43$); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

Age

Another relevant variable, as it pertains to the personal background of the teachers, was their age (see VH2). For the Australian sample, the analysis of age-differences turned out to be significant for all four attitude dimensions ($V=0.25$, $F(12, 420)=3.19$, $p<.01$). A closer look at the different effects revealed that age was significantly related to the general practices dimension (see Table 16). According to the means of the different age groups, the youngest teachers were most favourable, while teachers in the 31-40 and the 41-50 years groups were not as positive, and the teachers, who were above fifty years old had the least positive attitudes. Post hoc tests demonstrated that only the comparison between the youngest and the oldest teachers was significant. The 95% confidence intervals confirm that the lower limit of the 30 and under years group (-0.02) is above the upper limit of the above 50 years group (-0.28).

Table 16. Differences of the attitudes amongst Australian teachers with regards to age

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	0.03	3	.99			
30 and under years				0.00	-0.09	0.08
31-40 years				0.01	-0.16	0.19
41-50 years				0.01	-0.20	0.21
above 50 years				-0.04	-0.31	0.23
DIF	0.57	3	.64			
30 and under years				-0.01	-0.16	0.13
31-40 years				-0.11	-0.40	0.18
41-50 years				0.18	-0.16	0.51
above 50 years				0.02	-0.42	0.46
PRA	4.46**	3	.00			
30 and under years				0.18	-0.02	0.37
31-40 years				-0.19	-0.58	0.20
41-50 years				-0.17	-0.63	0.28
above 50 years				-0.88	-1.47	-0.28
SUP	5.98**	3	.00			
30 and under years				0.24	0.04	0.44
31-40 years				-0.36	-0.77	0.04
41-50 years				-0.41	-0.88	0.06
above 50 years				-0.75	-1.36	-0.15

Note: $V=0.25$, $F(12, 420)=3.19$, $p<.01$; $n=145$ (30 and under years $n=95$; 31-40 years $n=23$; 41-50 years $n=17$; above 50 years $n=10$); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

A similar result was found for the supports dimension of the attitudes (see Table 16). The differences of the central tendencies showed a similar pattern, and post hoc tests revealed the youngest and the oldest teachers to be significantly different with regards to their supports-related attitudes. The lower limit of the 95% confidence interval of the 30 and under years group (0.04) is above the upper limit of the oldest teacher group (-0.15). The confidence intervals also

demonstrated that the 31-40 years group is below the 30 and under years group (95% CI [-0.77, 0.04] versus 95% CI [0.04, 0.44]). Indeed, the more liberal Gabriel's procedure indicated in the post hoc analysis that the group 30 and under years is significantly different to all three other groups. Yet, Hochberg's GT2 procedure was not significant for the differences between the youngest teachers and the teachers between thirty-one up to fifty. While the finding that the youngest and the oldest teachers differ in their supports-related attitudes seems to have a solid evidence-base, the differences between the youngest and both middle groups might only be understood as a tendency.

Overall, the significant general effect of the Australian teachers' age on their attitudes towards inclusive education *for all* manifests particularly with regards to their views on general practices and supports. Their views on the general vision of inclusive education *for all* and differentiation, were not different across the age groups.

Pre-Service versus In-Service

A further aspect was, if the attitudes were different for pre-service and in-service teachers (see VH3). The teachers in the Australian sample were approached in two different kinds of institutions; namely, universities and schools. Hence, there might be differences of attitudes between pre-service teachers and in-service teachers, respectively. The analysis presented in Table 17, suggests that an overall effect existed ($V=0.17$, $F(4, 141)=7.12$, $p<.01$).

Table 17. Differences of the attitudes of Australian pre-service vs. in-service teachers

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	0.19	1	.66			
Pre-service				-0.01	-0.10	0.08
In-service				0.02	-0.09	0.13
DIF	4.36*	1	.04			
Pre-service				-0.10	-0.24	0.05
In-service				0.15	-0.03	0.33
PRA	4.95*	1	.03			
Pre-service				0.14	-0.06	0.34
In-service				-0.22	-0.47	0.03
SUP	7.92**	1	.01 ^b			
Pre-service				0.19	-0.02	0.40
In-service				-0.29	-0.55	-0.03

Note: $V=0.17$, $F(4, 141)=7.12$, $p<.01$; $n=146$ (pre-service $n=89$; in-service $n=57$); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

^b The actual p-value is 0,006.

The difference between pre-service and in-service teachers with regards to the supports dimension was significant on the one percent level. Pre-service teachers were more favourable towards supports-related aspects of inclusive education *for all*. The 95% confidence intervals were not overlapping, which confirmed the significance of the difference. On the five percent significance level, the general practices tended in a similar direction that the pre-service teachers were more in favour, while the in-service teachers had more negative attitudes. Yet, the overlapping confidence intervals raise doubts that this effect is substantial. Surprisingly, the differentiation dimension, which is significant on the five percent level, too, shows a contrary picture; the in-service teachers tended to be more in favour of differentiation. Yet, the lower limit of the 95% confidence interval of the in-service teachers (-0.03) is below the upper limit of the pre-service teachers (0.05); suggesting that there is a certain range, where the mean could be identical for both groups. Taken together, the overall effect seemed to arise to a large extent from the substantial difference between the more favourable views of pre-service teachers and the more unfavourable views of in-service teachers on supports-related issues. While pre-service teachers tended to emphasise the general practices, the in-service teachers tended to approve more the differentiation. Yet, although both of these teaching-related effects (DIF and PRA) were significant on the five percent level, the confidence intervals suggested that this result needs to be interpreted very carefully.

Table 18. Differences of the attitudes of Australian primary vs. secondary teachers

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	0.88	1	.35			
Primary				0.02	-0.07	0.11
Secondary				-0.05	-0.19	0.08
DIF	6.38*	1	.01 ^b			
Primary				0.11	-0.03	0.25
Secondary				-0.21	-0.43	0.00
PRA	0.16	1	.69			
Primary				-0.02	-0.22	0.18
Secondary				0.05	-0.25	0.36
SUP	0.58	1	.45			
Primary				-0.06	-0.28	0.15
Secondary				0.08	-0.24	0.40

Note: V=0.10, F(4, 133)=3.60, $p<.01$; n=138 (primary n=96; secondary n=42); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

^b The actual p-value is 0,008.

Primary versus Secondary

Australian teachers from the primary sector hold significantly different attitudes compared to teachers from the secondary sector (see VH4) ($V=0.10$, $F(4, 133)=3.60$, $p<.01$). The overall effect is significant on the one percent level. As Table 18 demonstrates, there seemed only one significant effect with regard to the differentiation-dimension. This effect is significant on the five percent level, and it shows that teachers in the primary sector were more positive (0.11) compared to their counterparts from the secondary sector (-0.21). The 95% confidence intervals of both groups were overlapping slightly, leaving some doubts if the effect can be considered substantial.

Years of Teaching Experience

The teachers were asked to indicate their years of teaching experience (see VH5) in the survey. The results of the analysis showed for the Australian teachers no overall effect of years of teaching experience on attitudes ($V=0.10$, $F(8, 112)=0.71$, $p=.68$). The absence of an overall effect is mirrored in none of the four dimensions having significant differences according to the different categories of teaching experiences in years (see Table 19). Accordingly, the evidence suggests that there is no effect of years of teaching experience on the teachers attitudes towards inclusive education *for all* in the Australian sample.

Table 19. Differences of the attitudes of Australian teachers with regards to years of teaching experience

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	0.20	2	.82			
Up to 5 years				-0.02	-0.20	0.16
6-11 years				0.04	-0.16	0.24
12 and more years				-0.04	-0.22	0.14
DIF	0.94	2	.40			
Up to 5 years				-0.06	-0.35	0.24
6-11 years				0.24	-0.09	0.58
12 and more years				0.10	-0.19	0.39
PRA	0.74	2	.48			
Up to 5 years				-0.13	-0.58	0.31
6-11 years				-0.08	-0.58	0.43
12 and more years				-0.44	-0.89	0.00
SUP	0.60	2	.55			
Up to 5 years				-0.08	-0.57	0.40
6-11 years				-0.29	-0.84	0.26
12 and more years				-0.46	-0.94	0.03

Note: $V=0.10$, $F(8, 112)=0.71$, $p>.05$; $n=61$ (up to 5 years $n=22$; 6-11 years $n=17$; 12 and more years $n=22$); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

Higher Degree Qualification

Whether teachers hold a higher degree qualification or not (see VH6), is significantly related to their overall attitudes ($V=0.07$, $F(4, 140)=2.61$, $p<.05$). The result is depicted in Table 20.

Table 20. Differences of the attitudes of Australian teachers with regards to holding a postgraduate degree/diploma or not

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	4.37*	1	.04			
Postgrad.				0.15	-0.01	0.31
No postgrad.				-0.04	-0.12	0.04
DIF	2.01	1	.16			
Postgrad.				0.16	-0.10	0.42
No postgrad.				-0.05	-0.17	0.08
PRA	0.00	1	.96			
Postgrad.				0.01	-0.36	0.38
No postgrad.				0.00	-0.18	0.18
SUP	3.27	1	.07			
Postgrad.				-0.31	-0.69	0.07
No postgrad.				0.07	-0.11	0.26

Note: $V=0.07$, $F(4, 140)=2.61$, $p<.05$; $n=145$ (postgraduate degree/diploma $n=28$; no postgraduate degree/diploma $n=117$); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

With regards to the vision that inclusive education *for all* might lead to positive outcomes, those holding a postgraduate degree or diploma were more positive (0.15), compared to those holding no postgraduate degree or diploma (-0.04). With a p-value of .04, this difference can be considered significant on the five percent level. Yet, the examination of the 95% confidence intervals were overlapping. The lower level of the postgraduate degree group (-0.01) is below the upper level of the non-postgraduate group (0.04), which suggests uncertainty concerning the means being on all accounts different to each other.

Knowledge of the Local Legislation and/or Policy

The teachers were asked to indicate their level of knowledge of the local legislation and/or policy as it pertains to inclusive education *for all* (see VH7), and Table 21 depicts the results. According to Pillai's trace, there was no overall effect ($V=0.14$, $F(16, 560)=1.30$, $p>.05$). Yet, Roy's Largest Root test was significant on the five percent level, suggesting the presence of an effect of knowledge levels on attitudes ($\Theta=0.08$, $F(4, 140)=2.70$, $p<.05$). A closer examination of the four dimensions revealed no significant differences between any levels of knowledge, as Table 21 demonstrates. Although the more liberal Roy's Largest Root test indicated a

significant effect, the overall result seems to suggest that knowledge and attitudes were not related in the present data from Australia.

Table 21. Differences of the attitudes of Australian teachers with regards to their inclusion-related knowledge

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	2.43	4	.05			
Very good				0.13	-0.07	0.33
Good				0.15	-0.02	0.31
Average				-0.11	-0.21	-0.01
Poor				0.02	-0.16	0.19
None				0.06	-0.16	0.27
DIF	1.87	4	.12			
Very good				0.28	-0.06	0.61
Good				0.21	-0.06	0.48
Average				-0.10	-0.27	0.07
Poor				-0.08	-0.37	0.21
None				-0.17	-0.52	0.19
PRA	0.83	4	.51			
Very good				0.13	-0.34	0.60
Good				0.17	-0.22	0.56
Average				-0.16	-0.39	0.08
Poor				0.07	-0.34	0.49
None				0.18	-0.32	0.68
SUP	0.29	4	.88			
Very good				-0.16	-0.66	0.33
Good				0.04	-0.37	0.45
Average				0.02	-0.24	0.27
Poor				-0.10	-0.54	0.33
None				0.18	-0.34	0.71

Note: $V=0.14$, $F(16, 560)=1.30$, $p>.05$; yet, Roy's test is significant: $\Theta=0.08$, $F(4, 140)=2.70$, $p<.05$; $n=145$ (very good $n=17$; good $n=25$; average $n=66$; poor $n=22$; none $n=15$); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

Training in Inclusive Education for All

A similar picture can be found with regard to the amount of training in inclusive education *for all* (see VH8).

As Table 22 shows, none of the four attitude dimensions comprised substantial differences as related to the level of inclusive education-specific training. And while Roy's Largest Root was again significant with a p-value of .049 ($\Theta=0.07$, $F(4, 139)=2.45$, $p=.05$), the more conservative Pillai's trace indicated no overall effect ($V=0.07$, $F(8, 278)=1.29$, $p=.25$). The Australian teachers' attitudes seemed not to be related to their reported amount of training in inclusive education *for all*.

Table 22. Differences of the attitudes of Australian teachers with regards to their level of inclusion-related training

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	0.07	2	.93			
High				0.03	-0.16	0.22
Some				-0.01	-0.10	0.09
None				-0.01	-0.13	0.11
DIF	1.27	2	.28			
High				0.02	-0.29	0.34
Some				0.07	-0.09	0.23
None				-0.13	-0.33	0.06
PRA	0.32	2	.73			
High				-0.06	-0.49	0.38
Some				-0.04	-0.26	0.19
None				0.09	-0.18	0.37
SUP	1.87	2	.16			
High				-0.21	-0.66	0.24
Some				-0.10	-0.34	0.13
None				0.21	-0.07	0.50

Note: V=0.07, F(8, 278)=1.29, p=.25, yet, Roy's test is significant: Θ =0.07, F(4, 139)=2.45, p=.05; n=144 (high n=20; some n=75; none n=49); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

Experiences with Inclusive Classroom Settings

Another aspect was, if the teachers had experiences with inclusive classroom settings (see VH9). The teachers could indicate if they had such experiences or not; and if 'yes', they were able to indicate the quality of their experience. With regard to all four attitude dimensions, the analysis suggested the presence of a significant effect ($V=0.24$, F(12, 402)=2.86, p<.01).

Table 23 indicates that all attitude dimensions showed significant differences. With regards to the vision dimension, post hoc tests revealed that those teachers, who reported negative experience with inclusive classrooms were significantly more unfavourable compared to those with positive experiences and those with no experiences at all. The 95% confidence intervals confirm this finding. Concerning the differentiation dimension, the post hoc tests showed those, who experienced positive examples of inclusive classrooms were more favourable as opposed to those with negative and those with no experiences. This is in line with what the 95% confidence intervals would suggest. With regards to the principled view on general practices, negative experiences seemed to have a particular lowering effect, compared to positive experiences and no experiences at all. Yet, the difference between the 'positive' and the 'neutral' group is significant within the post hoc tests, too. Like in the previous findings, an examination of the confidence intervals would lead to a similar conclusion. Although there is a

significant effect with regards to the supports dimension, no significant differences were found through the post hoc testing. The 95% confidence intervals show that the ‘neutral’ and the ‘negative’ group are both considerably lower compared to the ‘positive’ group or the group with no experiences. Yet, this finding might need to be carefully interpreted due to the non-significant post hoc tests.

Table 23. Differences of the attitudes of Australian teachers with regards to their inclusive education-related experiences

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	6.11**	3	.00			
Positive				0.13	0.02	0.24
Neutral				-0.09	-0.24	0.06
Negative				-0.52	-0.83	-0.21
None				-0.03	-0.15	0.09
DIF	7.10**	3	.00			
Positive				0.27	0.09	0.44
Neutral				-0.07	-0.31	0.17
Negative				-0.74	-1.23	-0.25
None				-0.17	-0.36	0.02
PRA	6.15**	3	.00			
Positive				0.26	0.01	0.51
Neutral				-0.35	-0.70	-0.01
Negative				-1.09	-1.79	-0.39
None				0.11	-0.16	0.38
SUP	3.61*	3	.02			
Positive				0.16	-0.10	0.42
Neutral				-0.37	-0.73	0.00
Negative				-0.68	-1.42	0.06
None				0.21	-0.07	0.50

Note: V=0.24, F(12, 402)=2.86, p<.01; n=139 (positive n=56; neutral n=29; negative n=7; none n=47); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

Taken together, negative experiences go together with attitudes that were more unfavourable. Interestingly, there was considerable evidence that in most dimensions, it would be even better to have no experiences at all, as opposed to negative experiences. Only with regards to the differentiation, solely the positive experiences were associated with more positive attitudes.

Teachers’ Self-Efficacy as it Pertains to Inclusive Education

Besides the variety of teachers’ background variables, the literature review showed the teachers’ self-efficacy to carry out inclusive practices (see VH10) as being of particular importance for explaining the teachers’ attitudes. In accordance to previous findings, the hypothesis was formulated that more positive attitudes go together with stronger self-efficacy beliefs. Many investigations in the field of inclusive education-related self-efficacy draw on the

so-called TEIP scale (Sharma et al., 2012), as discussed previously. Hence, this section on teachers' self-efficacy is organised with regards to the three TEIP dimensions; the self-efficacy in managing behaviour, in collaboration, and in using inclusive instruction. The TEIP scales' dimensions (see Appendix E for further information) were analysed as they related to the attitude dimensions in a multivariate linear regression analysis (using the GLM in SPSS).

Table 24. Differences of the attitudes of Australian teachers with regards to their self-efficacy in managing behaviour

	F	df	p	b	95% CI	
					LL	UL
VIS	3.03	1	.08			
Intercept				0.00	-0.07	0.07
Slope				0.08	-0.01	0.17
DIF	11.84**	1	.00			
Intercept				0.00	-0.11	0.12
Slope				0.25	0.11	0.39
PRA	0.87	1	.35			
Intercept				-0.01	-0.17	0.16
Slope				0.10	-0.11	0.30
SUP	0.02	1	.90			
Intercept				-0.01	-0.18	0.16
Slope				0.01	-0.20	0.23

Note: V=0.09, $F(4, 137)=3.55$, $p<.01$; $n=142$; CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

Within the Australian sample, an overall effect of the self-efficacy in managing behaviour on all four attitude dimensions was present ($V=0.09$, $F(4, 137)=3.55$, $p<.01$). A closer look at the effects for the four dimensions (see Table 24) revealed that only the differentiation dimension was significantly related to the self-efficacy in managing behaviour. The positive slope (0.25) indicated a positive relationship of both constructs in the way that more positive self-efficacy in managing behaviour were associated with more positive differentiation-related attitudes. The 95% confidence intervals of the regression slope ([0.11, 0.39]; see Table 24) suggests that the coefficient is with considerably certainty positive. As the confidence intervals of the other attitude aspects (VIS [-0.01, 0.17]; PRA [-0.11, 0.30]; SUP [-0.20, 0.23]; see Table 24) suggest, there is some chance that the slopes were also negative. This confirms the picture as it was obtained from examining the p-values that the differentiation dimension is the only dimension which is associated with the self-efficacy in managing behaviour in a substantial way.

For the self-efficacy in collaboration, an effect for all four attitude dimension was found, too ($V=0.16$, $F(4, 137)=6.45$, $p<.01$). Similar to the self-efficacy in managing behaviour, the self-efficacy in collaboration is positively associated (see slope of 0.41 in Table 25) with the differentiation-related attitudes. As Table 25 shows, besides the differentiation dimension, the

vision of particular outcomes is significant, too. With a slope of 0.19, this aspect of attitudes is not as strong as the differentiation perspective, yet, it is significant on the one percent level. The 95% confidence intervals confirm this view that there is some certainty that in case of the vision and the differentiation the slope is positive (VIS [0.08, 0.30]; DIF [0.23, 0.59]).

Table 25. Differences of the attitudes of Australian teachers with regards to their self-efficacy in collaboration

	F	df	p	b	95% CI	
					LL	UL
VIS	11.12**	1	.00			
Intercept				0.00	-0.07	0.07
Slope				0.19	0.08	0.30
DIF	19.92**	1	.00			
Intercept				0.00	-0.11	0.11
Slope				0.41	0.23	0.59
PRA	1.62	1	.21			
Intercept				-0.01	-0.17	0.15
Slope				0.17	-0.09	0.44
SUP	0.07	1	.80			
Intercept				-0.01	-0.18	0.16
Slope				-0.04	-0.32	0.24

Note: V=0.16, F(4, 137)=6.45, $p<.01$; n=142; CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

A similar result was found for the self-efficacy in using inclusive instruction (see Table 26). The overall effect of this aspect of the teachers' self-efficacy was significant (V=0.13, F(4, 137)=5.17, $p<.01$).

Table 26. Differences of the attitudes of Australian teachers with regards to their self-efficacy in using inclusive instruction

	F	df	p	b	95% CI	
					LL	UL
VIS	6.61*	1	.01			
Intercept				0.00	-0.07	0.07
Slope				0.12	0.03	0.21
DIF	16.41**	1	.00			
Intercept				0.00	-0.11	0.11
Slope				0.30	0.15	0.45
PRA	0.95	1	.33			
Intercept				-0.01	-0.17	0.16
Slope				0.11	-0.11	0.32
SUP	0.07	1	.80			
Intercept				-0.01	-0.18	0.16
Slope				-0.03	-0.26	0.20

Note: V=0.13, F(4, 137)=5.17, $p<.01$; n=142; CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

With regards to the different dimensions, the vision and the differentiation were significant on the five percent level and on the one percent level, respectively. As Table 26 demonstrates, for the differentiation-related attitudes, the slope is 0.30, suggesting a positive association with the self-efficacy in using inclusive instruction. The vision of particular outcomes and the self-efficacy in using inclusive instruction had a slope of 0.12. Hence, a positive relation is also present amongst these two constructs. Comparable to both of the previous self-efficacy dimensions, the examination of the confidence intervals confirms the results.

4.6.2 German Teachers' Attitudes, Self-Efficacy, and Background

The results for the German teacher sample are organised comparably to the previously presented results for the Australian sample. The order goes along the order of the validation hypotheses (VHs) as they were presented in Section 3.7.3.

Gender

The analysis of gender (see VH1) differences showed no significant overall effect ($V=0.03$, $F(4, 231)=2.01$, $p>.05$). Accordingly, none of the four attitude dimensions was significantly related to the teacher being female or male (see Table 27).

Table 27. Differences of the attitudes amongst German teachers with regards to gender

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	3.72	1	.05			
Female				0.06	-0.07	0.19
Male				-0.22	-0.48	0.04
DIF	1.44	1	.23			
Female				0.04	-0.08	0.15
Male				-0.12	-0.36	0.11
PRA	0.77	1	.38			
Female				0.03	-0.11	0.17
Male				-0.11	-0.39	0.17
SUP	0.40	1	.53			
Female				0.01	-0.09	0.12
Male				-0.06	-0.27	0.15

Note: $V=0.03$, $F(4, 231)=2.01$, $p>.05$; $n=238$ (female $n=189$; male $n=47$); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

Age

The age of the teachers (see VH2) played a role for their overall attitudes ($V=0.13$, $F(12, 693)=2.67$, $p<.01$). In Table 28 the results are depicted.

Table 28. Differences of the attitudes amongst German teachers with regards to age

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	0.35	3	.79			
30 and under years				-0.02	-0.16	0.11
31-40 years				0.01	-0.33	0.35
41-50 years				0.10	-0.32	0.52
above 50 years				0.17	-0.24	0.58
DIF	0.39	3	.76			
30 and under years				-0.02	-0.15	0.10
31-40 years				0.00	-0.31	0.31
41-50 years				0.16	-0.22	0.54
above 50 years				0.11	-0.26	0.48
PRA	0.15	3	.93			
30 and under years				0.02	-0.13	0.16
31-40 years				0.00	-0.37	0.38
41-50 years				-0.14	-0.60	0.31
above 50 years				0.03	-0.41	0.47
SUP	3.08*	3	.03			
30 and under years				0.08	-0.03	0.19
31-40 years				-0.11	-0.38	0.16
41-50 years				-0.34	-0.67	-0.01
above 50 years				-0.26	-0.58	0.07

Note: V=0.13, F(12, 693)=2.67, $p<.01$; n=236 (30 and under years n=172; 31-40 years n=27; 41-50 years n=18; above 50 years n=19); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

A closer examination of Table 28 showed that only the supports dimension comprised significant differences between the age groups. With regards to the central tendencies, it is noticeable that the 30 and under years group is the only group with a positive mean, while all the other groups had negative means. The post hoc tests revealed that the only significant difference is between the 30 and under years group (0.08) and the 41-50 years group (-0.34). This significance is only indicated by Gabriel's procedure, yet, not by Hochberg's GT2. The 95% confidence intervals, confirm that the confidence interval of the 31-40 years group covers almost the full range of the confidence interval of the youngest teachers; and the confidence interval of the group of the above 50 years old teachers has also considerable overlaps with the youngest group's confidence interval. Notably, the upper limit of the 41-50 years group (-0.01) is slightly higher than the lower limit of the 30 and under years group (-0.03), which means that the significance of the mean difference needs to be interpreted with caution.

Pre-Service versus In-Service

According to Pillai's trace, a significant effect was found of being a pre-service or an in-service teacher (see VH3) with regards to inclusive education-related attitudes (V=0.10, F(4, 233)=6.36, $p<.01$).

Table 29. Differences of the attitudes of German pre-service vs. in-service teachers

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	0.00	1	.95			
Pre-service				0.00	-0.13	0.14
In-service				-0.01	-0.22	0.21
DIF	0.53	1	.47			
Pre-service				-0.02	-0.15	0.10
In-service				0.06	-0.13	0.25
PRA	0.87	1	.35			
Pre-service				0.04	-0.11	0.19
In-service				-0.09	-0.32	0.14
SUP	10.34**	1	.00			
Pre-service				0.09	-0.01	0.20
In-service				-0.23	-0.40	-0.06

Note: V=0.10, F(4, 233)=6.36, $p<.01$; $n=238$ (pre-service $n=169$; in-service $n=69$); Box's test is significant; Levene's test is significant for SUP; CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

As discussed before, Box's test and Levene's test should not be overemphasised, especially in cases where the overall number of respondents is great and the groups differ considerably in number (see for further elaboration in this respect Field, 2013). Yet, notably, Box's test was significant in this case, indicating the assumption of equality of covariance matrices was violated; and Levene's test was significant for the supports variable, indicating problems with the assumption of homogeneity of variances. Hence, the result of this test needs to be interpreted with some caution. As Table 29 demonstrates, out of the four attitude dimensions, the supports dimension is significantly different between pre- and in-service teachers. The central tendencies and the 95% confidence intervals reveal that the pre-service teachers are more positive in terms of supports-related attitudes, compared to their in-service counterparts.

Primary versus Secondary

The VH4 suggested that attitudes might be more positive for teachers from the primary sector, compared to teachers from the secondary school sector. The analysis resulted in a significant overall effect ($V=0.08$, $F(4, 222)=4.75$, $p<.01$). Table 30 draws a clear picture that teachers in the primary sector were generally more positive compared to teachers in the secondary school sector. These effects were statistically significant for the vision, the differentiation, and the general practices dimensions; yet, no significant effect was found for the support dimension. The 95% confidence intervals support the strong effect. For the dimensions that were significant according to their p-value, the confidence intervals were clearly in the positive range for the primary teacher group and in the negative range for the secondary teacher group.

Table 30. Differences of the attitudes of German primary vs. secondary teachers

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	13.76**	1	.00			
Primary				0.19	0.04	0.34
Secondary				-0.25	-0.42	-0.07
DIF	13.48**	1	.00			
Primary				0.18	0.04	0.31
Secondary				-0.22	-0.38	-0.06
PRA	8.83**	1	.00			
Primary				0.18	0.01	0.34
Secondary				-0.21	-0.40	-0.01
SUP	2.17	1	.14			
Primary				0.08	-0.05	0.20
Secondary				-0.07	-0.21	0.08

Note: V=0.08, F(4, 222)=4.75, p<.01; n=227 (primary n=131; secondary n=96); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

Years of Teaching Experience

For the teaching experience in years (see VH5), a slight overall effect was found (V=0.18, F(8, 164)=2.06, p<.05).

Table 31. Differences of the attitudes of German teachers with regards to years of teaching experience

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	1.26	2	.29			
Up to 5 years				0.01	-0.33	0.36
6-11 years				-0.35	-0.86	0.17
12 and more years				0.15	-0.20	0.49
DIF	0.19	2	.83			
Up to 5 years				0.01	-0.31	0.32
6-11 years				0.05	-0.41	0.51
12 and more years				0.14	-0.17	0.45
PRA	0.07	2	.94			
Up to 5 years				-0.01	-0.39	0.37
6-11 years				-0.13	-0.69	0.43
12 and more years				-0.02	-0.40	0.35
SUP	1.41	2	.25			
Up to 5 years				0.05	-0.25	0.34
6-11 years				-0.21	-0.65	0.22
12 and more years				-0.30	-0.59	-0.01

Note: V=0.18, F(8, 164)=2.06, p<.05; n=87 (up to 5 years n=35; 6-11 years n=16; 12 and more years n=36); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

Table 31 shows clearly that, according to the p-values, none of the four attitude dimensions reached significance. This finding is mirrored in the ninety-five confidence intervals, because all confidence intervals were overlapping considerably and nearly all were in both the positive and negative range. Hence, there seemed to be a good chance that the differences that can be obtained from the column with the model estimated means ('M' in Table 31) were only different by chance, and not because of differences between the age groups.

Higher Degree Qualification

If teachers indicated to hold a higher degree qualification or not (see VH6) seemed to make no difference with regards to their attitudes towards inclusive education *for all*. No overall effect was present ($V=0.02$, $F(4, 224)=1.26$, $p>.05$), and, as depicted in Table 32, none of the four aspects of the teachers' attitudes turned out to be significant. The 95% confidence intervals confirm this interpretation.

Table 32. Differences of the attitudes of German teachers with regards to holding a postgraduate degree/diploma or not

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	0.13	1	.72			
Postgrad.				0.09	-0.38	0.57
No postgrad.				0.00	-0.12	0.12
DIF	0.30	1	.58			
Postgrad.				0.12	-0.31	0.55
No postgrad.				-0.01	-0.12	0.10
PRA	0.01	1	.94			
Postgrad.				-0.01	-0.53	0.50
No postgrad.				0.01	-0.13	0.14
SUP	2.29	1	.13			
Postgrad.				-0.28	-0.66	0.10
No postgrad.				0.02	-0.07	0.12

Note: $V=0.02$, $F(4, 224)=1.26$, $p>.05$; $n=229$ (postgraduate degree/diploma $n=14$; no postgraduate degree/diploma $n=215$); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

Knowledge of the Local Legislation and/or Policy

An overall effect on the different dimensions of their attitudes ($V=0.17$, $F(16, 908)=2.55$, $p<.01$) was found with regards to the teachers' knowledge of the local legislation and/or policy as it pertains to inclusive education *for all* (see VH7). First of all, in Table 33 it is conspicuous that the 'very good' group has in all four dimensions a much larger central tendency compared to all other knowledge categories. Notably, within the German teacher sample, only three teachers

indicated to have ‘very good’ knowledge. Hence, within the post hoc analysis, only Hochberg’s GT2 test was interpreted. As discussed previously, this post hoc test is recommended in cases when the sample sizes are very different between groups (Field, 2013). Nevertheless, the result should still be interpreted with caution, because of the considerably low number of respondents in the high knowledge group.

Table 33. Differences of the attitudes of German teachers with regards to their inclusive education-related knowledge

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	1.83	4	.12			
<i>Very good</i>				1.25	0.22	2.28
Good				-0.01	-0.28	0.26
Average				0.07	-0.14	0.28
Poor				-0.10	-0.28	0.09
None				0.05	-0.36	0.45
DIF	3.41**	4	.01			
<i>Very good</i>				1.21	0.29	2.12
Good				0.13	-0.11	0.37
Average				0.09	-0.10	0.27
Poor				-0.18	-0.35	-0.02
None				0.04	-0.32	0.40
PRA	2.27	4	.06			
<i>Very good</i>				1.35	0.25	2.46
Good				0.05	-0.24	0.33
Average				0.11	-0.12	0.34
Poor				-0.15	-0.35	0.05
None				0.02	-0.42	0.46
SUP	3.80**	4	.01			
<i>Very good</i>				1.35	0.54	2.16
Good				-0.14	-0.35	0.07
Average				0.12	-0.04	0.29
Poor				-0.02	-0.16	0.12
None				-0.12	-0.45	0.20

Note: V=0.17, F(16, 908)=2.55, $p<.01$; $n=232$ (very good $n=3$ [italics because of low number of respondents in this group]; good $n=44$; average $n=71$; poor $n=95$; none $n=19$); Levene’s test is significant for SUP; CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

The significant overall effect seems to emanate to a great extent from knowledge differences with regard to the differentiation and the supports dimension, as can be found in Table 33. Concerning the differentiation that teachers were willing to undertake, the knowledge groups consisting of ‘very good’ and ‘poor’ were significantly different to each other. The 95% confidence intervals confirm that those with very good knowledge were by far more favourable with regards to differentiate the teaching, compared to those with poor knowledge. Concerning the supports dimension, the Levene’s test is significant, which suggest to interpret the results with caution. Comparable to the differentiation, those with very good knowledge were

significantly more positive with regards to supports related attitudes. The post hoc tests indicated very good being significantly different from all the other categories. The 95% confidence intervals confirm this result.

Training in Inclusive Education for All

An overall effect on the five percent level was found for the amount of training focusing on inclusive education of all students (see VH8) ($V=0.08$, $F(8, 444)=2.37$, $p<.05$). Especially with regards to the differentiation dimension (teachers seemed to differ significantly according to their amount of training (see Table 34). The examination of the post hoc tests suggests that those who had no inclusive education-related training had significantly less favourable attitudes compared to those with some training or a high amount of training. The additional examination of the 95% confidence intervals confirmed at least that the lower level of the high group (-0.01) was above the upper level of the 'none' group (-0.02), which suggests that both groups were indeed significantly different. Yet, although the means were different between high and some training, the confidence interval of the high amount of training group was relatively wide, included most of the range of the confidence interval of the 'some'-training group. This suggests that the found difference has to be interpreted with some caution, because the difference between both might be found by chance.

Table 34. Differences of the attitudes of German teachers with regards to their level of inclusion-related training

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	1.72	2	.18			
High				0.26	-0.25	0.77
Some				0.10	-0.11	0.30
None				-0.10	-0.25	0.05
DIF	5.45**	2	.00			
High				0.44	-0.01	0.90
Some				0.15	-0.04	0.33
None				-0.15	-0.28	-0.02
PRA	2.47	2	.09			
High				0.28	-0.26	0.83
Some				0.13	-0.09	0.35
None				-0.13	-0.29	0.03
SUP	0.28	2	.76			
High				-0.03	-0.44	0.39
Some				0.04	-0.12	0.21
None				-0.03	-0.15	0.09

Note: $V=0.08$, $F(8, 444)=2.37$, $p<.05$; $n=227$ (high $n=12$; some $n=75$; none $n=140$); CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

Experiences with Inclusive Classroom Settings

The teachers' experiences with inclusive classroom settings (see VH9) had an overall effect on their attitudes ($V=0.35$, $F(12, 675)=7.52$, $p<.01$). This effect was present with regards to all four dimensions (see Table 35). For all four attitude aspects, the post hoc analysis and the examination of the confidence intervals demonstrates that negative experiences were associated with substantially unfavourable attitudes. For the supports dimension those with negative experiences in inclusive classroom settings had less favourable attitudes compared to all other groups. For the vision, differentiation, and general practices, another differentiation was found; namely that teachers with negative experiences had most negative attitudes, those with neutral and no inclusive experiences had middling attitudes, and those with positive experiences had more positive attitudes. These differences were also significant and substantial (according to the 95% confidence intervals). The overall picture suggests that negative experiences are worse for attitudes and that even no experience at all are even better. Positive experiences have extra effects for vision, differentiation, and general practices; yet, not for supports-related attitudes.

Table 35. Differences of the attitudes of German teachers with regards to their inclusive education-related experiences

	F	df	p	M ^a	95% CI	
					LL ^a	UL ^a
VIS	26.32**	3	.00			
Positive				0.57	0.36	0.79
Neutral				0.04	-0.18	0.27
Negative				-1.06	-1.36	-0.77
None				0.02	-0.13	0.17
DIF	23.74**	3	.00			
Positive				0.54	0.34	0.74
Neutral				0.13	-0.07	0.34
Negative				-0.86	-1.13	-0.60
None				-0.06	-0.20	0.07
PRA	26.69**	3	.00			
Positive				0.61	0.38	0.85
Neutral				0.11	-0.13	0.36
Negative				-1.16	-1.47	-0.84
None				0.00	-0.17	0.16
SUP	10.89**	3	.00			
Positive				0.27	0.08	0.46
Neutral				0.13	-0.07	0.33
Negative				-0.63	-0.89	-0.37
None				0.00	-0.13	0.13

Note: $V=0.35$, $F(12, 675)=7.52$, $p<.01$; $n=230$ (positive $n=50$; neutral $n=47$; negative $n=28$; none $n=105$); Box's test is significant; Levene's test is significant for VIS, ADJ, and PRA; CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

^a Model-estimated standardised marginal means and confidence intervals.

Teachers' Self-Efficacy as it Pertains to Inclusive Education

Comparable to the Australian teachers, for the German teachers the effects of self-efficacy (see VH10) on the different attitude dimensions were analysed separately for the three aspects of self-efficacy: self-efficacy in managing behaviour, in collaboration, and in using inclusive instruction. In the following section, multivariate linear regression analyses (using the GLM in SPSS) are presented.

For the overall model of the self-efficacy in managing behaviour, a significant effect was indicated by the Pillai's trace test ($V=0.06$, $F(4, 231)=3.46$, $p<.01$). Yet, a closer examination of the results in Table 36 demonstrate that there is none of the four attitude aspects significantly associated with the teachers' self-efficacy in managing behaviour.

Table 36. Differences of the attitudes of German teachers with regards to their self-efficacy in managing behaviour

	F	df	p	b	95% CI	
					LL	UL
VIS	0.17	1	.68			
Intercept				0.01	-0.11	0.12
Slope				0.03	-0.12	0.18
DIF	3.34	1	.07			
Intercept				0.00	-0.10	0.11
Slope				0.13	-0.01	0.26
PRA	0.52	1	.47			
Intercept				0.00	-0.12	0.13
Slope				0.06	-0.10	0.22
SUP	0.34	1	.56			
Intercept				0.00	-0.09	0.09
Slope				-0.04	-0.16	0.09

Note: $V=0.06$, $F(4, 231)=3.46$, $p<.01$; $n=236$; CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

Concerning the teachers' self-efficacy in collaboration, a significant effect was found for all four attitude dimensions ($V=0.14$, $F(4, 231)=9.18$, $p<.01$). The results for the four different dependent variables were all significant, as it is depicted in Table 37. The dimensions vision, differentiation, and general practices were all significant on the one percent level, while the support dimension was significant on the five percent level. This finding is represented in the high slopes of the three dimensions (0.46 for vision, 0.55 for differentiation, and 0.54 for general practices), while the slope for the support dimension is lower, yet still significant (0.22). For all four dimensions, the 95% confidence intervals confirms that the slope is very likely to be positive. For the vision, differentiation, and general practices dimensions, the lower level is relatively high at 0.25, 0.36, and 0.31, respectively; suggesting a strong relationship between these variables.

Table 37. Differences of the attitudes of German teachers with regards to their self-efficacy in collaboration

	F	df	p	b	95% CI	
					LL	UL
VIS	17.64**	1	.00			
Intercept				0.01	-0.11	0.12
Slope				0.46	0.25	0.68
DIF	31.75**	1	.00			
Intercept				0.00	-0.09	0.10
Slope				0.55	0.36	0.74
PRA	20.82**	1	.00			
Intercept				0.00	-0.12	0.12
Slope				0.54	0.31	0.77
SUP	5.60*	1	.02			
Intercept				0.00	-0.09	0.09
Slope				0.22	0.04	0.40

Note: $V=0.14$, $F(4, 231)=9.18$, $p<.01$; $n=236$; CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

The German teachers' self-efficacy in using inclusive instruction has a significant overall effect on the different aspects of the teachers' attitudes ($V=0.11$, $F(4, 231)=6.99$, $p<.01$). An examination of the different attitude dimensions reveals that this effect is to a large extent based on the significant effect with regards to the differentiation dimension and the general practices dimension. For both dimensions, a positive slope was found, indicating that higher self-efficacy in using inclusive instruction goes together with a more favourable differentiation and general practices-related attitude. The 95% confidence intervals confirm the positive relationship between these variables.

Table 38. Differences of the attitudes of German teachers with regards to their self-efficacy in using inclusive instruction

	F	df	p	b	95% CI	
					LL	UL
VIS	3.52	1	.06			
Intercept				0.01	-0.11	0.12
Slope				0.16	-0.01	0.32
DIF	13.21**	1	.00			
Intercept				0.00	-0.10	0.11
Slope				0.27	0.12	0.41
PRA	4.74*	1	.03			
Intercept				0.00	-0.12	0.13
Slope				0.19	0.02	0.37
SUP	0.00	1	.97			
Intercept				0.00	-0.09	0.09
Slope				0.00	-0.14	0.13

Note: $V=0.11$, $F(4, 231)=6.99$, $p<.01$; $n=236$; CI=Confidence Interval; LL=lower limit; UL=upper limit.

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

4.6.3 Synopsis of the Results for the Australian and the German sample

In the previous sections, a large number of results were presented as they were found for the Australian and the German sample. Due to the richness of the findings, it remained difficult to present an overview of all of the results as they pertained to the associations between the attitude dimensions and conceptually related aspects and constructs. Hence, before discussing the obtained results in the next chapter, a synopsis of the empirical results is presented to gain a comprehensive view of the associations, as they were found in the Australian and German data of the present study. Comparable to the findings presented so far, the order of this summary is comparable to the order of the validation hypotheses as they were presented in Section 3.7.3.

Concerning gender (see VH1), there was generally no significant effect on the teachers' attitudes towards inclusive education *for all* in both the Australian and the German sample. Concerning the age of the teachers (see VH2), in both samples a strong overall significant effect was apparent. This effect was especially present with regard to the support dimension. Although not perfectly linear, the general trend was that younger teachers tended to be more positive towards the supports-related aspect of inclusive education *for all*, compared to the older generations of teachers. Specifically in the Australian data, there was also a significant effect with regards to the general practices dimension. This effect was not present in the German data.

A significant overall effect was found for Australian and German teachers between pre-service and in-service teachers (see VH3). Comparable to the previously mentioned age-effect, this effect was apparent primarily with regards to the support dimension. In addition, a minor effect was present in the Australian data with regards to the general practices dimension. The general tendency was that teachers in the pre-service sector hold more positive attitudes compared to their in-service counterparts. Another aspect was, if the teachers were teaching/or in training to teach in primary or secondary schools (see VH4), which was clearly related to the teachers' attitudes, as significant overall effects in the Australian and the German sample demonstrated. Concerning the differentiation dimension this effect was present in both samples (in the German sample the effect was stronger, compared to the Australian sample). In addition, in the German sample, strong effects were found with regards to the vision dimension and the general practices dimension. Generally, the direction of the effect was stable, suggesting that in the primary sector, teachers had more positive attitudes compared to secondary sector teachers. The teaching experiences (in years) (see VH5) was another aspect. Yet, no overall significant effects were found in this regard, in the Australian sample, and only a relatively weak effect in the German sample. The absence of substantial overall effects is mirrored by the absence of any effect concerning the different attitude dimensions. Similarly, there was no strong evidence that attitude-differences were present between teachers, who were holding a

higher degree qualification and those who did not have any (see VH6). Only a slight overall effect was found in the Australian data, yet, the corresponding significant result on the vision dimension was not substantial.

The inclusive education-related background was investigated in the present study with regard to the knowledge, training, and experiences, as they pertained to inclusive education *for all*. Concerning the knowledge with regard to inclusive education policies and legislation (see VH7), a significant association with the attitudes (particularly the differentiation and the support dimensions) was only present in the German case. Yet, as reported previously, this result needs to be interpreted with some caution. Similarly, if the teachers had received training with regards to inclusive education *for all* (see VH8) there were also significant differences, however this was only the case in the German data. Here, the effect for German teachers is only present with regards to the differentiation dimension. A clear picture was suggested with regards to the inclusive classroom-related experiences of the teachers (see VH9). Both samples demonstrated overall effects of experiences on attitudes. This effect was clearly present in all four attitude dimensions in both samples. The overall message of these analyses was that most unfavourable attitudes can be found among those teachers who reported negative experiences with inclusive settings. A little bit more favourable were those with no experiences with inclusive settings, at all. The most favourable attitudes reported were those teachers who had experienced positive inclusive settings.

For the three self-efficacy dimensions concerning managing behaviour, collaboration, and using inclusive instruction (see VH10), an overall effect was clearly present for both the Australian and the German sample. As opposed to the significant overall effect for self-efficacy in managing behaviour, there were no substantial effects for each of the attitude dimensions. There was only one significant association with the differentiation dimension in the Australian sample. The absence of more specific effects, left some doubts, if the overall effect could actually be considered substantial. Self-efficacy in collaboration resulted in significant overall effects on the attitudes in both samples. Stronger self-efficacy to carry out collaboration was clearly associated with more positive attitudes on the vision and the differentiation dimensions. Only for the German sample, additional associations were found for the general practices and the supports dimensions. For the self-efficacy in using inclusive instruction, a significant overall effect was found for both the Australian and the German sample. This kind of self-efficacy seemed to be particularly related to the differentiation dimension of the attitudes. An effect was present in the vision dimension for Australia, and an effect was present in the general practices dimension for Germany. If all self-efficacy dimensions were interpreted together, the

tendency appeared that the self-efficacy was particularly related to differentiation-related attitudes.

4.7 Chapter Summary

This chapter comprised the results as they pertained to the study's purpose to develop a new and sound instrument to measure the teachers' attitudes towards inclusive education *for all*. Relevant indicators in English language were found through an extensive literature search and the revision of potential questionnaire items. The outcome of the translation and adaptation of these statements in German mirrored the attempt to create the new instrument being robust in cross-cultural settings. For both versions, face-validity was established through in-depth examination of each items' content and additional feedback from teachers. The empirical data, which was drawn from Australian and German teachers, was used to examine the internal structure of the new instrument to measure the teachers' attitudes towards inclusive education *for all*, which resulted in four dimensions. These were the vision, the differentiation, the general practices, and the supports as they pertain to inclusive education *for all*. How these four aspects of the new instrument were related to other teacher variables was examined through statistical analyses. The presentation of the results was concentrated on describing the outcome of the different analyses. All interpretations and discussions of the obtained results will be presented in the next chapter.

Chapter 5 · Discussion

5.1 Introduction

The present study's overall purpose was to develop and substantiate a new instrument to measure teachers' attitudes towards inclusive education *for all*. This purpose informed the methods of the present study and the obtained results clearly reflected the steps which were undertaken to achieve this ambitious goal. The Discussion Chapter reflects the quality of the new instrument in different regards. In a first section, the obtained results are interpreted. This section starts with examining the individual indicators, then, the dimensions are discussed and, finally, conclusions on the validation hypotheses are presented. After discussing a variety of limitations of the present study, implications of the present study's findings for practice and further research are discussed. Concluding remarks on the whole present investigation are given at the end of this chapter.

5.2 Interpretation of the Obtained Results

The literature review of the present study uncovered the need to develop a new instrument to measure the teachers' attitudes with regards to their views on inclusive education *for all*. As opposed to previous instruments that measured the teachers' thinking with regards to inclusive education *for some*, the present research study attempted to find a robust and sound instrument for measuring the teachers' attitudes towards inclusive education *for all*. As will be discussed in the following sections, the obtained results demonstrated considerable evidence that this purpose was achieved.

5.2.1 Indicators of Inclusive Education for All

The process of developing items that could be utilised for measuring the teachers' attitudes towards inclusive education *for all* resulted initially in 38 statements which were obtained from 12 studies (see Table 6 on p. 120). These items were translated and adapted for use in the German-speaking context.

As can be found in the previous presentation of the results and for example in the original questionnaire (see Appendix J and Appendix K), each of the 38 items were generally more focused on inclusive education *for all*, rather than on inclusive education *for some*. Hence, it can be concluded that the procedures, which were utilised for developing the new item pool turned out to be efficient.

As discussed previously with regards to methodological considerations (Bühner, 2011; DeVellis, 2011; Jonkisz et al., 2012; Rost, 2004), there was a trade-off between the need for a completely new operationalisation of the new construct (which would have suggested formulating completely new items), and the issue that the development of effective new questionnaire items is considerably difficult (which would have suggested using as many established items as possible). It was decided to start by systematically reviewing the literature for potential indicators. This decision comprised the assumption that appropriate indicators for the teachers' attitudes towards inclusive education *for all* would already exist in former scales, but that they would be mixed in the former scales with a larger number of narrow-focused indicators of attitudes towards inclusive education *for some*. The procedures utilised in the present study, then, assumed it would be possible to disentangle these indicators and to use only those indicators with a wider understanding of inclusive education *for all*. In research on teachers' attitudes towards inclusive education, such a strategy for developing a new scale does not seem to be unusual. Such a strategy was carried out by other researchers, who recently developed new scales on inclusive education *for some*, such as De Boer and colleagues (De Boer et al., 2011; De Boer et al., 2012), who utilised items from Stoiber et al. (1998), Avramidis et al. (2000b), and Mahat (2008), or for example Boyle et al. (2013), who utilised items from Van Reusen et al. (2001), Wilczenski (1992), Villa, Thousand, Meyers, and Nevin (1996) and Avramidis et al. (2000a).

Yet, to adapt former items for the operationalisation might actually be part of the problem. Even if conceptual understandings of inclusive education *for all* were generally acknowledged in current empirical studies, these studies selected and adapted indicators from a variety of former studies that utilised questionnaire statements, which represented ideas of mainstreaming and integration (which was demonstrated by the most recent literature reviews, such as Nilholm & Göransson, 2017; Ruberg & Porsch, 2017). In this way, the general idea of 'mainstreaming' is clearly represented by the instrument developed by Larrivee and Cook (1979), which was an innovative tool in the 1970s in the United States. Yet, the basic logic, which is inscribed into the statements of this instrument, did not change due to updating just some terms, such as for example 'mainstreaming' into 'integration' (Antonak & Larrivee, 1995), and later into 'inclusion' (Monsen et al., 2015; Taylor & Ringlaben, 2012). Similarly, in the 1990s,

Wilczenski (1992, 1995) concentrated on the feasibility of particular students (requiring special physical, academic, behavioural, or social accommodations) being placed in regular classrooms together with regular students, which resulted in the Attitudes Toward Inclusive Education Scale (ATIES). This placement logic of the ATIES, which assumed that there were special and regular students, which could be placed either in special or regular education, was not overcome by researchers who were building their new scale on the works of Wilczenski (1992, 1995), such as Forlin et al. (2011) who developed the Sentiments, Attitudes, and Concerns about Inclusive Education (SACIE) scale, which operationalised its attitude dimension similar to the ATIES. Even the most recent studies that used the ATIES (Sharma et al., 2015; Tsakiridou & Polyzopoulou, 2014) or the SACIE (Hecht et al., 2016; Montgomery & Mirenda, 2014; Yada & Savolainen, 2017) were not reflecting this underlying placement logic of the measurement instrument. Notably, to try explicitly not to adapt a previous instrument, but to develop a new one, does not prevent one from falling into this trap. Authors who attempted to completely write new items, such as Saloviita (2015), stayed in the mainstreaming/integration tradition of former empirical studies.

Taken together, empirical research in itself demands to acknowledge former research and to systematically review former operationalisations; yet, from time to time it seems worth reflecting the part that empirical research worldwide plays in the ‘re-runs of old theatre’; an expression that Slee and Allan (2001) coined, in order to point to the need to deconstructing traditional thinking on inclusive education before being able to establish more inclusive practices *for all*. In this way, the results obtained throughout the present study were considerably difficult to achieve, because of the tensions that arouse from acknowledging the efforts of all of the previous research on teachers’ attitudes towards inclusive education (*for some*) on the one hand, and from trying to put the innovativeness of the present study (towards a *for all*-related understanding) to the forefront on the other hand.

As demonstrated in the beginning of the present study, all of the empirical studies on teachers’ attitudes that were cited throughout the present study, were primarily focused in their empirical methods on the placement of particular students into the regular system and/or they were focused on students with special educational needs and/or disabilities (SEND). Notably, one exception is the relatively loose collection of quite different attitude items, which was recently published in the study of Dizdarevic, Mujezinovic, and Memisevic (2017). In this study, some aspects of the teachers’ attitudes towards a wider understanding of inclusion has been operationalised for Bosnia and Herzegovina and the European Union. Yet, Dizdarevic et al. (2017) leave many methodological questions unanswered and the quality of all of the items utilised in their study needs to be considered as being not sufficient. For example, all of the

items have only little variance (such as ‘All new students are helped to settle into the school’; Dizdarevic et al., 2017). In the present study, this was the case for item four, too, which was too broadly worded, hence, approved by many respondents, and accordingly it was discarded, because of insufficient statistical characteristics.

All in all, there is still a clear research gap with regards to the teachers’ attitudes towards inclusive education *for all* (at least in the English-/German-speaking literature). The most recent literature reviews (Nilholm & Göransson, 2017; Ruberg & Porsch, 2017) started to recognise and problematise this fact. Hence, to reach a number of 38 items with (nearly) no emphasis on inclusive education *for some*, can be considered a success, and a direct response to the most recent problematisation of this issue.

After all statistical analyses, which were presented previously, from the original 38 items, only 12 items remained, which were thought to have the best quality and which were considered being relevant and valid indicators in both contexts. Each of the 12 indicators can be considered to have sufficient content validity. The items are in line with the most recent thinking on inclusive education *for all*, and it can be assumed that utilising these indicators in empirical studies would obtain relevant information from teachers about how they view the 12 different aspects as they pertain to inclusive education *for all*.

5.2.2 Internal Structure of the New Instrument

The examination of how the items were related to each other resulted in a structure of four dimensions, comprising three items for each dimension. These four dimensions represented the vision, the differentiation, the general practices and the supports as they pertain to inclusive education *for all*. Hence, the purpose of the present study to develop a sound new instrument that might comprise certain dimensions, can be considered fulfilled.

Vision of Inclusive Education for All

The vision of inclusive education *for all* dimension comprised notions in what way inclusive education *for all* has certain benefits. According to the indicators of this dimension, the positive results that might be achieved through inclusive education *for all* were appropriate behaviour for all students, understanding of differences among students, and social inclusion. All the aspects that relate to this dimension have the general form that inclusive education *for all* leads to short-term and long-term outcomes. In the sense of there being a certain vision of what can be achieved through inclusive education *for all*, this dimension was called ‘vision’ or ‘vision of inclusive education *for all*’.

Other scales comprised dimensions, which were comparable to the vision dimension, too. One relatively obvious example is Taylor and Ringlaben (2012), whose item six ('Inclusion offers mixed group interaction which will foster understanding and acceptance of differences') was adapted for the item 35 of the present study. Taylor and Ringlaben (2012) put this item together with other items that all represent the idea of 'diversity acceptance'. This aspect has large overlaps with the scales that Taylor and Ringlaben (2012) have utilised in their study; namely the Opinions Relative to Integration scale (ORI; Antonak & Larrivee, 1995) and the Opinions Relative to Mainstreaming scale (ORM; Larrivee, 1982; Larrivee & Cook, 1979). The according dimension in the ORM is called the 'general philosophy' of mainstreaming, while the ORI termed it the 'benefits of integration'. These dimensions seem to be generally comparable to the vision dimension postulated in the present study. Notably, the original ORM from the 1970s (Larrivee & Cook, 1979) was developed into the ORI (Antonak & Larrivee, 1995), which was further adapted to be used by Taylor and Ringlaben (2012) and others (Avramidis et al., 2000a, 2000b; Dupoux, Wolman, & Estrada, 2005; Monsen et al., 2014). Moreover, the benefits of inclusion dimension are still present in one of the most recent and sound updates of the ORM; namely, the Teacher Attitude towards Inclusion Scale (TAIS; Monsen et al., 2015). However, the crucial difference compared to the vision dimension postulated in the present study is that the benefits dimension of the TAIS clearly addressed benefits for students with and without SEND, while the present study's vision dimension was focused on benefits *for all*.

Other studies postulated attitude dimensions with a similar general logic, too. T. Bennett et al. (1997) found an attitude dimension called 'general attitudes toward inclusion', which comprised notions of inclusion benefitting the students without SEND, and inclusion being a positive change in general. In addition, another indicator focussed on better support for students with SEND, and another item was just prompting that every student with SEND should be in the inclusive classroom. Hence, to a large extent, this dimension comprised notions of positive outcomes of inclusive education. Although the focus of this dimension on general attitudes toward inclusion (T. Bennett et al., 1997) was focused on disabilities and general/special education, the basic idea seemed similar to the focus on outcomes of the vision dimension in the present study. Another perspective on outcomes of inclusive education is presented in the study of Bosse and Spörer (2014); one of their dimensions was called the 'attitudes concerning the effects of inclusive classes'. A related dimension was postulated by Moberg (2003), who found different outcome-related items to be associated in a factor that was called 'social justice'. In the study of Stoiber et al. (1998) such a dimension was found, too, which the authors called 'expected outcomes'. All of these dimensions were clearly related to benefits, outcomes,

effects, etc. solely with regard to students with or without SEND. Nevertheless, the referenced studies provide certain evidence that an instrument that is supposed to measure the teachers' attitudes with regards to inclusive education (for all) might need at least one dimension that asks the teachers to indicate if they think that inclusive education (for all) has certain benefits and positive outcomes (for all).

Differentiation as it Pertains to Inclusive Education for All

The differentiation dimension comprised notions, which were related to teaching, which teachers carry out in order to cater for all of the students. The indicators of this dimension comprised to carry out differentiated adjustments, and to carry out adaptations of the curriculum, and of the assessment. These indicators have in common to refer to the differentiation of teaching that teachers were willing to carry out; accordingly, the dimension was termed 'differentiation as it pertains to inclusive education *for all*'.

One of the dimensions proposed by T. Bennett et al. (1997) was about the 'feasibility of inclusion'. This dimension seemed in a way relatively comparable to the differentiation dimension of the present study, because it comprised an evaluation of certain teaching practices as they pertained to ideas of inclusive education *for some*. This dimension comprised indicators on classroom practices in inclusive settings, where children with disabilities were present in the general classroom. All of these items comprise negative examples of practices, such as finding aides distracting, and finding individualised instruction, including of students with SEND, and meeting the needs of students with SEND difficult to be carried out.

Other dimensions that were in a way related to the differentiation dimension are the classroom management dimension of the ORI (Antonak & Larrivee, 1995) and the teaching practice dimension of the TAIS (Monsen et al., 2015). Yet, the indicators of these dimensions actually lack a clear statement about the personal intent of a respondent to carry out certain practices. Hence, the indicators utilised by authors such as Antonak and Larrivee (1995) and Monsen et al. (2015) are more about practices in principle. In this way, De Boer et al. (2011) noted in their review of attitude instruments that the TAIS (Monsen et al., 2015), which was also used in a comparable form in an earlier study of Monsen and Frederickson (2004), only touched upon the cognitive dimension of attitudes, but not upon the behavioural or the affective.

In the literature review of the present study, as part of the definition of inclusive education *for all*, it was emphasised that exclusion needs to be tackled. Accordingly, the teachers need to play a key role in carrying out inclusive education *for all*. In this way, the differentiation dimension, which comprised the feasibility of differentiated adjustments and the willingness to adapt the curriculum and the assessment, can be considered to give emphasis to the active part

that teachers need to play in carrying out teaching practices, which are inclusive *for all*. Hence, the differentiation dimension seems to fit well with the general scope of inclusive education *for all*.

General Practices as they Pertain to Inclusive Education for All

Another dimension was about general practices as they pertained to inclusive education *for all*. This dimension focussed on the subjective view of the teachers on *possibilities* of adaptations of practices *under certain circumstances*. The indicators of this dimension comprised conditional statements, which stated that classes are suitable for all, *if* organised appropriately; any student can learn, *if* the curriculum is adapted; and all students can be taught, and *if* teaching is differentiated. Compared to the differentiation dimension, which focussed on concrete differentiation that the teachers were actually willing to carry out, the general practices dimension was more focussed on a general view on the feasibility of certain practices, if the circumstances would allow these practices.

The general practices dimension seemed to be not very common in other studies. The dimension comprises conditional thinking; namely, *if* X is the case, *then* Y would be possible. Items with such a conditional character can be found for example in the scale developed by Moberg (1997). There is a range of items suggesting different outcomes, if students with SEND were placed full-time in the regular classrooms (such as ‘Students with mild disabilities would experience more academic failure if they were placed full-time in the regular classrooms’). Yet, clearly, these items do not refer to actual practices of the teachers; but solely to the placement of a student with SEND.

Notably, the statements as they pertain to the general practices dimension are in line with inclusive education *for all*-related thinking; hence, it seemed to be a valid dimension within an instrument that this supposed to measure different aspects of inclusive education *for all*.

Support as it Pertains to Inclusive Education for All

The supports as they pertain to inclusive education *for all* dimension comprised different notions of support for the teacher to carry out more inclusive teaching practices. The aspects of support were adequate personnel from outside school, adequate personnel within school, and adequate resources. All the indicators, as they were utilised in the present study, emphasised the goal to ‘address the unique educational needs of all students’ (wording from the questionnaire), and, in order to achieve this, specific supports were needed. In this way, the dimension was called support/s.

Supports as they are perceived by the teachers seem to be a relevant dimension, because they are present in different former scales on the teachers' inclusion-related attitudes. One of these instruments was developed by Al Zyoudi et al. (2011). Besides the general beliefs and the teacher preparation, one dimension was concerned with the availability of resources. These resources comprised specialists (such as special education teachers, speech and language specialists, etc.), but also, more generally, the schools' facilities. Notably, Al Zyoudi et al. (2011) focussed to a large extent on students with SEND; accordingly the notion of particular resources pertained primarily to the special education of students with SEND. In the study of Al Zyoudi et al. (2011), the authors asked the teachers how they judged the availability of resources as they pertained to inclusive education (for some). If inclusive education *for all* is taken as a basis, these resources might need to be interpreted as pertaining to all students, and not to some specifically.

Cullen, Gregory, and Noto (2010) investigated how teachers saw students with mild/moderate disabilities and found a dimension that they termed 'perception of professional roles and functions'. The idea of this dimension was that respondents indicated in a general way how positive their view was with regards to team teaching and collaborating with special education teachers. Contrary to this, in the present study, much emphasis is not only given to evaluating the supports, but also to if these supports are playing a crucial role for teachers to be able to address the needs of all. Similarly, T. Bennett et al. (1997) emphasised the 'confidence in the ability to carry out inclusion' in one dimension. This dimension referred to the help of other team members, to sufficiently support staff, and also the teachers' skills, availability of information, and related training. Compared to the present study, this aspect of the teachers' attitude is relatively similar to the supports dimension. In the study of T. Bennett et al. (1997) the focus was on aspects as they pertained to include students with SEND, while in the present study the focus was on catering *for all students*. Besides this difference, the notions of supports were relatively comparable. Yet, T. Bennett et al. (1997) had included more aspects into this dimension; namely, aspects as they are related to the professional development of the teachers (the teachers' skills, availability of information and related training). These aspects are comparable to what was called 'teacher preparation' as one aspect of the teachers' attitudes in the study of Al-Zyoudi (2006). These aspects were not present in the supports dimension of the present study. On the one hand, they would complement the obtained picture; but on the other hand, it seems open to debate if skills, availability of information and related training can actually be considered to be an attitude. In this way, McGhie-Richmond, Barber, Lupart, and Loreman (2009) called only one of their dimensions 'attitude', while other dimensions, such as 'support and training' were not called attitude. For example, besides the attitudes, Monsen et

al. (2014) utilised the Adequacy of Support questionnaire, which comprised the perceived adequacy of support available to them. In Monsen et al.'s study the teachers' attitudes and the teachers' judgement of the adequacy of support were two different constructs (2014). Yet, in the present study, the dimension has a slightly different direction. Namely, the supports dimension comprised the teachers' views on how *s/he personally feels supported* by other persons or by resources in order to carry out teaching practices that allows to address the needs of all. In this way, the present supports dimension seems to refer more to the personal attitudes towards inclusive education *for all* that a teacher might hold than to a completely different kind of thinking solely about the supports.

The confirmatory factor analysis of the four-factor model (see Figure 5 on p. 148) demonstrated the supports being strongly related to the general practices dimension, while being not so strongly related to the vision and the differentiation. This was in a way surprising, because the formulation of the items pointed clearly towards supporting the teacher in carrying out more inclusive practices; which would have suggested that the supports and the differentiation might be closer related to each other. On the other hand, from the previously discussed studies it seems evident that the association of resources (such as support) and the actual teaching practices (such as differentiation) is not as strong as one would suggest (see the weak association between resources, such as class size etc., and teaching-quality, as discussed in Section 2.3.1). Future investigations might need to further examine this point.

Taken together, the support appears as an important and valid aspect of the teachers' attitude towards inclusive education *for all*.

Discussion of the Structure of the New Scale

Although a theoretical discussion of teachers' attitudes (or even the development of a model or theory of teachers' attitudes towards inclusive education *for all*) goes way beyond the scope of the present study, a discussion of the attitude structure as it was obtained in the present study might inform future studies in carrying out more foundational research on attitude theory in the area of inclusive education *for all*.

Some of the empirical studies on inclusive education *for some* have argued that teachers' attitudes have three components; namely, affective, behavioural and cognitive (De Boer et al., 2012; Gregory & Noto, 2012; Mahat, 2008). This perspective was discussed earlier in the present study (see discussion in Section 2.3.1), and it was noted that such a perspective has a certain justification. Yet, it represents to a large extent a social-psychological perspective on the sheer functioning of attitudes, and not on the content of the attitudes. Other empirical researchers in the field of inclusive education *for some* have emphasised that particular kinds

of SEND (such as physical, academic, behavioural, or social) correspond with different dimensions of the teachers' attitudes (Tsakiridou & Polyzopoulou, 2014; Van der Veen et al., 2010; Wilczenski, 1992, 1995). As it was discussed throughout the present study, many studies existed that postulated more content-related dimensions of the teachers' attitudes. In the present study, such content-related attitude dimensions were found. If the content of these dimensions are related to one another, an interesting picture emerges. As noted previously, the support dimension comprised personnel and resources that support the teacher so that s/he can carry out more inclusive practices. The differentiation dimension makes suggestions about the teachers' willingness to carry out such kind of teaching; in other words, the differentiation dimension links the teacher and the teaching. The general practices dimension relates *potential* teaching practices with particular outcomes under certain circumstances. Hence, inclusive education *for all*, as teachers see it, starts with particular resources (and a particular context), that the teacher uses to carry out inclusive teaching practices, in order to reach certain outcomes. The general ideas, as they are represented in the four dimensions of the new scale, relate to established models of teaching, such as the Offer and Use Model of Classroom Effectiveness (Helmke, 2009), which assumes that teaching starts with the teacher, who influences the teaching quality. Teaching ultimately leads (through learning activities of the students) to student outcomes. The context and resources influence all aspects, including the teacher, the teaching, the learning activities of the students, and the student outcomes. In this way, the teachers' thinking, as it is represented in the new scale, seems to be comparable to models of (effective) classroom teaching. Hence, it might be justified to consider the dimensional structure as being valid and sound.

5.2.3 Conclusions on the Validation Hypotheses

The development of the scale and the examination of its internal structure demonstrated certain strengths of the procedures, which were followed in the present study. This section discusses the evidence regarding relationships with conceptually related constructs (see Standard 1.16 of the AERA et al., 2014). Hence, in the following section, the new scale is substantiated further by discussing the obtained relations of the attitudes with other teacher aspects against the background of the proposed validation hypotheses (VHs; see Section 3.7.3).

Conclusion Regarding the VH1: Gender of the Teachers

In Section 3.7.3, the following validation hypothesis was proposed:

- VH1: Gender is not assumed to have a substantial effect on the attitudes towards inclusive education *for all*.

Because of the contradicting evidence from former studies, a strong effect of *gender* in a particular direction was not expected. In a variety of studies, males were found to be more positive (Ahmad, 2012; Ahmmed et al., 2012; Bhatnagar & Das, 2014; Sharma et al., 2015); and in other studies it was found that female teachers tended to hold more positive attitudes (Alghazo & Gaad, 2004; Avramidis et al., 2000a; Boyle et al., 2013; Saloviita & Schaffus, 2016; Tsakiridou & Polyzopoulou, 2014). In the present study, gender played no significant role with regards to attitudes towards inclusive education *for all*. The multivariate analysis of variance found no overall effect in the Australian sample, nor in the German sample. Despite this, a minor effect appeared in the Australian sample that the female teachers tended to be more in favour of differentiating their teaching. Yet, an examination of the data revealed the evidence being considerably weak.

A closer look at the most recent studies reveals indeed that gender differences were relatively weak. The gender difference in favour of female teachers, reported by Saloviita and Schaffus (2016) was relatively weak for their Finnish sample, and not present for their German sample. Comparable to this result, in the Greek sample, analysed by Tsakiridou and Polyzopoulou (2014), the gender differences with regards to ‘behaviour problem’-related attitudes were also only significant on the five percent level. For the studies that reported males holding more favourable attitudes, similarly weak evidence was found, such as in the studies of Sharma et al. (2015) and Bhatnagar and Das (2014).

Taken together, from a critical perspective, the evidence presented in recent attitude studies with regards gender seems not to suggest any consistent effect of gender in any direction. In addition, from a relatively neutral position, there would be no reason to assume gender differences with regards to inclusive education *for all*. Taken together, the absence of any effect in the present study seems to be at least not unusual. Hence, it might be justified to carefully interpret the results as a hint that the instrument can be considered valid in this regard. Yet, it is known from statistical methodology that the ‘absence of evidence is not evidence of absence’ (Quertemont, 2011); meaning that the non-significant effect does not necessarily mean that gender plays no role in explaining attitudes towards inclusive education *for all*. Further investigations in this area might first of all clarify what theoretical bases there is to assume certain gender differences, and then might apply the most recent methodology in order to obtain stable effects and discourage over-interpretation (see Standard 1.15 in AERA et al., 2014).

Conclusion Regarding the VH2: Age of the Teachers

In Section 3.7.3, the following validation hypothesis was proposed:

- VH2: Younger teachers hold more positive attitudes towards inclusive education *for all* compared to older teachers.

The literature review revealed that age was a relevant variable and younger teachers tended to have more positive attitudes. This result appeared particularly in studies with a sole focus on in-service teachers in different contexts such as India (Bhatnagar & Das, 2014), England (Monsen et al., 2014), and Germany (Saloviita & Schaffus, 2016). In the present empirical study, an effect on the attitudes towards inclusive education *for all* was found. The multivariate analysis of variance demonstrated an overall effect of age on attitudes for both the sample from Australia and from Germany. For the German sample, the youngest age group tended to be more positive with regards to the support dimension-related attitude; especially compared to those teachers in the 41-50 years group. Yet, as discussed previously, this effect was relatively uncertain, due to overlapping confidence intervals. The analysis of the Australian sample suggested a more stable picture. For Australian teachers, it was found that the youngest age group had most positive attitudes with regards to the general practices dimension and the support dimension. The youngest teachers tended to be more positive compared to teachers between 31 and 50; and the youngest differed completely with the age group above 50 years.

The results obtained in the present study and the literature, fit well together. Hence, the empirical evidence suggests that the new instrument can be considered valid. It seems that – albeit the focus was broadened in this study from students with disabilities and/or special educational need to all students – the age effect seems to be persistent. Yet, notably, the studies that reported attitude differences with regard to age were solely sampling a particular group of the teachers; namely, in-service teachers in schools. The present study included both pre-service and in-service teachers. The vast majority (yet, not all) of the pre-service teachers were in the 30 and under years group. In the Australian sample, 89% of the pre-service teachers were 30 and under years old; while 29% of the in-service teachers were in the same age category. For the German sample, even 97% of the pre-service teachers ticked 30 and under years in the survey, while only 13% of the in-service teachers were amongst the youngest group in the sample. In other words, it seemed difficult to disentangle effects related to age from other effects, such as being pre-service or in-service teacher, or such as years of teaching experience for example. This issue was already discussed in Section 2.3.2. In order to statistically control for these aspects, a more complex model would have been needed that goes beyond the scope of this thesis; yet, in future studies on attitudes towards inclusive education *for all*, a focus might be on drawing a clearer picture of these effects.

A possible explanation for younger teachers tending to be more positive was already given within the literature review. Young teachers might at least be confronted in their initial training

with new kinds of thinking. As demonstrated by Woodcock and Hardy (2017), training in special education was detrimental to moving forwards towards an inclusive pedagogy *for all*. On the one hand, it seems unlikely that the younger teachers might come across some of the new concepts in their pre-service phase, which directly lead to them having the most positive attitudes towards inclusive education *for all*. Given the marginal importance of the concept of inclusive education *for all*, and its non-existence in the present empirical research (as discussed in the beginning of the present study, and as also recently noted e.g. by Nilholm & Göransson, 2017), it seemed unlikely that this would be the sole explanation. Yet, on the other hand, the effects were considerably stronger for the general practices and support dimensions in the Australian sample, compared to the slight effect reported for the support dimension in the German sample. This could indeed be an effect of pre-service training in Australia being some steps ahead in terms of embracing a *for all* perspective. Textbooks for teaching purposes on inclusive education *for all* are available for the English-speaking world (such as Carrington & Macarthur, 2012), yet, the German language text books in this area are still negotiating the opposites of and boundaries between regular and special education. A most recent example is the text book by Werning, Amrhein, Lütje-Klose, and Riecke-Baulecke (forthcoming) with the title “basics for teacher training: inclusion in school and classes – essentials in special education” (Ger. ‘Basiswissen Lehrerbildung: Inklusion in Schule und Unterricht – Grundlagen in der Sonderpädagogik’), which addressed all teachers, by offering experts’ knowledge from special education. Yet, this interpretation is built on a relatively thin empirical basis, suggesting that future research (across countries) might shed further light on the age effect with regards to the teachers’ attitudes towards inclusive education *for all*.

Conclusion Regarding the VH3: Pre-Service or In-Service Teachers

In Section 3.7.3, the following validation hypothesis was proposed:

- VH3: Pre-service teachers hold more positive attitudes towards inclusive education *for all* compared to in-service teachers.

The evidence as it was obtained from the present study is in line with the validation hypothesis, suggesting that the instrument seems to be valid in this regard. The present investigation is relatively unique in its approach to consider teachers as one group at different stages in their professional careers, and not follow the most common divide into teachers as learners in their initial training phase (*‘pre-service’ teachers*), and teachers as lifelong learners in their post-initial training phase (*‘in-service’ teachers*). In order to obtain information, if there were considerable differences between pre-service and in-service teachers, the hypothesis was formulated that pre-service teachers might hold more positive attitudes compared to in-service

teachers. If teachers were in their pre-service or their in-service phase had a significant overall effect on their attitudes towards inclusive education *for all*, which was suggesting that the new instrument might be considered valid in this regard. This result was found for both samples from Australia and Germany. The differential view on the different dimensions of the attitudes revealed that in both samples, the overall effect seemed to be to a large extent attributable to the substantial effect concerning the support dimension. Pre-service teachers were significantly more in favour on the support dimension, compared to their in-service counterparts. Some minor effects were found for the Australian pre-service teachers in comparison to in-service teachers, suggesting that pre-service teachers were more positive with regards to the general practices dimension, while more negative attitudes were found towards the differentiation dimension. Yet, as noted before, these minor effects were significant on the five percent level; however, the examination of the confidence intervals left doubts that the effect could be considered substantial.

The finding appears to be relatively similar to the result obtained with regards to the teachers' age. The support-related differences for the Australian and the German sample, and the general practices-related differences only for the Australian sample were comparable between the analysis of age and the analysis of pre- vs. in-service teachers. This suggests to some extent that the previously reported age-effect is related to the younger generation being actually closer to their initial pre-service training, where they learn about the most contemporary concepts, and hence, might tend to hold more positive attitudes towards them. Yet, the minor effect of the differentiation being judged by the pre-service teachers in a more unfavourable way, compared to their in-service counterparts, was not significant in the age-and-attitudes model. That pre-service teachers tend to be more positive towards general practices, while being more negative towards differentiation might be explained in part by the gap between expectations and realities of teaching (for example, as explained by Cole & Knowles, 1993). Despite considerable efforts to avoiding the practice shock for teachers (e.g. Delamarter, 2015), teachers' expectations of 'real-world' teaching practices may still be considered as being idealistic and/or unrealistic. This was best expressed in the words of Wanzare (2007), who noted that "beginning teachers often have varying strengths and vulnerabilities and their idealistic expectations usually become unrealistic as they are overwhelmed by difficult and pressing challenges in the workplace" (p. 349). Hence, the difference between the attitude score on the general practices dimension and the attitude score on the differentiation dimension might have some potential to illuminate some aspects of the expectations-to-realities gap.

Conclusion Regarding the VH4: Primary vs. Secondary School Teachers

In Section 3.7.3, the following validation hypothesis was proposed:

- VH4: Teachers from the primary school sector hold more positive attitudes towards inclusive education *for all* compared to teachers from the secondary school sector.

The hypothesis was drawn from the literature that respondents who were in training to teach or who were currently teaching in *primary* schools would have more positive attitudes compared to respondents from the *secondary* field. In this way, McGhie-Richmond et al. (2013) found in their study in Canada that amongst in-service teachers, those working in primary schools were holding more positive attitudes compared to other in-service teachers working in secondary schools. In the present study, a significant overall effect was found concerning differences between primary and secondary teachers. For the Australian sample, relatively weak evidence was obtained from the analysis, which suggested that teachers in the primary sector tended to be more favourable only towards differentiation. For the German sample, the overall effect was manifested in substantial effects with regards to the vision, the differentiation and the general practices. In all these dimensions German pre- and in-service teachers from the primary sector had significantly more positive attitudes, compared to their secondary sector counterparts, which suggests the scale being a valid measure.

There were considerable differences between the Australian and the German sample; namely, none of the secondary schools in Sydney participated in the present study. Hence, those respondents from the Australian secondary school sector were solely pre-service teachers, which makes it difficult to interpret the findings. In addition, the divide into primary and secondary was not comparable between both countries in terms of children's age. While the primary sector in Sydney comprised grades K-6, the primary sector in Giessen comprised grades one to four, as discussed in Section 3.3.2. Hence, the results obtained here, might be compared between the Australian and the German context with some caution.

Although certain developments took place in recent years, especially in the German case, the secondary school system is still divided at least into two separate tracks, where one track (so-called 'Gymnasium') is defined by only higher-achieving children are allowed to attend; which is clearly mirrored in different secondary teacher training for 'Gymnasium' or non-'Gymnasium' accordingly (although considerable developments are ongoing in this respect, there is still a separate track for special education and a separate teacher training in special education). In this light, the primary sector always had to deal with the broadest variety of children; while all the other sectors had more of a specialised sub-set of particular children to cater for. Hence, it is not surprising that the results for the German sample was suggesting that respondents who were in training to teach or who were teaching in primary schools had

substantially more positive attitudes towards inclusive education *for all*, compared to their secondary school counterparts.

While the primary versus secondary contexts had a substantial effect on three of the four dimensions in the German sample, this was not the case for the support dimension. Different interpretations are possible in this regards. On the one hand, the actual developments in Germany towards merging the general and special school system makes supports important for both sectors, and due to the covariation, no differential effect was found. On the other hand, it might be the case that a collaborative approach towards teaching (including different supports) is not yet present in the teachers' minds in Germany; hence, due to nearly random variation, no differential effect was found. Another interpretation would be that the adequacy of the supports depends much on the context; hence multilevel analysis would need to have shed some more light on effects on an individual and on an institutional level.

Conclusion Regarding the VH5: Teaching Experience in Years

In Section 3.7.3, the following validation hypothesis was proposed:

- VH5: Teaching experiences (in years) is not assumed to have a substantial effect on the attitudes towards inclusive education *for all*.

With regards to the *teaching experience in years*, a directed hypothesis was not assumed. Although a number of studies reported years of teaching experience to be generally a significant predictor of the teachers' inclusive education-related attitudes, there was not a clear picture in previous studies in regards to how many years of experience would go together with more favourable attitudes.

The results obtained in the present study showed that there was no substantial effect regarding years of teaching experiences on the teachers' attitudes towards inclusive education *for all*. In this way, the results suggest the new instrument being a valid measure. Notably, it was not feasible with the present data to disentangle possible effects concerning the developments of attitudes over time from effects related to certain events and to generations of teachers (e.g. teacher training in the 1980s was different, compared to teacher training more recently). Hence, interpretations of the non-existence of any substantial effect need to be made carefully and further research in this direction might shed some more light on teaching experience and attitudes.

Conclusion Regarding the VH6: Higher Degree Qualification

In Section 3.7.3, the following validation hypothesis was proposed:

- VH6: A higher degree qualification is not assumed to have a substantial effect on the attitudes towards inclusive education *for all*.

Because only one study reported significant negative effects, a specific hypothesis concerning the *higher degree qualification* was not proposed. Yet, this aspect was included in the survey to identify how respondents might differ in their attitudes according to their answer, if they were holding a higher degree qualification or not. In the literature, it was Ahmmed et al. (2012) who reported that those with a lower qualification were more positive in terms of attitudes, compared to those with a higher qualification. The results of the present study showed a significant overall effect for the Australian sample, yet, no effect for the German sample. For the Australian sample, the overall effect was significant but weak. An examination of the four dimensions revealed that only the vision dimension had a significant effect. Respondents with a postgraduate degree or diploma reported significantly more positive attitudes on this dimension. Yet, if the overlapping confidence intervals were considered, this effect turned out to be relatively weak. Although it would be plausible to assume that more educated teachers would be aware of the positive effects of inclusive education *for all* (which is captured by the vision dimension), this finding needs to be interpreted with certain caution. Especially because the empirical evidence presented by Ahmmed et al. (2012) pointed in a different direction.

An issue could be that teachers understand this item in a variety of ways. As discussed earlier (see Section 3.3.2), the study programs in both contexts were not identical. The teachers were asked in the questionnaire to give more details; and their responses give some more insights into how they understood the question. In the Australian context, the question a ‘postgraduate degree/diploma’ was understood for example as ‘Masters of Teaching’, or ‘Mteach’. Some have also written ‘MEd’, ‘Grad Dip Exp + Perf Arts’ or ‘M. Ed. Post Grad Dip’. The responses in the German sample comprised for example ‘Biologist’, ‘Music Therapist’, but also ‘Diploma in Pedagogy’, or ‘Diploma in Social Pedagogy’. In both samples there were also notions of ‘Bachelor of Education’ (in the Australian sample), or for example ‘BA Language, Literature, Culture’ (in the German sample). There seemed to be a particular need for further clarification in regards to the higher degree qualification of the teachers as it relates to their attitudes.

Conclusion Regarding the VH7: Knowledge as it Pertains to Inclusive Education for All

In Section 3.7.3, the following validation hypothesis was proposed:

- VH7: Teachers with a higher knowledge of the inclusive education *for all*-related legislation and/or policy hold more positive attitudes compared to teachers with less knowledge.

With regards to the teachers' *knowledge* of the inclusive education-related legislation and/or policy, the hypothesis was formulated that more knowledge is associated with more positive attitudes. This hypothesis was built to a large extent on the findings of Loreman, Forlin, et al. (2007) and Forlin et al. (2010). From the present analysis it was obtained that for the Australian sample there was no overall effect of the knowledge on the teachers' attitudes. For the German sample, an overall effect was found. Those respondents indicating very good knowledge of the local legislation and/or policy as it pertained to inclusive education *for all* were significantly more positive on the differentiation and on the support dimension of attitudes. Considering the confidence intervals, this result was found to be substantial.

Comparing the results with the hypothesis, it was surprising to find no effect for the Australian sample. Contrary to the Australian non-significant finding, the analysis of the German sample confirmed the proposed direction of the effect. Yet, the result for the German sample needs to be interpreted with some caution, because those teachers indicating very good knowledge were significantly different compared to the others, yet, this group only comprised three individuals (compared to the 12% of those with very good knowledge within the Australian sample). This could be interpreted in the direction of a certain need to provide more inclusive education *for all*-related knowledge for teaching practitioners in Germany.

In addition, the way 'knowledge' is operationalised in studies seemed to play a role, as mentioned before (see p. 37): Sucuoğlu et al. (2013, 2014) did not find any association between the knowledge and the attitudes, when they used a knowledge test instead of a single item. Hence, the validation hypothesis might depend in a way on which studies one gives more emphasis to. Accordingly, there seems to be a need for further clarification in this regard, and qualitative empirical research might help to understand certain kinds of knowledge that teachers might have in relation to their attitudes. This post-hoc explanation suggests that the measurement instrument, as it was developed in the present study, could still be considered valid.

Conclusion Regarding the VH8: Training as it Pertains to Inclusive Education for All

In Section 3.7.3, the following validation hypothesis was proposed:

- VH8: Teachers with more training in inclusive education *for all* hold more positive attitudes towards inclusive education *for all* compared to teachers with less training.

The amount of training focusing on inclusive education *for all* students was thought to be related to the teachers' attitudes; more training was supposed to go together with more positive attitudes. This hypothesis was drawn from a large number of studies, which were confirming this relation (Avramidis et al., 2000b; Avramidis & Kalyva, 2007; Bhatnagar & Das, 2014;

Boyle et al., 2013; Forlin et al., 2010; Sokal & Sharma, 2013; Tsakiridou & Polyzopoulou, 2014). Only one study by Sharma et al. (2015) presented a contradicting significant result.

The results of the present study showed no effect of training on attitudes amongst Australian teachers. If the teachers in Australia had a high amount of training focusing on inclusive education of all students, or only some or even none made no difference in terms of their attitudes towards inclusive education *for all*. Within the German sample of teachers, a different result was obtained. There was a significant overall effect of training on attitudes. Specifically the differentiation dimension of attitudes seemed to be significantly related to the training level. Significant differences were found between teachers with no training at all and those with some and a high amount of training. Using the confidence intervals, only the difference between the ‘none’ and the ‘high’ group turned out being substantial.

Similar to the knowledge-related finding, which was discussed previously, there was no effect in the Australian sample with regards to the training in inclusive education *for all*. Previous research seemed to draw a clear picture of this relation; yet, it seemed to make a difference if the attention of the respondents was focused on training and attitudes as they pertain to students with SEND, or if they pertain to all students. On the other hand, the German case confirms the proposed relation. That training and differentiation were related amongst German teachers seems plausible, because training specifically on inclusive education *for all* is probably mostly concerned with how to carry out more inclusive practices. Yet, if this would be a valid explanation, a similar effect would have been suggested amongst Australian teachers. Hence, further research needs to clarify this relation. As mentioned before, it would be of particular importance to include information about the content of the training (Woodcock & Hardy, 2017), in order to control for a possible interference of special educational needs-oriented content with the goals to achieve more positive views on catering *for all*.

In addition, in some ways the result obtained for knowledge and for training seemed similar. Both were overall non-significant in Australia and overall significant in Germany. Moreover, in both analyses, for the German sample the differentiation dimension played a significant role. Hence, it would be a question for further research, if there would be any effects across the concepts of knowledge and training, or if interactions between both concepts might play a role for the attitudes. To introduce more variables into the models would have gone beyond the scope of this thesis. Yet, if it would be carried out in future studies, there would be a certain potential to further understand the underlying effects that trigger certain attitude towards inclusive education *for all*. A sample that might be more balanced and which would be extended in size might be needed, to provide enough empirical information for estimating a larger number paths within one model (Kline, 2011).

It is known from recent research that the kinds of training (for some or for all) and the kinds of attitudes (for some or for all) converge (Woodcock & Hardy, 2017); in this way, inclusive education *for all* is only likely to be related to training in inclusive education *for all*. Although the obtained results and the validation hypotheses were not in line with each other, a valid post-hoc hypothesis seemed to be that the teachers might have had training on inclusion *for some*, rather than *for all*, which they did not reflect when completing the demographics part of the questionnaire.

Conclusion Regarding the VH9: Experiences with Inclusive Classroom Settings

In Section 3.7.3, the following validation hypothesis was proposed:

- VH9: Teachers having positive experiences with inclusive classroom settings hold more positive attitudes towards inclusive education *for all* compared to those with less positive or no experiences.

Concerning the *teachers' experiences with inclusive classroom settings* the hypothesis was formulated that teachers who experienced inclusive educational settings in a positive way would hold more positive attitudes towards inclusive education. Corresponding findings were known from former research that 'teaching students with disabilities' would go together with having more positive attitudes towards including students with special educational needs and/or disabilities (Ahmmed et al., 2012; Batsiou et al., 2008; Bhatnagar & Das, 2014; Forlin et al., 2010; Hellmich & Görel, 2014; Malinen et al., 2012; Tsakiridou & Polyzopoulou, 2014). More general inclusive education-related experiences, such as 'active experience of inclusion' (Avramidis et al., 2000b) or 'working in schools with integration units' (Avramidis & Kalyva, 2007), were also found to be positively associated with the teachers' attitudes.

The results in the present study demonstrated that for the Australian and the German sample a significant overall effect of the experiences on the attitudes could be found. This effect demonstrated substantial differences for all attitude dimensions. In this way, the validation hypothesis was confirmed, which suggests that the measurement was valid. As discussed previously, the effects for all of the dimensions in both samples were slightly different. Generally, the tendency was that positive experiences were associated with more positive attitudes and negative experiences were associated with more negative attitudes. What differed was the effect concerning 'neutral' experiences or no experiences at all. The strongest differences were found for the vision, differentiation, and the general practices dimensions amongst German teachers, because in these cases the positive experiences were associated with more positive attitudes compared to neutral and no experiences, which were again more positive

compared to negative experiences. Taken together, strong evidence suggested the presence of the proposed influence of experiences on attitudes.

It seems important to point out the existence of the association between experiencing negative examples of inclusive education *for all* and more negative attitudes. In this way it seems important to try to foster not only experiences of inclusive classrooms when trying to develop more positive attitudes amongst teachers; but also to take care that these experiences are perceived as positive examples of inclusive education *for all*.

Conclusion Regarding the VH10: Teachers' Self-Efficacy

In Section 3.7.3, the following validation hypothesis was proposed:

- VH10: Teachers with a higher self-efficacy to carry out inclusive practices hold more positive attitudes towards inclusive education *for all* compared to teachers with weaker self-efficacy.

In accordance with former studies, the hypothesis was formulated that the *teachers' self-efficacy in carrying out inclusive practices* was positively related to the teachers' attitudes towards inclusive education *for all*. When teachers reported being confident to teach students with special educational needs and/or disabilities, they also reported more positive attitudes (Forlin et al., 2010; Sokal & Sharma, 2013; Tsakiridou & Polyzopoulou, 2014). Similar positive results were found, if the studies utilised multi-item instruments to measure the teachers' self-efficacy (Bosse et al., 2016; Hecht et al., 2016; Hellmich & Görel, 2014; Malinen et al., 2012; Montgomery & Mirenda, 2014; Savolainen et al., 2012; Weisel & Dror, 2006; Yada & Savolainen, 2017). Some studies, such as Hecht et al. (2016) or Yada and Savolainen (2017) reported differential effects of the dimensions of self-efficacy; in the former study, the self-efficacy in collaboration and the self-efficacy in using inclusive instructions were emphasised as being particularly associated with attitudes, and in the latter study, the self-efficacy in collaboration was also underlined; yet the self-efficacy in managing behaviour was highlighted as a second important dimension.

In the present study, a significant overall effect of the teachers' self-efficacy and the teachers' attitude towards inclusive education *for all* was found for both samples and with regards to all three self-efficacy dimensions. The validation hypothesis and the results as they were obtained in the present study fit well together, suggesting a valid measure. Concerning the self-efficacy in managing behaviour, amongst the Australian teachers, only a significant association to the differentiation dimension of the attitudes was found. In other words, Australian teachers who were willing to differentiate in order to provide inclusive education *for all* students, tended to also think that they had a certain capability to coping effectively with

difficult student behaviour. Despite a significant overall effect of the managing behaviour dimension in the German sample, none of the attitude dimensions were particularly associated with this aspect of self-efficacy. Concerning the self-efficacy in collaboration, a significant association was found amongst Australian teachers with regards to the vision and the adjustment dimension of the attitudes. For German teachers, all four attitude dimensions were positively related to this aspect of self-efficacy. In the Australian sample, the self-efficacy in using inclusive instruction was associated with the vision and the differentiation dimension of the attitudes, while in the German sample, this aspect of self-efficacy was associated with the differentiation and the general practices. All these findings turned out to be substantial, according to the examination of the confidence intervals.

The results of the present study support the strong empirical evidence from previous studies. Although the focus of the attitudes was widened from a narrow view on students with special educational needs and/or disabilities to inclusive education *for all*, the relation of attitudes to self-efficacy seems to still be considerable (notably, the wording of the self-efficacy instrument was revised, too, in order to avoiding SEND-related expressions). The strong association of the differentiation dimension with the teachers' self-efficacy is relatively obvious, because the 'I can'-perspective of the teachers self-efficacy items is a continuation of the 'I am willing' perspective of the differentiation dimension. A large number of the studies, cited in the literature review, actually understood the relation between both as one construct influencing the other (accordingly, the present study also calculated a multivariate linear regression analysis of the self-efficacy *on* the attitudes, suggesting that self-efficacy predicts attitudes). Yet, from a theoretical point of view (Ajzen, 1988, 1991; Fishbein & Ajzen, 1975), and from empirical studies (Bosse et al., 2016), it seems more to be the case that attitudes and self-efficacy need to be understood as 'side-by-side', and not as 'one-after-another'. Hence, future research might analyse the relation of the self-efficacy and the attitudes in more depth. A further step of analysis of the present data could be to examine discriminant validity, by comparing nested models that include all attitude and all self-efficacy items, to investigate if the dimensions could be established in contrast to each other. If the theory of planned behaviour (Ajzen, 1991) would be assumed as an underlying model, it would be one of the next steps to develop a measure of the subjective norms as they pertain to inclusive education *for all*, in order to complement a comprehensive measurement instrument of the independent variables within a model of planned behaviour. As this study showed, a large number of dimensions (four attitude dimensions, three self-efficacy dimensions, plus potential subjective norms dimensions) could be involved in such a model.

A substantial association between self-efficacy in collaboration and the support dimension of attitudes was only found for the German teachers, but not for the Australian teachers. This finding was surprising, because, generally, it was expected that these two perspectives were relatively similar. Yet, the collaboration dimension comprised notions that the teachers feel competent to work jointly with others, while the support dimension comprised notions that the teachers feel that there are others who support the teacher. It was interesting to find in the data that the German teachers did not have differential views on these two perspectives, while their Australian counterparts seemed to differentiate between working jointly with others versus receiving support from others. In a way, this finding mirrors the ongoing discussions in Germany about establishing successful collaboration between different practitioners (the body of research in the area in Germany has grown considerably in recent years; examples are Breuer & Reh, 2010; Buchna, Coelen, Dollinger, & Rother, 2016; Dizinger, Fussangel, & Böhm-Kasper, 2011; Kielblock et al., 2017; Reh & Breuer, 2012; Speck, Olk, & Stimpel, 2011). Research demonstrated that collaboration is in fact understood (not only by teachers) as a teacher being in charge and others need to assist the teacher, which was recently called ‘normalised hierarchy’ (Buchna et al., 2016) between regular teachers and other staff at schools. More research is needed to understand how to tap the full potential of supporting teaching practices that are meant to be *for all students*.

Overview on all Validation Hypotheses

All of the examined relationships of the attitude dimensions to related other teacher aspects revealed a relatively clear picture; namely, that the new instrument seemed to measure the teachers’ attitudes in an appropriate way. As discussed throughout this section, some validation hypotheses could not be confirmed. Yet, post hoc explanations were found that illuminated the absence of effects and pointed to more research that needs to be carried out. These implications for further research will be brought together in the end of the present study (see Section 5.4.2).

In both contexts, the evidence was relatively strong that the new instrument relates more positive attitudes towards inclusive education *for all* to younger teachers, pre-service teachers, teachers from the primary sector, teachers with positive experiences in inclusive settings and teachers with higher levels of self-efficacy. These relationships are not only supported by previous studies (which were only available with regards to attitudes towards inclusive education *for some*), but these relationships can also be considered plausible. In this way, the instrument that the present study attempted to develop has resulted in a valid measure of the teachers’ attitude towards inclusive education *for all*.

5.3 Limitations of the Study

Although a considerable amount of effort was invested in making informed decisions in selecting most current and relevant methodologies and utilising most recent standards for all methods applied, a range of limitations need to be considered. These limitations are important to recognise, first, in order to discourage over-interpretation of the present findings (see Standard 1.15 in AERA et al., 2014), and, second, in order to point to certain gaps and issues that future studies might be able to elaborate on further.

A first limitation pertained to the English questionnaire, which was measuring attitudes directly, as opposed to an indirect measurement. If respondents give information about how they feel concerning certain aspects that pertain to their attitudes, there is a chance that these responses do not actually converge with their actual attitudes. As discussed earlier in this study (see Section 3.4.3), there might be some attitudes actually ‘hidden’ or being implicit that might only be brought to light using an indirect attitude measurement. To rely on self-reports of teachers (hence, direct attitude measurement) might be pragmatic and well-justified, as discussed in the Methodology Chapter. Yet, the critique that there might be an implicit part of the attitudes might lead to other research on attitudes towards inclusive education *for all* using implicit measurement, in order to complement the present study’s results.

The present study was focused on Sydney (New South Wales, Australia) and Giessen (Hesse, Germany). Although, arguments were given, why these contexts were adequate for the development of the new instrument (see Section 3.3), the samples of teachers did not consider the broader population, because the present study was thought to explore new ways of measuring attitudes towards inclusive education *for all*, and not to draw a representative picture of the wider population of teachers in these contexts. Although the samples were pulled randomly, in order to not pick particular schools or units (subjects), this random selection did not follow a random sample design. A first and most obvious limitation is that the random selection did not select individuals randomly, but schools/units. Hence, the results cannot be generalised on teachers in general (in Sydney and Giessen). And second, if a selected school/unit declined to participate, another school/unit was selected. In a fully random study, the random sample design would need to clarify, which school/unit needs to be asked instead (second order random selected school/unit, etc.). Hence, the results need to be generalised with some caution on the wider teacher population.

To argue that the successful measurement of the teachers’ attitudes towards inclusive education *for all* (in terms of finding appropriate items and finding a plausible and sound statistical solution for examining these items), suggest that the teachers’ attitudes towards

inclusive education *for all* must actually *exist*, was coined by Kane (2015) the “*reification fallacy*” [Ger. Verdinglichungs-Fehlschluss]. Reification fallacy means to wrongly claim (without any other justification) that a construct, which was measured, actually existed. In this way, the innovative and exploratory character of the present study is a limitation in itself; the present study was only able to start putting together some first pieces of the bigger picture of studying how teachers relate to inclusive education *for all*. Yet, many further studies will be needed to understand the teachers’ attitudes in more depth. Not only empirical developments are needed to move forward, but also (and maybe even more urgent) theoretical developments in research on teachers and inclusive education *for all*.

5.4 Implications of the Findings and Conclusion

The findings of the present study have implications not only for policy and practice but also for further research. The variety of implications are discussed, before an overall conclusion of the present study is reached.

5.4.1 Implications for Policy and Practice

The present study’s findings seem to be promising for strengthening the *for all* perspective in inclusive education policy and practice. The findings have emphasised the importance of thinking about inclusive education as being meant *for all* students, rather than *for some*.

The findings have implications for teachers and their teaching practices. Across the globe, teachers and teaching practices are generally focussed on ideas that are related to inclusive education *for some*. For individual teachers, who might be thinking that all of their students were average (and that they were trained on educating the average student), but that inclusive education now brings exceptional students into the formerly homogeneous setting, inclusive education must be quite a shock. Contrary to such a view, inclusive education *for all* would focus the need of the teacher and the teaching to responding to issues as they pertain to the presence, participation and achievement of all of the students. To change the personal views towards a *for all*-related perspective reminds the teacher of each individual student being an individual learner with certain strengths and capabilities. In this way, the concept of inclusive education *for all*, which was defined and consistently used in the present study, might help teachers (and others, e.g. principles, parents, policy makers) to understand that current thinking about inclusive education *for some* is likely to even increase exclusiveness (Slee, 2013), to

acknowledge the vision of inclusive education *for all* and adapt their practices accordingly, to be able to cater for all students.

As it was pointed out in Section 2.2.2, a truly inclusive school is mindful of exclusionary dynamics and finds continuously creative ways to foster the prospering for all students. Developing schools and communities in this direction is a complex and very difficult undertaking (Booth & Ainscow, 2011; first edition published in 2000). A relevant aspect, as it pertains to the schools, is that inclusive schools need a clear guiding philosophy or mission (Peters, 2004). The notion of inclusive education *for all* might support school leaders in developing such a vision that is compatible with inclusive values. For adopting such a guiding philosophy in school and teaching practices and for creating a positive school climate, positive teacher attitudes are crucial, as Peters (2004) pointed out. In this way, the present study is not only relevant for individual teachers but also for the schools' teaching staff as a whole. An advantage of the quantitative approach, which the present study utilised, is that it would be possible to gain at least some insights into the thinking of the whole teaching staff of a school. Besides the limitations of such an approach (as they were discussed in Section 5.3), a school would gain valuable insights into how positive the teachers think about the vision, differentiation, general practices and supports as they pertain to inclusive education *for all*. According to this evidence, specific school developments could be initiated and the new instrument could even be used to monitor if these interventions have the desired effects. In this sense, the new instrument could be used in schools to promote interventions that facilitate the development of inclusive thinking and practices.

For the Education Departments/Boards and for the institutions that provide pre-service teacher training (such as universities) the results of the present study suggest to change and re-develop the policies and programs so that they are in line with inclusive education *for all*. The instrument that the present study developed might be used to collect information from teachers what impact these changes are making. In accordance, further changes could be initiated and the effects could again be monitored by using the instrument, which was developed in the present study.

5.4.2 Implications for Further Research

As it was emphasised in the literature review of the present study, there is a great amount of recent studies on teachers' attitudes towards inclusive education *for some*. Ruberg and Porsch (2017) found for the recent couple of years 24 German-speaking studies on the teachers' attitudes towards different aspects of inclusive education *for some*. This is a considerable number, in view of the fact that inclusive education is a relatively new term in Germany and a

relatively new area of research. Accordingly, all of these teacher studies (and all the other international studies) on attitudes towards inclusive education *for some* remind the teachers repeatedly that there would be a need to think about particular students differently compared to all of the other students that are considered to be normal. A most recent study in the area of school's evaluation practices demonstrated clearly that teachers' surveys can be considered a vehicle for delivering interventions (Gehlbach, Robinson, Finefter-Rosenbluh, Benshoof, & Schneider, 2017); in other words, the content of a questionnaire impacts on the thinking of the teachers. With regards to the present study, this might mean that all the attitudes towards inclusive education *for some*-related questionnaires might deflect the teachers' thinking from inclusive education *for all*, and might draw their attention repeatedly on issues as they pertain to *some students* particularly. The findings of the present study clearly indicate the issue that current empirical research on teachers' attitudes is focused on inclusive education *for some*, rather than *for all*, and that there is now a way to measure the teachers' attitudes towards inclusive education *for all*, which implies that future research on teachers' attitudes would utilise the new instrument, which was developed throughout the present study.

In the literature, a vast amount of instruments are available for measuring inclusive education *for some*. Although authors such as Antonak and Livneh (1988) emphasised that instead of creating new scales, there would be a certain value in refining established scales, it seems that many studies created their own set of inclusive education *for some*-related items. In this way, there was no agreement on what dimensions of an attitude scale in this area should comprise (affective, behavioural, cognitive vs. certain kinds of disabilities vs. other content areas). In this sense there seems to be a lack of more theoretical discussions around the teachers' attitudes. The absence of any comprehensive theory on the teachers' views on inclusive education (*for some* or *for all*) was one of the core issues of the present study. It was difficult to providing a completely new view on the issue, and, at the same time, emphasising the connectedness of the study to previous research. This is, where theory normally comes into play, combining what is known about a certain topic and presenting it in an abstract form, so that it is possible to use or adapt the theory even if the paradigm has changed. In this way, more theoretical reflections might be needed in future research, in order to provide more of a model (or theory) of teachers' attitudes towards inclusive education *for all*. The obtained dimensions of the new measurement instrument might give some first hints in this direction.

5.4.3 Overall Conclusion

The present study defined and consistently utilised inclusive education *for all* as a new concept in education, which disassociates itself from inclusive education *for some*. The main part of the

present study was focused on developing a sound and robust instrument to measure the teachers' attitudes towards inclusive education *for all*. If the purposes of the present study are revisited (namely, that the new instrument should be sound, robust, and multidimensional), it can be concluded that the aims were achieved. The new instrument comprised four dimensions; namely, the vision, the differentiation, the general practices, and the supports as they pertain to inclusive education *for all*. These dimensions were stable across certain contexts and through data analysis the quality of the measurement could be established.

Inclusive education has a certain history, and although great steps were taken, sometimes it seems that nothing has changed and nothing has been achieved in the past few decades. Comparably, as early as 1966, Coleman et al. (1966) found that in terms of student achievement the schools and other factors were not as important as the quality of the teachers. Fifty years later, Goldhaber (2016) noted that according to what is known today, these findings are still valid; namely that the way to improve student outcomes would be through improving teachers, yet, it is still not clear how exactly to achieve this (Alvunger, Sundberg, & Wahlström, 2017; Levin, 2017). This picture seems to be relatively similar in inclusive education; hence, it seems to be imperative to gain more empirical knowledge about inclusive education *for all*, and to gain insights into fostering more inclusive teaching practices that are effective for the learning of *all students*. The present study might lay the foundation to proceed some steps in a desirable direction towards a more just and equitable future *for all*.

References

- AERA, APA, & NCME. (2014). *Standards for Educational and Psychological Testing*. Washington, DC: AERA.
- Ahmad, L. (2012). Development of instrument for measuring the attitude of teachers, administrators and policy makers towards inclusive education. *Interdisciplinary Journal of Contemporary Research in Business*, 4(5), 86-106.
- Ahmed, L. (2012). *Attitude of teachers, administrators and policy makers towards inclusive education*. (PhD thesis), University of the Punjab, Retrieved from <http://pr.hec.gov.pk/Thesis/2853S.pdf>
- Ahmmed, M., Sharma, U., & Deppeler, J. (2012). Variables affecting teachers' attitudes towards inclusive education in Bangladesh. *Journal of Research in Special Educational Needs*, 12(3), 132-140. doi:10.1111/j.1471-3802.2011.01226.x
- Ainscow, M. (Ed.) (1991). *Effective schools for all*. London: David Fulton.
- Ainscow, M., Booth, T., & Dyson, A. (2006). *Improving schools, developing inclusion*. London: Routledge.
- Ainscow, M., Farrell, P., & Tweddle, D. (2000). Developing policies for inclusive education: A study of the role of local education authorities. *International Journal of Inclusive Education*, 4(3), 211-229. doi:10.1080/13603110050059150
- Ainscow, M., & Miles, S. (2008). Making Education for All inclusive: Where next? *Prospects*, 38(1), 15-34. doi:10.1007/s11125-008-9055-0
- Ajzen, I. (1988). *Attitudes, personality, and behavior*. Milton Keynes: Open University Press.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211. doi:10.1016/0749-5978(91)90020-T
- Al Zyoudi, M., Al Sartwai, A., & Dodin, H. (2011). Attitudes of pre-service teachers towards inclusive education in UAE and Jordan (a comparative study). *International Journal of Disability, Community & Rehabilitation*, 10(1).
- Al-Zyoudi, M. (2006). Teachers' attitudes towards inclusive education in Jordanian schools. *International Journal of Special Education*, 21(2), 55-62.
- Albarracín, D., Johnson, B. T., & Zanna, M. P. (2005). *The handbook of attitudes*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Alghazo, E. M., & Gaad, E. E. N. (2004). General education teachers in the United Arab Emirates and their acceptance of the inclusion of students with disabilities. *British Journal of Special Education*, 31(2), 94-99. doi:10.1111/j.0952-3383.2004.00335.x
- Allport, G. W. (1935). Attitudes. In C. Murchison (Ed.), *A handbook of social psychology* (pp. 798-844). Worcester, MA: Clark University Press.
- Alvunger, D., Sundberg, D., & Wahlström, N. (2017). Teachers matter – but how? *Journal of Curriculum Studies*, 49(1), 1-6. doi:10.1080/00220272.2016.1205140
- Alwin, D. F. (2007). *Margins of Error: A study of reliability in survey measurement*. Hoboken, NJ: John Wiley & Sons.
- Anderson, J., & Boyle, C. (2015). Inclusive education in Australia: Rhetoric, reality and the road ahead. *Support for Learning*, 30(1), 4-22. doi:10.1111/1467-9604.12074
- Andrews, S., & Clementson, J. J. (1997). *Active learning's effect upon preservice teachers' attitudes toward inclusion*. ERIC Document Reproduction Service No. ED 410 217.
- Antonak, R. F., & Larrivee, B. (1995). Psychometric analysis and revision of the opinions relative to mainstreaming scale. *Exceptional Children*, 62(2), 139-149. doi:10.1177/001440299506200204
- Antonak, R. F., & Livneh, H. (1988). *The measurement of attitudes toward people with disabilities: Methods, psychometrics, and scales*. Springfield, IL: Charles C. Thomas.

- Antonak, R. F., & Livneh, H. (1995). Direct and indirect methods to measure attitudes toward persons with disabilities, with an exegesis of the error-choice test method. *Rehabilitation Psychology, 40*(1), 3-24. doi:10.1037/0090-5550.40.1.3
- Arduin, S. (2015). A review of the values that underpin the structure of an education system and its approach to disability and inclusion. *Oxford Review of Education, 41*(1), 105-121. doi:10.1080/03054985.2015.1006614
- Armstrong, D., Armstrong, A. C., & Spandagou, I. (2011). Inclusion: By choice or by chance? *International Journal of Inclusive Education, 15*(1), 29-39. doi:10.1080/13603116.2010.496192
- Avramidis, E., Bayliss, P., & Burden, R. (2000a). Student teachers' attitudes towards the inclusion of children with special educational needs in the ordinary school. *Teaching and Teacher Education, 16*(3), 277-293. doi:10.1016/S0742-051X(99)00062-1
- Avramidis, E., Bayliss, P., & Burden, R. (2000b). A survey into mainstream teachers' attitudes towards the inclusion of children with special educational needs in the ordinary school in one local education authority. *Educational Psychology: An International Journal of Experimental Educational Psychology, 20*(2), 191-211. doi:10.1080/713663717
- Avramidis, E., & Kalyva, E. (2007). The influence of teaching experience and professional development on Greek teachers' attitudes towards inclusion. *European Journal of Special Needs Education, 22*(4), 367-389. doi:10.1080/08856250701649989
- Avramidis, E., & Norwich, B. (2002). Teachers' attitudes towards integration/inclusion: A review of the literature. *European Journal of Special Needs Education, 17*(2), 129-147. doi:10.1080/08856250210129056
- Bailey, J. (2004). The validation of a scale to measure school principals' attitudes toward the inclusion of students with disabilities in regular schools. *Australian Psychologist, 39*(1), 76-87. doi:10.1080/00050060410001660371
- Ballard, K. (2012). Inclusion and social justice: Teachers as agents of change. In S. Carrington & J. MacArthur (Eds.), *Teaching in inclusive school communities* (pp. 65-87). Milton: John Wiley & Sons.
- Banafsche, M. (2013). Kinder und Jugendliche mit Behinderung zwischen SGB VIII und SGB XII – im Lichte der UN-Behindertenrechtskonvention. [Children and youths with disabilities between SGB VIII and SGBXII – in light of the UN Convention on the Rights of Persons with Disabilities]. In C. Dorrance & C. Dannenbeck (Eds.), *Doing Inclusion: Inklusion in einer nicht inklusiven Gesellschaft* (pp. 42-57). Bad Heilbrunn: Julius Klinkhardt.
- Barnett, C., & Monda-Amaya. (1998). Principals' knowledge of and attitudes toward inclusion. *Remedial and Special Education, 19*(3), 181-192. doi:10.1177/074193259801900306
- Batsiou, S., Bebetos, E., Panteli, P., & Antoniou, P. (2008). Attitudes and intention of Greek and Cypriot primary education teachers towards teaching pupils with special educational needs in mainstream schools. *International Journal of Inclusive Education, 12*(2), 201-219. doi:10.1080/13603110600855739
- Beacham, N., & Rouse, M. (2012). Student teachers' attitudes and beliefs about inclusion and inclusive practice. *Journal of Research in Special Educational Needs, 12*(1), 3-11. doi:10.1111/j.1471-3802.2010.01194.x
- Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health, 25*(5), 464-469. doi:10.1111/j.1467-842x.2001.tb00294.x
- Bennett, T., Deluca, D., & Bruns, D. (1997). Putting inclusion into practice: Perspectives of teachers and parents. *Exceptional Children, 64*(1), 115-131. doi:10.1177/001440299706400108

- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588-606. doi:10.1037//0033-2909.88.3.588
- Bergman, M. M. (1998). A theoretical note on the differences between attitudes, opinions, and values. *Swiss Political Science Review*, 4(2), 81-93. doi:10.1002/j.1662-6370.1998.tb00239.x
- Berkemeyer, N., Bos, W., Hermstein, B., & Abendroth, S. (2017). *Chancenspiegel – eine Zwischenbilanz: Zur Chancengerechtigkeit und Leistungsfähigkeit der deutschen Schulsysteme seit 2002. [Mirror of opportunities – interim results: Concerning equal opportunities and performance of the German school systems since 2002]*. Gütersloh: Bertelsmann Stiftung.
- Berlach, R. G., & Chambers, D. J. (2011). Interpreting inclusivity: An endeavour of great proportions. *International Journal of Inclusive Education*, 15(5), 529-539. doi:10.1080/13603110903159300
- Bhatnagar, N., & Das, A. (2014). Attitudes of secondary school teachers towards inclusive education in New Delhi, India. *Journal of Research in Special Educational Needs*, 14(4), 255-263. doi:10.1111/1471-3802.12016
- Black-Hawkins, K., & Florian, L. (2012). Classroom teachers' craft knowledge of their inclusive practice. *Teachers and Teaching*, 18(5), 567-584. doi:10.1080/13540602.2012.709732
- Blasberg, J. S., Hewitt, P. L., Flett, G. L., Sherry, S. B., & Chen, C. (2016). The importance of item wording: The distinction between measuring high standards versus measuring perfectionism and why it matters. *Journal of Psychoeducational Assessment*, 34(7), 702-717. doi:10.1177/0734282916653701
- Blömeke, S., Olsen, R. V., & Suhl, U. (2016). Relation of student achievement to the quality of their teachers and instructional quality. In T. Nilsen & J.-E. Gustafsson (Eds.), *Teacher quality, instructional quality and student outcomes: Relationships across countries, cohorts and time* (pp. 21-50): Springer Open.
- Bohner, G., & Dickel, N. (2011). Attitudes and attitude change. *Annual Review of Psychology*, 62(1), 391-417. doi:10.1146/annurev.psych.121208.131609
- Bohner, G., & Wänke, M. (2002). *Attitudes and attitude change*. Hove: Psychology Press.
- Booth, T. (1995). Mapping inclusion and exclusion: Concepts for all? In C. Clark, A. Dyson, & A. Millward (Eds.), *Towards inclusive schools?* (pp. 96-108). London: David Fulton.
- Booth, T., & Ainscow, M. (2011). *Index for inclusion: Developing learning and participation in schools* (3 ed.). Bristol: Centre for Studies on Inclusive Education (CSIE).
- Bosse, S., Henke, T., Jäntschi, C., Lambrecht, J., Vock, M., & Spörer, N. (2016). Die Entwicklung der Einstellung zum inklusiven Lernen und der Selbstwirksamkeit von Grundschullehrkräften. [The development of inclusive attitudes and self-efficacy of primary school teachers]. *Empirische Sonderpädagogik*, 8(1), 103-116.
- Bosse, S., & Spörer, N. (2014). Erfassung der Einstellung und der Selbstwirksamkeit von Lehramtsstudierenden zum inklusiven Unterricht. [Assessment of attitudes and self-efficacy of pre-service teachers towards inclusive education]. *Empirische Sonderpädagogik*, 6(4), 279-299.
- Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of Public Health*, 27(3), 281-291. doi:10.1093/pubmed/fdi031
- Boyle, C., Topping, K., & Jindal-Snape, D. (2013). Teachers' attitudes towards inclusion in high schools. *Teachers and Teaching*, 19(5), 527-542. doi:10.1080/13540602.2013.827361
- Bracken, B. A., & Barona, A. (1991). State of the art procedures for translating, validating and using psychoeducational tests in cross-cultural assessment. *School Psychology International*, 12(1), 119-132. doi:10.1177/0143034391121010

- Braeken, J., & van Assen, M. A. L. M. (2017). An empirical Kaiser criterion. *Psychological Methods*, 22(3), 450-466. doi:10.1037/met0000074
- Bray, M., Adamson, B., & Mason, M. (2007). Different models, different emphases different insights. In M. Bray, B. Adamson, & M. Mason (Eds.), *Comparative education research approaches and methods* (pp. 363-379).
- Bray, M., & Thomas, R. M. (1995). Levels of comparison in educational studies: Different insights from different literatures and the value of multilevel analyses. *Harvard Educational Review*, 65(3), 472-491. doi:10.17763/haer.65.3.g3228437224v4877
- Breuer, A., & Reh, S. (2010). Zwei ungleiche Professionen? Wie LehrerInnen und ErzieherInnen in Teams zusammenarbeiten. [Unequal professions? The cooperation of teachers and child care worker]. *Soziale Passagen*, 2(1), 29-46. doi:10.1007/s12592-010-0043-x
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1(3), 185-216. doi:10.1177/135910457000100301
- Brislin, R. W. (1980). Translation and content analysis of oral and written materials. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology: Methodology (Volume 2)* (pp. 389-444). Boston: Allyn and Bacon.
- Brislin, R. W., Lonner, W. J., & Thorndike, R. M. (1973). *Cross-cultural research methods*. New York: John Wiley & Sons.
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2 ed.). New York: Guilford.
- Buchna, J., Coelen, T., Dollinger, B., & Rother, P. (2016). Normalisierte Hierarchie in Ganztagsgrundschulen: Empirische Befunde zur innerorganisationalen Zusammenarbeit von Lehrkräften und weiterem pädagogisch tätigen Personal. [Normalised hierarchy in primary all-day schools: Empirical findings on intra-organisational teamwork among teachers and additional educational staff]. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 36(3), 281-297.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion. [Introduction to test- and questionnaire development]*. München: Pearson Studium.
- Burnett, N. (2008). *What sort of UNESCO for what sort of education: Gaitskell Lecture*, University of Nottingham 22 May 2008. Retrieved from <http://unesdoc.unesco.org/images/0016/001600/160090e.pdf>
- Buyse, V., Wesley, P. W., & Keyes, L. (1998). Implementing early childhood inclusion: Barrier and support factors. *Early Childhood Research Quarterly*, 13(1), 169-184. doi:10.1016/S0885-2006(99)80031-3
- Carrington, S., & Macarthur, J. (Eds.). (2012). *Teaching in inclusive school communities*. Milton: John Wiley & Sons.
- Carrington, S., MacArthur, J., Kearney, A., Kimber, M., Mercer, L., Morton, M., & Rutherford, G. (2012). Towards an inclusive education for all. In S. Carrington & J. Macarthur (Eds.), *Teaching in inclusive school communities* (pp. 3-38). Milton: John Wiley & Sons.
- Chaiklin, H. (2011). Attitudes, behavior, and social practice. *Journal of Sociology & Social Welfare*, 38(1), 31-54.
- Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, 36(4), 462-494. doi:10.1177/0049124108314720
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9), 2633-2679. doi:10.3386/w19424
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. London: Routledge.

- Cohen, P. S. (1966). Social attitudes and sociological enquiry. *The British Journal of Sociology*, 17(4), 341-352. doi:10.2307/589181
- Coldron, J., Crawford, M., Jones, S., & Simkins, T. (2015). The positions of primary and secondary schools in the English school field: a case of durable inequality. *Journal of Education Policy*, 30(5), 671-687. doi:10.1080/02680939.2014.972989
- Cole, A. L., & Knowles, J. G. (1993). Shattered images: Understanding expectations and realities of field experiences. *Teaching and Teacher Education*, 9(5-6), 457-471. doi:10.1016/0742-051x(93)90030-k
- Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, D.C.: U.S. Government Printing Office.
- Colina, S., Marrone, N., Ingram, M., & Sánchez, D. (2016). Translation quality assessment in health research: A functionalist alternative to back-translation. *Evaluation & the Health Professions*(Epub ahead of print), 1-27. doi:10.1177/0163278716648191
- Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research*, 12(3), 229-238. doi:10.1023/A:1023254226592
- Costello, S., & Boyle, C. (2013). Pre-service secondary teachers' attitudes towards inclusive education. *Australian Journal of Teacher Education*, 38(4), 129-143. doi:10.14221/ajte.2013v38n4.8
- Cowen, E. L., Rockway, A. M., Bobrove, P. H., & Stevenson, J. (1967). Development and evaluation of an attitude to deafness scale. *Journal of Personality and Social Psychology*, 6(2), 183-191. doi:10.1037/h0024552
- Cullen, J. P., Gregory, J. L., & Noto, L. A. (2010). *The Teacher Attitudes Toward Inclusion Scale (TATIS): Technical report*. Paper presented at the Annual Meeting of the Eastern Educational Research Association.
<http://files.eric.ed.gov/fulltext/ED509930.pdf>
- Da Mota Falcão, D., Ciconelli, R. M., & Ferraz, M. B. (2003). Translation and cultural adaptation of quality of life questionnaires: An evaluation of methodology. *Journal of Rheumatology*, 30(2), 379-385.
- Daniels, H. (2006). The dangers of corruption in special needs education. *British Journal of Special Education*, 33(1), 4-9. doi:10.1111/j.1467-8578.2006.00405.x
- Davis, R. E., Couper, M. P., Janz, N. K., Caldwell, C. H., & Resnicow, K. (2010). Interviewer effects in public health surveys. *Health Education Research*, 25(1), 14-26. doi:10.1093/her/cyp046
- Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, 50(1), 61-77.
- De Boer, A., Pijl, S. J., & Minnaert, A. (2011). Regular primary schoolteachers' attitudes towards inclusive education: A review of the literature. *International Journal of Inclusive Education*, 15(3), 331-353. doi:10.1080/13603110903030089
- De Boer, A., Timmerman, M., Pijl, S. J., & Minnaert, A. (2012). The psychometric evaluation of a questionnaire to measure attitudes towards inclusive education. *European Journal of Psychology of Education*, 27(4), 573-589. doi:10.1007/s10212-011-0096-z
- De Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, 44(2), 147-181. doi:10.1080/00273170902794206
- Delamarter, J. (2015). Avoiding practice shock: Using teacher movies to realign pre-service teachers' expectations of teaching. *Australian Journal of Teacher Education*, 40(2). doi:10.14221/ajte.2015v40n2.1
- DeVellis, R. F. (2011). *Scale development: Theory and applications* (Vol. 3). Thousand Oaks, CA: Sage.

- Dizdarevic, A., Mujezinovic, A., & Memisevic, H. (2017). Comparison of teachers' attitudes towards inclusive education in Bosnia and Herzegovina and European Union. *Journal of Special Education and Rehabilitation*, 18(1-2), 92-108. doi:10.19057/jser.2017.21
- Dizinger, V., Fussangel, K., & Böhm-Kasper, O. (2011). Interprofessionelle Kooperation an Ganztagschulen aus der Perspektive der Lehrkräfte. Wie lässt sie sich erfassen und wie wird sie bewertet? [Interprofessional collaboration at all-day schools from the perspective of the teachers. How to measure and to evaluate it?]. In K. Speck, T. Olk, O. Böhm-Kasper, H.-J. Stolz, & C. Wiezorek (Eds.), *Ganztagschulische Kooperation und Professionsentwicklung. Studien zu multiprofessionellen Teams und sozialräumlicher Vernetzung* (pp. 114-127). Weinheim: Juventa.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation. [Research methods and evaluation]*. Berlin: Springer.
- Dudley-Marling, C., & Burns, M. B. (2014). Two perspectives on inclusion in the United States. *Global Education Review*, 1(1), 14-31.
- Dupoux, E., Wolman, C., & Estrada, E. (2005). Teachers' attitudes toward integration of students with disabilities in Haïti and the United States. *International Journal of Disability, Development and Education*, 52(1), 43-58. doi:10.1080/10349120500071894
- Dyson, A. (2004). Inclusive education: A global agenda? *Japanese Journal of Special Education*, 41(6), 613-625.
- Dyson, A. (2014). A response to Göransson and Nilholm. *European Journal of Special Needs Education*, 29(3), 281-282. doi:10.1080/08856257.2014.933542
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Orlando, FL: Harcourt Brace Jovanovich.
- Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden. [Statistics and research methods]*. Weinheim: Beltz.
- Elliott, S. N., Davies, M., & Kettler, R. J. (2012). Australian students with disabilities accessing NAPLAN: Lessons from a decade of inclusive assessment in the United States. *International Journal of Disability, Development and Education*, 59(1), 7-19. doi:10.1080/1034912X.2012.654934
- Ellis, B. B., & Mead, A. D. (2004). Item analysis. Theory and practice using classical and modern test theory. In S. G. Rogelberg (Ed.), *Handbook of Research Methods in Industrial and Organizational Psychology* (pp. 324-343). Malden, MA: Blackwell.
- Engsig, T. T., & Johnstone, C. J. (2015). Is there something rotten in the state of Denmark? The paradoxical policies of inclusive education: Lessons from Denmark. *International Journal of Inclusive Education*, 19(5), 469-486. doi:10.1080/13603116.2014.940068
- Epstein, J., Osborne, R. H., Elsworth, G. R., Beaton, D. E., & Guillemin, F. (2015). Cross-cultural adaptation of the Health Education Impact Questionnaire: Experimental study showed expert committee, not back-translation, added value. *Journal of Clinical Epidemiology*, 68(4), 360-369. doi:10.1016/j.jclinepi.2013.07.013
- Epstein, J., Santob, R. M., & Guillemina, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *Journal of Clinical Epidemiology*, 68(4), 435-441. doi:10.1016/j.jclinepi.2014.11.021
- Ercikan, K., Arim, R., Law, D., Domene, J., Gagnon, F., & Lacroix, S. (2010). Application of think aloud protocols for examining and confirming sources of differential item functioning identified by expert reviews. *Educational Measurement: Issues & Practice*, 29(2), 24-35. doi:10.1111/j.1745-3992.2010.00173.x
- Ercikan, K., & Lyons-Thomas, J. (2013). Adapting tests for use in other languages and cultures. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. C. Hansen, N. R. Kuncel, S. P. Reise, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, Vol. 3: Testing and assessment in school psychology and education* (pp. 545-569). Washington, DC: APA.

- Eremenco, S. L., Cella, D., & Arnold, B. J. (2005). A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation & the Health Professions, 28*(2), 212-232. doi:10.1177/0163278705275342
- Erkut, S. (2010). Developing multiple language versions of instruments for intercultural research. *Child Development Perspectives, 4*(1), 19-24. doi:10.1111/j.1750-8606.2009.00111.x
- Erkut, S., Alarcón, O., García Coll, C., Tropp, L. R., & Vázquez García, H. A. (1999). The dual-focus approach to creating bilingual measures. *Journal of Cross-Cultural Psychology, 30*(2), 206-218. doi:10.1177/0022022199030002004
- Erten, O., & Savage, R. S. (2012). Moving forward in inclusive education research. *International Journal of Inclusive Education, 16*(2), 221-233. doi:10.1080/13603111003777496
- European Social Survey. (2016). *ESS round 8 interviewer briefing: NC manual*. London: ESS ERIC Headquarters.
- Evans, W. N., Morrill, M. S., & Parente, S. T. (2010). Measuring inappropriate medical diagnosis and treatment in survey data: The case of ADHD among school-age children. *Journal of Health Economics, 29*(5), 657-673. doi:10.1016/j.jhealeco.2010.07.005
- Fachinetti, A. (2015). A short personal and political history of NAPLAN. *Education Today, 15*(4), 20-22.
- Favazza, P. C., & Odom, S. L. (1997). Promoting positive attitudes of kindergarten-age children toward people with disabilities. *Exceptional Children, 63*(3), 405-418. doi:10.1177/001440299706300308
- Ferguson, E., & Cox, T. (1993). Exploratory factor analysis: A users' guide. *International Journal of Selection and Assessment, 1*(2), 84-94. doi:10.1111/j.1468-2389.1993.tb00092.x
- Field, A. (2013). *Discovering statistics using IBM SPSS Statistics* (4 ed.). London: Sage.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.
- Florian, L., & Black-Hawkins, K. (2011). Exploring inclusive pedagogy. *British Educational Research Journal, 37*(5), 813-828. doi:10.1080/01411926.2010.501096
- Florian, L., & Linklater, H. (2010). Preparing teachers for inclusive education: Using inclusive pedagogy to enhance teaching and learning for all. *Cambridge Journal of Education, 40*(4), 369-386. doi:10.1080/0305764X.2010.526588
- Forlin, C., Earle, C., Loreman, T., & Sharma, U. (2011). The Sentiments, Attitudes, and Concerns about Inclusive Education Revised (SACIE-R) scale for measuring pre-service perceptions teachers' perception about inclusion. *Exceptionality Education International, 21*(3), 50-65.
- Forlin, C., García Cedillo, I., Romero-Contreras, S., Fletcher, T., & Rodríguez Hernández, H. J. (2010). Inclusion in Mexico: Ensuring supportive attitudes by newly graduated teachers. *International Journal of Inclusive Education, 14*(7), 723-739. doi:10.1080/13603111003778569
- Forlin, C., & Sin, K.-F. (2010). Developing support for inclusion: A professional learning approach for teachers in Hong Kong. *International Journal of Whole Schooling, 6*(1), 7-26.
- Fourie, J., & Feinauer, I. (2005). The quality of translated medical research questionnaires. *Southern African Linguistics and Applied Language Studies, 23*(4), 349-367. doi:10.2989/16073610509486395
- Fraij, A., Maschke, S., & Stecher, L. (2015). Die Scholarisierung der Jugendphase – ein Zeitvergleich. [Scholarisation of adolescence – a time comparison]. *Diskurs Kindheits- und Jugendforschung, 10*(2), 167-182.

- Fraser, B. J., Anderson, G. J., & Walberg, H. J. (1982). *Assessment of learning environments: Manual for Learning Environment Inventory (LEI) and My Class Inventory (MCI)*. Perth: Australian Institute of Technology.
- Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. London: Sage.
- Gehlbach, H., Robinson, C. D., Finefter-Rosenbluh, I., Benshoof, C., & Schneider, J. (2017). Questionnaires as interventions: Can taking a survey increase teachers' openness to student feedback surveys? *Educational Psychology*. doi:10.1080/01443410.2017.1349876
- Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, 6(4), 304-312. doi:10.1037/1040-3590.6.4.304
- Geisinger, K. F. (2003). Testing and assessment in cross-cultural psychology. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of Psychology, Vol. 10: Assessment Psychology* (pp. 95-117). Hoboken, NJ: John Wiley & Sons, Inc.
- Gerbing, D. W., & Hamilton, J. G. (1996). Viability of exploratory factor analysis as a precursor to confirmatory factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 3(1), 62-72. doi:10.1080/10705519609540030
- Glasman, L. R., & Albarracín, D. (2006). Forming attitudes that predict future behavior: A meta-analysis of the attitude-behavior relation. *Psychological Bulletin*, 132(5), 778-822. doi:10.1037/0033-2909.132.5.778
- Glazzard, J. (2013). A critical interrogation of the contemporary discourses associated with inclusive education in England. *Journal of Research in Special Educational Needs*, 13(3), 182-188. doi:10.1111/1471-3802.12018
- Goldhaber, D. (2016). In schools, teacher quality matters most. *Education Next*, 16(2), 56-62.
- Göransson, K., & Nilholm, C. (2014). Conceptual diversities and empirical shortcomings – a critical analysis of research on inclusive education. *European Journal of Special Needs Education*, 29(3), 265-280. doi:10.1080/08856257.2014.933545
- Graham, J. W. (2012). *Missing data: Analysis and design*. New York: Springer.
- Graham, L. J., & Jahnukainen, M. (2011). Wherefore art thou, inclusion? Analysing the development of inclusive education in New South Wales, Alberta and Finland. *Journal of Education Policy*, 26(2), 263-288. doi:10.1080/02680939.2010.493230
- Graham, L. J., & Macartney, B. (2012). Naming or creating a problem? The mis/use of labels in schools. In S. Carrington & J. Macarthur (Eds.), *Teaching in inclusive school communities* (pp. 189-208). Milton: John Wiley & Sons.
- Graham, L. J., & Slee, R. (2008). An illusory interiority: Interrogating the discourse/s of inclusion. *Educational Philosophy and Theory*, 40(2), 277-293. doi:10.1111/j.1469-5812.2007.00331.x
- Graham, L. J., & Sweller, N. (2011). The inclusion lottery: Who's in and who's out? Tracking inclusion and exclusion in New South Wales government schools. *International Journal of Inclusive Education*, 15(9), 941-953. doi:10.1080/13603110903470046
- Gregory, J. L., & Noto, L. A. (2012). *Technical manual for Attitudes Towards Teaching All Students (ATTAS-mm) instrument*. Paper presented at the Council for Exceptional Children/Teacher Education Division Annual Conference, Grand Rapids, MI. <http://files.eric.ed.gov/fulltext/ED509930.pdf>
- Guillemin, F., Bombardier, C., & Beaton, D. E. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology*, 46(12), 1417-1432. doi:10.1016/0895-4356(93)90142-N
- Guo, S. (2014). Revisioning Education for All in the age of migration: Global challenges and opportunities for lifelong learning. *International Review of Education*, 60(4), 481-497. doi:10.1007/s11159-014-9441-1

- Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9(2), 139-150. doi:10.2307/2086306
- Haddad, W. D., Colletta, N. J., Fisher, N., Lakin, M., & Sutton, M. (1990). *Meeting basic learning needs: A vision for the 1990s*. New York: UNICEF House.
- Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hambleton, R. K., & Patsula, L. (1998). Adapting tests for use in multiple languages and cultures. *Social Indicators Research*, 45(1), 153-171. doi:10.1023/A:1006941729637
- Hardy, I., & Woodcock, S. (2015). Inclusive education policies: Discourses of difference, diversity and deficit. *International Journal of Inclusive Education*, 19(2), 141-164. doi:10.1080/13603116.2014.908965
- Harkness, J. A., Edwards, B., Hansen, S. E., Miller, D. R., & Villar, A. (2010). Designing questionnaires for multipopulation research. In J. A. Harkness, M. Braun, B. Edwards, T. P. Johnson, L. Lyberg, P. P. Mohler, B.-E. Pennell, & T. W. Smith (Eds.), *Survey methods in multinational, multiregional, and multicultural contexts* (pp. 33-57). Hoboken, NJ: John Wiley & Sons.
- Harrington, I., Kastirke, N., & Holtbrink, L. (Eds.). (2016). *Inklusion in Deutschland und Australien. [Inclusion in Germany and Australia]*. Wiesbaden: Springer VS.
- Hattie, J. A. C. (2003). *Teachers make a difference, what is the research evidence?* Paper presented at the Building teacher quality: What does the research tell us, ACER Research Conference, Melbourne, Australia.
http://research.acer.edu.au/research_conference_2003/4/
- Hattie, J. A. C. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Haug, P. (2014). Empirical shortcomings? A comment on Kerstin Göransson and Claes Nilholm, 'Conceptual diversities and empirical shortcomings – a critical analysis of research on inclusive education'. *European Journal of Special Needs Education*, 29(3), 283-285. doi:10.1080/08856257.2014.933548
- Haug, P. (2017). Understanding inclusive education: Ideals and reality. *Scandinavian Journal of Disability Research*, 19(3), 206-217. doi:10.1080/15017419.2016.1224778
- Hebborn, K. (2014). Inklusion im Schulbereich: Handlungsmöglichkeiten und Gelingensbedingungen aus kommunaler Sicht. [Inclusion in schools: Opportunities for action and conditions for success from a municipality's point of view]. In S. Kroworsch (Ed.), *Inklusion im deutschen Schulsystem: Barrieren und Lösungswege* (pp. 56-67). Berlin: Verlag des Deutschen Vereins für öffentliche und private Fürsorge e.V.
- Hecht, P., Niedermair, C., & Feyerer, E. (2016). Einstellungen und inklusionsbezogene Selbstwirksamkeitsüberzeugungen von Lehramtsstudierenden und Lehrpersonen im Berufseinstieg – Messverfahren und Befunde aus einem Mixed-Methods-Design. *Empirische Sonderpädagogik*, 8(1), 86-102.
- Hellmich, F., & Görel, G. (2014). Erklärungsfaktoren für Einstellungen von Lehrerinnen und Lehrern zum inklusiven Unterricht in der Grundschule. [Prerequisites of primary school teachers' attitudes toward inclusion]. *Zeitschrift für Bildungsforschung*, 4(3), 227-240. doi:10.1007/s35834-014-0102-z
- Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität: Diagnose, Evaluation und Verbesserung des Unterrichts. [Quality teaching and professionalism of teachers: Diagnostics, evaluation and improvement of teaching]*. Seelze-Velber: Kallmeyer.
- Henry-Huthmacher, C. (2015). Auf dem Prüfstand: Inklusion im deutschen Schulsystem. *Analysen & Argumente*, 185, 1-11.

- Herdman, M., Fox-Rushby, J., & Badia, X. (1997). 'Equivalence' and the translation and adaptation of health-related quality of life questionnaires. *Quality of Life Research*, 6(3), 237-247. doi:10.1023/A:1026410721664
- Hinz, A. (2015). Inclusion as a 'North Star' and prospects for everyday life: Considerations about concerns, transformations and necessities of inclusion in schools in Germany. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 18(1), 17-33. doi:10.6018/reifop.18.1.214321
- Hitlin, S., & Pinkston, K. (2013). Values, attitudes, and ideologies: Explicit and implicit constructs shaping perception and action. In J. DeLamater & A. Ward (Eds.), *Handbook of social psychology* (pp. 319-339). Dordrecht: Springer.
- Hogarty, K. Y., Hines, C. V., Kromrey, J. D., Ferron, J. M., & Mumford, K. R. (2005). The quality of factor solutions in exploratory factor analysis: The influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement*, 65(2), 202-226. doi:10.1177/0013164404267287
- Horne, P. E., Timmons, V., & Adamowycz, R. (2008). Identified teacher supports for inclusive practice. *Exceptionality Education Canada*, 18(3), 82-94.
- Hsieh, W.-Y., & Hsieh, C.-M. (2012). Urban early childhood teachers' attitudes towards inclusive education. *Early Child Development and Care*, 182(9), 1167-1184. doi:10.1080/03004430.2011.602191
- Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. doi:10.1080/10705519909540118
- Huang, H.-H., & Diamond, K. E. (2009). Early childhood teachers' ideas about including children with disabilities in programmes designed for typically developing children. *International Journal of Disability Development and Education*, 56(2), 169-182. doi:10.1080/10349120902868632
- Husbands, C., & Pearce, J. (2012). *What makes great pedagogy? Nine claims from research*. Retrieved from Nottingham:
https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/329746/what-makes-great-pedagogy-nine-claims-from-research.pdf
- ITC. (2010). *International Test Commission guidelines for translating and adapting tests*. Retrieved from <http://www.intestcom.org>
- Jackson, D. L., Gillasp, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14(1), 6-23. doi:10.1037/a0014694
- Jha, M. M. (2008). *School without walls: Inclusive education for all*. New Delhi: Dorling Kindersley.
- Johnston, J. (2017). Australian NAPLAN testing: In what ways is this a 'wicked' problem? *Improving Schools*, 20(1), 18-34. doi:10.1177/1365480216673170
- Jones, P. S., Lee, J. W., Phillips, L. R., Zhang, X. E., & Jaceldo, K. B. (2001). An adaptation of Brislin's translation model for cross-cultural research. *Nursing Research*, 50(5), 300-304.
- Jones, W. P., & Loe, S. A. (2013). Optimal number of questionnaire response categories: More may not be better. *SAGE Open*, 3(2), 1-10. doi:10.1177/2158244013489691
- Jonkisz, E., Moosbrugger, H., & Brandt, H. (2012). Planung und Entwicklung von Tests und Fragebogen. [Planning and developing tests and questionnaires]. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 27-74). Berlin: Springer.
- Kane, M. (2015). Validation strategies. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 64-80). New York: Routledge.
- Katz, D. (1960). The functional approach to the study of attitudes. *Public Opinion Quarterly*, 24(2), 163-204. doi:10.1086/266945

- Kielblock, S., Gaiser, J. M., & Stecher, L. (2017). Multiprofessionelle Kooperation als Fundament der inklusiven Ganztagschule. [Multiprofessional collaboration as a foundation for inclusive all-day schools]. *Gemeinsam leben*, 25(3), 140-148.
- Kielblock, S., & Lange, A. (2013). Das problemzentrierte Interview: Grundlagen und Forschungspraxis. [The problem-centred interview: Principles and practice]. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 33(4), 439-448.
- Kiuppis, F. (2014). Why (not) associate the principle of inclusion with disability? Tracing connections from the start of the 'Salamanca Process'. *International Journal of Inclusive Education*, 18(7), 746-761. doi:10.1080/13603116.2013.826289
- Klemm, K. (2012). *Zusätzliche Ausgaben für ein inklusives Schulsystem in Deutschland*. [Additional expenses for an inclusive school system in Germany]. Retrieved from Gütersloh: https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/Zusaetzl_Ausgaben_inkl_Schulsystem_in_D_Mrz_12.pdf
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3 ed.). New York: Guilford Press.
- KMK. (2011). *Inklusive Bildung von Kindern und Jugendlichen mit Behinderungen in Schulen (Beschluss der Kultusministerkonferenz vom 20.10.2011)*. [Inclusive education of children and youths with disabilities in schools (Decision of the Standing Conference of the Ministers of Education and Cultural Affairs from 20/10/2011)]. Retrieved from http://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2011/2011_10_20-Inklusive-Bildung.pdf
- Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30(3), 607-610. doi:10.1177/001316447003000308
- Kruse, S., & Dederig, K. (2017). The idea of inclusion: Conceptual and empirical diversities in Germany. *Improving Schools*(OnlineFirst), 1-13. doi:10.1177/1365480217707835
- Kuyini, A. B., & Desai, I. (2007). Principals' and teachers' attitudes and knowledge of inclusive education as predictors of effective teaching practices in Ghana. *Journal of Research in Special Educational Needs*, 7(2), 104-113. doi:10.1111/j.1471-3802.2007.00086.x
- Lane, S., Raymond, M. R., Haladyna, T. M., & Downing, S. M. (2015). Test development process. In S. Lane, M. R. Raymond, & T. M. Haladyna (Eds.), *Handbook of Test Development* (pp. 3-18). New York: Routledge.
- Larrivee, B. (1982). Factors underlying regular classroom teachers' attitude toward mainstreaming. *Psychology in the Schools*, 19(3), 374-379. doi:10.1002/1520-6807(198207)19:3<374::aid-pits2310190319>3.0.co;2-p
- Larrivee, B., & Cook, L. (1979). Mainstreaming: A study of the variables affecting teacher attitude. *The Journal of Special Education*, 13(3), 315-324. doi:10.1177/002246697901300310
- Lee, J., & Paek, I. (2014). In search of the optimal number of response categories in a rating scale. *Journal of Psychoeducational Assessment*, 32(7), 663-673. doi:10.1177/0734282914522200
- Leigh, A. (2010). Estimating teacher effectiveness from two-year changes in students' test scores. *Economics of Education Review*, 29(3), 480-488. doi:10.1016/j.econedurev.2009.10.010
- Lenzer, T., Neuert, C., & Otto, W. (2015). Kognitives Pretesting. [Cognitive pre-testing]. *GESIS Survey Guidelines*. doi:10.15465/gesis-sg_010
- Levin, Henry M. (2017). Empirical puzzles on effective teachers: U.S. research. In M. Rosén, K. Yang Hansen, & U. Wolff (Eds.), *Cognitive abilities and educational outcomes: A festschrift in honour of Jan-Eric Gustafsson* (pp. 189-204). Cham: Springer.

- Lietz, P. (2010). Research into questionnaire design: A summary of the literature. *International Journal of Market Research*, 52(2), 249-272. doi:10.2501/s147078530920120x
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, No. 140.
- Lonner, W. J. (1968). The SVIB visits German, Austrian, and Swiss psychologists. *American Psychologist*, 23(4), 164-179. doi:10.1037/h0025668
- Loreman, T. (2014). Measuring inclusive education outcomes in Alberta, Canada. *International Journal of Inclusive Education*, 18(5), 459-483. doi:10.1080/13603116.2013.788223
- Loreman, T., Earle, C., Sharma, U., & Forlin, C. (2007). The development of an instrument for measuring pre-service teachers' sentiments, attitudes, and concerns about inclusive education. *International Journal of Special Education*, 22(2), 150-159.
- Loreman, T., Forlin, C., & Sharma, U. (2007). An international comparison of pre-service teacher attitudes towards inclusive education. *Disability Studies Quarterly*, 27(4). doi:10.18061/dsq.v27i4.53
- Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology*, 4(2), 73-79. doi:10.1027/1614-2241.4.2.73
- Lüke, T., & Grosche, M. (2017). Implicitly measuring attitudes towards inclusive education: a new attitude test based on single-target implicit associations. *European Journal of Special Needs Education*. doi:10.1080/08856257.2017.1334432
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4(1), 84-99. doi:10.1037/1082-989X.4.1.84
- MacFarlane, K., & Woolfson, L. M. (2013). Teacher attitudes and behavior toward the inclusion of children with social, emotional and behavioral difficulties in mainstream schools: An application of the theory of planned behavior. *Teaching and Teacher Education*, 29(1), 46-52. doi:10.1016/j.tate.2012.08.006
- Mahat, M. (2008). The development of a psychometrically-sound instrument to measure teachers' multidimensional attitudes toward inclusive education. *International Journal of Special Education*, 23(1), 82-92.
- Maio, G. R., & Haddock, G. (2009). *The psychology of attitudes and attitude change*. London: SAGE.
- Malinen, O.-P., Savolainen, H., & Xu, J. (2012). Beijing in-service teachers' self-efficacy and attitudes towards inclusive education. *Teaching and Teacher Education*, 28(4), 526-534. doi:10.1016/j.tate.2011.12.004
- Maneesriwongul, W., & Dixon, J. K. (2004). Instrument translation process: A methods review. *Methodological Issues in Nursing Research*, 48(2), 175-186. doi:10.1111/j.1365-2648.2004.03185.x
- Mansfield, R. K. (2015). Teacher quality and student inequality. *Journal of Labor Economics*, 33(3), 751-788. doi:10.1086/679683
- Marom, M., Cohen, D., & Naon, D. (2007). Changing disability-related attitudes and self-efficacy of Israeli children via the partners to inclusion programme. *International Journal of Disability, Development and Education*, 54(1), 113-127. doi:10.1080/10349120601149821
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(3), 320-341. doi:10.1207/s15328007sem1103_2
- Matsumoto, D. (2005). Cross-cultural research. In S. F. Davis (Ed.), *Handbook of research methods in experimental psychology* (pp. 189-208). Malden, MA: Blackwell.

- Maul, A. (2013). Method effects and the meaning of measurement. *Frontiers in Psychology*, 4, 1-13. doi:10.3389/fpsyg.2013.00169
- McGhie-Richmond, D., Barber, J., Lupart, J., & Loreman, T. (2009). The development of a Canadian instrument for measuring teacher views of their inclusive school environment in a rural context: The Teacher Perceptions of Inclusion in Rural Canada (TPIRC) scale. *International Journal of Special Education*, 24(2), 103-108.
- McGhie-Richmond, D., Irvine, A., Loreman, T., Cizman, J. L., & Lupart, J. (2013). Teacher perspectives on inclusive education in rural Alberta, Canada. *Canadian Journal of Education*, 36(1), 195-239.
- Messiou, K. (2017). Research in the field of inclusive education: Time for a rethink? *International Journal of Inclusive Education*, 21(2), 146-159. doi:10.1080/13603116.2016.1223184
- Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept*. Cambridge: Cambridge University Press.
- Miles, S., Lene, D., & Merumeru, L. (2014). Making sense of inclusive education in the Pacific region: Networking as a way forward. *Childhood*, 21(3), 339-353. doi:10.1177/0907568214524458
- Miles, S., & Singal, N. (2010). The Education for All and inclusive education debate: Conflict, contradiction or opportunity? *International Journal of Inclusive Education*, 14(1), 1-15. doi:10.1080/13603110802265125
- Miville, M. L., Gelso, C. J., Pannu, R., Liu, W., Touradji, P., Holloway, P., & Fuertes, J. (1999). Appreciating similarities and valuing differences: The Miville-Guzman universality-diversity scale. *Journal of Counseling Psychology*, 46(3), 291-307. doi:10.1037/0022-0167.46.3.291
- Moberg, S. (1997). Inclusive educational practices as perceived by prospective special education teachers in Estonia, Finland, and the United States. *International Journal of Rehabilitation Research*, 20(1), 29-40.
- Moberg, S. (2003). Education for all in the North and the South: Teachers' Attitudes Towards Inclusive Education in Finland and Zambia. *Education and Training in Developmental Disabilities*, 38(4), 417-428.
- Monsen, J. J., Ewing, D. L., & Boyle, J. (2015). Psychometric properties of the revised teachers' attitude toward inclusion scale. *International Journal of School & Educational Psychology*, 3(1), 64-71. doi:10.1080/21683603.2014.938383
- Monsen, J. J., Ewing, D. L., & Kwoka, M. (2014). Teachers' attitudes towards inclusion, perceived adequacy of support and classroom learning environment. *Learning Environments Research*, 17(1), 113-126. doi:10.1007/s10984-013-9144-8
- Monsen, J. J., & Frederickson, N. (2004). Teachers' attitudes towards mainstreaming and their pupils' perceptions of their classroom learning environment. *Learning Environments Research*, 7(2), 129-142. doi:10.1023/B:LERI.0000037196.62475.32
- Montgomery, A., & Mirenda, P. (2014). Teachers' self-efficacy, sentiments, attitudes, and concerns about the inclusion of students with developmental disabilities. *Exceptionality Education International*, 24, 18-32.
- Mundy, K. (2016). "Leaning in" on Education for All. *Comparative Education Review*, 60(1), 1-26. doi:10.1086/684434
- Muñiz, J., García-Cueto, E., & Lozano, L. M. (2005). Item format and the psychometric properties of the Eysenck Personality Questionnaire. *Personality and Individual Differences*, 38(1), 61-69. doi:10.1016/j.paid.2004.03.021
- National Health and Medical Research Council. (2015). *National statement on ethical conduct in human research 2007 (updated May 2015)*. Retrieved from www.nhmrc.gov.au/guidelines/publications/e72

- Nes, K., & Strømstad, M. (2006). Strengthened adapted education for all – no more special education? *International Journal of Inclusive Education*, 10(4-5), 363-378. doi:10.1080/13603110500264610
- Nilholm, C., Almqvist, L., Göransson, K., & Lindqvist, G. (2013). Is it possible to get away from disability-based classifications in education? An empirical investigation of the Swedish system. *Scandinavian Journal of Disability Research*, 15(4), 379-391. doi:10.1080/15017419.2012.735201
- Nilholm, C., & Göransson, K. (2017). What is meant by inclusion? An analysis of European and North American journal articles with high impact. *European Journal of Special Needs Education*, 32(3), 437-451. doi:10.1080/08856257.2017.1295638
- Norwegian Ministry of Education and Research. (2014). *Act of 17 July 1998 no. 61 relating to Primary and Secondary Education and Training (the Education Act)*. Oslo: Department of Education and Training.
- NSW Department of Education. (2015a). *Gender analysis of school teachers*. Retrieved from www.dec.nsw.gov.au/documents/15060385/15385042/gender-analysis-school-teacher.pdf
- NSW Department of Education. (2015b). *State Education Research Applications Process (SERAP): Guidelines 2015*. Retrieved from www.serap.det.nsw.edu.au
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers*, 32(3), 396-402. doi:10.3758/BF03200807
- OECD. (2005). *Teachers matter: Attracting, developing and retaining effective teachers*. Paris: OECD.
- OECD. (2007). *Education at a glance 2007: OECD indicators*. Paris: OECD.
- OECD. (2017). *Education at a glance 2017: OECD indicators*. Paris: OECD Publishing.
- Opertti, R., & Brady, J. (2011). Developing inclusive teachers from an inclusive curricular perspective. *Prospects*, 41(3), 459-472. doi:10.1007/s11125-011-9205-7
- Opertti, R., Brady, J., & Duncombe, L. (2009). Moving forward: Inclusive education as the core of Education for All. *Prospects*, 39(3), 205-214. doi:10.1007/s11125-009-9112-3
- Osgood, C. E. (1952). The nature and measurement of meaning. *Psychological Bulletin*, 49(3), 197-237. doi:10.1037/h0055737
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Pantić, N. (2015). A model for study of teacher agency for social justice. *Teachers and Teaching*, 21(6), 759-778. doi:10.1080/13540602.2015.1044332
- Pantić, N., & Florian, L. (2015). Developing teachers as agents of inclusion and social justice. *Education Inquiry*, 6(3), 333-351. doi:10.3402/edui.v6.27311
- Penfield, R. D. (2013). Item analysis. In K. F. Geisinger (Ed.), *APA Handbook of Testing and Assessment in Psychology: Vol. 1. Test Theory and Testing and Assessment in Industrial and Organizational Psychology* (pp. 121-138). Washington, DC: American Psychological Association.
- Peters, S. J. (2004). *Inclusive Education: An EFA strategy for all children*. Washington, D.C.: World Bank.
- Peters, S. J. (2007). "Education for All?" A historical analysis of international inclusive education policy and individuals with disabilities. *Journal of Disability Policy Studies*, 18(2), 98-108. doi:10.1177/10442073070180020601
- Phillips, H. P. (1959). Problems of translation and meaning in field work. *Human Organization*, 18(4), 184-192. doi:10.17730/humo.18.4.27454x0233213823
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1-15. doi:10.1016/S0001-6918(99)00050-5

- Procter, M. (2008). Measuring Attitudes. In N. Gilbert (Ed.), *Researching Social Life* (3 ed., pp. 206-225). London: SAGE.
- Quertemont, E. (2011). How to statistically show the absence of an effect. *Psychologica Belgica*, 51(2), 109-127. doi:10.5334/pb-51-2-109
- Reh, S., & Breuer, A. (2012). Positionierungen in interprofessionellen Teams – Kooperationspraktiken an Ganztagschulen. [Positioning in interprofessional teams – Practices of collaboration at all-day schools]. In S. G. Huber & F. Ahlgrimm (Eds.), *Kooperation: Aktuelle Forschung zur Kooperation in und zwischen Schulen sowie mit anderen Partnern* (pp. 185-201). Münster: Waxmann.
- Revilla, M. A., Saris, W. E., & Krosnick, J. A. (2014). Choosing the number of categories in agree–disagree scales. *Sociological Methods & Research*, 43(1), 73-97. doi:10.1177/0049124113509605
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417-458. doi:10.1111/j.1468-0262.2005.00584.x
- Roche, S. (2016). Education for all: Exploring the principle and process of inclusive education. *International Review of Education*, 62(2), 131-137. doi:10.1007/s11159-016-9556-7
- Rödler, P. (2012). Inklusion und Standards: Eine verständliche, aber paradoxe Forderung. [Inclusion and standards: A reasonable, but paradoxical claim]. *Gemeinsam leben*, 20(1), 33-40.
- Rosenbaum, P. L., Armstrong, R. W., & King, S. M. (1986). Children's attitudes toward disabled peers: A self-report measure. *Journal of Pediatric Psychology*, 11(4), 517-530. doi:10.1093/jpepsy/11.4.517
- Rosenbaum, P. L., Armstrong, R. W., & King, S. M. (1987). Parental attitudes toward children with handicaps: New perspectives with a new measure. *Journal of Developmental and Behavioral Pediatrics*, 8(6), 327-334. doi:10.1097/00004703-198712000-00004
- Rosenberg, M. J., & Hovland, C. I. (1960). Cognitive, affective, and behavioral components of attitudes. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson, & J. W. Brehm (Eds.), *Attitude organization and change: An analysis of consistency among attitude components* (pp. 1-14). New Haven, CT: Yale University Press.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1-36. doi:10.18637/jss.v048.i02
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion*. [Textbook test theory – test development]. Bern: Hans Huber.
- Ruberg, C., & Porsch, R. (2017). Einstellungen von Lehramtsstudierenden und Lehrkräften zur schulischen Inklusion: Ein systematisches Review deutschsprachiger Forschungsarbeiten. [Attitudes of pre-service and in-service teachers towards school inclusion: A systematic review of German studies]. *Zeitschrift für Pädagogik*, 63(4), 393-415.
- Saloviita, T. (2015). Measuring pre-service teachers' attitudes towards inclusive education: Psychometric properties of the TAIS scale. *Teaching and Teacher Education*, 52, 66-72. doi:10.1016/j.tate.2015.09.003
- Saloviita, T., & Schaffus, T. (2016). Teacher attitudes towards inclusive education in Finland and Brandenburg, Germany and the issue of extra work. *European Journal of Special Needs Education*, 31(4), 458-471. doi:10.1080/08856257.2016.1194569
- Savin-Baden, M., & Major, C. H. (2013). *Qualitative research: The essential guide to theory and practice*. London: Routledge.
- Savolainen, H., Engelbrecht, P., Nel, M., & Malinen, O.-P. (2012). Understanding teachers' attitudes and self-efficacy in inclusive education: Implications for pre-service and in-service teacher education. *European Journal of Special Needs Education*, 27(1), 51-68. doi:10.1080/08856257.2011.613603

- Schneider, C. (2015). Social participation of children and youth with disabilities in Canada, France and Germany. *International Journal of Inclusive Education*, 19(10), 1068-1079. doi:10.1080/13603116.2015.1037867
- Schnell, R., Hill, P. B., & Esser, E. (2011). *Methoden der empirischen Sozialforschung. [Methods of empirical social research]*. München: Oldenbourg.
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323-338. doi:10.3200/JOER.99.6.323-338
- Schwarz, N. (2015). Attitude measurement. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences (Vol. 2)* (2 ed., pp. 178-182). Oxford, UK: Elsevier.
- Sharma, U., & Desai, I. (2002). Measuring concerns about integrated education in India. *Asia and Pacific Journal of Disability*, 5(1), 2-14.
- Sharma, U., & George, S. (2016). Understanding teacher self-efficacy to teach in inclusive classrooms. In S. Garvis & D. Pendergast (Eds.), *Asia-pacific perspectives on teacher self-efficacy* (pp. 37-51). Rotterdam: Sense.
- Sharma, U., Loreman, T., & Forlin, C. (2012). Measuring teacher efficacy to implement inclusive practices. *Journal of Research in Special Educational Needs*, 12(1), 12-21. doi:10.1111/j.1471-3802.2011.01200.x
- Sharma, U., Shaukat, S., & Furlonger, B. (2015). Attitudes and self-efficacy of pre-service teachers towards inclusion in Pakistan. *Journal of Research in Special Educational Needs*, 15(2), 97-105. doi:10.1111/1471-3802.12071
- Sharma, U., & Sokal, L. (2016). Can teachers' self-reported efficacy, concerns, and attitudes toward inclusion scores predict their actual inclusive classroom practices? *Australasian Journal of Special Education*, 40(1), 21-38. doi:jse.2015.14
- Shoemaker, D. M. (1975). Toward a framework for achievement testing. *Review of Educational Research*, 45(1), 127-147. doi:10.3102/00346543045001127
- Shyman, E. (2015). Toward a globally sensitive definition of inclusive education based in social justice. *International Journal of Disability, Development and Education*, 62(4), 351-362. doi:10.1080/1034912X.2015.1025715
- Slee, R. (2001a). Driven to the margins: Disabled students, inclusive schooling and the politics of possibility. *Cambridge Journal of Education*, 31(3), 385-397. doi:10.1080/03057640120086620
- Slee, R. (2001b). Social justice and the changing directions in educational research: The case of inclusive education. *International Journal of Inclusive Education*, 5(2/3), 167-177. doi:10.1080/13603110010035832
- Slee, R. (2008). Beyond special and regular schooling? An inclusive education reform agenda. *International Studies in Sociology of Education*, 18(2), 99-116. doi:10.1080/09620210802351342
- Slee, R. (2011). *The irregular school: Exclusion, schooling and inclusive education*. London: Routledge.
- Slee, R. (2013). How do we make inclusive education happen when exclusion is a political predisposition? *International Journal of Inclusive Education*, 17(8), 895-907. doi:10.1080/13603116.2011.602534
- Slee, R., & Allan, J. (2001). Excluding the included: A reconsideration of inclusive education. *International Studies in Sociology of Education*, 11(2), 173-192. doi:10.1080/09620210100200073
- Sokal, L., & Sharma, U. (2013). Canadian in-service teachers' concerns, efficacy, and attitudes about inclusive teaching. *Exceptionality Education International*, 23(1), 59-71.
- Solano-Flores, G., Backhoff, E., & Contreras-Niño, L. Á. (2009). Theory of test translation error *International Journal of Testing*, 9(2), 78-91. doi:10.1080/15305050902880835

- Solano-Flores, G., Contreras-Niño, L. Á., & Backhoff, E. (2013). The measurement of translation error in PISA-2006 items: An application of the theory of test translation error. In M. Prenzel, M. Kobarg, K. Schöps, & S. Rönnebeck (Eds.), *Research on PISA. Research outcomes of the PISA research conference 2009* (pp. 71-85). Dordrecht: Springer.
- Sousa, V. D., & Rojjanasiriat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice*, 17(2), 268-274. doi:10.1111/j.1365-2753.2010.01434.x
- Speck, K., Olk, T., & Stimpel, T. (2011). Professionelle Kooperation unterschiedlicher Berufskulturen an Ganztagschulen – zwischen Anspruch und Wirklichkeit. [Professional collaboration of different professional cultures at all-day schools – between ambition and reality]. In K. Speck, T. Olk, O. Böhm-Kasper, H.-J. Stolz, & C. Wiezorek (Eds.), *Ganztags schulische Kooperation und Professionsentwicklung. Studien zu multiprofessionellen Teams und sozialräumlicher Vernetzung* (pp. 69-84). Weinheim: Juventa.
- Statistische Ämter des Bundes und der Länder. (2017). Lehrkräfte an allgemein bildenden Schulen nach Beschäftigungsumfang und Geschlecht. Schuljahr 2015/2016. [Teachers at regular schools, divided into volume of employment and gender: School year 2015/2016]. Retrieved from http://www.statistik-portal.de/Statistik-Portal/de_jb04_jahrtab24sch.asp
- Stecher, L. (2005). Quantitative und qualitative Auswahlverfahren – Widerspruch oder Ergänzung?. [Quantitative and qualitative sampling procedures – contradiction or complement?]. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 25(4), 440-448.
- Stoiber, K. C., Gettinger, M., & Goetz, D. (1998). Exploring factors influencing parents' and early childhood practitioners' beliefs about inclusion. *Early Childhood Research Quarterly*, 13(1), 107-124. doi:10.1016/S0885-2006(99)80028-3
- Sucuoğlu, B., Bakkaloğlu, H., İşcen Karasu, F., Demir, Ş., & Akalın, S. (2013). Inclusive preschool teachers: Their attitudes and knowledge about inclusion. *International Journal of Early Childhood Special Education*, 5(2), 107-128. doi:10.20489/intjces.107929
- Sucuoğlu, B., Bakkaloğlu, H., İşcen Karasu, F., Demir, Ş., & Akalın, S. (2014). Preschool teachers' knowledge levels about inclusion. *Educational Sciences: Theory & Practice*, 14(4), 1477-1483. doi:10.12738/estp.2014.4.2078
- Taber, K. S. (2017). The use of Cronbach's Alpha when developing and reporting research instruments in science education. *Research in Science Education*. doi:10.1007/s11165-016-9602-2
- Taylor, R. W., & Ringlaben, R. P. (2012). Impacting pre-service teachers' attitudes toward inclusion. *Higher Education Studies*, 2(3), 16-23. doi:10.5539/hes.v2n3p16
- Thomas, G. (2013). A review of thinking and research about inclusive education policy, with suggestions for a new kind of inclusive thinking. *British Educational Research Journal*, 39(3), 473-490. doi:10.1080/01411926.2011.652070
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology*, 33(4), 529-554. doi:10.1086/214483
- Thurstone, L. L., & Chave, E. J. (1929). *The measurement of attitude*. Chicago: University of Chicago Press.
- Tomlinson, S. (2012). The irresistible rise of the SEN industry. *Oxford Review of Education*, 38(3), 267-286. doi:10.1080/03054985.2012.692055
- Tomlinson, S. (2015). Is a sociology of special and inclusive education possible? *Educational Review*, 67(3), 273-281. doi:10.1080/00131911.2015.1021764

- Tomlinson, S. (2017). *A sociology of special and inclusive education: Exploring the manufacture of inability*. London: Routledge.
- Tota, P. M. (2014). Filling the gaps: The role and impact of international non-governmental organisations in 'Education for All'. *Globalisation, Societies and Education*, 12(1), 92-109. doi:10.1080/14767724.2013.858988
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Triandis, H. C. (1971). *Attitude and attitude change*. New York: John Wiley & Sons.
- Tsakiridou, H., & Polyzopoulou, K. (2014). Greek Teachers' Attitudes toward the Inclusion of Students with Special Educational Needs. *American Journal of Educational Research*, 2(4), 208-218. doi:10.12691/education-2-4-6
- Tyupa, S. (2011). A theoretical framework for back-translation as a quality assessment tool. *New Voices in Translation Studies*, 7, 35-46.
- UN. (1948). Universal Declaration of Human Rights. In. Paris: United Nations.
- UN. (2006). *Convention on the Rights of Persons with Disabilities*. Retrieved from <http://www.un.org/disabilities/documents/convention/convoptprot-e.pdf>
- UN. (2013). *Committee on the rights of persons with disabilities: Concluding observations on the initial report of Australia* (CRPD/C/AUS/CO/1). Retrieved from <http://www.refworld.org/docid/5280b5cb4.html>
- UN. (2015). *Committee on the rights of persons with disabilities: Concluding observations on the initial report of Germany* (CRPD/C/DEU/CO/1). Retrieved from http://www.institut-fuer-menschenrechte.de/fileadmin/user_upload/PDF-Dateien/Pakte_Konventionen/CRPD_behindertenrechtskonvention/crpd_state_report_germany_1_2011_ConObs_2015_en.pdf
- UNDP. (2016). *Human development report 2016: Human development for everyone*. New York, NY: United Nations Development Programme.
- UNESCO. (1990). *World declaration on Education for All*. Paris: UNESCO.
- UNESCO. (1994). *The Salamanca Statement*. Paris: UNESCO.
- UNESCO. (2000). The Dakar Framework for Action. In. Paris: UNESCO.
- UNESCO. (2005). Guidelines for inclusion: Ensuring access to education for all. In. Paris: UNESCO.
- UNESCO. (2014). Basic texts: 2014 edition. In. Paris: United Nations Educational, Scientific and Cultural Organization.
- UNESCO. (2015). Incheon declaration. Education 2030: Towards inclusive and equitable quality education and lifelong learning for all. In. Paris: UNESCO.
- UNESCO, & UNDP. (1995). *The education for all teacher-training package (vol. 1)*. United Kingdom: UNESCO.
- Unianua, E. M. (2012). Teachers' attitudes towards inclusive education. *Procedia – Social and Behavioral Sciences*, 33, 900-904. doi:10.1016/j.sbspro.2012.01.252
- Van de Vijver, F. J. R. (2015a). Cross-cultural research methods in psychology. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 318-321): Elsevier.
- Van de Vijver, F. J. R. (2015b). Cross-cultural research methods in sociology. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 322-326): Elsevier.
- Van de Vijver, F. J. R., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks: Sage.
- Van der Veen, I., Smeets, E., & Derriks, M. (2010). Children with special educational needs in the Netherlands: Number, characteristics and school career. *Educational Research*, 52(1), 15-43. doi:10.1080/00131881003588147

- Van Prooijen, J.-W., & Van der Kloot, W. A. (2001). Confirmatory analysis of exploratively obtained factor structures. *Educational and Psychological Measurement*, 61(5), 777-792. doi:10.1177/00131640121971518
- Van Reusen, A. K., Shoho, A. R., & Barker, K. S. (2001). High school teacher attitudes toward inclusion. *The High School Journal*, 84(2), 7-20.
- Vanderfaellie, J., De Fever, F., & Lombaerts, K. (2003). First-year university students of educational sciences on inclusive education: Attitudes and convictions in Flanders. *European Journal of Teacher Education*, 26(2), 265-277. doi:10.1080/0261976032000088783
- Vignes, C., Coley, N., Grandjean, H., Godeau, E., & Arnaud, C. (2008). Measuring children's attitudes towards peers with disabilities: A review of instruments. *Developmental Medicine & Child Neurology*, 50(3), 182-189. doi:10.1111/j.1469-8749.2008.02032.x
- Villa, R. A., Thousand, J. S., Meyers, H., & Nevin, A. (1996). Teacher and administrator perceptions of heterogeneous education. *Exceptional Children*, 63(1), 29-45. doi:10.1177/001440299606300103
- Voas, D. (2014). Towards a Sociology of Attitudes. *Sociological Research Online*, 19(1). doi:10.5153/sro.3289
- Waitoller, F. R., & Artiles, A. J. (2013). A decade of professional development research for inclusive education: A critical review and notes for a research program. *Review of Educational Research*, 83(3), 319-356. doi:10.3102/0034654313483905
- Wanzare, Z. O. (2007). The transition process: The early years of being a teacher. In T. Townsend & R. Bates (Eds.), *Handbook of teacher education: Globalization, standards and professionalism in times of change* (pp. 343-364). Dordrecht: Springer.
- Weeks, A., Swerissen, H., & Belfrage, J. (2007). Issues, challenges, and solutions in translating study instruments. *Evaluation Review*, 31(2), 153-165. doi:10.1177/0193841X06294184
- Weiber, R., & Mülhhaus, D. (2010). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS*. [Structural equation modelling: A practical introduction to causal analysis using AMOS, SmartPLS, and SPSS]. Berlin: Springer.
- Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27(3), 236-247. doi:10.1016/j.ijresmar.2010.02.004
- Weisel, A., & Dror, O. (2006). School climate, sense of efficacy and Israeli teachers' attitudes toward inclusion of students with special needs. *Education, Citizenship and Social Justice*, 1(2), 157-174. doi:10.1177/1746197906064677
- Werner, C., & Schermelleh-Engel, K. (2010). *Deciding between competing models: Chi-square difference tests*. Retrieved from http://www.psychologie.uzh.ch/fachrichtungen/methoden/team/christinawerner/sem/chisquare_diff_en.pdf
- Werner, O., & Campbell, D. T. (1970). Translating, working through interpreters, and the problem of decentering. In R. Naroll & R. Cohen (Eds.), *A handbook of method in cultural anthropology* (pp. 398-420). New York: Columbia University Press.
- Werning, R., Amrhein, B., Lütje-Klose, B., & Riecke-Baulecke, T. (Eds.). (forthcoming). *Basiswissen Lehrerbildung: Inklusion in Schule und Unterricht – Grundlagen in der Sonderpädagogik*. [Basic knowledge for teacher training: Inclusion in school and classrooms – Essentials in special education]. Münster: Waxmann.
- WHO. (n.y.). Process of translation and adaptation of instruments. Retrieved from www.who.int/substance_abuse/research_tools/translation/en

- Wilczenski, F. L. (1992). Measuring attitudes toward inclusive education. *Psychology in the Schools*, 29(4), 306-312. doi:10.1002/1520-6807(199210)29:4<306::aid-pits2310290403>3.0.co;2-1
- Wilczenski, F. L. (1995). Development of a scale to measure attitudes toward inclusive education. *Educational and Psychological Measurement*, 55(2), 291-299. doi:10.1177/0013164495055002013
- Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health*, 8(2), 94-104. doi:10.1111/j.1524-4733.2005.04054.x
- Willis, G. B. (2005). *Cognitive interviewing*. Thousand Oaks: Sage.
- Witzel, A., & Reiter, H. (2012). *The problem-centred interview*. Los Angeles: Sage.
- Woodcock, S., & Hardy, I. (2017). Probing and problematizing teacher professional development for inclusion. *International Journal of Educational Research*, 83, 43-54. doi:10.1016/j.ijer.2017.02.008
- Yada, A., & Savolainen, H. (2017). Japanese in-service teachers' attitudes toward inclusive education and self-efficacy for inclusive practices. *Teaching and Teacher Education*, 64, 222-229. doi:10.1016/j.tate.2017.02.005

Appendices

Appendix A · Ethics Approval

RE: HS Ethics Application - Approved (5201600738)(Con/Met)

Kay Bowes-Tseng [kay.bowes-tseng@mq.edu.au] im Auftrag von FHS Ethics [fhs.ethics@mq.edu.au]

Gesendet: Freitag, 4. November 2016 04:05

An: Stuart Woodcock [stuart.woodcock@mq.edu.au]

Cc: Kielblock, Stephan

Dear Dr Woodcock,

Re: "Attitudes towards inclusive education for all. The construction and substantiation of a quantitative measuring instrument" (5201600738)

Thank you very much for your response. Your response has addressed the issues raised by the Faculty of Human Sciences Human Research Ethics Sub-Committee and approval has been granted, effective 3rd November 2016. This email constitutes ethical approval only.

This research meets the requirements of the National Statement on Ethical Conduct in Human Research (2007). The National Statement is available at the following web site:

<https://www.nhmrc.gov.au/book/national-statement-ethical-conduct-human-research>

The following personnel are authorised to conduct this research:

Dr Stuart Woodcock
Mr Stephan Kielblock

Appendix B · Approved SERAP



Mr Stephan Kielblock
C3A Macquarie University
Macquarie Park NSW 2109

DOC17/91283
SERAP 2016598

Dear Mr Kielblock

I refer to your application to conduct a research project in NSW government schools entitled *Attitudes towards inclusive education for all*. I am pleased to inform you that your application has been approved.

You may contact principals of the nominated schools to seek their participation. **You should include a copy of this letter with the documents you send to principals.**

This approval will remain valid until 02-Feb-2018.

As this research does not involve face-to-face contact with children, no researchers or research assistants have been screened to interact with or observe children.

I draw your attention to the following requirements for all researchers in NSW government schools:

- The privacy of participants is to be protected as per the NSW Privacy and Personal Information Protection Act 1998.
- School principals have the right to withdraw the school from the study at any time. The approval of the principal for the specific method of gathering information must also be sought.
- The privacy of the school and the students is to be protected.
- The participation of teachers and students must be voluntary and must be at the school's convenience.
- Any proposal to publish the outcomes of the study should be discussed with the research approvals officer before publication proceeds.
- All conditions attached to the approval must be complied with.

When your study is completed please email your report to: serap@det.nsw.edu.au
You may also be asked to present on the findings of your research.

I wish you every success with your research.

Yours sincerely

A handwritten signature in black ink, appearing to read "R. Stevens".

Dr Robert Stevens
Manager, Research
2 February 2017



Appendix C · Approval of Data Collection at the JLU

JUSTUS-LIEBIG-



UNIVERSITÄT
GIESSEN

PRÄSIDENT

Justus-Liebig-Universität Gießen, Postfach 11 14 40, 35359 Gießen

Herrn
Stephan Kielblock
Institut für Erziehungswissenschaft
Rathenaustraße 8
35394 Gießen

Der Datenschutzbeauftragte

Sachbearbeitung: Dr. Robert Pfeffer
Ludwigstraße 23
35390 Gießen
Telefon: 0641 / 99 – 1 22 50
Fax: 0641 / 99 – 1 22 29
E-Mail: Robert.Pfeffer@admin.uni-giessen.de
Az.: B 1.5 – DS
08.11.2016

**Ihr Forschungsvorhaben „Attitudes towards Inclusive Education for All.
The Construction and Substantiation of a Quantitative Measuring Instrument“**

Sehr geehrter Herr Kielblock,

hiermit bestätige ich Ihnen, dass Ihre im Rahmen des o.g. Forschungsvorhabens für das Wintersemester 2016/17 geplante Umfrage unter Lehramtsstudierenden mittels Fragebogen keinen datenschutzrechtlichen Bedenken begegnet.

Mit freundlichem Gruß

In Vertretung

Dr. Pfeffer

Appendix D · Approval of Conduct of the Study at Schools in Hesse, Germany

Hessisches Kultusministerium

HESSEN



Hessisches Kultusministerium Postfach 3160 65021 Wiesbaden
 Justus-Liebig-Universität Gießen
 z.Hd. Herrn Stephan Kielblock
 FB 03 Institut für Erziehungswissenschaften
 Karl-Glöckner-str. 21B

Geschäftszeichen 660.003.000-00767
 Bearbeiter/in Yasar Karakas
 Durchwahl 2734
 Ihr Zeichen 14.11.2016
 Ihre Nachricht
 Datum 13.06.2017

35394 Gießen

GWU 589 - Inklusives Bildungssystem auf allen Ebenen

Sehr geehrter Herr Kielblock,

hiermit genehmige ich die Durchführung der von Ihnen beantragten o.g. Untersuchung an folgenden Schulen:

Dst.-Nr	Name	Straße	PLZ	Ort

Die o.a. Genehmigung Ihrer Untersuchung wird verbunden mit der Beachtung der folgenden, generell geltenden Auflagen:

1. Die an der Untersuchung beteiligten Personen sind vorab über Zielsetzung und Inhalt der Untersuchung zu informieren, sowie über die Art ihrer Beteiligung an der Untersuchung und der Verarbeitung der erhobenen Daten aufzuklären.
2. Jeder vom Forschungsvorhaben Betroffene muss darauf hingewiesen werden, dass eine Teilnahme freiwillig ist und ihm keine Nachteile entstehen, wenn er an der Untersuchung nicht teilnimmt. Von Schulleiterinnen und Schulleitern, Lehrerinnen und Lehrern sowie Schülerinnen und Schülern, die in die Untersuchung einbezogen werden, ist eine schriftliche Einverständniserklärung einzuholen. Bei minderjähri-

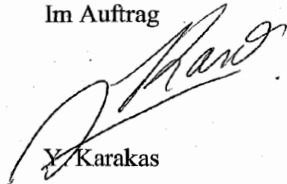
-2-

gen Schülerinnen und Schülern ist das Einverständnis der Erziehungsberechtigten erforderlich.

3. Die Untersuchung ist anonym durchzuführen. Sofern Merkmale festgehalten werden, mit denen ein Personenbezug hergestellt werden kann, sind diese von den übrigen Erhebungsdaten und -materialien schon während der Auswertungen zu trennen und gesondert zu speichern. Sie sind spätestens nach der Auswertung der Erhebungsunterlagen mit diesen zu vernichten. Da auch Video- und Tonaufnahmen personenbezogene Daten sind, müssen hier die entsprechenden Einschränkungen bei der Teilnahme und späteren Auswertung der Mitschnitte beachtet werden.
4. Die Erhebungsunterlagen dürfen nur für den angegebenen Zweck ausgewertet werden. Die Weitergabe an Dritte ist untersagt.
5. Die Durchführung der Untersuchung ist mit der Schulleitung so zu regeln, dass der Unterrichtsbetrieb möglichst wenig gestört wird.
6. Auf die Regelungen des Hessischen Datenschutzgesetzes, insbesondere die §§ 7, 10, 13 und 33, verweise ich.
7. Nach Fertigstellung Ihrer Untersuchungsergebnisse übersenden Sie mir bitte ein Exemplar zwecks Information der hiesigen Fachreferate.

Der geplanten Untersuchung wünsche ich einen guten Verlauf und Ihnen viel Erfolg bei Ihrer wissenschaftlichen Arbeit.

Mit freundlichen Grüßen
Im Auftrag



Y. Karakas

Appendix E · Analysis of the TEIP scale

The eighteen self-efficacy items were tested in a multi-group confirmatory factor analysis. The model was specified in accordance with the structure as proposed by Sharma et al. (2012), which was confirmed in many studies (Forlin et al., 2010; Hecht et al., 2016; Malinen et al., 2012; Montgomery & Mirenda, 2014; Sharma et al., 2015; Sharma & Sokal, 2016).

Table E.1. Analysis of the self-efficacy scale. Full and reduced model compared

Self-efficacy in...	Full model (3*6 items)		Reduced model (3*3 items)	
	Australia	Germany	Australia	Germany
1) Managing behaviour (MAB)				
1 Make expectations clear	.66	.49	-	-
3 Calm disruptive student	.66	.79	-	-
5 Prevent disruptive behaviour	.74	.80	.70	.76
10 Get children follow rules	.85	.75	.89	.80
13 Control disruptive behaviour	.84	.83	.84	.83
16 Deal with physic. aggressive	.58	.62	-	-
2) Collaboration (COL)				
4 Make parents coming to school	.71	.62	-	-
8 Assist families to help child	.73	.66	.74	.54
9 Inform others about policies	.55	.59	-	-
14 Get parents involved	.61	.43	-	-
15 Collaborate with others	.72	.63	.74	.78
18 Work jointly with others	.74	.65	.76	.79
3) Using inclusive instruction (UII)				
2 Gauge student comprehension	.67	.54	-	-
6 Design learning tasks for all	.64	.55	-	-
7 Give alternative explanation	.75	.61	.80	.71
11 Vary assessment strategy	.75	.57	.77	.62
12 Get students work together	.72	.51	.73	.52
17 Challenge capable students	.62	.57	-	-
Covariances				
1 (MAB) vs. 2 (COL)	.85	.60	.79	.38
1 (MAB) vs. 3 (UII)	.89	.75	.86	.67
2 (COL) vs. 3 (UII)	.94	.85	.92	.72
Chi-square	800.88		106.45	
df	264		48	
n	382		378	
CFI	.82		.96	
TLI	.80		.94	
RMSEA	.10		.08	
90% CI	.10, .11		.06, .10	
SRMR	.08		.05	

Note: Multiple-group confirmatory factor analyses.

Full model: Two cases had missings on all variables (...they had stopped completing the survey) That is why the number of items is reduced. (GER: n=236 [total would have been 238]; AUS: n=146).

Reduced model: Six cases had missings on all variables. That is why the number of items is reduced (GER: n=236 [total would have been 238]; AUS: n=142 [total would have been 146]).

Appendix F · Univariate Analysis of the Attitude Items

Appendix F comprises different tables as they pertain to the univariate analysis of the attitude items. Table F.1 includes information about missing values. The distribution of each item is examined further using measures for the central tendency, skewness, and kurtosis (see Table F.2 for the Australian sample and Table F.3 for the German sample). Both tables are systematically compared in Table F.4 and informed decisions are indicated.

Table F.1. Missing values for all attitude items

	Australia		Germany	
	n	%	n	%
1 Inclusion facilitates socially appropriate behaviour	3	2.1%	0	0.0%
2 Teachers are able to meet the needs of all children	0	0.0%	0	0.0%
3 (rec.) I get overwhelmed when I have to differentiate	3	2.1%	2	0.8%
4 Education is a right that should be available to all	1	0.7%	0	0.0%
5 There is personnel from outside school to support me	3	2.1%	0	0.0%
6 All will receive appropriate education and services	1	0.7%	3	1.3%
7 There is support from the Education Department/Board	1	0.7%	2	0.8%
8 It is possible to organise classes suitable for all	2	1.4%	2	0.8%
9 There are personnel within school to support me	3	2.1%	2	0.8%
10 Inclusion will foster acceptance of differences	1	0.7%	0	0.0%
11 (rec.) Labelling is necessary for quality education	0	0.0%	4	1.7%
12 (rec.) Differentiated practices cannot be achieved	6	4.1%	4	1.7%
13 All should be educated in the inclusive classroom	0	0.0%	2	0.8%
14 All are capable of learning in inclusive settings	0	0.0%	1	0.4%
15 (rec.) Inclusion represents a negative change	0	0.0%	3	1.3%
16 With the right supports in place inclusion can work	0	0.0%	0	0.0%
17 Inclusive education is a practical idea in my country	0	0.0%	0	0.0%
18 Separating students is not necessary	0	0.0%	2	0.8%
19 Any student can learn if the curriculum is adapted	0	0.0%	1	0.4%
20 (rec.) External support services are a waste of time	1	0.7%	2	0.8%
21 (rec.) Inclusion cannot be implemented	0	0.0%	1	0.4%
22 The Education Department/Board supports efforts	1	0.7%	3	1.3%
23 (rec.) Frustrated when I have to adapt the curriculum	3	2.1%	2	0.8%
24 Willing to adapt the curriculum of all students	2	1.4%	1	0.4%
25 Differentiated adjustments can be carried out	0	0.0%	1	0.4%
26 (rec.) _recoded I do not need support for inclusive practice	3	2.1%	1	0.4%
27 (rec.) _recoded Too difficult to accommodate all differences	2	1.4%	4	1.7%
28 I am willing to adapt the assessment	0	0.0%	1	0.4%
29 I feel that the community is supportive of inclusion	3	2.1%	1	0.4%
30 Inclusion is the best way to meet the needs of all	1	0.7%	4	1.7%
31 Good teachers can differentiate their practices	0	0.0%	4	1.7%
32 Diversity enriches the learning environment	0	0.0%	5	2.1%
33 There are adequate resources to support me	3	2.1%	2	0.8%
34 (rec.) _recoded Parents hinder successful inclusive education	0	0.0%	2	0.8%
35 Inclusion will foster understanding of differences	0	0.0%	2	0.8%
36 Inclusion is a valuable experience for all children	0	0.0%	1	0.4%
37 Working collaboratively with parents is important	0	0.0%	2	0.8%
38 Inclusive education leads to social inclusion	0	0.0%	3	1.3%

Note: The absolute and relative number of missing values for all Australian cases (n=146) and for all German cases (n=238) are presented in this table.

Table F.2. Univariate analysis of the attitude items for the Australian sample

No	Initial examination			Treat outlier	Second examination		
	Mean	Skew.	Kurt.		Mean	Skew.	Kurt.
2	-	-7.79	7.13	2	-	-6.41	4.46
4	6.94	-31.39	112.41	3	6.94	-24.80	68.39
8	-	-4.87	-	-	-	-4.87	-
10	6.23	-9.13	11.12	3	6.25	-5.40	-
13	-	-3.30	-	-	-	-3.30	-
14	-	-3.46	-	-	-	-3.46	-
15	-	-4.87	-	1	-	-4.14	-
16	6.05	-5.36	-	-	6.03	-5.36	-
17	-	-6.27	4.41	-	-	-6.27	4.41
19	-	-3.58	-	-	-	-3.58	-
20	6.27	-9.25	11.79	2	6.30	-5.70	-
24	-	-3.59	-	-	-	-3.59	-
25	-	-	-	1	-	-	-
26	-	-5.08	-	-	-	-5.08	-
28	-	-5.28	4.16	1	-	-4.09	-
31	-	-7.32	6.94	2	-	-6.56	5.54
32	6.27	-6.40	3.40	2	6.28	-5.93	-
35	6.04	-4.90	3.39	1	-	-3.67	-
36	-	-4.73	-	2	-	-3.41	-
37	6.38	-5.98	-	1	6.34	-5.28	-
38	-	-5.43	-	-	-	-5.43	-

Note: Means are included if unusually low (below 2.0) or high (above 6.0). Skewness, kurtosis and outliers are included if significant ($p < .001$). Skewness and kurtosis values are z-values. Outliers are treated by setting the individual outlying value missing.

Table F.3. Univariate analysis of the attitude items for the German sample

No	Initial examination			Treat outlier	Second examination		
	Mean	Skew.	Kurt.		Mean	Skew.	Kurt.
1	-	-7.35	4.96	-	-	-7.35	4.96
4	6.87	-27.80	67.14	6	6.92	-32.85	113.41
8	-	-3.96	-	-	-	-3.96	-
10	-	-8.77	7.49	4	-	-6.38	-
11	-	-3.80	-	-	-	-3.80	-
16	-	-8.89	8.04	2	-	-8.29	6.95
17	-	-4.94	-	-	-	-4.94	-
19	-	-4.22	-	-	-	-4.22	-
20	-	-6.92	-	2	-	-6.92	-
24	-	-4.23	-	-	-	-4.23	-
25	-	-5.84	-	-	-	-5.84	-
26	-	-9.95	9.26	4	6.06	-6.79	-
28	-	-4.58	-	-	-	-4.58	-
32	-	-6.85	3.57	-	-	-6.85	3.57
35	-	-6.75	4.33	5	-	-4.25	-
36	-	-6.04	-	-	-	-6.04	-
37	-	-4.65	-	1	-	-3.63	-
38	-	-5.16	-	-	-	-5.16	-

Note: Means are included if unusually low (below 2.0) or high (above 6.0). Skewness, kurtosis and outliers are included if significant ($p < .001$). Skewness and kurtosis values are z-values. Outliers are treated by setting the individual outlying value missing.

Table F.4. Comparison of the items' univariate distributions for both samples

No	Australian sample			German sample			Decision
	Mean	Skew.	Kurt.	Mean	Skew.	Kurt.	
1	-	-	-	-	-7.35	4.96	Retain
2	-	-6.41	4.46	-	-	-	Retain
4	6.94	-24.8	68.39	6.92	-32.85	113.41	Discard from further analysis
8	-	-4.87	-	-	-3.96	-	Retain
10	6.25	-5.4	-	-	-6.38	-	Discard from further analysis
11	-	-	-	-	-3.8	-	Retain
13	-	-3.3	-	-	-	-	Retain
14	-	-3.46	-	-	-	-	Retain
15	-	-4.14	-	-	-	-	Retain
16	6.03	-5.36	-	-	-8.29	6.95	Discard from further analysis
17	-	-6.27	4.41	-	-4.94	-	Discard from further analysis
19	-	-3.58	-	-	-4.22	-	Retain
20	6.3	-5.7	-	-	-6.92	-	Discard from further analysis
24	-	-3.59	-	-	-4.23	-	Retain
25	-	-	-	-	-5.84	-	Retain
26	-	-5.08	-	6.06	-6.79	-	Discard from further analysis
28	-	-4.09	-	-	-4.58	-	Retain
31	-	-6.56	5.54	-	-	-	Retain
32	6.28	-5.93	-	-	-6.85	3.57	Discard from further analysis
35	-	-3.67	-	-	-4.25	-	Retain
36	-	-3.41	-	-	-6.04	-	Retain
37	6.34	-5.28	-	-	-3.63	-	Discard from further analysis
38	-	-5.43	-	-	-5.16	-	Retain

Note: Ferguson and Cox (1993) reported that it acceptable to retain some items with skew and/or kurtosis. Hence, decisions were made in a way to remain as many items as possible, and only to discard those items with highly unusual distribution in both samples. To inform this decision, the three indicators mean, skewness, kurtosis were used. Those items that were problematic with regards to at least two indicators in one sample and at the same time at least one indicator in the other sample, were discarded from further analysis. All other items were retained.

Appendix G · Correlative Analysis of the Attitude Items

Appendix G includes information about the association amongst the variables. Table G.1 lists the items that were not discarded due to non-sufficient univariate properties. For each item, the absolute and relative number of substantial correlations ($r > .3$; see Field, 2013) is presented. Items without any or with only one substantial correlation were deleted step-wise. In addition to this approach to test if enough correlations are present, another prerequisite for conducting factor analyses is that not too many and too high correlations are present. Accordingly, Table G.2 presents the results of the multicollinearity analysis.

Table G.1. Deletion of items without sufficient substantive correlations

No	Australian sample		German sample		
	Step 1	Step 2	Step 1	Step 2	Step 3
1	8 (26.7%)	8 (27.6%)	14 (46.7%)	14 (58.3%)	14 (63.6%)
2	10 (33.3%)	10 (34.5%)	6 (20.0%)	6 (25.0%)	6 (27.3%)
3	3 (10.0%)	2 (6.9%)	0 (0.0%)	[deleted]	[deleted]
5	5 (16.7%)	5 (17.2%)	1 (3.3%)	[deleted]	[deleted]
6	18 (60.0%)	18 (62.1%)	13 (43.3%)	13 (54.2%)	12 (54.5%)
7	9 (30.0%)	9 (31.0%)	1 (3.3%)	[deleted]	[deleted]
8	20 (66.7%)	20 (69.0%)	18 (60.0%)	18 (75.0%)	18 (81.8%)
9	7 (23.3%)	7 (24.1%)	4 (13.3%)	3 (12.5%)	3 (13.6%)
11	1 (3.3%)	[deleted]	2 (6.7%)	1 (4.2%)	[deleted]
12	4 (13.3%)	4 (13.8%)	12 (40.0%)	13 (54.2%)	13 (59.1%)
13	13 (43.3%)	13 (44.8%)	17 (56.7%)	18 (75.0%)	18 (81.8%)
14	16 (53.3%)	16 (55.2%)	16 (53.3%)	17 (70.8%)	17 (77.3%)
15	17 (56.7%)	17 (58.6%)	20 (66.7%)	21 (87.5%)	20 (90.9%)
18	8 (26.7%)	8 (27.6%)	6 (20.0%)	6 (25.0%)	6 (27.3%)
19	12 (40.0%)	12 (41.4%)	16 (53.3%)	16 (66.7%)	15 (68.2%)
21	13 (43.3%)	13 (44.8%)	18 (60.0%)	17 (70.8%)	17 (77.3%)
22	13 (43.3%)	13 (44.8%)	2 (6.7%)	1 (4.2%)	[deleted]
23	7 (23.3%)	7 (24.1%)	1 (3.3%)	[deleted]	[deleted]
24	11 (36.7%)	11 (37.9%)	16 (53.3%)	15 (62.5%)	15 (68.2%)
25	17 (56.7%)	17 (58.6%)	16 (53.3%)	15 (62.5%)	15 (68.2%)
27	9 (30.0%)	9 (31.0%)	14 (46.7%)	14 (58.3%)	14 (63.6%)
28	6 (20.0%)	6 (20.7%)	9 (30.0%)	8 (33.3%)	8 (36.4%)
29	3 (10.0%)	3 (10.3%)	0 (0.0%)	[deleted]	[deleted]
30	23 (76.7%)	23 (79.3%)	19 (63.3%)	19 (79.2%)	19 (86.4%)
31	15 (50.0%)	15 (51.7%)	18 (60.0%)	18 (75.0%)	18 (81.8%)
33	8 (26.7%)	8 (27.6%)	7 (23.3%)	10 (41.7%)	9 (40.9%)
34	2 (6.7%)	2 (6.9%)	0 (0.0%)	[deleted]	[deleted]
35	14 (46.7%)	14 (48.3%)	13 (43.3%)	13 (54.2%)	13 (59.1%)
36	18 (60.0%)	18 (62.1%)	19 (63.3%)	20 (83.3%)	20 (90.9%)
38	14 (46.7%)	14 (48.3%)	16 (53.3%)	16 (66.7%)	16 (72.7%)
n	120	120	210	212	215

Note: For each item the absolute and relative number of substantial correlations ($r > .3$; see Field, 2013) with other items is given. If an item has none or only one substantial correlation to other items, it is deleted and the correlations were calculated again. This was repeated until all items had a minimum of three substantial correlations. In anticipation that SPSS was used for calculating the exploratory factor analysis, listwise deletion was used. The numbers of cases for each step are indicated in the last row of the table.

Table G.2. Step-wise deletion of items with highest VIF to mitigate multicollinearity

	Initial	Step 1	Step 2	Step 3
Australia				
Determinant	0.000000232	0.0000008865	0.000003483	0.00001069
Interpretation	multicollinearity	multicollinearity	multicollinearity	sufficient data
Deleted item		Item 13	Item 36	Item 30
Variance inflation factor		3.821	3.388	3.069
Germany				
Determinant	0.00001149			
Interpretation	sufficient data			

Note: The table reports the determinant as a measure of multicollinearity. The absence of multicollinearity is considered to be an important prerequisite for factor analysis. An indicator for non-multicollinearity is the determinant being greater 0.00001 (Field, 2013). The Variance Inflation Factor (VIF) was calculated for each variable for the Australian sample, using SPSS. Within each step, the item with the highest VIF value was discarded, and with this new set of items, the determinant and also the VIFs were calculated again. This was carried out iteratively. (Notably, even the highest VIF values, as reported in the table, are relatively low.)

Appendix H · Structure Matrices

The pattern matrices of the exploratory factor analyses, as they were presented in Section 4.5.2, contain factor loadings, which do not take into account that the oblique rotation allowed the factors to be correlated with each other. The *structure matrices* that are depicted in Table H.1 (for the Australian sample) and Table H.2 (for the German sample) include similar loadings but with taking into account the correlations amongst the factors. As Field (2013) recommends, to examine the structure matrix in addition to the pattern matrix is a ‘useful double-check’.

Table H.1 Structure matrix for the Australian sample

	Factor						
	1	2	3	4	5	6	7
38 · Inclusive education leads	.75	-.28	.07	.40	-.19	.28	-.38
35 · Inclusion will foster	.67	-.14	.10	.42	-.32	.28	-.42
1 · Inclusion facilitates socially	.59	-.32	.09	.38	-.47	.30	-.31
15 (rec.) · Inclusion represents	.54	-.25	.14	.27	-.17	.17	-.30
7 · There is support from the	.26	-.74	.09	.29	-.18	.11	-.23
9 · There are personnel within	.16	-.72	.10	.14	-.08	.19	-.33
5 · There is personnel from	-.03	-.69	.11	.08	-.06	.24	-.32
33 · There are adequate	.21	-.66	.02	.37	-.21	.12	-.33
22 · The Education Department	.21	-.66	.10	.09	-.06	.12	-.17
6 · All will receive appropriate	.39	-.50	-.06	.37	-.36	.36	-.35
23 (rec.) · Frustrated when I	.28	-.04	.76	.29	-.25	.25	-.11
3 (rec.) · I get overwhelmed	-.01	-.21	.59	.07	-.23	.14	-.05
28 · I am willing to adapt the	.30	-.03	.01	.75	-.24	.22	-.16
24 · Willing to adapt the	.27	-.29	.19	.71	-.15	.29	-.37
25 · Differentiated adjustments	.45	-.27	.34	.65	-.17	.32	-.36
21 (rec.) · Inclusion cannot be	.38	-.12	.22	.29	-.74	.26	-.28
12 (rec.) · Differentiated	.15	-.08	.30	.20	-.66	.24	-.11
27 (rec.) · Too difficult to	.24	-.15	.54	.25	-.61	.32	-.29
34 (rec.) · Parents hinder	.16	-.05	.11	.16	-.26	.71	-.10
29 · I feel that the community is	.15	-.26	.18	.30	-.01	.53	-.26
19 · Any student can learn if the	.49	-.26	.05	.24	-.08	.24	-.79
8 · It is possible to organise	.25	-.37	.10	.40	-.43	.15	-.66
14 · All are capable of learning	.52	-.32	-.18	.38	-.26	.27	-.61
2 · Teachers are able to meet	.28	-.38	.12	.41	-.30	.30	-.51
18 · Separating students is not	.13	-.36	-.08	.32	-.40	.20	-.48
31 · Good teachers can	.33	-.08	.13	.33	-.14	.37	-.44

Note: Structure matrix (which takes into account the relationship between factors; see Field, 2013) of the exploratory factor analysis of the Australian sample. This table provides additional information to the pattern matrix, which is depicted and interpreted in Section 4.5.2. Highest loadings appear in bold.

Table H.2. Structure matrix for the German sample

	Factor		
	1	2	3
a30 Inclusion is the best way to meet the needs of all (Item 30)	.79	.36	.57
a21_recoded Inclusion cannot be implemented (Item 21 [rec.])	.75	.46	.54
a15_recoded Inclusion represents a negative change (Item 15 [rec.])	.74	.41	.65
a13 All should be educated in the inclusive classroom (Item 13)	.69	.22	.63
a31 Good teachers can differentiate their practices (Item 31)	.64	.42	.44
a19 Any student can learn if the curriculum is adapted (Item 19)	.63	.42	.57
a8 It is possible to organise classes suitable for all (Item 8)	.62	.51	.33
a6 All will receive appropriate education and services (Item 6)	.60	.16	.43
a27_recoded Too difficult to accommodate all differences (Item 27 [rec.])	.57	.46	.33
a24 Willing to adapt the curriculum of all students (Item 24)	.50	.72	.40
a25 Differentiated adjustments can be carried out (Item 25)	.55	.67	.52
a28 I am willing to adapt the assessment (Item 28)	.40	.42	.40
a35 Inclusion will foster understanding of differences (Item 35)	.48	.31	.74
a38 Inclusive education leads to social inclusion (Item 38)	.58	.33	.72
a1 Inclusion facilitates socially appropriate behaviour (Item 1)	.50	.24	.67
a18 Separating students is not necessary (Item 18)	.40	.19	.42

Note: Structure matrix (which takes into account the relationship between factors; see Field, 2013) of the exploratory factor analysis of the German sample. This table provides additional information to the pattern matrix, which is depicted and interpreted in Section 4.5.2. Highest loadings appear in bold.

*Appendix I · Additional Information as they pertain to the Confirmatory Factor Analysis**Table I.1. Item properties of the 4-factor model*

	Australian sample			German sample		
	Intercept	Variance	R-square	Intercept	Variance	R-square
<i>Vision of outcomes of inclusive education for all</i>						
Item 1	6.24	0.75	0.25	4.01	0.45	0.55
Item 35	6.34	0.32	0.68	4.46	0.42	0.58
Item 38	4.66	0.43	0.57	3.31	0.40	0.60
<i>Differentiation as it pertains to inclusive education for all</i>						
Item 24	6.07	0.37	0.63	3.84	0.54	0.46
Item 25	5.76	0.46	0.54	3.75	0.34	0.66
Item 28	5.14	0.59	0.41	3.42	0.68	0.32
<i>General practices of inclusive education for all</i>						
Item 8	3.38	0.48	0.52	2.84	0.58	0.42
Item 19	3.79	0.58	0.42	3.33	0.50	0.50
Item 31	5.00	0.57	0.43	2.73	0.62	0.38
<i>Supports as they pertain to inclusive education for all</i>						
Item 5	2.75	0.47	0.53	2.84	0.70	0.30
Item 9	3.00	0.43	0.57	2.77	0.53	0.47
Item 33	2.48	0.42	0.58	2.82	0.66	0.34

Note: Item properties as they were obtained through a multiple-group confirmatory factor analysis with four factors (see Section 4.5.3). $\chi^2(96, n=384)=181.493$, CFI=0.94, TLI=0.92, RMSEA=0.07, SRMR=0.06. Full information maximum likelihood (FIML) estimation. Groups are Australian teachers (n=146) vs. German teachers (n=238).

Attitudes towards Inclusive Education for All

Questionnaire in English Language (IEfA-EN)

[illegible]

Attitudes towards Inclusive Education for All

Questionnaire in English Language (IEfA-EN)

Section Two – Efficacy to Implement Inclusive Education for All
The following statements address your efficacy to implement inclusive education for all. Please indicate below how strongly you disagree or agree with the different statements.

-3	=	very strongly disagree
-2	=	strongly disagree
-1	=	disagree
0	=	neither disagree nor agree
+1	=	agree
+2	=	strongly agree
+3	=	very strongly agree

[illegible]

Attitudes towards Inclusive Education for All

Questionnaire in English Language (IEfA-EN)

Section Three – Personal Information and Background

Finally, these questions are about you personally and your background. Please remember that this questionnaire is completely anonymous. No individuals or their results will be identified.

1 I am:

☐ female☐ male

2 My age:

☐ up to 30 years☐ 31-40 years☐ 41-50 years☐ 51-60 years☐ above 60 years

3 Training to teach/currently teaching:

☐ primary/elementary☐ secondary/high school☐ other *If other give more detail:*

_____ 

4 If you are currently teaching, how many years experience have you had?

_____ years

☐ N/A

5 Do you hold a postgraduate degree/diploma?

☐ no☐ yes *If yes give more detail:*

_____ 

6 My knowledge of the local legislation and/or policy as it pertains to inclusive education for all:

☐ very good☐ good☐ average☐ poor☐ none

7 I have had the following amount of training focusing on inclusive education of all students:

☐ high (at least 40hrs)☐ some☐ none

8 Do you have experience with inclusive classroom settings (teaching experience, professional experience, teaching assistance, voluntary assistance etc.)?

☐ no☐ yes *If yes...*

Quality of past experience with an inclusive setting:

☐ positive☐ neutral☐ negative

Thank you very much for your support!

If you have any questions, please contact the researcher.

[illegible]

Einstellungen zu inklusiver Bildung für alle

Fragebogen in deutscher Sprache (IEfA-de)

lehne sehr
stark ab

stimme sehr
stark zu

-3 -2 -1 0 +1 +2 +3

[illegible]

Einstellungen zu inklusiver Bildung für alle

Fragebogen in deutscher Sprache (IEfA-de)

Abschnitt zwei – Wirksamkeit, inklusive Bildung für alle umzusetzen
Die folgenden Aussagen betreffen Ihre Selbstwirksamkeit, inklusive Bildung für alle umsetzen zu können. Bitte kreuzen Sie unten an, in welchem Maße Sie den verschiedenen Aussagen zustimmen oder wie stark Sie diese ablehnen.

	+3 = stimme sehr stark zu
--	---------------------------

-3	=	lehne sehr stark ab
-2	=	lehne stark ab
-1	=	lehne ab
0	=	weder noch
+1	=	stimme zu
+2	=	stimme stark zu
+3	=	stimme sehr stark zu

lehne sehr
stark **ab** **stimme** sehr
stark **zu**

-3 -2 -1 0 +1 +2 +3

[illegible]

Einstellungen zu inklusiver Bildung für alle

Fragebogen in deutscher Sprache (IEfA-de)

Abschnitt drei – Persönliche Hintergrundinformationen

Abschließend stehen hier noch Fragen zu Ihnen persönlich und zu Ihrem Hintergrund. Bitte seien Sie nochmals versichert, dass diese Befragung vollständig anonym ist und keine Identifizierung von Personen vorgenommen wird.

1	Ich bin:	<input type="checkbox"/> weiblich <input type="checkbox"/> männlich
2	Mein Alter:	<input type="checkbox"/> bis 30 Jahre <input type="checkbox"/> 31-40 Jahre <input type="checkbox"/> 41-50 Jahre <input type="checkbox"/> 51-60 Jahre <input type="checkbox"/> über 60 Jahre
3	Ich bin in der Lehramtsausbildung für bzw. ich unterrichte zurzeit:	<input type="checkbox"/> Grundschule/Primarbereich <input type="checkbox"/> Sekundarbereich/Gymnasium <input type="checkbox"/> Andere <i>Falls andere, bitte erläutern:</i> <div style="border-bottom: 1px solid black; width: 150px; margin-top: 5px;"></div>
4	Falls Sie zurzeit unterrichten, wie viele Jahre Unterrichtserfahrung haben Sie?	<div style="display: flex; align-items: center;"> <div style="border-bottom: 1px solid black; width: 80px; margin-right: 10px;"></div> <div>Jahr/e</div> <div style="margin-left: 20px;"> <input type="checkbox"/> nicht zutreffend </div> </div>
5	Verfügen Sie über einen weiterführenden Studienabschluss?	<input type="checkbox"/> nein <input type="checkbox"/> ja <i>Falls ja, bitte erläutern:</i> <div style="border-bottom: 1px solid black; width: 150px; margin-top: 5px;"></div>
6	Meine Kenntnis der regionalen Gesetzgebung und/oder Politik in Bezug auf „inklusive Bildung für alle“ ist:	<input type="checkbox"/> sehr gut <input type="checkbox"/> gut <input type="checkbox"/> mittel <input type="checkbox"/> wenig <input type="checkbox"/> nicht vorhanden
7	Ich habe folgenden Umfang an Aus-/Fortbildung in Bezug auf „inklusive Bildung für alle Schüler/innen“:	<input type="checkbox"/> hoch (mindestens 40 Stunden) <input type="checkbox"/> wenig <input type="checkbox"/> nicht vorhanden
8	Haben Sie Erfahrung mit inklusiven Unterrichtssettings (Unterrichtserfahrung, Berufserfahrung, Unterrichtspraktikum, Volontärarbeit usw.)?	<input type="checkbox"/> nein <input type="checkbox"/> ja <i>Falls ja, ...</i> <div style="margin-top: 10px;"> <i>Qualität der Erfahrung mit einem inklusiven Setting:</i> <input type="checkbox"/> positiv <input type="checkbox"/> neutral <input type="checkbox"/> negativ </div>

Vielen Dank für Ihre Unterstützung!

Wenn Sie noch Fragen haben, setzen Sie sich bitte mit dem Wissenschaftler in Verbindung.

Appendix L · Original Items and Revised Items

The following table compares the 48 statements as they were extracted from the literature with the 38 statements as they were used for the pre-testing in English language.

Original items (from selected studies)	Revised items (after revision of wording)
Ahmad (2012)	
Inclusive education is [a] practicable idea in Pakistan.	[17] I feel that inclusive education is a practical idea in my country.
Inclusive education supports the belief of equality of mankind.	/
Inclusive education affects teacher student interaction.	/
Inclusive education ultimately leads to social inclusion.	[38] Inclusive education ultimately leads to social inclusion.
Inclusive education is a cost effective system of education.	/
Adaptations in methods and techniques of assessment and evaluation are required for inclusive classroom[s].	/
Al Zyoudi et al. (2011)	
I believe all children are capable to learn in inclusive setting[s].	[14] All children are capable to learn in inclusive settings.
I am aware that the[re are] individual capabilities of students.	/
I expect the best from all students in the classroom and I am aware of their capabilities.	/
I think it is impossible to try and accommodate too many differences in one classroom.	[27] It is too difficult to accommodate all students' differences in an inclusive classroom.
Andrews & Clementson (1997)	
I am familiar with inclusion.	/
The primary motivation behind inclusion is to save money.	/
Separating and labelling students is not necessary to provide a quality education to them.	[11] Labelling students (e.g. gender, race, ethnicity, disability, language, socio-economic status) is necessary to provide a quality education to them. [18] Separating students is not necessary to provide a quality education to them.
Good teachers can teach all students.	[31] Good teachers can differentiate their practices so that they can teach all students in their class/es.
Only minor adjustments will be needed to teach all students in the regular classroom.	[25] I feel all differentiated adjustments in an inclusive classroom can be done.
Barnett & Monda-Amaya (1998)	
All children should be educated in the general education classroom.	[13] All children should be educated in the inclusive classroom.
Our school currently is working toward becoming a more inclusive school.	/
I feel that inclusion can work in my school.	[16] I believe that with the right supports in place inclusion can work.
I feel that the school community is supportive of the implementation of inclusion in our school.	[29] From my experience, I feel that the community is supportive of the implementation of inclusion.
Beacham & Rouse (2012)	
Schools can help to build an inclusive society.	/
Teachers should be responsible for the learning of all children in the classes they teach.	[2] Effective Teachers are able to meet the needs of all children in the classes they teach.
Education is a right that should be available to all children.	[4] Education is a right that should be available to all children.
Bennett et al. (1997)	
Inclusion represents a positive change in our education system.	[15] Inclusion represents a negative change in our education system.
Bosse & Spörer (2014)	
Classes can be organised in a way that is suitable for all children. (orig. in German)	[8] It is possible to organise classes in a way that is suitable for all children.

(continued)

Original items (from selected studies)	Revised items (after revision of wording)
Horne et al. (2008) I believe that inclusion is the best way to meet the needs of all students. Parents of students in my school are willing to accept the philosophy of including all students. My school board supports efforts at including all students into the classroom. Diversity within the classroom enriches the learning environment. There is support for inclusion from the Department of Education. As a result of inclusion, parents will be more satisfied with their child's education.	[30] Inclusion is the best way to meet the needs of all students. [37] Working collaboratively with parents play a major part in the success of inclusion. [22] I feel from my experience that the Education Department/Board supports efforts at including all students into the classroom. [32] Diversity within the classroom enriches the learning environment. [7] I feel from my experience that there is support for inclusion from the Education Department/Board. /
Hsieh & Hsieh (2012) I feel it is a valuable experience for all children to be educated in inclusive classrooms. I feel the strengths of implementing inclusion at preschool outweigh the weaknesses. I feel inclusion is a good idea.	[36] It is a valuable experience for all children to be educated in inclusive classrooms. / /
Mahat (2008) I believe that an inclusive school is one that permits academic progression of all students regardless of their ability. I believe that inclusion facilitates socially appropriate behaviour amongst all students. I believe that any student can learn in the regular curriculum of the school if the curriculum is adapted to meet their individual needs. I get frustrated when I have to adapt the curriculum to meet the individual needs of all students. I am willing to adapt the curriculum to meet the individual needs of all students regardless of their ability. I am willing to adapt the assessment of individual students in order for inclusive education to take place.	/ [1] Inclusion facilitates socially appropriate behaviour amongst all students. [19] I believe that any student can learn in an inclusive school if the curriculum is adapted to meet their individual needs. [23] I get frustrated when I have to adapt the curriculum to meet the individual needs of all students. [24] I am willing to adapt the curriculum to meet the individual needs of all students within inclusive classrooms. [28] I am willing to adapt the assessment of individual students in order for inclusive education to take place.
Moberg (1997) All students will receive appropriate education and related services in regular education. Regular education has the resources and personnel to address the unique educational needs of all students.	[6] All students will receive appropriate education and related services in inclusive education. [5] I feel there are adequate personnel from outside school to support me to address the unique educational needs of all students. [9] I feel there are adequate personnel within school to support me to address the unique educational needs of all students. [33] I feel there are adequate resources to support me to address the unique educational needs of all students.
Stoiber et al. (1998) We must learn more about the effects of inclusive classrooms before inclusive classrooms take place on a large scale basis. The best way to begin educating children in inclusive settings is just to do it.	/ /
Taylor & Ringlaben (2012) Inclusion offers mixed group interaction which will foster understanding and acceptance of differences.	[10] Inclusion will foster acceptance of differences among students. [35] Inclusion will foster understanding of differences among students.
Vanderfaellie et al. (2003) Children in inclusive classes integrate better into society. I prefer an inclusive school for my own child.	/ /

(continued)

Original items (from selected studies)	Revised items (after revision of wording)
Newly developed	<p>[3] I get overwhelmed when I have to differentiate to cater for all of the students' needs in my classroom.</p> <p>[12] The differentiated practices that inclusive education would require cannot be achieved.</p> <p>[20] I feel that external support services are a waste of time.</p> <p>[21] The philosophy of inclusion cannot be implemented in 'real world' practices.</p> <p>[26] I do not need any support to put inclusive education into practice.</p> <p>[34] Parents hinder the successful implementation of inclusive education.</p>

Note: Numbers given in the right column indicate the number of the item in the final questionnaire. Statements in the right column are the revised items as they were used for pre-testing (they are not the final items).

Appendix M · All Versions Compared

The following table compares all different versions of the items. In the two left columns, the final English items (as they were used for data collection) and the professional English back-translation are depicted. The middle column and the two columns on the right comprise the professional translation to German, the more literal translation to German and the final German items (as they were used for data collection), respectively.

English		German		
Final items (EN)	Back-translation	Prof. translation	Literal translation	Final items (DE)
Section One – Attitudes Towards Inclusive Education for All				
[1] Inclusion facilitates socially appropriate behaviour for all students.	Inclusive education encourages the socially appropriate behaviour of all pupils.	Inklusion fördert das sozial angemessene Verhalten aller Schüler.	Inklusion fördert ein angemessenes Sozialverhalten bei allen Schüler/innen.	[1] Inklusion fördert das sozial angemessene Verhalten aller Schüler/innen.
[2] Effective teachers are able to meet the needs of all children in the classes they teach.	Good teachers can meet the needs of all of the pupils in their class.	Gute Lehrer können die Bedürfnisse aller Schüler ihrer Klasse erfüllen.	Effektive Lehrer/innen können auf die Bedürfnisse aller Schüler/innen in ihrer Klasse eingehen.	[2] Gute Lehrer/innen können die Bedürfnisse aller Schüler/innen ihrer Klasse erfüllen.
[3] I get overwhelmed when I have to differentiate to cater for all of the students' needs in my classroom.	I find it difficult to cope if I have to differentiate my teaching style in order to address the needs of all of my pupils.	Ich fühle mich überfordert, wenn ich im Unterricht differenzieren muss, um den Bedürfnissen aller meiner Schüler gerecht zu werden.	Es überfordert mich, differenziert zu unterrichten, um auf alle Bedürfnisse der Schüler/innen in meiner Klasse einzugehen.	[3] Es überfordert mich, wenn ich im Unterricht differenzieren muss, um den Bedürfnissen aller meiner Schüler/innen gerecht zu werden.
[4] Education is a right that should be available to all children.	Education is a right that should be available to all children.	Bildung ist ein Recht, das allen Kindern offenstehen muss.	Bildung ist ein Recht, das allen Kindern gleichermaßen zustehen sollte.	[4] Bildung ist ein Recht, das allen Kindern offenstehen muss.
[5] I feel there are adequate personnel from outside school to support me to address the unique educational needs of all students.	There are suitable external staff who help me to accommodate the special educational needs of all pupils.	Es gibt geeignetes externes Personal, das mir hilft, auf die speziellen Bildungsbedürfnisse aller Schüler einzugehen.	Ich habe das Gefühl, dass adäquates außerschulisches Personal mich dabei unterstützt, den einzigartigen Bildungsbedürfnissen aller Schüler/innen nachzukommen.	[5] Meinem Gefühl nach gibt es geeignetes externes Personal, das mir hilft, auf die einzigartigen Bildungsbedürfnisse aller Schüler/innen einzugehen.
[6] All students will receive appropriate education and related services in inclusive education.	In an inclusive education system, every pupil receives the education appropriate for him/her and the services that are associated with this.	In einem inklusiven Bildungssystem erhält jeder Schüler die für ihn angemessene Bildung und die damit zusammenhängenden Leistungen.	Inklusive Bildung bedeutet, dass alle Schüler/innen angemessene Bildung und damit zusammenhängende Betreuung erhalten.	[6] Durch Inklusion wird jede/r Schüler/in die für sie/ihn angemessene Bildung und die damit zusammenhängenden Unterstützungsleistungen erhalten.
[7] I feel from my experience that there is support for inclusion from the Education Department/Board.	In my experience, inclusive education is supported by the Department of Education/Board of Education.	Meiner Erfahrung nach wird Inklusion vom Bildungsministerium/Bildungsausschuss unterstützt.	Aus meiner Erfahrung heraus habe ich das Gefühl, dass das Bildungsministerium Inklusion unterstützt.	[7] Meiner Erfahrung nach denke ich, dass Inklusion vom Bildungs-/Kultusministerium unterstützt wird.
[8] It is possible to organise classes in a way that is suitable for all children.	It is possible to organise classes so that the lesson is suitable for all children.	Es ist möglich, Klassen so zu organisieren, dass der Unterricht für alle Kinder geeignet ist.	Es ist möglich, den Unterricht so zu gestalten, dass er für alle Kinder angemessen ist.	[8] Es ist möglich, Klassen so zu organisieren, dass der Unterricht für alle Kinder geeignet ist.
[9] I feel there are adequate personnel within school to support me to address the unique educational needs of all students.	There are appropriate internal staff who help me to respond to the special educational needs of all pupils.	Es gibt geeignetes internes Personal, das mir hilft, auf die speziellen Bildungsbedürfnisse aller Schüler einzugehen.	Ich habe das Gefühl, dass adäquates innerschulisches Personal mich dabei unterstützt, den einzigartigen Bedürfnissen aller Schüler/innen gerecht zu werden.	[9] Ich denke es gibt geeignetes internes Personal, das mir hilft, auf die einzigartigen Bildungsbedürfnisse aller Schüler/innen einzugehen.
[10] Inclusion will foster acceptance of differences among students.	Inclusive education facilitates the acceptance of differences between pupils.	Inklusion fördert die Akzeptanz von Unterschieden zwischen den Schülern.	Inklusion wird die Akzeptanz von Unterschieden zwischen Schüler/innen fördern.	[10] Inklusion fördert die Akzeptanz von Unterschieden zwischen den Schüler/innen.

(continued)

English		German		
Final items (EN)	Back-translation	Prof. translation	Literal translation	Final items (DE)
[11] Labelling students (e.g. gender, race, ethnicity, disability, language, socio-economic status) is necessary to provide a quality education to them.	Grouping pupils according to gender, background, ethnicity, disability, language, socio-economic background, etc. is necessary in order to provide them with high-quality education.	Die Einteilung der Schüler nach Geschlecht, Abstammung, ethnischer Zugehörigkeit, Behinderung, Sprache, sozioökonomischem Hintergrund etc. ist notwendig, um ihnen qualitativ hochwertige Bildung zu bieten.	Schüler/innen mit einem Label zu versehen (z. B. Geschlecht, Migrationshintergrund, Behinderung, Sprache, sozioökonomischer Status) ist notwendig um ihnen qualitativ hochwertige Bildung zuteilwerden zu lassen.	[11] Die Einteilung der Schüler/innen nach Geschlecht, Migrationshintergrund, Behinderung, Sprache, sozioökonomischem Hintergrund etc. ist notwendig, um ihnen qualitativ hochwertige Bildung zu bieten.
[12] The differentiated practices that inclusive education would require cannot be achieved.	It is not possible to provide differentiated instruction, as is required for inclusive education.	Ein differenzierter Unterricht, wie er für inklusive Bildung nötig ist, kann nicht geleistet werden.	Differenzierte Praktiken, wie sie für inklusive Bildung notwendig wären, können nicht umgesetzt werden.	[12] Ein differenzierter Unterricht, wie er für inklusive Bildung nötig ist, kann nicht geleistet werden.
[13] All children should be educated in the inclusive classroom.	All children should receive an inclusive education.	Alle Kinder sollten inklusiv unterrichtet werden.	Alle Kinder sollten in inklusiven Klassen unterrichtet werden.	[13] Alle Kinder sollten inklusiv unterrichtet werden.
[14] All children are capable of learning in inclusive settings.	All children can learn in an inclusive environment.	Alle Kinder können in einem inklusiven Umfeld lernen.	Alle Kinder sind in der Lage, in inklusiven Settings zu lernen.	[14] Alle Kinder können in einem inklusiven Umfeld lernen.
[15] Inclusion represents a negative change in our education system.	Inclusive education represents a change for the worse in our education system.	Inklusion ist eine negative Veränderung in unserem Bildungssystem.	Inklusion ist eine negative Entwicklung unseres Bildungssystems.	[15] Inklusion ist eine negative Veränderung in unserem Bildungssystem.
[16] I believe that with the right supports in place inclusion can work.	I believe that inclusive education can work if appropriate support is in place.	Ich glaube, dass Inklusion funktionieren kann, wenn entsprechende Hilfestellung geleistet wird.	Ich glaube, dass Inklusion funktionieren kann, wenn die richtigen Unterstützungssysteme etabliert sind.	[16] Ich glaube, dass Inklusion funktionieren kann, wenn die richtige Hilfestellung geleistet wird.
[17] I feel that inclusive education is a practical idea in my country.	In my view, it would be possible to implement inclusive education in my country.	Meiner Meinung nach ist inklusive Bildung in meinem Land realisierbar.	Meinem Gefühl nach ist inklusive Bildung eine in meinem Land praktisch umsetzbare Idee.	[17] Meiner Meinung nach ist inklusive Bildung in meinem Land realisierbar.
[18] Separating students is not necessary to provide a quality education to them.	It is not necessary to separate pupils in order to provide them with high-quality education.	Es ist nicht nötig, Schüler zu trennen, um ihnen qualitativ hochwertige Bildung zu bieten.	Schüler/innen zu separieren ist nicht notwendig um ihnen qualitativ hochwertige Bildung zuteilwerden zu lassen.	[18] Es ist nicht nötig, Schüler/innen voneinander zu trennen, um ihnen qualitativ hochwertige Bildung zu bieten.
[19] I believe that any student can learn in an inclusive school if the curriculum is adapted to meet their individual needs.	I believe that every pupil is capable of learning at an inclusive school if the syllabus is adapted to individual needs.	Ich glaube, dass jeder Schüler an einer inklusiven Schule lernen kann, wenn der Lehrplan auf die individuellen Bedürfnisse abgestimmt wird.	Ich glaube, dass jede/r Schüler/in in einer inklusiven Schule lernen kann, wenn das Curriculum ihren individuellen Bedürfnissen angepasst ist.	[19] Ich glaube, dass jede/r Schüler/in an einer inklusiven Schule lernen kann, wenn das Curriculum auf die individuellen Bedürfnisse abgestimmt wird.
[20] I feel that external support services are a waste of time.	In my view, external support services are a waste of time.	Meiner Meinung nach sind externe Unterstützungsdienste Zeitverschwendung.	Ich habe das Gefühl, dass externe Unterstützungsangebote reine Zeitverschwendung sind.	[20] Meiner Meinung nach sind externe Unterstützungsdienste Zeitverschwendung.
[21] The philosophy of inclusion cannot be implemented in 'real world' practices.	Inclusive education is a concept that cannot be implemented in the "real world".	Inklusion ist eine Idee, die in der „echten Welt“ nicht umsetzbar ist.	The grundsätzliche Philosophie, die hinter Inklusion steckt, kann in der 'wahren Welt' nicht umgesetzt werden.	[21] Inklusion ist eine Idee, die in der Praxis nicht umsetzbar ist.
[22] I feel from my experience that the Education Department/Board supports efforts at including all students into the classroom.	In my experience, the Department of Education/Board of Education supports efforts to include all pupils in lessons.	Meiner Erfahrung nach unterstützt das Bildungsministerium/ der Bildungsausschuss Bemühungen, alle Schüler in den Unterricht einzubeziehen.	Aus Erfahrung habe ich das Gefühl, dass das Bildungsministerium Bemühungen, alle Schüler/innen zu inkludieren, unterstützt.	[22] Aus meiner Erfahrung heraus denke ich, dass das Bildungs-/ Kultusministerium Bemühungen, alle Schüler/innen in den Unterricht einzubeziehen unterstützt.
[23] I get frustrated when I have to adapt the curriculum to meet the individual needs of all students.	I find it frustrating when I have to adapt my lesson plan to the individual needs of all of the pupils.	Es frustriert mich, wenn ich den Lehrplan auf die individuellen Bedürfnisse aller Schüler abstimmen muss.	Es frustriert mich, Anpassungen am Lehrplan vornehmen zu müssen, um den individuellen Bedürfnissen aller Schüler/innen gerecht zu werden.	[23] Es frustriert mich, das Curriculum auf die individuellen Bedürfnisse aller Schüler/innen hin umarbeiten zu müssen.

(continued)

English		German		
Final items (EN)	Back-translation	Prof. translation	Literal translation	Final items (DE)
[24] I am willing to adapt the curriculum to meet the individual needs of all students within inclusive classrooms.	I am prepared to adapt my lesson plan to the individual needs of all of the pupils in an inclusive class.	Ich bin bereit, den Lehrplan auf die individuellen Bedürfnisse aller Schüler einer inklusiven Klasse abzustimmen.	Ich bin bereit, den Lehrplan an die individuellen Bedürfnisse der Schüler/innen in inklusiven Klassen anzupassen.	[24] Ich bin bereit, in inklusiven Klassen das Curriculum auf die individuellen Bedürfnisse aller Schüler/innen hin anzupassen.
[25] I feel differentiated adjustments can be carried out in an inclusive classroom.	I believe that differentiation is possible in an inclusive class.	Ich bin der Meinung, dass Differenzierung in einer inklusiven Klasse möglich ist.	Ich habe das Gefühl, dass differenzierte Anpassungen im inklusiven Unterricht vorgenommen werden können.	[25] Ich bin der Meinung, dass Differenzierung in einer inklusiven Klasse möglich ist.
[26] I do not need any support to put inclusive education into practice.	I do not need any support to put inclusive education into practice.	Ich brauche keine Unterstützung, um inklusive Bildung in die Praxis umzusetzen.	Um Inklusion umsetzen zu können, brauche ich keinerlei Unterstützung.	[26] Ich brauche keine Unterstützung, um inklusive Bildung in die Praxis umzusetzen.
[27] It is too difficult to accommodate all students' differences in an inclusive classroom.	It is too difficult to take all of the differences between the pupils in an inclusive class into consideration.	Es ist zu schwierig, alle Unterschiede der Schüler einer inklusiven Klasse zu berücksichtigen.	Es ist zu kompliziert allen Unterschieden der Schüler/innen in inklusiven Klassen gerecht zu werden.	[27] Es ist zu schwierig, alle Unterschiede der Schüler/innen einer inklusiven Klasse zu berücksichtigen.
[28] I am willing to adapt the assessment of individual students in order for inclusive education to take place.	I am prepared to adapt the assessment of individual pupils in order to facilitate inclusive education.	Ich bin bereit, die Bewertung einzelner Schüler anzupassen, um inklusive Bildung zu ermöglichen.	Ich bin bereit, die Prüfungen auf die individuellen Schüler/innen anzupassen, damit inklusive Bildung umgesetzt werden kann.	[28] Ich bin bereit, die Notengebung einzelner Schüler/innen anzupassen, um inklusive Bildung zu ermöglichen.
[29] From my experience, I feel that the community is supportive of the implementation of inclusion.	In my experience, the local community supports the implementation of inclusive education.	Meiner Erfahrung nach unterstützt die Gemeinde die Umsetzung von Inklusion.	Aus Erfahrung habe ich das Gefühl, dass die Gesellschaft die Umsetzung von Inklusion unterstützt.	[29] Aus meiner Erfahrung heraus denke ich, dass die Gesellschaft die Umsetzung von Inklusion unterstützt.
[30] Inclusion is the best way to meet the needs of all students.	Inclusive education is the best way to fulfil the needs of all pupils.	Inklusion ist der beste Weg, um die Bedürfnisse aller Schüler zu erfüllen.	Inklusion ist der beste Weg den Bedürfnissen aller Schüler/innen gerecht zu werden.	[30] Inklusion ist der beste Weg, um die Bedürfnisse aller Schüler/innen zu erfüllen.
[31] Good teachers can differentiate their practices so that they can teach all students in their class/es.	Good teachers can provide differentiated instruction and thus address the needs of all of the pupils in their class(es).	Gute Lehrer können differenziert unterrichten und so allen Schülern ihrer Klasse/n gerecht werden.	Gute Lehrer/innen können ihre Praktiken differenzieren sodass sie alle Schüler/innen in ihrer Klasse unterrichten können.	[31] Gute Lehrer/innen können differenziert unterrichten und so allen Schüler/innen ihrer Klasse/n gerecht werden.
[32] Diversity within the classroom enriches the learning environment.	Variety in the classroom enriches the learning environment.	Vielfalt im Klassenzimmer bereichert die Lernumgebung.	Unterschiede innerhalb einer Klasse bereichern die Lernumgebung.	[32] Vielfalt in der Klasse bereichert die Lernumgebung.
[33] I feel there are adequate resources to support me to address the unique educational needs of all students.	In my view, suitable resources are available that help me to respond to the individual learning requirements of all pupils.	Meiner Meinung nach gibt es geeignete Ressourcen, die mir helfen, auf die individuellen Lernbedürfnisse aller Schüler einzugehen.	Ich habe das Gefühl, dass mich adäquate Ressourcen dabei unterstützen, den einzigartigen Bedürfnissen aller Schüler/innen nachzukommen.	[33] Meiner Meinung nach gibt es geeignete Ressourcen, die mir helfen, auf die individuellen Lernbedürfnisse aller Schüler/innen einzugehen.
[34] Parents hinder the successful implementation of inclusive education.	Parents present an obstacle to the successful implementation of an inclusive education system.	Eltern behindern die erfolgreiche Umsetzung eines inklusiven Bildungssystems.	Eltern verhindern die erfolgreiche Umsetzung inklusiver Bildung.	[34] Eltern behindern die erfolgreiche Umsetzung von inklusiver Bildung.
[35] Inclusion will foster understanding of differences among students.	Inclusive education promotes understanding of differences between pupils.	Inklusion fördert das Verständnis für die Unterschiede zwischen Schülern.	Inklusion wird zum Verständnis für Unterschiede zwischen Schüler/innen beitragen.	[35] Inklusion fördert das Verständnis für die Unterschiede zwischen Schüler/innen.
[36] It is a valuable experience for all children to be educated in inclusive classrooms.	Being taught in an inclusive class is a valuable experience for all pupils.	In einer inklusiven Klasse unterrichtet zu werden, ist eine wertvolle Erfahrung für alle Schüler.	Es ist eine wertvolle Erfahrung für alle Schüler/innen in inklusiven Klassen unterrichtet zu werden.	[36] In einer inklusiven Klasse unterrichtet zu werden, ist eine wertvolle Erfahrung für alle Schüler/innen.
[37] Working collaboratively with parents plays a major part in the success of inclusion.	The involvement of parents is a decisive factor for the success of inclusive education.	Die Einbindung der Eltern ist entscheidend für den Erfolg von Inklusion.	Die Zusammenarbeit mit Eltern spielt eine große Rolle für den Erfolg von Inklusion.	[37] Die Zusammenarbeit mit den Eltern ist entscheidend für den Erfolg von Inklusion.
[38] Inclusive education ultimately leads to social inclusion.	Inclusive education ultimately results in social inclusion.	Inklusive Bildung führt letztendlich zu sozialer Inklusion.	Inklusive Bildung führt letztendlich zu sozialer Inklusion.	[38] Inklusive Bildung führt letztendlich zu sozialer Inklusion.

(continued)

English		German		
Final items (EN)	Back-translation	Prof. translation	Literal translation	Final items (DE)
Section Two – Efficacy to Implement Inclusive Education for All				
[1] I can make my expectations clear about student behaviour.	I can make my expectations regarding pupil behaviour clear.	Ich kann meine Erwartungen in Bezug auf das Schülerverhalten deutlich machen.	Meine Erwartungen, wie sich Schüler/innen verhalten sollten, kann ich klarmachen.	[1] Ich kann meine Erwartungen an das Schüler/innenverhalten deutlich machen.
[2] I can accurately gauge student comprehension of what I have taught.	I am perfectly capable of judging whether or not a pupil has understood what I have explained.	Ich kann genau beurteilen, ob ein Schüler verstanden hat, was ich erklärt habe.	Ich kann genau ermitteln, was die Schüler/innen von dem, was ich ihnen beigebracht habe, verstanden haben.	[2] Ich kann genau beurteilen, ob ein/e Schüler/in verstanden hat, was ich erklärt habe.
[3] I am able to calm a student who is disruptive or noisy.	I can calm a disruptive or loud pupil.	Ich kann einen störenden oder lauten Schüler beruhigen.	Ich bin in der Lage, eine/n Schüler/in runterzubringen, wenn er/sie stört oder laut ist.	[3] Ich bin in der Lage, eine/n störende/n oder laute/n Schüler/in zu beruhigen.
[4] I can make parents feel comfortable coming to school.	I can make coming to school a pleasurable experience for parents.	Ich kann dafür sorgen, dass Eltern gern zur Schule kommen.	Ich kann Eltern ein gutes Gefühl vermitteln, in der Schule zu erscheinen.	[4] Ich kann Eltern ein gutes Gefühl vermitteln, in der Schule zu erscheinen.
[5] I am confident in my ability to prevent disruptive behaviour in the classroom before it occurs.	I am confident that I can prevent disruptive behaviour in class.	Ich bin überzeugt, dass ich störendes Verhalten in der Klasse verhindern kann.	Ich bin zuversichtlich die Fähigkeit zu haben, einen störungspräventiven Unterricht zu machen.	[5] Ich bin überzeugt von meiner Fähigkeit, störendes Verhalten in der Klasse präventiv verhindern zu können.
[6] I am confident in designing learning tasks so that the individual needs of all students are accommodated.	I can set tasks so that the individual requirements of all pupils are taken into account.	Ich kann Aufgaben so stellen, dass die individuellen Bedürfnisse aller Schüler berücksichtigt werden.	Ich bin zuversichtlich, die Lernaufgaben so zu gestalten, dass sie den individuellen Bedürfnissen aller Schüler/innen gerecht werden.	[6] Ich kann Aufgaben so stellen, dass die individuellen Bedürfnisse aller Schüler/innen berücksichtigt werden.
[7] I am able to provide an alternate explanation or example when students are confused.	I can provide an alternative explanation or an example if pupils are confused.	Ich kann eine alternative Erklärung oder ein Beispiel geben, wenn die Schüler verwirrt sind.	Ich bin in der Lage, eine alternative Erklärung oder ein alternatives Beispiel zu liefern, wenn die Schüler/innen verwirrt sind.	[7] Ich kann eine alternative Erklärung oder ein Beispiel geben, wenn die Schüler/innen verwirrt sind.
[8] I can assist families in helping their children to do well in school.	I can help families to support their children so that they perform well in school.	Ich kann Familien helfen, ihre Kinder zu unterstützen, damit sie in der Schule gute Leistungen bringen.	Ich kann Familien dabei unterstützen, dass ihre Kinder in der Schule gut zurechtkommen.	[8] Ich kann Familien dabei unterstützen, ihren Kindern zu helfen, in der Schule gut zurechtkommen.
[9] I am confident in informing others who know little about laws and policies relating to the inclusion of all students.	I can provide information to others who do not know very much about the laws and guidelines regarding the inclusive education of all pupils.	Ich kann andere informieren, die wenig über Gesetze und Richtlinien zur Inklusion aller Schüler wissen.	Ich bin zuversichtlich, andere informieren zu können, die wenig über die Gesetze und die Politik zur Inklusion aller Schüler/innen wissen.	[9] Ich bin zuversichtlich, andere informieren zu können, die wenig über die Gesetze und die Politik hinsichtlich der Inklusion aller Schüler/innen wissen.
[10] I am able to get children to follow classroom rules.	I can teach pupils to observe classroom rules.	Ich kann die Schüler dazu anleiten, sich an Unterrichtsregeln zu halten.	Ich kann die Schüler/innen dazu bringen, die Regeln in der Klasse zu befolgen.	[10] Ich kann die Kinder dazu bringen, sich an Unterrichtsregeln zu halten.
[11] I can use a variety of assessment strategies (e.g., portfolio assessment, modified tests, performance-based assessment, etc.).	I can apply different assessment strategies (e.g. portfolio assessment, modified tests, results-orientated assessment, etc.).	Ich kann verschiedene Bewertungsstrategien anwenden (z. B. Portfoliobeurteilung, abgeänderte Tests, ergebnisorientierte Bewertung, etc.).	Ich kann eine Reihe unterschiedlicher Prüfungsformen anwenden (z. B. Portfolio, modifizierte Klassenarbeiten, Praxis-orientierte Prüfungsformen etc.).	[11] Ich kann eine Reihe unterschiedlicher Bewertungsstrategien anwenden (z. B. Portfoliobeurteilung, abgeänderte Tests, etc.).
[12] I am confident in my ability to get students to work together in pairs or in small groups.	I am confident that I can teach pupils to work in pairs or small groups.	Ich bin überzeugt, dass ich Schüler dazu anleiten kann, in Zweier- oder Kleingruppen zu arbeiten.	Ich bin zuversichtlich, dass ich die Fähigkeit besitze, Schüler/innen zur Partnerarbeit oder zur Arbeit in kleinen Gruppen zu bringen.	[12] Ich bin von meiner Fähigkeit überzeugt, dass ich Schüler/innen dazu bringen kann, in Zweier- oder Kleingruppen zusammen zu arbeiten.
[13] I can control disruptive behaviour in the classroom.	I can control disruptive behaviour during class.	Ich kann störendes Verhalten im Unterricht kontrollieren.	Ich kann störendes Verhalten im Klassenraum kontrollieren.	[13] Ich kann störendes Verhalten im Unterricht kontrollieren.
[14] I am confident in my ability to get parents of all students involved in school activities of their children.	I am confident that I can include the parents of all pupils in the school activities of their children.	Ich bin überzeugt, dass ich die Eltern aller Schüler in die Schulaktivitäten ihrer Kinder einbinden kann.	Ich bin zuversichtlich, dass ich die Fähigkeit habe, die Eltern aller Schüler/innen in die schulischen Aktivitäten ihrer Kinder involviert zu sein.	[14] Ich kann die Eltern aller Schüler/innen dazu bewegen, in die Schulaktivitäten ihrer Kinder eingebunden zu sein.

(continued)

English		German		
Final items (EN)	Back-translation	Prof. translation	Literal translation	Final items (DE)
[15] I can collaborate with other professionals (e.g., itinerant teachers or speech pathologists) in designing educational plans for all students.	I can work together with other specialists (e.g. special educational needs teachers or speech therapists) in order to produce education plans for all pupils.	Ich kann mit anderen Fachleuten (z. B. Sonderpädagogen oder Logopäden) zusammenarbeiten, um Bildungspläne für alle Schüler zu erstellen.	Ich kann mit anderen Fachkräften (z. B. Wanderlehrer oder Sprachheilpädagogen) zusammenarbeiten, um Bildungspläne für alle Schüler/-innen zu entwerfen.	[15] Ich kann mit anderen Fachleuten (z. B. Sonderpädagogen oder Logopäden) zusammenarbeiten, um Bildungspläne für alle Schüler/innen zu entwerfen.
[16] I am confident when dealing with students who are physically aggressive.	I am confident when dealing with physically aggressive pupils.	Ich bin selbstsicher im Umgang mit körperlich aggressiven Schülern.	Ich bin selbstsicher, wenn ich mit Schüler/innen umgehen muss, die körperlich aggressiv sind.	[16] Ich bin selbstsicher im Umgang mit körperlich aggressiven Schüler/innen.
[17] I can provide appropriate challenges for very capable students.	I can challenge gifted pupils appropriately.	Ich kann sehr gute Schüler entsprechend fördern.	Ich kann für besonders leistungsfähige Schüler/innen angemessene Herausforderungen bereitstellen.	[17] Ich kann besonders leistungsfähige Schüler/innen entsprechend fördern.
[18] I am able to work jointly with other professionals and staff (e.g., aides, other teachers) to teach all students in the classroom.	I can work together with other specialists and staff (e.g. assistants or other teachers) in order to teach all of the pupils in a class.	Ich kann mit anderen Fachleuten und Mitarbeitern (z. B. Hilfskräften oder anderen Lehrern) zusammenarbeiten, um alle Schüler der Klasse zu unterrichten.	Ich kann mit anderen Fachkräften und anderem Personal (z. B. Schulbegleiter, andere Lehrer/innen) zusammenarbeiten um alle Schüler/innen zu unterrichten.	[18] Ich kann mit anderen Fachleuten und Mitarbeiter/innen (z. B. weiteres pädagogisches Personal oder andere Lehrer/innen) zusammenarbeiten, um alle Schüler/innen der Klasse zu unterrichten.
Section Three – Personal Information and Background				
[1] I am: female / male	I am: male / female	Ich bin: weiblich / männlich	Ich bin: weiblich / männlich	[1] Ich bin: weiblich / männlich
[2] My age: up to 30 years / 31-40 years / 41-50 years / 51-60 years / above 60 years	My age: 30 or below / 31-40 / 41-50 / 51-60 / over 60	Mein Alter: bis 30 Jahre / 31-40 Jahre / 41-50 Jahre / 51-60 Jahre / über 60 Jahre	Mein Alter: bis 30 Jahre / 31-40 Jahre / 41-50 Jahre / 51-60 Jahre / über 60 Jahre	[2] Mein Alter: bis 30 Jahre / 31-40 Jahre / 41-50 Jahre / 51-60 Jahre / über 60 Jahre
[3] Training to teach/currently teaching: primary/elementary / secondary/high school / other. If other give more detail:	I am training to teach / I am currently teaching: primary school / secondary school / grammar school / other. If other, please give details.	In der Lehrerausbildung für/Ich unterrichte zurzeit: Grundschule / Realschule / Gymnasium / andere. Falls andere, bitte näher erläutern.	Schulform für die Sie ausgebildet werden, bzw. an der Sie aktuell unterrichten: Primar / Sekundar / Andere. Wenn andere, bitte geben Sie mehr Details:	[3] Ich bin in der Lehramtsausbildung für bzw. ich unterrichte zurzeit: Grundschule/Primarbereich / Sekundarbereich / Gymnasium / Andere Falls andere, bitte erläutern:
[4] If you are currently teaching, how many years experience have you had? ___ years / N/A	If you currently teach, how much teaching experience do you have? ___ years / n/a	Falls Sie zurzeit unterrichten, wie viel Unterrichtserfahrung haben Sie? ___ Jahre / entfällt	Falls Sie aktuell unterrichten, wie lange Erfahrung haben Sie als Lehrer/in: ___ Jahre / nicht zutreffend	[4] Falls Sie zurzeit unterrichten, wie viele Jahre Unterrichtserfahrung haben Sie? ___ Jahr/e / nicht zutreffend
[5] Do you hold a postgraduate degree/diploma? no / yes. If yes give more detail:	Do you have an advanced degree? No / Yes. If yes, please give details.	Verfügen Sie über einen weiterführenden Studienabschluss? Nein / Ja. Wenn ja, bitte näher erläutern.	Haben Sie einen weiterführenden Hochschulabschluss? nein / ja. Falls ja, bitte erläutern	[5] Verfügen Sie über einen weiterführenden Studienabschluss? nein / ja. Falls ja, bitte erläutern:
[6] My knowledge of the local legislation and/or policy as it pertains to inclusive education for all: very good / good / average / poor / none	My knowledge of regional legislation and/or guidelines regarding inclusive education for all is: very good / good / average / poor / no knowledge	Meine Kenntnis der regionalen Gesetzgebung und/oder Richtlinien in Bezug auf inklusive Bildung für alle: sehr gut / gut / mittelmäßig / schlecht / keine	Mein Wissen zur aktuellen Gesetzgebung und/oder Politik zur inklusiven Bildung für alle ist: sehr gut / gut / durchschnittlich / schlecht / nicht vorhanden	[6] Meine Kenntnis der regionalen Gesetzgebung und/oder Politik in Bezug auf „inklusive Bildung für alle“ ist: sehr gut / gut / mittel / wenig / nicht vorhanden
[7] I have had the following amount of training focusing on inclusive education of all students: high (at least 40hrs) / some / none	I have the following level of training in inclusive education for all pupils: high (at least 40 hours) / low / none	Ich habe folgendes Ausbildungsniveau in Bezug auf inklusive Bildung für alle Schüler: hoch (mindestens 40 Stunden) / niedrig / keines	Ich hatte das folgende Ausmaß an Aus-/bzw. Fortbildung zum Thema inklusive Bildung für alle Schüler/innen: hoch (mindestens 40 Stunden) / ein bisschen / keine	[7] Ich habe folgenden Umfang an Aus-/Fortbildung in Bezug auf „inklusive Bildung für alle Schüler/innen“: hoch (mindestens 40 Stunden) / wenig / nicht vorhanden
[8] Do you have experience with inclusive classroom settings (teaching experience, professional experience, teaching assistance, voluntary assistance etc.)? no / yes. If yes... Quality of past experience with an inclusive setting: positive / neutral / negative	Do you have experience with inclusive education (teaching experience, professional experience, teaching internship, voluntary work, etc.)? No / Yes. If yes, please give details. How would you describe your experience with inclusive education? Positive / neutral / negative.	Haben Sie Erfahrung mit inklusivem Unterricht (Unterrichtserfahrung, Berufserfahrung, Unterrichtspraktikum, Volontärsarbeit usw.)? Nein / Ja. Wenn ja, bitte näher erläutern: Art der Erfahrung mit inklusivem Unterricht: positiv / neutral / negativ.	Haben Sie Erfahrung mit inklusiven Klassensettings (unterrichtet, im Team unterrichtet, Praktikum etc.)? Ja / Nein. Wenn ja, geben Sie mehr Details: Wie bewerten Sie Ihre Erfahrungen mit inklusiven Settings: positiv / neutral / negativ	[8] Haben Sie Erfahrung mit inklusiven Unterrichtssettings (Unterrichtserfahrung, Berufserfahrung, Unterrichtspraktikum, Volontärsarbeit usw.)? nein / ja. Falls ja, ... Qualität der Erfahrung mit einem inklusiven Setting: positiv / neutral / negativ