

**Fachbereich Erziehungswissenschaft und Psychologie
der Freien Universität Berlin**

Evaluation of a teacher preparation program using the development of teacher
self-efficacy as an outcome - a longitudinal study

Dissertation
zur Erlangung des akademischen Grades
Doktorin der Philosophie (Dr. phil.)

vorgelegt von Dipl.-Psych. Franziska Pfitzner-Eden
Berlin, Mai 2015

Erstgutachterin: Prof. Dr. Felicitas Thiel, Freie Universität Berlin

Zweitgutachter: Prof. Dr. Ralf Schwarzer, Freie Universität Berlin

Tag der Disputation: 14.07.2015

Abstract

Teacher self-efficacy (TSE) is a central predictor for a multitude of sought after outcomes for teachers and students alike. However, it is less well known how TSE develops in preservice teachers. Addressing this gap in research, the purpose of this thesis was to explore four research questions: (1) Does TSE develop in preservice teachers during their university preparation? (2) If such changes are found, do they have any impact on preservice teachers' changes in their intention to quit the teaching degree? (3) Can changes in TSE during the practical phases of a teacher preparation program be predicted by the four sources of information assumed to be influential in forming teacher self-efficacy beliefs? (4) How do the four sources interact in predicting changes in teacher self-efficacy? In order to address these questions, a new instrument for measuring the three proposed dimensions of TSE (i.e., instructional strategies, classroom management, student engagement) in preservice teachers was developed on the basis of a thorough analysis of previous TSE instruments and Bandura's self-efficacy theory (Chapter 1). This process and first results from a validation study are presented in Chapter 2 (Study 1). Participants in this study were preservice teachers in three diverse samples from Germany and New Zealand, at two different stages of teacher education. Confirmatory factor analyses and invariance tests confirmed a good fit for the intended three-factor model and equivalent measurement properties for all samples. In addition, first evidence in favor of good construct validity was gathered by examining convergent validity with *general self-efficacy* and *study satisfaction*. Research question 1 and 2 were then explored in Chapter 3 (Study 2). For the purpose of this study, data was collected at three times for two samples of preservice teachers. Beginning preservice teachers in Sample 1 were tested at the start of their first year of teacher education, at the end of their first year before their first practicum at school, and at the beginning of their second year after their practicum. Advanced preservice teachers in Sample 2 were at least third year students (80% Master's students), and were tested at the beginning of the university term, at the end of that term before a teaching practicum at school, and at the beginning of the next term after their practicum. Changes were analyzed using latent-true-change modeling, an improvement over previous analysis methods employed in this area. Results showed that depending on TSE dimension, TSE decreased or did not change during the university year for beginning preservice teachers, and increased during the practicum. For the advanced preservice teachers, again depending on TSE dimension, TSE increased both during the term and during the practicum. In both cohorts, increases in TSE during the practicum were associated with

decreases in preservice teachers' intention to quit the teaching degree. In Chapter 4 (Study 3) research question 3 and 4 were addressed. For the purpose of this study, a new measure of the four sources hypothesized to influence TSE was developed using a deductive process to maximize alignment with self-efficacy theory. Changes in TSE that occurred during the practicum in beginning and advanced preservice teachers were significantly predicted by all sources as expected. Three theoretically plausible models of how the sources interact in predicting TSE changes were then explored. The evaluation of the different models provided some first evidence in favor of the hypothesis that preservice teachers' TSE beliefs are significantly influenced by mastery experiences made during the practicum, which are largely informed to varying degrees by the other three sources. Specifically, the results highlighted the role of the mentor teacher's feedback and preservice teachers' negative physiological and affective responses to the practicum experience. In the final chapter, the central findings are discussed, a range of limitations to the studies is detailed, several implications for theory and practice are outlined, and potentially fruitful future avenues for further exploration are suggested.

Key words:

teacher self-efficacy – preservice teachers – sources of self-efficacy – latent true change

Zusammenfassung

LehrerSelbstwirksamkeit (LSW) wird als ein zentraler Prädiktor für eine Vielzahl von erwünschten Effekten für Lehrkräfte sowie für Schülerinnen und Schüler anerkannt. Weniger erforscht ist allerdings, wie sich die LSW im Laufe des Lehramtsstudiums entwickelt. Um sich diesem Desiderat empirisch anzunähern, wurden im Rahmen dieser Arbeit vier Forschungsfragen untersucht: (1) Verändert sich die LSW bei Lehramtsstudierenden während des Lehramtsstudiums? (2) Wenn sich Veränderungen der LSW zeigen: haben diese einen Einfluss auf eine Veränderung der Studienabbruchneigung von Lehramtsstudierenden? (3) Können Veränderungen der LSW, die während der praktischen Phasen des Lehramtsstudiums auftreten, durch vier Informationsquellen, von denen angenommen wird, dass sie die LSW beeinflussen, erklärt werden? (4) In welcher Beziehung zueinander stehen die Quellen bei der Vorhersage von Veränderungen der LSW? Um diese Fragen zu untersuchen, wurde zunächst auf der Basis einer systematischen Analyse bisheriger Instrumente zur Erfassung der LSW und Bandura`s einschlägiger Theorie zur Selbstwirksamkeit (Kapitel 1) ein neues Instrument zur Messung der drei postulierten Dimensionen der LSW (Instruktion, Klassenmanagement und Schülermotivierung) bei Lehramtsstudierenden entwickelt. Dieser Prozess sowie erste Ergebnisse einer Validierungsstudie werden im zweiten Kapitel vorgestellt (Studie 1). Untersucht wurden insgesamt drei Stichproben von Lehramtsstudierenden aus Deutschland und Neuseeland, die sich in unterschiedlichen Abschnitten des Studiums befanden. Konfirmatorische Faktorenanalysen und Invarianztestungen belegten einen guten Fit des intendierten 3-Faktorenmodells und äquivalente Messeigenschaften für alle Stichproben. Zusätzlich konnten durch die Untersuchung von konvergenter Validität zu den Kriterien *Generelle Selbstwirksamkeit* und *Studienzufriedenheit* erste Belege für eine gute Konstruktvalidität gesammelt werden. In Kapitel 3 (Studie 2) wurde dann der ersten und zweiten Forschungsfrage nachgegangen. Dafür wurden Daten zu drei Messzeitpunkten an zwei Stichproben von Lehramtsstudierenden erhoben. Die erste Stichprobe wurde direkt am Beginn des ersten Studienjahres, am Ende des ersten Studienjahres vor einem ersten Schulpraktikum, und am Anfang des nächsten Studienjahres nach dem Praktikum befragt. Die zweite Stichprobe setzte sich aus fortgeschrittenen Lehramtsstudierenden (mindestens 3. Studienjahr; 80% Masterstudierende) zusammen und wurde am Beginn des Semesters, am Ende des Semesters vor einem Unterrichtspraktikum, und am Anfang des nächsten Semesters nach dem Praktikum befragt. Unterschiede in der LSW und Studienabbruchneigung wurden mit Hilfe von *latent-true-change* Modellen analysiert, was eine Verbesserung bisheriger

Analysemethoden in diesem Bereich darstellt. Die Ergebnisse konnten zeigen, dass, in Abhängigkeit von der LSW-Dimension, die LSW während des ersten Studienjahres zurückging oder sich nicht veränderte und während des Praktikums anstieg. In der Gruppe der fortgeschrittenen Studierenden zeigten sich, ebenfalls abhängig von der untersuchten LSW-Dimension, Zuwächse während des Semesters und des Praktikums. In beiden Kohorten war ein Anstieg der LSW mit einem Rückgang der Studienabbruchneigung assoziiert. In Kapitel 4 (Studie 3) wurde die dritte und vierte Forschungsfrage bearbeitet. Im Rahmen dieser Studie wurde zunächst auf der Basis eines deduktiven Ansatzes ein Instrument zur Messung der vier postulierten Quellen der LSW entwickelt. Die Wahl des deduktiven Ansatzes trug dazu bei, eine höchstmögliche Übereinstimmung mit Bandura`s Selbstwirksamkeitstheorie zu erzielen. Veränderungen der LSW, die während des Praktikums bei Studienanfängerinnen und -anfängern sowie bei fortgeschrittenen Studierenden auftraten, konnten durch alle vier Quellen erwartungsgemäß vorhersagt werden. Daraufhin wurden drei theoretisch plausible Modelle verglichen, die unterschiedliche Beziehungen zwischen den Quellen annehmen. Der Modellvergleich lieferte erste Hinweise darauf, dass Erfolgserfahrungen während des Praktikums die Veränderungen in der LSW maßgeblich beeinflussen und, dass die Einschätzung der Erfolgserfahrungen wiederum zu unterschiedlichen Anteilen von den anderen drei Quellen abhängt. Im Einzelnen stellten die Ergebnisse die bedeutende Rolle von Mentorenfeedback heraus, sowie den Einfluss von negativen physiologischen und affektiven Reaktionen der Lehramtsstudierenden auf das Praktikum. Im abschließenden Kapitel der Arbeit werden die zentralen Ergebnisse integriert diskutiert, die Grenzen der vorliegenden Arbeit beleuchtet, Implikationen für Theorie und Praxis abgeleitet, sowie ein Ausblick auf weitere Forschung gegeben.

Schlagwörter:

Lehrerselbstwirksamkeit – Lehramtsstudierende – Quellen der Selbstwirksamkeit – Latentes Veränderungsmodell

Table of Contents

Chapter 1	Introduction	2
1.1	How is teacher self-efficacy represented in the field of teaching and teacher education?	4
1.2	What is teacher self-efficacy (and what is it not)?	10
1.3	How can teacher self-efficacy be measured?	15
1.4	Why does teacher self-efficacy matter?	21
1.5	How does teacher self-efficacy develop?	24
1.6	Overview of the thesis and research aims	31
1.7	References	38
Chapter 2	An adapted measure of teacher self-efficacy for preservice teachers: Exploring its validity across two countries	50
2.1	Introduction	52
2.2	Method	55
2.3	Results	59
2.4	Discussion	63
2.5	References	67
2.6	Acknowledgements	71
Chapter 3	I feel less confident so I quit? True changes in teacher self-efficacy predict changes in preservice teachers' intention to quit their teaching degree	74
3.1	Introduction	75
3.2	Method	82
3.3	Results	88
3.4	Discussion	94
3.5	References	100
3.6	Acknowledgments	106

Chapter 4	Why do I feel more confident? Bandura's sources predict preservice teachers' latent changes in teacher self-efficacy	108
4.1	Introduction	109
4.2	Method	115
4.3	Results	123
4.4	Discussion	132
4.5	Acknowledgments	143
Chapter 5	General Discussion	146
5.1	Summary and discussion of results	147
5.2	Limitations and strengths	157
5.3	Implications	162
5.4	Future directions	165
5.5	Conclusions	167
5.6	References	168
Danksagung	173
Erklärung	175
Curriculum Vitae	176

1

Introduction

1 Introduction

Teaching is a demanding, complex profession (e.g., Darling-Hammond & Bransford, 2005; Kunter, 2014). Imagine the difference between having to do a challenging job every day that you do not feel you are capable of and doing a challenging job every day that you do feel capable of. Such is the difference that self-efficacy makes. Self-efficacy defined as "...beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments." (Bandura, 1997, p. 3) is a powerful predictor of human behavior (e.g., Bandura, 2001; Bandura & Locke, 2003). More specifically, *teacher* self-efficacy (TSE), or the beliefs that teachers hold regarding their teaching capabilities, has been shown to be a reliable predictor for numerous sought after outcomes for teachers and even students. In contrast, if teachers do not feel that they are capable of doing what needs doing on a daily basis, a downward spiral can ensue. Numerous studies have consistently linked burnout symptoms in teachers as well as teachers' attrition and turnover to TSE (Chesnut & Burley, 2015; Coladarci, 1992; Klassen & Chiu, 2011; Schwarzer & Hallum, 2008; Skaalvik & Skaalvik, 2007, 2008, 2010).

Teachers' attrition and turnover is a serious problem in many countries (e.g., Lindqvist, Nordänger, & Carlsson, 2014; Wang, Hall, & Rahimi, 2015). Particularly among early career teachers attrition rates are high (OECD¹, 2005), with reported rates within the first five years of teaching ranging between 30% and 50% in Australia, the UK², and the USA³ (Wang et al., 2015). While attrition rates in Germany are seen as unproblematic and relatively low (under 5%) (European Commission, 2012), premature retirement of teachers is common (OECD, 2005). In fact, 94% of teachers in Germany retire early, on average at the age of 59 (OECD, 2005). The leading cause for this early retirement is burnout syndrome (Bauer et al., 2006). In addition, the European Commission (2012) reports Germany, Italy and Austria to be the EU⁴ countries with the oldest teacher workforce, so teacher shortage is an imminent problem in these countries. Teacher shortage for certain subject areas is already a serious problem in Germany, as it is in Belgium, Luxemburg and Turkey (European Commission, 2012). Drawing on PISA⁵ 2009 results, Schleicher (2012) reports severe teacher shortages in Germany particularly with regard to qualified Science teachers (just over 40% of surveyed principals reported shortages, compared to just under 20% OECD average), and qualified

¹ Organization for Economic Co-operation and Development

² United Kingdom

³ United States of America

⁴ European Union

⁵ Programme for International Student Assessment

Mathematics teachers (just under 30% of principals reported shortages, compared to just under 20% OECD average). Less is known about attrition or retention rates during teacher education or during the transition between formal university teacher education and entering the profession. Wang et al. (2015) report that in several European countries, between 30% and 40% of preservice teachers who graduate, do not enter the teaching career.

Taken together, this data paints a rather bleak picture. Great individual and societal costs are incurred when preservice teachers do not transition into a teaching career, or teachers prematurely leave the career (OECD, 2005). So, if TSE is one of the factors that can potentially mitigate teacher attrition, act as a protective factor against the development of burnout syndrome in teachers, and predict positive outcomes for students, one might ask: How do you develop such a belief in the first place? And, what role does teacher education play in its development? These are the overarching questions addressed by this thesis. Specifically, the purpose of this thesis was to explore the following four research questions:

1. Does TSE develop in preservice teachers during their university preparation (Chapter 3)?
2. If TSE changes are found, do they have any impact on preservice teachers' changes in their intention to quit the teaching degree (Chapter 3)?
3. Can changes in TSE during the practical phases of a teacher preparation program be predicted by the four sources of information assumed to be influential in forming TSE beliefs (Chapter 4)?
4. How do the four sources interact in predicting changes in TSE (Chapter 4)?

Since none of the existing instruments commonly used to assess TSE was suited to the assessment of changes in preservice teachers' TSE, it was a first aim of this undertaking to develop an instrument that was fit for this task (Chapter 2).

In this introduction (Chapter 1) I address five questions, which are relevant to the purpose of this thesis. First, I explore if and how TSE plays a role in the field of teaching and teacher education. Second, the conceptual foundations of TSE are examined. Third, I determine how TSE is commonly measured and how it should be measured for the purpose of this thesis. Fourth, I review why TSE is of any relevance to the field of teaching and teacher education. And fifth, I describe the theory that underpins my understanding of how TSE develops. In the last section, I provide an overview of the overall study design and research aims.

1.1 How is teacher self-efficacy represented in the field of teaching and teacher education?

This section explores if and how TSE is conceptualized in frameworks developed to guide research and evaluation of teachers and teacher education. Since it would go far beyond the scope of this introduction to review all frameworks pertaining to teaching and teacher education nationally and internationally, I will limit my examination to a small number of influential standards, competence models, and conceptual frameworks of international large-scale studies.

Standards are one major class of frameworks that guide the evaluation of teachers and teacher education, as well as the design of teacher preparation programs. Standards reflect an output-oriented approach, focusing on a definition of optimal *behavior* that can be observed and measured (Frey, 2014). They provide a scaling for competencies that allows for evaluating how well a person can perform the behavior in question (Terhart, 2005). In the German-speaking countries, standards have been on the rise since controversies about quality assurance and a reform of teacher education in the late nineties and early 2000s (Baumert & Kunter, 2006), following sobering results from international large-scale comparative studies such as TIMMS⁶ and PISA (Klieme et al., 2003). Keeping the output-oriented focus of standards in mind, a construct like TSE should likely find no mention in any standards-based approach, as TSE is not a form of behavior in itself, but rather precedes behavior. A look at three influential standards proves this assumption to be true. TSE does not feature either in Oser's (1997) proposed list of 88 standards for teacher education, nor in the German standards for teacher education passed by the Standing Conference of the Ministers of Education and Cultural Affairs of the German States (KMK, 2004), or the five core propositions developed by the US National Board for Professional Teaching Standards (NBPTS), which function as standards for National Board Certification of teachers in the US (2002).

Competence-models are another class of frameworks often used to guide research on teaching and teacher education. Competence models typically focus on a classification of skills and knowledge that behavior is based on (Frey, 2014). According to Piwovar (2013), German researchers generally follow Weinert's (2001) rather comprehensive definition of competence as comprising individuals' cognitive skills and abilities for problem-solving, as well as their motivational, volitional, and social dispositions, which allow for the implementation of problem-solving strategies in different situations. A key feature of competencies, as opposed to strictly cognitive dispositions, is that they can be acquired

⁶ Trends in International Mathematics and Science study

through experience and learning (Klieme & Leutner, 2006), for example during teacher education. This fact makes competency models attractive to teacher educators.

In Germany, one of the most prominent frameworks outlining professional competencies of teachers is the COACTIV model first introduced by Baumert and Kunter in 2006, and later elaborated on by Baumert and Kunter in 2011. At the highest level the authors distinguish between four aspects of professional competence (see Figure 1): (a) beliefs and value commitments, (b) motivational orientations, (c) self-regulation, and (d) professional knowledge. The focus of the COACTIV project was on the professional knowledge aspect of professional competence, thus this aspect was further divided into five areas of knowledge, which are in part based on Shulman's (1986; 1987) popular distinction of content knowledge, general pedagogical knowledge, curriculum knowledge, pedagogical content knowledge, knowledge of learners and their characteristics, knowledge of educational contexts, and knowledge of educational ends, purposes, and values. While TSE is not explicitly displayed in the COACTIV model, the authors (Baumert & Kunter, 2006; Baumert & Kunter, 2011) refer to it in the description of their model as part of motivational orientations. While the COACTIV model first guided research on inservice teachers, it has since also been used for studying the German formal induction program (i.e., *Referendariat*) in the COACTIV-R study, and has been adopted (with some adaptations) for a recently implemented first comprehensive longitudinal study on teacher education programs in Germany: PaLea⁷ (Bauer et al., 2010). The PaLea model features the same aspects of teacher competence, whereby TSE is also classified as part of motivational orientations in preservice teachers (Bauer et al., 2010).

⁷ *Panel zum Lehramtsstudium*/Panel study of teacher education

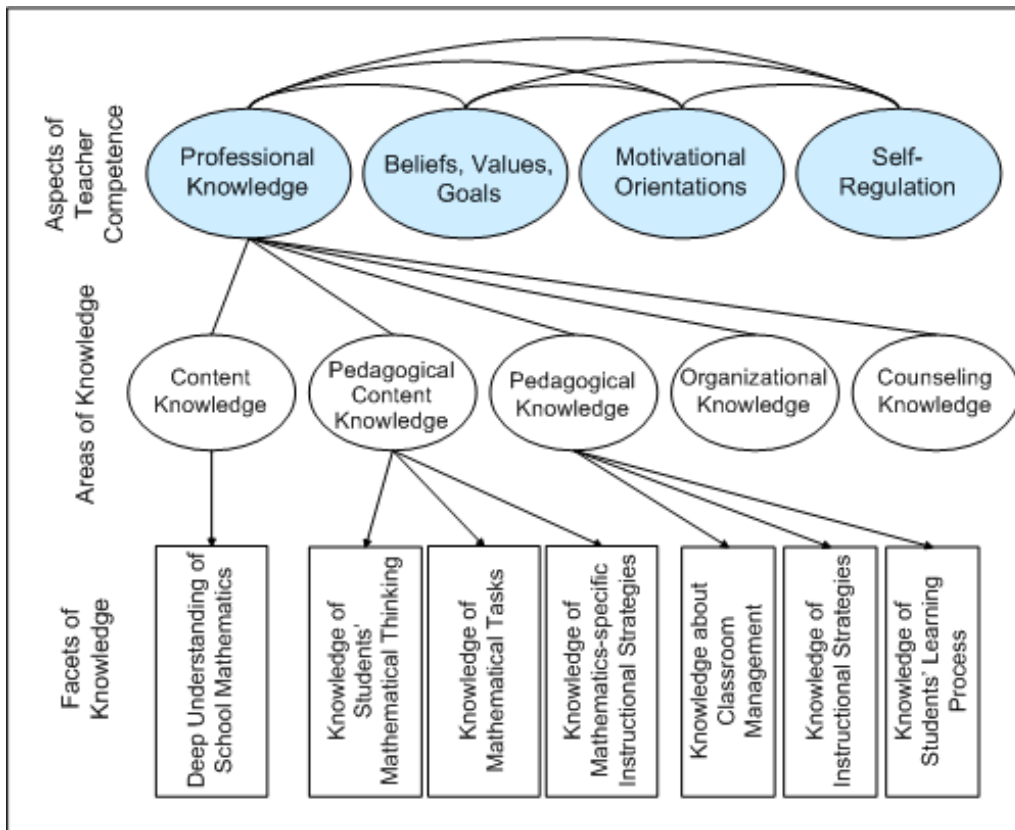


Figure 1. COACTIV model of teachers' professional competence (Source: <https://www.mpib-berlin.mpg.de/coactivr/englisch/research-questions/index.php>)

The 2005 book *“Preparing Teachers for a Changing World”* commissioned by the US National Academy of Education is considered to be a major milestone in the international discussion about the professional competence of teachers (Baumert & Kunter, 2006). In this book, Bransford, Darling-Hammond, and LePage (2005) propose a framework for understanding teaching and learning (see Figure 2). Also drawing on aspects of Shulman's (e.g., 1986; 1987) work, this framework focuses on different areas of knowledge: knowledge of learners and their development in social contexts, knowledge of subject matter and curriculum goals, knowledge of teaching (Bransford et al., 2005). In contrast to the COACTIV model, TSE, or other aspects of teacher competence for that matter, find no mention.

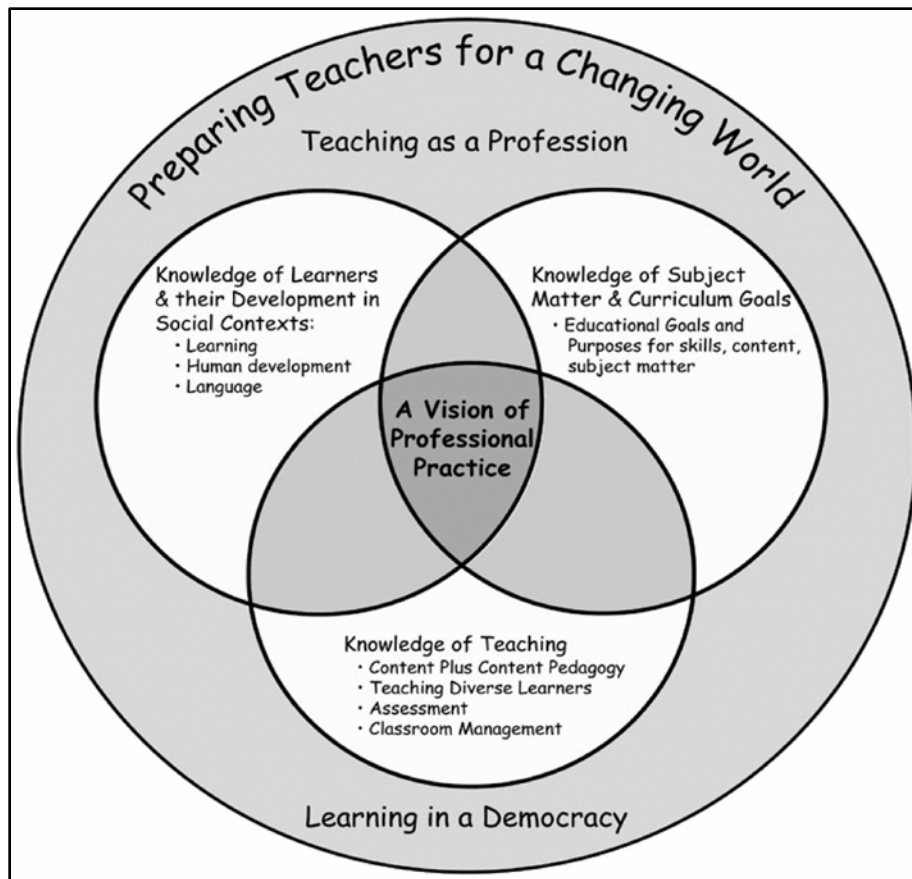


Figure 2. Framework for the analysis of teaching practices and beliefs (Source: Bransford et al., 2005, as illustrated in Darling-Hammond, 2006, p. 5)

As examples of international large-scale studies, I examine the conceptual frameworks of the TALIS⁸ survey of inservice teachers conducted by the OECD, and the TEDS-M⁹ study (Blömeke, Felberich, & Müller, 2008) of teacher education. Both studies are guided by conceptual frameworks, which comprise some aspects of competence models, but go beyond teacher competencies, taking a broader perspective.

The TALIS survey, conducted by the OECD (e.g., OECD, 2014) in 2008 with 24 participating countries, and in 2013 with 34 participating countries (Germany does not participate), offers an insight into teachers' working conditions and practices as seen by secondary teachers and school leaders. The TALIS framework (see Figure 3) explicitly features teachers' self-efficacy. Here, TSE is categorized as a job-related attitude, rather than being viewed as an aspect of professional competence, which also features in this model.

⁸ Teaching And Learning International Survey

⁹ Teacher Education and Development Study in Mathematics

Interestingly, a three-dimensional approach to TSE, which was first introduced by Tschannen-Moran and Woolfolk Hoy (2001)¹⁰, has been adopted in the TALIS model.

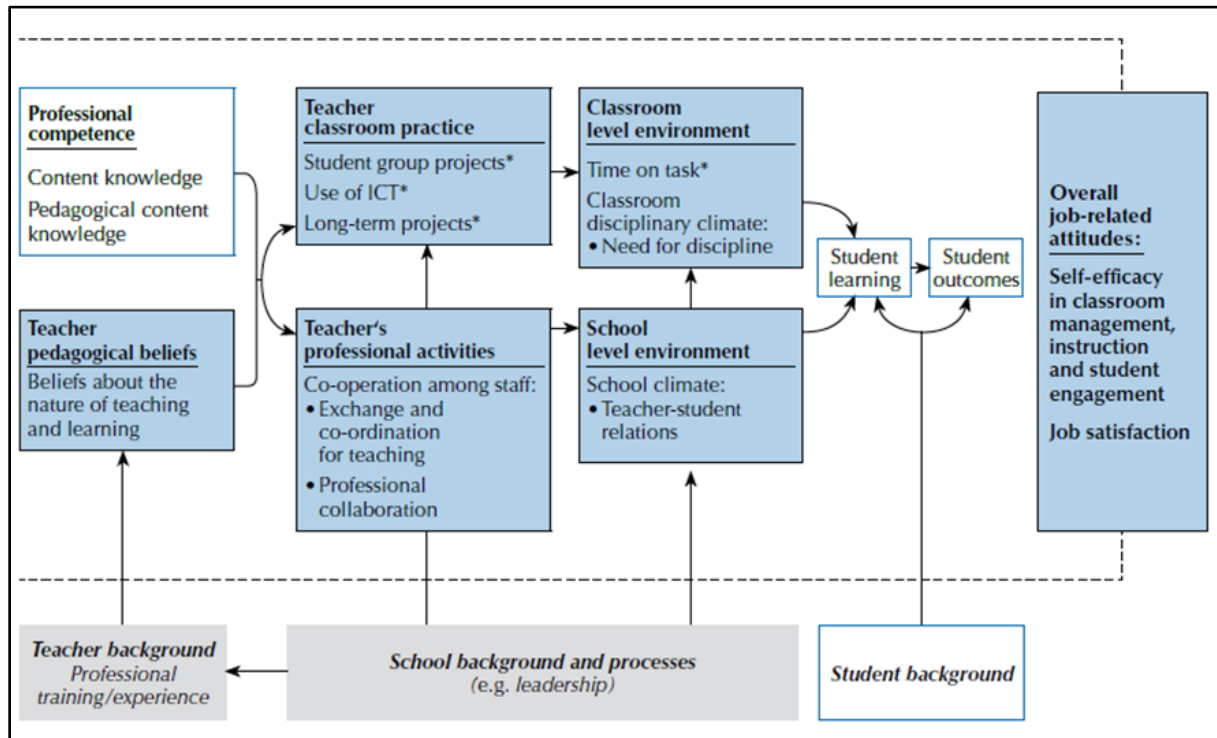


Figure 3. Framework for the analysis of teaching practices and beliefs (Source: OECD, 2014, p. 151). Note: Constructs that TALIS assessed are highlighted in darker grey; the asterisk (*) indicates single-item measures.

The international TEDS-M study, conducted in 17 countries (Botswana, Canada, Chile, Chinese Taipei, Georgia, Germany, Malaysia, Norway, Oman, Philippines, Poland, Russian Federation, Singapore, Spain, Switzerland, Thailand, and the United States) represents a first comprehensive attempt to study teacher preparation around the world (Tatto et al., 2008). This study focuses on how different policies and institutional practices shape the outcomes of preservice teachers. In the TEDS-M framework (see Figure 4) self-efficacy of preservice teachers is also explicitly featured: as a learning outcome of teacher preparation, classified as a teacher belief.

¹⁰ This is discussed in greater detail in section 1.3 of this introduction.

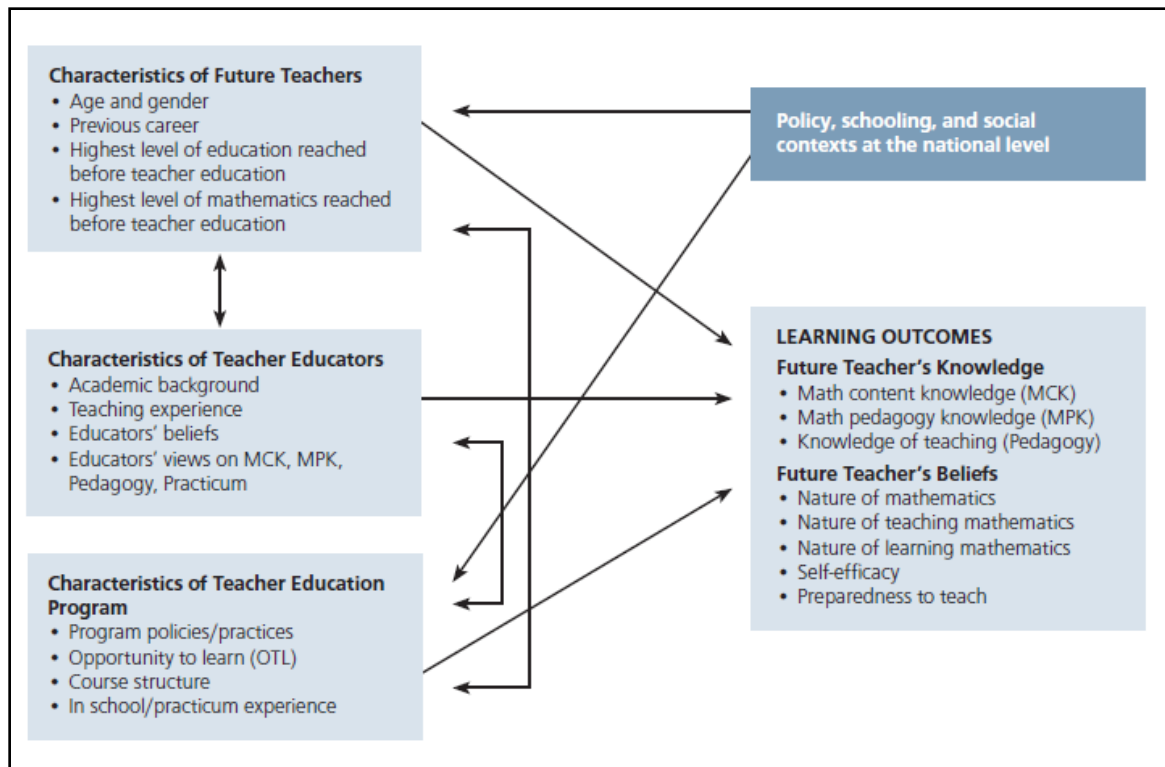


Figure 4. TEDS-M conceptual framework (Source: Tatto et al., 2008, p.14)

In sum, the reviewed frameworks indicate that within the field of teaching and teacher education, TSE has either been neglected, or has been regarded as part of three distinct constructs: (a) motivation, (b) attitudes, and (c) beliefs. There seems to be little consensus in the way the same construct is viewed by different educational researchers. In order to understand why researchers have categorized TSE in each of these three categories, I will briefly touch on definitions of each.

According to Kunter (2014) there is no clear definition of motivation. Already over thirty years ago, Kleinginna and Kleinginna (1981) proved just how diverse definitions of motivation were then, by listing and classifying 102 definitions. Kunter (2014) describes motivation as a psychological aggregate construct that subsumes processes as well as attributes of individuals, which generally present reasons for initiating, directing, executing, and maintaining behavior. She further offers a taxonomy of motivational constructs relevant to teachers (cf., p. 700). Here, she categorizes TSE (as well as control beliefs and attributions) as self-referent cognitions. At the highest level of classification Kunter (2014) distinguishes between two types of motivational constructs, one is responsible for the intensity, duration and quality (i.e., execution) of behavior, and the other is responsible for the initiation and direction of behavior. Self-referent cognitions, and thus TSE, are classified as the latter.

According to Eagly and Chaiken (2007) an attitude can be understood as “a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor” (p. 598). However, in his attempt to “clean up a messy construct” namely teacher beliefs, Pajares (1992) identifies attitudes as an “alias” (p. 309) of beliefs. Furthermore, he categorizes self-efficacy as a “self-belief” (p. 308). Instead of offering a succinct definition of teacher beliefs though (which might be an impossible endeavor), based on his comprehensive review he derives a list of 16 attributes that can be used to describe the nature of teacher beliefs¹¹. Pajares’ view of TSE as a self-belief has been shared by other prominent TSE researchers (e.g., Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998; Woolfolk Hoy, Davis, & Pape, 2006). In their ecological model of teachers’ knowledge and beliefs, Woolfolk Hoy et al. (2006) also categorize TSE as a belief about the self, whereby they treat beliefs and knowledge as overlapping, inseparable constructs. Furthermore, the recently published International Handbook of Research on Teachers’ Beliefs edited by Fives and Gill (2015) also categorizes TSE as a belief, and features a chapter on why TSE beliefs matter to the career development of preservice and inservice teachers (Siwatu & Chesnut, 2015).

Definitional issues seem to plague each and every one of the three (or rather two) constructs to which educational researchers have assigned TSE. In fact, these issues might be one of the reasons for why TSE has been placed under so many umbrella terms. In the next section, I explore what constitutes the psychological origins of TSE in greater detail in order to shed some more light on the nature of this construct.

1.2 What is teacher self-efficacy (and what is it not)?

Self-efficacy and similar constructs

The construct of TSE, as understood throughout this thesis, is based on Bandura’s (1977; 1986; 1997; 2006) conceptualization of self-efficacy as: “...beliefs in one’s capabilities to organize and execute the courses of action required to produce given attainments.” (Bandura, 1997, p. 3). In an effort to postulate a comprehensive psychological theory of human behavior, Bandura (1986, 1997, 2001) put forth his social cognitive theory. In labelling this theory, Bandura acknowledges the social origins of people’s cognitions and actions, as well as the causal contribution of people’s thought processes to their motivation, affect, and subsequently action. The construct of self-efficacy plays a pivotal role in his theory. Bandura (e.g., 1977, 1986, 1997, 2001, 2006) routinely refers to self-efficacy as *beliefs*, and classifies

¹¹ Thus, I will not relay any of the manifold definitions of teacher beliefs, and the interested reader is referred to the review by Pajares.

them as “self-referent thought” (1986, p. 390), a cognitive construct which mediates the relationship between knowledge and behavior. Accordingly, knowledge and skills are a necessary but not a sufficient requirement for accomplishments. Evidence from experimental studies shows that perceived self-efficacy contributes significantly to performance over and above skill level (e.g., Bandura, 1986, 1997; Bandura & Locke, 2003). Self-efficacy influences behavior via cognitive, motivational, affective, and selective processes (e.g., Bandura, 2001). For example, self-efficacy beliefs shape attributional thinking (cognitive), govern how much effort people invest and how long they persevere when encountering obstacles (motivational), influence people’s emotional reactions in demanding situations (affective), and determine if they choose or avoid challenging activities and environments (selective).

Bandura (2006) stresses that self-efficacy needs to be distinguished from other constructs that often get confused, particularly “self-esteem, locus of control and outcome expectancies” (p. 309). This is especially important to a valid measurement of self-efficacy beliefs (the focus in the next section of this introduction). While both self-efficacy and self-esteem are self-judgements, the former refers to one’s capabilities, whereas the latter refers to one’s worth (Bandura, 2006). Locus of control (a theoretical perspective introduced by Rotter in 1966) is concerned with whether outcomes are contingent on one’s own behavior (i.e., internal locus of control) or on factors outside of one’s control (i.e., external locus of control). Outcome expectancy is another construct that plays a role in social cognitive theory. While self-efficacy is concerned with one’s perceived capability to execute a certain performance, outcome expectations are concerned with the consequences (i.e., outcomes) of these performances (see Figure 5 for a graphic distinction). Bandura (e.g., 1986) argues that outcome expectations have a low predictive validity with regard to behavior, because if people do not feel like they are capable of executing a certain action, they will not attempt it even though they are convinced that desirable outcomes will flow from that action. On the other hand, one could argue that people who are not convinced that desirable outcomes will flow from their action, will also not engage in that action, regardless of whether they believe they can execute it or not. However, Bandura (e.g., 1986) reasons that in many instances the outcomes an individual expects are largely dependent on their self-efficacy judgments, and thus outcome expectancies do not add much incremental value, in addition to self-efficacy, to predicting behavior. Outcomes which are objectively not or only weakly related to performance quality or level are the exception. Teaching performances, for example, overall have been found to be significantly related to students’ achievement (e.g., Hattie, 2009). In

light of this result and Bandura's reasoning described above, it is justified to assume that self-efficacy beliefs of teachers play a greater role than their outcome expectations in the prediction of their performances¹².

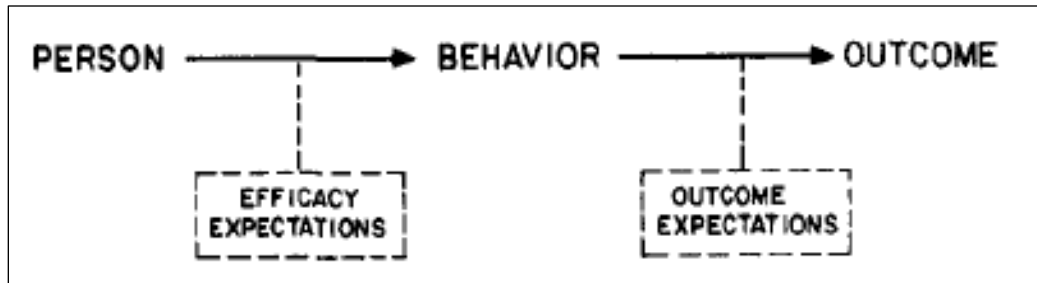


Figure 5. Depiction of the difference between self-efficacy beliefs and outcome expectations (Source: Bandura, 1977, p. 193)

Other researchers have also pointed to the difficult task of separating out a host of similarly conceptualized expectancy constructs (e.g., Bong, 2006; Pajares, 1996, 1997), control constructs (e.g., Skinner, 1996), or self-beliefs (Schunk & Zimmermann, 2006). Particularly one expectancy construct is often confused with self-efficacy: self-concept (Bong, 2006). Self-concept has been defined rather globally as “an individuals’ perception of himself” (Shavelson, Hubner, & Stanton, 1976, p. 435). This global conceptualization of self-concept can be equated to self-esteem (Marsh & O’Mara, 2008), which, as mentioned above, Bandura already succinctly differentiated from self-efficacy. Nonetheless, it might become increasingly more difficult to distinguish between these two constructs, when self-concept is conceptualized with a narrower focus on a particular domain and includes a competence judgment (Bong, 2006). What still distinguishes self-concept from self-efficacy is the evaluative component and the broader focus of self-concept (Bong & Skaalvik, 2003; Pajares, 1996). For example, a teacher’s self-concept item could be: *I am good at classroom management*. Whereas a teacher’s self-efficacy item would have a greater task-specific focus without an evaluative component, for instance: *I am confident that I can prevent noisy students from ruining my lesson*.

In an attempt to bring order to a vast amount of control constructs, Skinner (1996) introduced an “agents, means, and ends of control” (p. 551) distinction. In this conceptualization, agents are individuals or a group of individuals exerting control, means comprise the ways through which control can be exerted, and ends refer to outcomes, which

¹² For the sake of completeness, it should be noted that efficacy beliefs operate at the individual (i.e., self-efficacy) as well as at the collective level (e.g., Bandura, 2003). The focus of this thesis however is entirely on the individual level.

might be desired or undesired. Skinner (1996) categorizes self-efficacy as a belief about the agent-means relation, whereas outcome expectancies refer to means-end relations. In contrast, Schunk and Zimmerman (2008) classify self-efficacy as a competence belief and outcome expectancy as a control belief. Thus, other researchers do not seem to agree on the type of belief that self-efficacy represents, but they do agree on the separation between self-efficacy and outcome expectancies, even if they use different terminology.

Teacher self-efficacy

In order to maximize predictive validity, Bandura (e.g., 1997) recommends a tailoring of self-efficacy to the domain of functioning that is of interest. This is due to the fact that self-efficacy beliefs are not stable across different domains of action. People can simultaneously have low self-efficacy in the domain of romantic relationships for example, but high self-efficacy for work-related behavior. *Teacher self-efficacy* (TSE) is a domain-specific self-efficacy belief that has many definitions. While similar, these definitions are by all means not identical, as the compiled selection below indicates:

- “...the confidence teachers hold about their [...] capability to influence student learning.” Klassen, Tze, Betts, and Gordon (2011, p. 21)
- “...teachers’ belief in their ability to influence valued student outcomes.” (Wheatley, 2005, p. 748)
- “...the teacher's belief in his or her capability to organize and execute courses of action required to successfully accomplish a specific teaching task in a particular context.” (Tschannen-Moran et al., 1998, p. 233)
- “...the belief that one can bring about desired outcomes in one's students” (Soodak and Podell, 1996, p. 401)
- “...teachers’ belief that their efforts [...] will bring about student learning.” (Ross, 1998, p. 49-50)
- “...teachers' belief or conviction that they can influence how well students learn, even those who may be difficult or unmotivated” (Guskey and Passaro, 1994, p. 4)
- “...teachers’ situation-specific expectation that they can help students learn” (Ashton and Webb, 1986, p. 3).

Interestingly, with the exception of the definition by Tschannen-Moran et al. (1998), all definitions include a reference to student outcomes. Thus, most definitions of TSE make reference to outcome expectations, a construct supposed to be distinguished from self-efficacy

(as mentioned earlier). But why is that? It might have to do with the use of “attainment” in Bandura’s self-efficacy (1997) definition, and with the use of “outcomes” in Bandura’s (1977) first definition of self-efficacy as: “the conviction that one can successfully execute the behavior required to produce the outcomes” (p.193). Especially in the 1977 definition, the use of outcomes is unfortunate and has potentially misled TSE researchers to include student outcomes in their TSE definitions. However, in Bandura’s (e.g., 1986, 1997) further elaborations on self-efficacy it becomes clear that he simply means to specify a certain level of performance on the part of the agent. Thereby, the agent judges whether he or she can execute an action to a certain level of performance that would theoretically be required to produce a certain outcome. In contrast, self-efficacy is not a judgment about executing a behavior to just any standard, as in this case the self-efficacy judgment would be rendered meaningless. Whether or not one believes that the attained level of behavior would then actually lead to a certain outcome, is the outcome expectation. For the purpose of this thesis I would like to offer a definition which extends to preservice teachers:

Teacher self-efficacy refers to preservice and inservice teachers’ beliefs in their capabilities to organize and execute the courses of action required to produce given teaching attainments.

With regard to the issue of what type of construct TSE represents, for the purpose of this thesis I will follow Bandura’s (e.g., 1997), Pajares’ (e.g., 1992), Woolfolk Hoy et al.’s (2006), and Fives and Gill’s (2015) tradition of classifying TSE as a self-belief. While Baumert and Kunter (2006) might have been pragmatic in their focus on the motivational function of TSE, classifying TSE as a belief in the context of this thesis will serve as a reminder that TSE does not only influence behavior via motivation but also through other pathways (e.g., affective). This might have significant implications for the design of a study on TSE development, the interpretation of its results, and recommendations for further avenues of investigation.

Historically, TSE research however has been influenced by two theoretical perspectives: Bandura’s social cognitive theory, and Rotter’s (1966) locus of control theory. In fact, TSE made its first appearance under a slightly different label: teacher efficacy. Prominent researchers in the field of TSE (Henson, 2002; Klassen et al., 2011; Tschannen-Moran et al., 1998) agree that two items in a research project conducted by the RAND corporation (Armor et al., 1976) mark the dawn of this construct in educational research. The authors state that both items are based on Rotter’s (1966) locus of control theory and together measure teacher efficacy, defined as “...the extent to which the teacher believes he or she has the capacity to produce an effect on the learning of students.” (Armor et al., 1976, p. 23). One item assessed

teachers' external locus of control: "When it comes right down to it, a teacher really can't do much (because) most of a student's motivation and performance depends on his or her home environment" (Armor et al., 1976, p. 23). The other item assessed teachers' internal locus of control: "If I really try hard, I can get through to even the most difficult or unmotivated students" (Armor et al., 1976, p. 23). For the next 20 odd years, TSE researchers expanded on this first approach by the RAND Corporation and reinterpreted teacher efficacy in terms of Bandura's social cognitive theory. Commonly, self-efficacy was interpreted as being reflected in the internal locus of control dimension of teacher efficacy termed *personal teaching efficacy*, and outcome expectancies were assumed to be reflected in the external locus of control dimension termed *general teaching efficacy* (Gibson & Dembo, 1984; Tschannen-Moran et al., 1998; Woolfolk & Hoy, 1990). It is only since the nineties¹³ that researchers have started using the term teacher *self*-efficacy. This difference in label may (or may not) indicate a shift in the conceptual underpinnings of TSE research towards Bandura's construct of self-efficacy, away from Rotter's locus of control theory. However, labeling has been inconsistent, with researchers using teacher efficacy and teacher self-efficacy seemingly interchangeably. Over the last decade, more and more reviews (Henson, 2002; Klassen et al., 2011; Wheatley, 2005; Wyatt, 2013) have drawn attention to the fact that due to TSE being rooted in two conceptual strands, its conceptual integrity and consequently its measurement had been severely compromised. The issue of how to measure TSE is explored in greater detail in the next section.

1.3 How can teacher self-efficacy be measured?

This question has been the bane of TSE research. As mentioned above, the conceptual confusion surrounding TSE has had serious adverse impacts on its valid measurement. Tschannen-Moran et al. (1998) were the first to demand greater congruence of the conceptualization and measurement of TSE with Bandura's construct of self-efficacy. These calls were soon echoed by other reviewers (Henson, 2002; Klassen et al., 2011; Wheatley, 2002). In 2006, Bandura published a guide for constructing self-efficacy scales. In this guide, he emphasizes that in order to ensure content validity, efficacy items should be phrased "in terms of *can do* rather than *will do*" (p. 308), as the former signifies capability, while the latter signifies intent. He further stresses the conceptual distinction between self-efficacy and self-

¹³ A quick search of the "Psycinfo" database using "teacher efficacy" as a key word produced 328 hits published between 1979 and 2015. Using "teacher self-efficacy" as a key word produced 100 hits, with the first publication year being 1990.

esteem, locus of control, and outcome expectancies (as mentioned earlier), which needs to be kept in mind when designing self-efficacy scales. Another aspect of measurement that Bandura (2006) focuses on is the specification of the domain of action, which the self-efficacy scale is designed to measure. For this purpose he recommends a sound analysis of the relevant domain. In the domain of teaching for example, a comprehensive analysis of what teaching encompasses would be a prerequisite to any TSE measure. Further, Bandura recommends designing self-efficacy items only with regard to those behavioral aspects of the relevant domain that are somewhat within people's control. For example, in the case of TSE, an item that refers to an act that is within the teacher's control would be: *I am confident that I can keep students' attention focused on my instruction during lessons*. An example for an act outside the teacher's control would be: *I am confident that I can keep students' attention focused on their home work when they are at home*. The latter is entirely outside the teacher's control and would be a poor predictor of performance. Similarly, adopting a focus of the domain of action that is too narrow or too broad will also entail limited predictive validity and usefulness of the scale. A last point Bandura emphasizes, which needs to be considered in the construction of self-efficacy items, is the degree of challenge that the items reflect. Activities that are being judged should not be too easy, otherwise the scale will not discriminate between individuals with different self-efficacy levels. What constitutes a challenge is again dependent on the domain of action, and could be represented "in terms of level of ingenuity, exertion, accuracy, productivity, threat, or self-regulation" (Bandura, 2006, p. 311).

Overall, there have been two major issues with the assessment of TSE. One issue has been the confusion of the underlying self-efficacy construct with other constructs (e.g., Henson, 2002). The other has been casting the domain of action at the right level (e.g., Tschannen-Moran et al., 1998).

Tschannen-Moran et al. (1998) focused on the latter issue in a comprehensive review of TSE measures and their conceptual underpinnings. While the authors acknowledge the confusion created by the two conceptual strands that TSE has been based on, they emphasize issues with the specificity of the teaching domain of previous measures, leading them to conclude: "Researchers have tried very simple, general measures, as well as long, complex vignettes. None of the measures currently in use seems to have found the proper balance between specificity and generality." (p. 219). Consequently they focused on providing such balance with the development of their own measure: the TSES¹⁴ (Tschannen-Moran & Woolfolk Hoy, 2001). The TSES comprises three task areas: instructional strategies,

¹⁴ TSES = Teachers' Sense of Efficacy Scale

classroom management, and student engagement. Each dimension is measured by six items in the long form and four items in the short form of the TSES. Other authors promote different approaches to specificity. Schwarzer, Schmitz, and Daytner (1999), for example, created a one-dimensional measure with a more general focus. As compared to the TSES, their 10-item measure nonetheless comprises a broader range of activities, reflecting four task areas: job accomplishment; skill development on the job; social interaction with students, parents, and colleagues; and coping with job stress. Skaalvik and Skaalvik (2007; 2008) on the other hand criticize the TSES for having too narrow a focus. Similarly, they criticize the teacher self-efficacy scale by Schwarzer et al. (1999) for being limited to only one dimension. Instead, Skaalvik and Skaalvik (2007; 2008) argue for a broader focus, since teachers face tasks and demands that cannot all be represented by one or three dimensions. They suggest the following six dimensions, featured in their 24-item Norwegian Teacher Self-Efficacy Scale: “instruction, adapting education to individual students’ needs, motivating students, keeping discipline, cooperating with colleagues and parents, and coping with changes and challenges” (Skaalvik & Skaalvik, 2007, p. 620). Emmer and Hickman (1991) in contrast, suggested a 14-item scale specifically for TSE in classroom management. And these are just examples of subject-unspecific TSE measures. There is also a host of subject-specific TSE measures (e.g., an adaptation of the TSES for TSE in mathematics by Charalambous, Philippou, & Kyriakides, 2008; the Mathematics Teaching Efficacy Belief Instrument by Enochs, Smith, & Huinker, 2000; the elementary teachers’ Science Teaching Efficacy Belief Instrument by Riggs & Enochs, 1990). As can be seen from these examples, there is little agreement among researchers as to what constitutes the ‘right’ level of specificity in a TSE measure. Researchers who want to assess TSE have to decide for themselves what level of specificity is suited to their research purpose.

The second issue that has seriously hampered TSE research is the confusion between self-efficacy and other constructs, mainly created by the constructs’ historical rooting in two different theoretical frameworks, but also by the confusion between self-efficacy and outcome expectancies. In their review, Tschannen-Moran and Woolfolk Hoy (2001) identified a total of eight measures (incl. the RAND items) that had by then been developed to measure teacher efficacy. Nonetheless, only one of these instruments, namely the Gibson and Dembo Teacher Efficacy Scale, had dominated the research since its introduction in 1984 (Henson, 2002; Henson, Kogan, & Vacha-Haase, 2001). This scale extended the original RAND items into a two-dimensional 30-item instrument, which the authors interpreted to represent self-efficacy and outcome expectancies. However, this interpretation was later refuted by Guskey and

Passaro (1994), who demonstrated that the two scales represented rather an internal and external locus of control orientation of teachers. After the debut of the TSES (Tschannen-Moran & Woolfolk Hoy) in 2001, and possibly as a consequence of Henson's (2002) and Henson et al.'s (2001) strong critique of the Gibson and Dembo Scale, the TSES soon became the new standard measure of TSE (e.g., Duffin, French, & Patrick, 2012). However, I argue that the TSES also lacks a clear conceptual distinction between locus of control and self-efficacy¹⁵.

So, which instrument should be used for the purpose of this research? In order to study the development of TSE in preservice teachers over time, the instrument used to assess TSE should be:

- (a) subject-unspecific
- (b) suited to the German context but retain international appeal
- (c) suited specifically to preservice teachers
- (d) be cast at an appropriate level of specificity for observing change in TSE
- (e) conceptually sound and in line with Bandura's (2006) recommendations.

With regard to (a): the scope of this research goes beyond subjects and encompasses preservice teachers of any subject area. This is to examine if general patterns in the development of TSE, independent of subject area, can be discerned first. Regarding (b): data for this research is collected in Germany, from preservice teachers that attend a German teacher education program in the federal state of Berlin. Thus, the instrument needs to be appropriate for this context. In order to reach an international audience though, the instrument simultaneously needs to be applicable to a broader (i.e., English-speaking) context. With respect to (c): preservice teachers, especially early in their studies, have limited professional knowledge and experience. The teacher education program of the university of the study sample prepares teachers with a focus on pedagogical knowledge, content-knowledge, pedagogical-content knowledge, and practical teaching experiences. These are common aspects of the curricula of teacher preparation programs in many other countries (OECD, 2005; Totto et al., 2012). So, in order to tailor the instrument to preservice teachers, items should focus on teachers' actions which are within the realm of what is taught and practiced during the university program. Regarding (d): in order to be able to detect changes in preservice teachers' TSE judgments, the instrument needs to feature items and response scales that are conducive to that purpose. For example, it might be harder to detect intra-individual variance over time using a very general item (e.g., *I am certain that I can successfully teach*

¹⁵ Section 2.1 in Chapter 2 offers a more detailed elaboration of this point.

all students) as compared to a more specific item (e.g., *I am certain that I can provide an alternative example when students have trouble understanding*), which provides a more detailed frame of reference. Similarly restricting the response format to just a few categories (e.g., *1 = not certain, 2 = somewhat certain, 3 = certain*) also limits the intra-individual variability that could be detected over time. Bandura (2006) also recommends avoiding response scales with few categories as they are “less sensitive and less reliable” (p. 312). With regard to (e): in order to maximize the validity of the results from this study, the instrument used to assess TSE should be based on Bandura’s conceptualization of self-efficacy as outlined earlier in section 1.2. It should be further aligned closely with Bandura’s (2006) recommendations for constructing self-efficacy scales. A recent meta-analysis on the effect of TSE on commitment to the teaching profession by Chesnut and Burley (2015) provides empirical evidence for the importance of this requirement. The authors identified the accuracy of the TSE measure as a significant moderator. TSE measures that were deemed conceptually accurate with regard to Bandura’s (1997) definition of self-efficacy and his 2006 guidance on measurement produced higher effect sizes when compared to instruments deemed conceptually inaccurate.

None of the published TSE instruments¹⁶ fulfills all of the above-listed requirements. As argued earlier, the most commonly used subject-unspecific instruments (i.e., the Gibson and Dembo Scale and the TSES) do not fulfill requirement (e). The scale by Schwarzer et al. (1999) does not fulfill requirements (c) and (d), since their scale assesses a range of teachers’ tasks which are outside the sphere of what preservice teachers learn at university and would practice during a practicum, and it features a rather restricted response format of four categories. Similarly, the scale by Skaalvik and Skaalvik (2007) does not fulfill requirement (c), as two of the dimensions assessed (i.e., cooperating with colleagues and parents, and coping with changes and challenges) are outside the realm of preservice teachers’ university preparation and their tasks during practical experiences. Emmer and Hickman’s (1991) scale does not fulfill requirement (e), as it is partly based on the Gibson and Dembo Scale, and some items tap into knowledge (e.g., Item 5, 9) rather than self-efficacy. There is another TSE scale in Germany, published by Schulte, Bögeholz, and Watermann (2008). This scale though does not fulfill requirement (b), as it is based on the KMK (2004) standards for teacher

¹⁶ While great care was taken to identify all TSE scales (published in peer-reviewed journals) found by searching various databases (e.g., PsycINFO, ScienceDirect, SpringerLink), some TSE scales might nevertheless have been overlooked.

education and is thus too specific to the German context. Furthermore, some of its items also tap into knowledge (e.g., Item 4, 8, 9) rather than self-efficacy.

As there is no existing measure that qualifies for use in the proposed study, a new measure would have to be developed. But there is no need to reinvent the wheel. The TSES (Tschannen-Moran & Woolfolk Hoy, 2001) fulfills many of the requirements outlined above. Its three-dimensional approach with a focus on the teaching task of teachers bears the advantage of international comparability. While I agree with Skaalvik and Skaalviks' (2007, 2008) argument that teachers face a multitude of tasks, some of them outside the classroom, these tasks might vary considerably across nations (OECD, 2005). Teaching, however, is surely a core task that every teacher around the world would perform. And the TSES has, in fact, been adopted as the instrument most used for assessing TSE (Duffin et al., 2012), and has been found to be fit-for-purpose in diverse cultural settings (Klassen et al., 2009). But would the instrument also be appropriate to the German context? The three-dimensional distinction of the TSES shows parallels with the three-dimensional distinction of pedagogical knowledge that Baumert and Kunter (2006) conceptualized in their COACTIV model. These are areas of professional knowledge that can be expected to be taught at teacher preparation in German universities. Similarly, Ophardt and Thiel (2013) also describe these three dimensions (i.e., constructing knowledge = instructional strategies, motivation = student engagement, and classroom management) as core dimensions of teaching. The TSES has also often been used with preservice teachers, albeit without showing the three-dimensional structure for beginning preservice teachers (e.g., Duffin et al., 2012). Since these three dimensions play a role in teacher education, and particularly classroom management has been found to be of great concern to preservice teachers (e.g. Veenman, 1984), the three-dimensional assessment would be useful to pinpoint opportunities for further development or focus of teacher education programs. Furthermore, the items of the TSES might also be cast at an appropriate level of specificity, which would enable detecting change in TSE, as previous studies (e.g., Woolfolk Hoy & Burke Spero, 2005) have reported TSE changes using the TSES.

Thus, for the purpose of this study, the TSES was adapted to create greater coherence with Bandura's (1997, 2006) conceptualization of self-efficacy and to provide a three-dimensional structure when used with beginning preservice teachers. The instrument development process, as well as first validity tests and their results are outlined in Chapter 2.

But why should we even care how to measure TSE correctly? As mentioned before, a conceptually sound measurement will entail greater predictive validity of TSE. The next

section provides an overview of the many positive outcomes for which the predictive validity of TSE has been demonstrated.

1.4 Why does teacher self-efficacy matter?

TSE matters because of its meaningful associations with numerous positive outcomes for students and the teachers themselves, even during teacher training. This section provides an overview of the many results from research on the correlates of TSE, while not trying to be exhaustive, as this would go beyond the scope of this introduction. This section is rather intended to give an indication of the breadth of research that is available on this topic.

Outcomes for inservice teachers

A large body of research on the positive correlates of TSE for teachers has accumulated over the last four decades. Typically enquiry has focused on outcomes such as aspects of teacher motivation and teaching performance, stress and burnout, as well as job satisfaction and attrition from the teaching profession.

With regard to motivational aspects, positive outcomes of TSE are wide-ranging. Previous research has found, for example, that teachers with higher TSE, in comparison to teachers with lower TSE, show a reduced likelihood of referring problematic students to special needs education (Soodak & Podell, 1996), try different ways of teaching and are more prepared (e.g., Allinder, 1994), are more willing to implement new instructional practices following professional development (Guskey, 1988), exhibit greater extra-role behavior (Somech & Drach-Zahavy, 2000), and are more persistent in the face of incorrect answers given by students (Gibson & Dembo, 1984).

Regarding teaching performance, studies that use rigorous external measures rather than self-report are relatively scarce. In a recent meta-analysis of 41 studies published between 1985 and 2013, comprising over 9000 participants, Klassen and Tze (2014) found only nine studies in total that had reported teaching performance using external measures. After excluding three studies due to independence violations¹⁷, the authors determined a moderate association between TSE and teaching performance as rated either by teachers' students, supervisors, or principals (effect size $r = .28$). Interestingly, this compared favorably to a much smaller effect of $r = .08$ for personality variables of teachers (based on ten studies). While this meta-analysis gives a good indication of what can be expected for the average

¹⁷ The meta-analysis determined effects of TSE and personality on teaching effectiveness. Studies reporting effects for both factors (TSE and personality) were excluded from the overall effect size calculation for each factor.

association between TSE and teaching performance, it does not give any indication of the directionality of this relationship, as mostly research has been conducted on the basis of cross-sectional data. Recent evidence from a longitudinal study by Holzberger, Philipp, and Kunter (2013) points to a reverse relationship between TSE and the instructional quality of teachers, whereby students' ratings of cognitive activation, and students' and teachers' ratings of classroom management predicted teachers' subsequent TSE levels, instead of the other way around.

Several studies have examined TSE as a protective factor against stress and burnout. A predictive relationship between TSE and burnout is well documented. Brisse, von Hoover-Dempsey, and Bassler (1988) were the first to establish a link between TSE and burnout in teachers. However, the authors offered little information on the nature and psychometric properties of their TSE measure, so it is unclear how validly TSE was assessed. In 2000, Brouwers and Tomic demonstrated that TSE in classroom management predicted depersonalization (i.e., one factor in burnout) over a period of five months. Using a cross-lagged panel design, Schwarzer and Hallum (2008) could demonstrate that lower TSE levels predicted higher burnout levels (comprising both factors: emotional exhaustion and depersonalization) a year later, while earlier burnout levels did not affect later TSE levels. Furthermore, providing an explanatory route, this effect was mediated by job stress, and tended to be stronger for younger teachers. A recent review by Beltman, Mansfield, and Price (2011), focusing on resilience in early career teachers, determined TSE to be a key individual protective factor. Dicke et al. (2015) reported that low TSE predicted emotional exhaustion in early career teachers, which was mediated by classroom disturbances.

Another substantial amount of research has centered on teachers' job satisfaction and occupational commitment, as well as attrition from the teaching profession. Klassen and Chiu (2011), for example, found that TSE for instructional strategies was positively linked to occupational commitment, which was in turn negatively linked to teachers' intention to quit teaching. Some twenty years earlier, Coladarci (1992) had already demonstrated a strong link between TSE and commitment to teaching, albeit using a different measure of TSE (i.e., the Gibson and Dembo scale), more in accordance with locus-of-control theory. Most recently, Chesnut and Burley (2015) reported an overall effect size of $r = .32$ in their meta-analysis of the relationship between TSE and commitment to the teaching profession, which increases to $.35$ when only accurate measures of TSE are taken into account. Regarding job satisfaction, Caprara, Barbaranelli, Borgogni, and Steca (2003) identified TSE and collective efficacy beliefs of teachers as the main determinants, among a range of other factors that also

influenced job satisfaction (i.e., the principal, teacher colleagues, other school staff, students, and parents). Other researchers (e.g., Caprara, Steca, Barbaranelli, & Malone, 2006; Collie, Shapka, & Perry, 2012; Klassen & Chiu, 2010) have consistently found small to medium-sized effects between TSE and job satisfaction.

Outcomes for preservice teachers

Research on the outcomes associated with TSE for preservice teachers is a lot more limited in quantity. Most research on the topic of TSE with regard to preservice teachers has focused on factors that might influence TSE rather than examining outcomes of TSE. Nonetheless, there are a few studies that explored outcomes associated with preservice teachers' TSE. Fives, Hamman, and Olivarez (2007), for instance, documented a link between TSE and burnout for preservice teachers. The aforementioned study by Klassen and Chiu (2011) found that for preservice teachers TSE in classroom management, in contrast to instructional strategies for inservice teachers, was linked to occupational commitment, which was also in turn negatively linked to preservice teachers' intention to quit the teaching profession. In addition, the meta-analysis by Chesnut and Burley (2015) found no differences in effect sizes between preservice and inservice teachers, indicating that the relationship between TSE and commitment to the teaching profession is already as meaningful during teacher education as it is later on during service. In contrast to the positive effects on teaching performance found for inservice teachers, Jamil, Downer, and Pianta (2012) found no significant relationships between preservice teachers' TSE and a classroom assessment by university supervisors. Possibly, the mechanism by which TSE affects teaching performance might only operate later on, when teachers have finished their training. Furthermore, Schulte et al. (2008) did not find any relationship between professional knowledge (objective test) and TSE.

Outcomes for students

Following Armor et al.'s (1976) seminal results, documenting a positive effect of teacher efficacy on students' advances in reading achievement between the 5th and 6th grade, there have been a multitude of studies reporting linkages between TSE and student outcomes. However, in a review of TSE research between 1998 and 2009, Klassen et al. (2011) only identified two relevant studies¹⁸ and thus came to the conclusion that more evidence in favor of such links was needed. In contrast, most recently, Klassen and Tze (2014) conducted a much more comprehensive meta-analysis, identifying a respectable total of 24 studies conducted between 1985 and 2013, linking TSE and student achievement. Those studies

¹⁸ This low number might have resulted from their rather limited inclusion criteria.

represented a wide range of students' achievement data comprising different grades and subjects (e.g., Math, English, Sciences, Social Studies). Results of the meta-analysis showed an overall, albeit, small association between TSE and student achievement (effect size $r = .08$).

In addition to students' achievement, research has also found relations between students' motivation and their (academic) self-efficacy. For example, Midgley, Feldlaufer, and Eccles (1989) found TSE of mathematics teachers to be positively related to students' self- and task-related beliefs in math. Anderson, Greene, and Loewen (1988) showed that teachers' TSE beliefs were related to their students' academic self-efficacy beliefs. Ross, Hogaboam-Gray, and Hannay (2001) also showed a relationship between TSE and students' computer self-efficacy.

Concluding remarks

Since the dawn of TSE research in 1976, researchers have accumulated findings in favor of its importance regarding a wide range of positive outcomes. Unfortunately, this research has often been hampered by measurement issues and cross-sectional designs. Thus, there is little evidence with regard to the causal direction of effects. Measurement issues plague especially older results, which were derived using various instruments based on Rotter's rather than Bandura's theory. This attenuates the relationships between TSE and the positive outcomes (see Chesnut & Burley, 2015) that have been found for preservice and inservice teachers, as well as for their students. At any rate, the existing results generally show weaker positive effects at the student level (especially for achievement), and stronger effects at the teacher level. Despite the above-mentioned methodological drawbacks, there seems to be enough evidence to support the notion of many researchers (e.g., Klassen et al., 2009; Tschannen-Moran et al., 1998; Vieluf, Kunter, & van de Vijver, 2013) that TSE is one of the influential teacher characteristics in predicting positive outcomes for teachers and students alike. Thus, the endeavor of studying its development, which is the focus in Chapter 3 and 4, seems justified. The next section offers an insight into theoretical underpinnings as well as empirical findings regarding the development of TSE.

1.5 How does teacher self-efficacy develop?

Early on, when Bandura introduced his construct of self-efficacy in 1977, he already theorized that self-efficacy beliefs developed based on four principal sources of information. Later, Bandura (1997) elaborated on his theory of the sources in greater detail and put forward a collection of evidence in favor of their effects in a variety of self-efficacy domains, yet not

TSE. However, in the absence of a comprehensive theory on how specifically TSE develops, Bandura's theory of the sources of self-efficacy provide a sound theoretical basis. Therefore, in this section, first Bandura's (1997) account of the sources is introduced and described, and second, empirical findings relating to the development of TSE are briefly discussed.

Bandura's sources of self-efficacy¹⁹

According to Bandura's theory of the sources, self-efficacy beliefs are formed through the interpretation and integration of information from four channels: mastery experiences, vicarious experiences, verbal persuasion, and physiological and affective states. He stresses that cognitive processing of the information from each source is paramount to the formation of self-efficacy beliefs. There are two main mechanisms by which such cognitive processing operates: (1) selecting which type of information one attends to and thus uses as a source of self-efficacy, and (2) weighting and integrating the different pieces of information. Bandura further postulates the sources to be differentially effectual with regard to the development of self-efficacy. He considers enactive mastery experiences to be generally the most influential of the four sources, because they provide the most authentic information of people's capabilities.

Mastery experiences provide information about one's past successes or failures, and thus constitute indicators of people's capability. While successes raise self-efficacy beliefs, failures lower them. Failures are especially detrimental to self-efficacy if they occur at an early stage of self-efficacy formation, when beliefs are still in flux and have not yet been firmly established. During the cognitive processing stage, there are a host of factors, which govern the influence of mastery experiences on self-efficacy development. Bandura counts (a) preexisting self-knowledge, (b) task difficulty, (c) effort expenditure, (d) selective recall of experiences, and (e) attainment trajectories as influential factors at this stage. Each factor is briefly described in the following paragraphs.

(a) Pre-existing self-knowledge can act as a confirmatory bias in the interpretation of mastery experiences, whereby information that is incongruent with preexisting beliefs is discarded. Thus, if one already possesses strong self-efficacy beliefs, failures are more likely to be discounted and therefore do not affect self-efficacy beliefs. If one holds a strong belief in one's inefficacy on the other hand, successes tend to be discounted as products of "laborious effort" (Bandura, 1997, p. 82) rather than indicators of capability.

¹⁹ This entire section is based on one reference: Bandura (1997).

(b) Appraisal of task difficulty also influences the effect that mastery experiences exert over changes in self-efficacy. If task difficulty is perceived to be low, mastery is less likely to lead to a change in self-efficacy beliefs, as this carries no new information about one's capabilities. Conversely, mastering a task that is perceived to be difficult raises self-efficacy beliefs.

(c) How much effort people invest in task performances, in combination with task difficulty, will also determine the effect mastery experiences have on the formation of self-efficacy beliefs. For example, if task difficulty is low and the invested effort is high, then failure would have a disastrous effect on self-efficacy judgements. If the invested effort was also low on the other hand, effects on self-efficacy would be minimal. Successful performance on difficult tasks that was achieved with relatively little effort, would raise self-efficacy, while the same performance resulting from high effort and struggle, is less likely to increase beliefs in one's efficacy and might even lower them.

(d) Selectively recalling poor past performances is likely to lead to an underestimation of one's efficacy. In turn, self-efficacy can be enhanced by a selective focus on successful performances.

(e) Attainment trajectories especially play a role when the behavior that is appraised in terms of self-efficacy is complex and must thus be developed over a rather long period of time, as is the case with teaching for example. If people succeed consistently in their attainments, but then plateau, they might interpret this temporary leveling off in their prior rate of improvement as the limit of their capabilities. In contrast, people whose performance improvement over time is interrupted by failures are more likely to raise their efficacy beliefs, as they invest the time and effort needed to keep improving.

Vicarious experiences provide information about the attainments of others. This source will be particularly influential, when one has little experience with the capability to be judged. Vicarious experiences influence the formation of self-efficacy beliefs mainly through two mechanisms: (1) model learning, and (2) social comparison.

(1) Watching the performance of others conveys competencies, which actually aids people's skill development. It is particularly conducive to the model learning aspect of vicarious experience, if:

- (a) many different models are observed, since this allows for observing diverse skills,
- (b) models are observed overcoming difficulties, as they show effective strategies in overcoming obstacles and perseverance,

- (c) models are competent in the concerned field, because their performance demonstrates successful skills and strategies.

(2) Social comparison is particularly relevant to the judgment of self-efficacy, when the level of attainment cannot be easily judged, as would be the case with teaching. It would not be immediately clear to a preservice teacher in their practicum for example, if their teaching performance was successful, as they are inexperienced and there are few objective indicators. Comparing their own teaching performance with that of their fellow preservice teachers on the other hand, would help them gauge their own level of performance. Furthermore, social comparison have a greater impact on people's efficacy appraisals, if the person they compare themselves to, is perceived to be similar to them, as would be the case with the fellow preservice teacher.

Verbal persuasion, that is being persuaded by somebody that one has what it takes to succeed, can help increase people's self-efficacy. As with vicarious experiences, verbal persuasion carries more weight when one does not have the relevant knowledge or experience to appraise one's own level of ability. Again, Bandura postulates a range of variables that moderate or mediate the influence of verbal persuasion on self-efficacy: (a) the perceived credibility and knowledgeable-ness of the persuader, (b) the way the performance feedback is framed, and (c) the degree of disparity between one's own and the persuader's appraisal.

(a) The greater the person who is persuaded perceives the credibility and knowledgeable-ness of their persuader to be, then the greater will be the impact of the persuasion on their efficacy judgment. Persuaders will be perceived as credible and knowledgeable if they:

- are themselves competent in the activity to be judged,
- are experienced in judging many different aspirants,
- possess realistic knowledge regarding the task-related demands that aspirants will face.

(b) Performance feedback that highlights the progress someone has made increases self-efficacy beliefs, whereas feedback highlighting shortfalls from a certain standard undermines self-efficacy. Further, attributional feedback also plays a role. Generally, phrasing feedback with regard to ability rather than effort, increases efficacy beliefs to a greater degree.

(c) Performance appraisals by others need to be within realistic boundaries. Otherwise repeated failures will ensue, which would undermine one's self-efficacy beliefs as well as the credibility of the persuader. Bandura puts the optimal level of discrepancy at moderately beyond what the appraised person is currently able to do.

Physiological and affective states can also provide information that people use in judging their capabilities in a given domain. Such somatic indicators are particularly relevant, when the domain of functioning includes stressful or taxing situations, as would be the case for teaching. In these situations, people tend to interpret their physiological and affective responses as a sign of dysfunction. This can lead to a vicious cycle, whereby people who are experiencing high levels of arousal expect this to impact negatively on their performance, which in turn increases their level of distress, which then actually impacts negatively on their performance. As with the other three sources, the strength of the effect that physiological and affective states have on self-efficacy varies with several factors at the cognitive processing stage.

(a) The attribution of the perceived source of somatic activation plays an important role. If physiological or affective states are externally, rather than internally, attributed, they are of little influence with respect to self-efficacy. For example, if a preservice teacher experiences sweating, an increased heart rate, and wobbly knees in front of the class during their first teaching attempts, ascribing these somatic indicators to the three coffees they drank earlier in the day, would not affect their efficacy beliefs at all. If, on the other hand, the preservice teacher in this situation was to read these somatic indicators as a sign of coping deficiencies, the effect on self-efficacy would be negative.

(b) The level of emotional and physical activation is another factor that influences the impact of physiological and affective states on self-efficacy. Bandura recommends generally a moderate level of activation as being conducive to self-efficacy, since that affects attentiveness and the application of skills positively, while high levels of activation impact negatively on functioning.

(c) Construal biases also act as a significant factor with regard to the interpretation of somatic information. In Bandura's words: "...the problem is not the arousal per se but the view one takes of it." (1997, p. 109). According to Bandura, a somatic indicator in itself does not carry much information with regard to its meaning. Construal biases, for example, preexisting self-efficacy beliefs, vary from individual to individual and can act as a memory (e.g., selective recall), interpretive (e.g., arousal = challenge, or arousal = panic), and attentional (e.g., heightened sensitivity) bias when processing information stemming from physiological and affective states.

(d) Mood states can also impact perceived self-efficacy, whereby a positive mood elevates people's judgments of self-efficacy, while a negative mood lowers them. This impact of mood

operates in part through biases with regard to the interpretation and recall of past mastery experiences.

So, there are four principal sources of information, which get channeled through a multitude of cognitive processes, before an individual arrives at an efficacy judgment. But how do people integrate the information they receive into one coherent judgment of self-efficacy? According to Bandura, individuals assign weights to each type of efficacy information they receive. First and foremost, these weights differ with the domain of functioning. However, typically the weights people assign to different factors that carry efficacy information are not necessarily an accurate representation of the true efficacy value of that factor. People's descriptions of factors they think influenced their judgment, often put more emphasis on the less important factors and less emphasis on more important factors. This is due to the complex nature of the integration task, which is typically managed by using only a few salient factors and simplified heuristics for integrating them.

Empirical findings

Research regarding the sources of TSE is reviewed in detail in section 4.1 in Chapter 4. Research regarding changes in TSE is reviewed in detail in section 3.1 in Chapter 3. In order to minimize redundancies, this subsection presents only a high-level summary of research in this area. While there are empirical findings regarding other factors that might affect TSE, these factors fall outside the scope of this introduction, since they do not immediately pertain to the purpose of this thesis.

Overall, there is some research on changes in TSE, and very little research specifically on the sources of TSE. Since Bandura (e.g., 1997) emphasized the mastery experience source as being the most influential in the development of self-efficacy, nearly all research on changes in TSE has focused on changes that take place during some sort of practical phase of teacher education. Commonly this practical phase constitutes a practicum of varying lengths, situated in school, which involves first teaching experiences on the part of the preservice teacher. It is sometimes mentioned that these teaching experiences are supervised to some degree by a mentor teacher at the school. Findings from this line of enquiry have produced mixed results, whereby an increase in TSE was more often reported than decreases or no changes. However, in most cases the methodological quality of the studies was compromised in one way or another.

First and foremost, the majority of studies did not actually examine a change in TSE, as TSE pre-post scores were not analyzed at the intra-individual but rather at the group level (in Garvis, Pendergast, & Keogh, 2012; Hoy & Woolfolk, 1990; Liaw, 2009; Lin & Gorrell,

2001; Pendergast, Garvis, & Keogh, 2011; Swan, Wolf, & Cano, 2011). Mean group differences between two measurement occasions do not tell us anything about changes within individuals (e.g., Singer & Willett, 2003). As McArdle (2009) illustrates, there are a number of individual longitudinal change patterns that can explain the same cross-sectional pattern of data points across two measurement occasions. Thus, in fact, these studies do not tell us anything about TSE changes in preservice teachers.

Second, nearly all studies used manifest data analysis approaches. A recent study by Klassen and Durksen (2014), documenting increases in TSE, is the exception. All other identified studies analyzing changes in the TSE of preservice teachers (Fives et al., 2007; Knoblauch & Woolfolk Hoy, 2008; Mergler & Tangen, 2010; Woolfolk Hoy & Burke Spero, 2005) analyzed manifest data. Prominent researchers (e.g., Cronbach & Furby, 1970; McArdle, 2009; Singer & Willett, 2003) have highlighted serious methodological drawbacks associated with a manifest analysis of longitudinal data. One major issue of manifest analyses is that the true change in the underlying construct one is interested in studying is confounded with changes in the calibration of the measurement scale by participants, and changes in the conceptualization of the target construct by participants (e.g., Millsap & Hartog, 1988). In latent analysis, this issue is addressed in a test of longitudinal measurement invariance (e.g., Meade & Bauer, 2007), which serves as a prerequisite to any latent analysis of change (Steyer, Eid, & Schwenkmezger, 1997).

Third, in several studies (Hoy & Woolfolk, 1990; Liaw, 2009; Lin & Gorrell, 2001) TSE was assessed using the Gibson and Dembo scale, which (as argued in section 1.3 of this chapter) cannot be considered a valid measurement of TSE. Nearly all other studies (Fives et al., 2007; Garvis et al., 2012; Klassen & Durksen, 2014; Knoblauch & Woolfolk Hoy, 2008; Mergler & Tangen, 2010; Pendergast et al., 2011; Swan et al., 2011; Woolfolk Hoy & Burke Spero, 2005) employed the TSES (Tschannen-Moran & Woolfolk Hoy, 2001), which, as argued earlier (in section 1.3 of this chapter), is also lacking in validity. Further, only two studies (Fives et al., 2007; Pendergast et al., 2011) offered an examination of the three dimensions of TSE. This is most likely due to the confusion regarding the factor structure of the TSES when used with preservice teachers.

In addition to these methodological drawbacks, previous research on the changes in TSE in preservice teachers is limited in scope. Only one study (Garvis et al., 2012) examined preservice teachers early on in their preparation, and also only one study (Mergler & Tangen, 2010) examined changes in TSE that occurred during a course at university, rather than during a practicum at school.

What little research there is on the sources of TSE is characterized by a lack of instruments assessing the four sources. Only one instrument assessing sources of TSE has been published (Poulou, 2007). However, this instrument was derived using an inductive approach, which means there is little overlap with Bandura's (1997) description of the sources. It combines two sources into one for example. There are also a few studies which have reported findings on single sources using proxy indicators instead of validated measures (Moulding, Stewart, & Dunmeyer, 2014; Ruble, Usher, & McGrew, 2011; Tschannen-Moran & Woolfolk Hoy, 2007; Woolfolk Hoy & Burke Spero, 2005). In general, findings have been inconsistent and patchy, with no single study addressing all four sources. Furthermore, with the exception of one study by Woolfolk Hoy and Burke Spero (2005), no study has related the assessed sources to *changes* in TSE. All other studies predicted static levels of TSE, which may or may not be related to the sources. The study presented in Chapter 4 addresses these previous shortcomings in this area of TSE research by providing an assessment of all four sources, which are modeled to predict changes in TSE.

In the next section, I outline the specific aims of each study in this thesis and present the overall study design.

1.6 Overview of the thesis and research aims

In order to address the overarching question of how TSE develops during teacher preparation at university, a two-cohort longitudinal study with three waves of data collection was designed. A two-cohort design was chosen in order to gain first insights into the development of TSE at two different stages (i.e., beginning and advanced) of teacher preparation, while employing a relatively economical approach to data collection (i.e., a timeframe of one year instead of four to five years). Figure 6 provides an overview of the three waves of data collection and the different samples. Table 1 shows the basic study characteristics of the three studies presented in Chapter 2, 3, and 4 of this thesis. In the last chapter the results of each study are summarized and discussed with regard to their implications for theory and practice. Limitations and strengths of this thesis are also outlined. The last chapter finishes with recommendations for further research in this area.

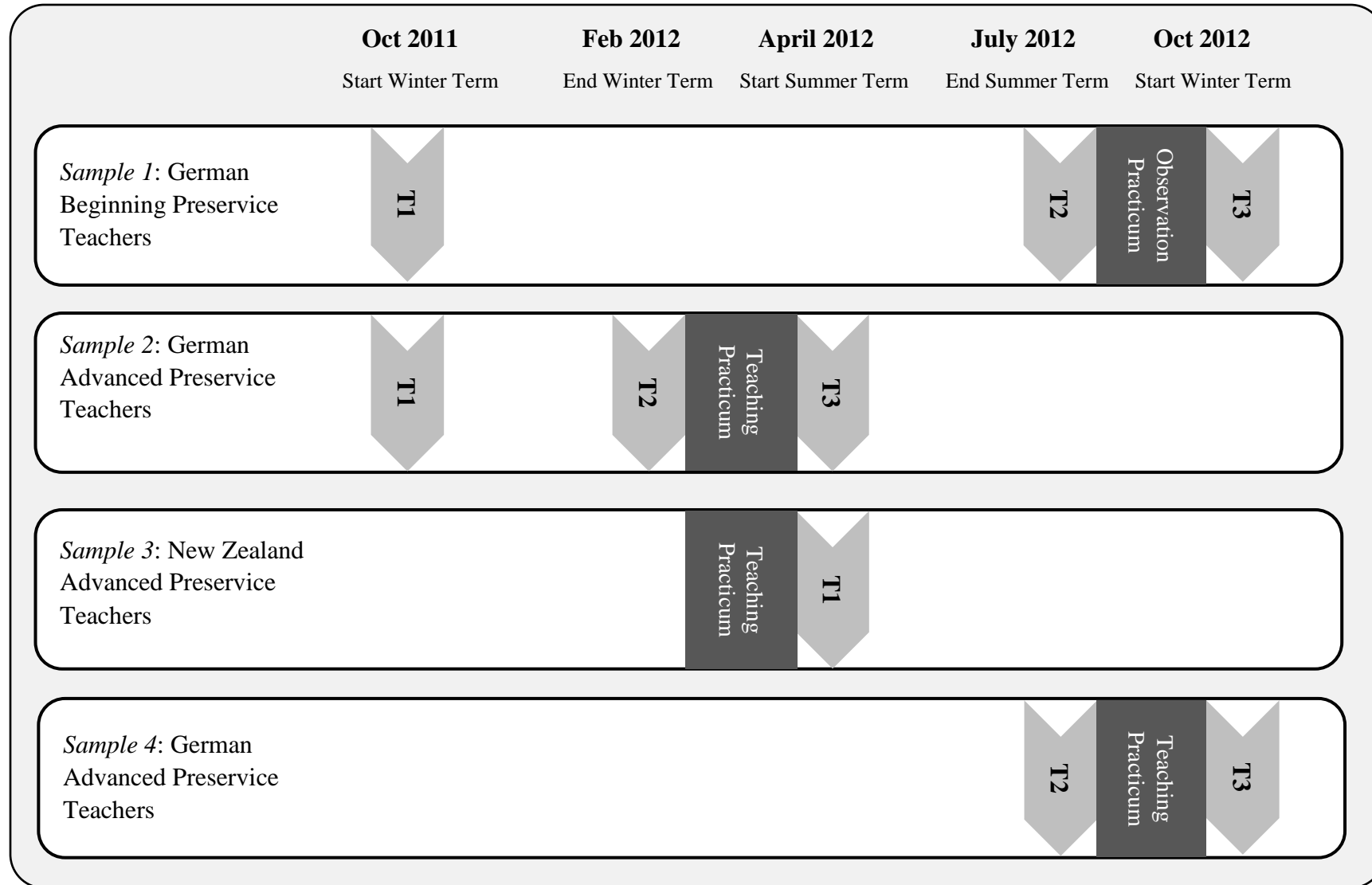


Figure 6. Overview of the multiple waves of data collection

Table 1: Overview of study characteristics

Chapter	Wave	Sample	N	Mode of Assessment	Assessed Variables of Interest
2 (Study 1)	T1	1	438	Paper-pencil	TSE
		2	282		GSE
		3	131		Study Satisfaction (only in Sample 1 and 2) 4 Sources of TSE (piloted in Sample 2 and 3) ²⁰
3 (Study 2)	T1, T2, T3	1	438 (T1)	Paper-pencil (T1 and T2)	TSE
		1	359 (T2)		
		1	226 (T3)	Online (T3)	ITQ
		2	274 (T1)		
		2	255 (T2)		
2	143 (T3)				
4 (Study 3)	T2, T3	1	359 (T2)	Paper-pencil (T2)	TSE
		1	226 (T3)		
		2	255 (T2)	Online (T3)	4 Sources of TSE
		2	143 (T3)		
		4	140 (T2)		
4	80 (T3)				

Note: TSE = teacher self-efficacy; GSE = general self-efficacy; ITQ = intention to quit the teaching degree.

²⁰ The results of the pilot study of the self-developed scales for assessing the four sources of TSE are not reported in Study 1 or any of the other studies. The pilot study is only briefly mentioned in Study 3.

Chapter 2 – Study 1

As outlined in section 1.3, there were no suitable TSE measures for the purpose of this study, but the TSES provided a good starting point. Consequently, it was a first aim of this thesis to adapt the TSES to ensure a valid and reliable three-dimensional assessment of TSE for preservice teachers at an early and an advanced stage of teacher preparation. This process is documented in Chapter 2. In order to gather first evidence of construct validity across different stages of teacher education, and in the German as well as the English-speaking context, measurement invariance was tested for three diverse samples of preservice teachers. Furthermore, the convergent validity with general self-efficacy and study satisfaction was tested. The following criteria were expected to be fulfilled in order to deem the adapted measure of TSE suitable for the longitudinal study:

- acceptable psychometric properties (internal consistency, factor loadings, interfactor correlations between the three TSE dimensions, construct reliability of the three latent TSE factors)
- metric level invariance for all three samples
- medium positive correlation with general self-efficacy
- positive correlation with study satisfaction.

Study 1 was also used to conduct a pilot test of a self-developed measure for assessing the four sources of TSE. This pilot study was conducted with the German and the New Zealand sample of advanced students, because the source measure is tailored to a practicum experience, which both samples had previously undergone.

Chapter 3 – Study 2

In section 1.4, it was demonstrated that TSE matters with regard to a variety of positive outcomes for preservice and inservice teachers, as well as students. As mentioned in the beginning of the introduction, burnout and attrition from the teaching career are a serious issue in this profession. While there are longitudinal studies that have linked changes in burnout and TSE, a longitudinal link between TSE and attrition of preservice teachers is missing. Furthermore, as shown in section 1.5, previous research on the development of TSE in preservice teachers has been hampered by methodological issues and an incomprehensive approach to changes that occur during teacher education. Addressing both gaps in previous research, the purpose of Study 2 was twofold.

First, changes in each dimension of TSE were systematically documented for two cohorts of preservice teachers at an early and an advanced stage of university teacher

preparation. Data was collected at three times, which were purposefully selected taking into account prior research and Bandura's (e.g., 1997) theory on the development of self-efficacy. Specifically, data was collected at the beginning of the university term, before and after a practicum at school. This choice of measurement occasions allowed for determining TSE changes that occurred during the theoretical phase at university and during the practical phase at school. Since previous research has heavily concentrated on observing TSE changes in advanced preservice teachers that occur during the practicum, this design should contribute to a better understanding of how TSE develops during teacher preparation.

Second, it was examined whether changes that occurred either during the term or during the practicum in each cohort impacted on changes in preservice teachers' intention to quit their teaching degree. Specifically, in Study 2 it was examined whether:

- on average changes occurred in each dimension of TSE during the term and during the practicum.
- changes in TSE and preservice teachers' intention to quit were dependent on their respective baseline levels.
- TSE and preservice teachers' intention to quit changed similarly across individuals.
- changes in each dimension of TSE predicted changes in preservice teachers' intention to quit.

Addressing methodological shortcomings of previous research, this study used advanced statistical analysis techniques. Specifically, latent true change modeling (e.g., Steyer et al., 1997) was employed. This technique allows for a measurement-error free estimation of change and has the advantage that the basic measurement assumptions necessary for the interpretation of longitudinal changes can be tested.

Chapter 4 – Study 3

Several reviewers (e.g., Henson, 2002; Klassen et al., 2011) have highlighted the need for research on the development of TSE with a particular focus on the sources. Study 3 was designed to address this call. As outlined in section 1.5, empirical results from quantitative studies are virtually non-existent, mostly due to a lack of measurement instruments for the four sources proposed by Bandura (e.g., 1997). There is some qualitative research on the sources of TSE (as detailed in section 4.1 in Chapter 4). However, as noted in the first part of section 1.5, because people tend to misjudge the relative importance of each source of efficacy, qualitative research into the weights of sources is of limited value. Thus, drawing on

a thorough analysis of Bandura's (e.g., 1997) description of the sources of self-efficacy (presented in section 1.5), an instrument assessing the four sources specifically for TSE was developed as part of this thesis. Instrument development included a pilot test of the newly developed scales, which was conducted at T1 in Sample 2 and 3 (see Table 1 and Figure 6).

The purpose of Study 3 was to gather first evidence of the psychometric properties and validity of the newly developed instrument. Simultaneously, it was explored how the four sources of TSE could interact in predicting TSE changes. Specifically, the source measure was examined with regard to the following reliability and validity criteria:

- good psychometric properties (internal consistency, factor loadings, interfactor correlations between the sources, construct reliability of the source factors)
- good fit of the confirmatory factor analysis reproducing the sources
- significant positive bivariate correlations between the mastery experience, vicarious experience, verbal persuasion sources and TSE changes, and negative bivariate correlations between the physiological and affective states (PAS) source and TSE changes
- among the four sources, mastery experiences show the strongest (i.e., bivariate) relationship to TSE changes
- a greater impact of mastery experiences on TSE changes during the teaching practicum (i.e., in Sample 2 and 4) as compared to the observation practicum (i.e., in Sample 1)
- a greater impact of vicarious experiences on TSE changes during the observation practicum in comparison to the teaching practicum.

Partly drawing on previous research and on Bandura's (e.g., 1997) (lack of) guidance with regard to the relationship between the sources (as detailed in section 4.1 in Chapter 4), three plausible models are proposed and tested in Study 3. The first model is the simplest model, in which all sources are directly predicting TSE change. In the second model, all sources predict TSE changes directly, with the exception of verbal persuasion, whose influence on TSE changes is mediated by mastery experiences. The third model is the most parsimonious model, in which only mastery experiences directly predict TSE changes, and the influence of all other sources on TSE changes is mediated by mastery experiences.

An overview of the relationships between constructs of interest examined in this thesis is depicted in Figure 7.

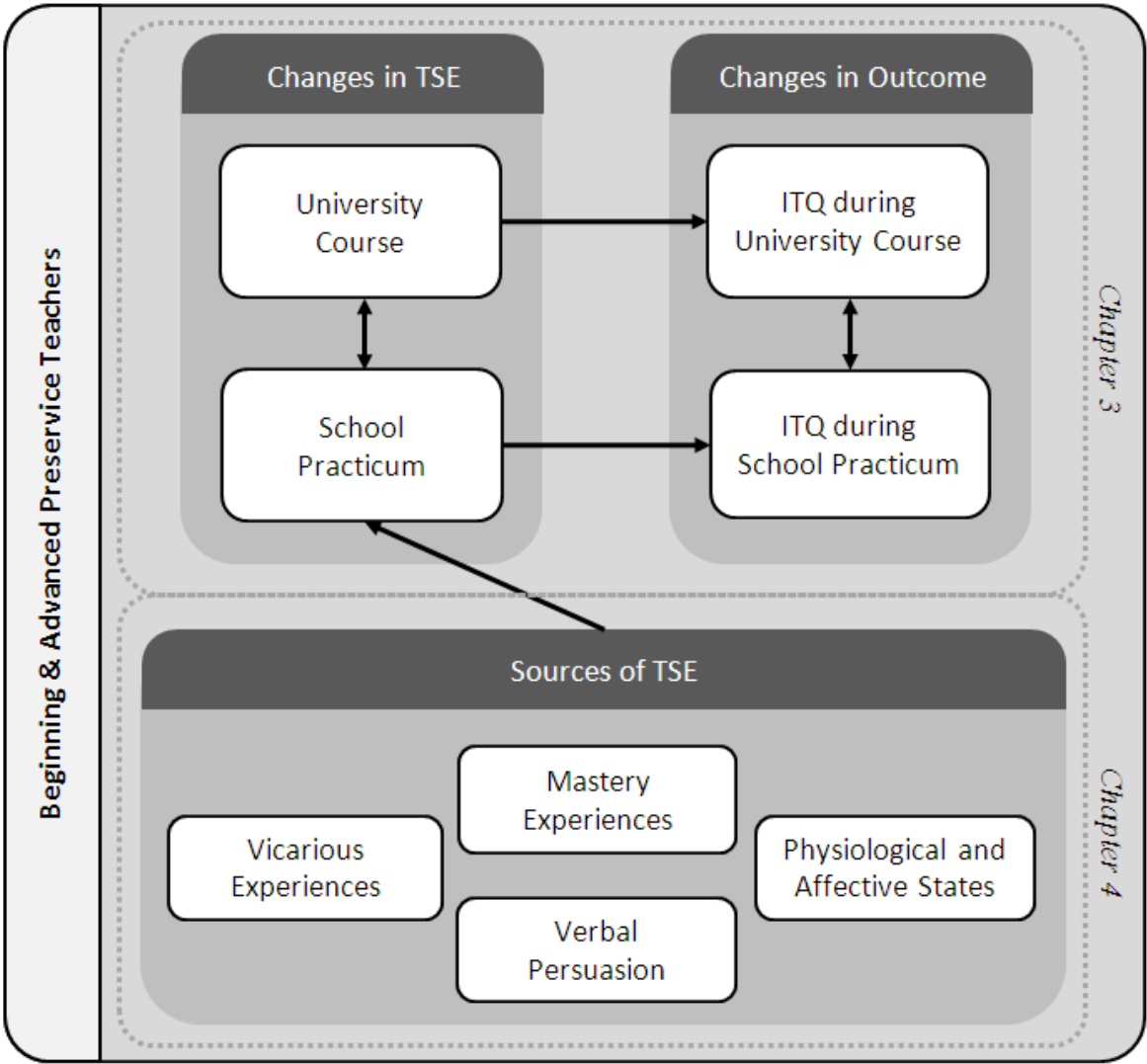


Figure 7. Overview of the examined relationships between study variables
Note: ITQ = Intention to Quit the teaching degree

1.7 References

- Allinder, R. M. (1994). The relationship between efficacy and the instructional practices of special education teachers and consultants. *Teacher Education and Special Education*, 17(2), 86–95.
- Anderson, R. N., Greene, M. L., & Loewen, P. S. (1988). Relationships among teachers' and students' thinking skills, sense of efficacy, and student. *The Alberta Journal of Educational Research*, 34(2), 148–165.
- Armor, D. J., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., . . . Zellman, G. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools* (Report No. R-2007-LAUDS). Santa Monica, CA.
- Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement*. New York, NY: Longman.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1–26.
- Bandura, A. (2006). Guide for creating self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich: Information Age Publishing.
- Bandura, A., & Locke, E. A. (2003). Negative self-efficacy and goal effects revisited. *Journal of Applied Psychology*, 88(1), 87–99.
- Bauer, J., Drechsel, B., Retelsdorf, J., Sporer, T., Rösler, L., Prenzel, M., & Möller, J. (2010). Panel zum Lehramtsstudium – PaLea: Entwicklungsverläufe zukünftiger Lehrkräfte im Kontext der Reform der Lehrerbildung. *Beiträge zur Hochschulforschung*, 32(2), 34–55.
- Bauer, J., Stamm, A., Virnich, K., Wissing, K., Müller, U., Wirsching, M., & Schaarschmidt, U. (2006). Correlation between burnout syndrome and psychological and psychosomatic symptoms among teachers. *International archives of occupational and environmental health*, 79(3), 199–204. doi:10.1007/s00420-005-0050-y
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.

- Baumert, J., & Kunter, M. (2011). Das Kompetenzmodell von COACTIV. In M. Kunter, J. Baumert, W. Blum, U. Klusmann, S. Krauss, & M. Neubrand (Eds.), *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV* (pp. 29–53). Münster: Waxmann.
- Beltman, S., Mansfield, C., & Price, A. (2011). Thriving not just surviving: A review of research on teacher resilience. *Educational Research Review*, 6(3), 185–207.
doi:10.1016/j.edurev.2011.09.001
- Blömeke, S., Felberich, A., & Müller, C. (2008). Theoretischer Rahmen und Untersuchungsdesign. In S. Blömeke (Ed.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerbildung* (pp. 15–48). Münster: Waxmann.
- Bong, M. (2006). Asking the right question: How confident are you that you could successfully perform these tasks? In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 287–306). Greenwich: Information Age Publishing.
- Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1), 1–40.
- Bransford, J., Darling-Hammond, L., & LePage, P. (2005). Introduction. In L. Darling-Hammond & J. Bransford (Eds.), *The Jossey-Bass education series. Preparing teachers for a changing world. What teachers should learn and be able to do* (1st ed.). San Francisco, CA: Jossey-Bass.
- Brisse, J. S., von Hoover-Dempsey, K., & Bassler, O. C. (1988). Individual, situational contributors to teacher burnout. *Journal of Educational Research*, 82(2), 106–112.
- Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, 16, 239–253.
- Caprara, G. V., Barbaranelli, C., Borgogni, L., & Steca, P. (2003). Efficacy beliefs as determinants of teachers' job satisfaction. *Journal of Educational Psychology*, 95(4), 821–832.
- Caprara, G. V., Steca, P., Barbaranelli, C., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology*, 44(6), 473–490.
doi:10.1016/j.jsp.2006.09.001

- Charalambous, C. Y., Philippou, G. N., & Kyriakides, L. (2008). Tracing the development of preservice teachers' efficacy beliefs in teaching mathematics during fieldwork. *Educational Studies in Mathematics*, *67*(2), 125–142. doi:10.1007/s10649-007-9084-2
- Chesnut, S. R., & Burley, H. (2015). Self-efficacy as a predictor of commitment to the teaching profession: A meta-analysis. *Educational Research Review*, *15*, 1–16. doi:10.1016/j.edurev.2015.02.001
- Coladarci, T. (1992). Teachers' sense of efficacy and commitment to teaching. *The Journal of Experimental Education*, *60*(4), 323–337. Retrieved from <http://www.jstor.org/stable/20152340?origin=JSTOR-pdf&>
- Collie, R. J., Shapka, J. D., & Perry, N. E. (2012). School climate and social–emotional learning: Predicting teacher stress, job satisfaction, and teaching efficacy. *Journal of Educational Psychology*, *104*(4), 1189–1204.
- Cronbach, L. J., & Furby, L. (1970). How should we measure "change" - or should we? *Psychological Bulletin*, *74*(1), 68–80.
- Darling-Hammond, L. (2006). Constructing 21st-century teacher education. *Journal of Teacher Education*, *57*(3), 300–314. doi:10.1177/0022487105285962
- Darling-Hammond, L., & Bransford, J. (Eds.). (2005). The Jossey-Bass education series. Preparing teachers for a changing world: What teachers should learn and be able to do (1st ed.). San Francisco, CA: Jossey-Bass.
- Dicke, T., Parker, P. D., Holzberger, D., Kunina-Habenicht, O., Kunter, M., & Leutner, D. (2015). Beginning teachers' efficacy and emotional exhaustion: Latent changes, reciprocity, and the influence of professional knowledge. *Contemporary Educational Psychology*, *41*, 62–72. doi:10.1016/j.cedpsych.2014.11.003
- Duffin, L. C., French, B. F., & Patrick, H. (2012). The teachers' sense of efficacy scale: Confirming the factor structure with beginning pre-service teachers. *Teaching and Teacher Education*, *28*(6), 827–834.
- Duffy, H., Bransford, J., Hammerness, K., Darling-Hammond, L., & LePage, P. (Eds.). (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do* (1st ed.). San Francisco, CA: Jossey-Bass.
- Eagly, A. H., & Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition*, *25*(5), 582–602. doi:10.1521/soco.2007.25.5.582
- Emmer, E. T., & Hickman, J. (1991). Teacher efficacy in classroom management and discipline. *Educational and Psychological Measurement*, *51*, 755–765. Retrieved from <http://epm.sagepub.com/content/51/3/755>

- Enochs, L. G., Smith, P. L., & Huinker, D. (2000). Establishing factorial validity of the mathematics teaching efficacy beliefs instrument. *School Science and Mathematics, 100*(4), 194–202.
- European Commission. (2012). *Key data on education in Europe 2012*. Brussels.
- Fives, H., & Gill, M. G. (Eds.). (2015). Educational psychology handbook series. International handbook of research on teachers' beliefs. New York, NY: Routledge.
- Fives, H., Hamman, D., & Olivarez, A. (2007). Does burnout begin with student-teaching? Analyzing efficacy, burnout, and support during the student-teaching semester. *Teaching and Teacher Education, 23*, 916–934.
- Frey, A. (2014). Kompetenzmodelle und Standards in der Lehrerbildung und im Lehrerberuf. In E. Terhart, H. Brennewitz, & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (2nd ed., pp. 712–744). München, New York: Waxmann.
- Garvis, S., Pendergast, D., & Keogh, J. (2012). Changes in teacher self-efficacy in the first year of primary school teacher education study. *The Journal of the World Universities Forum, 5*(1).
- Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*(4), 669–682.
- Guskey, T. R. (1988). Teacher efficacy, self-concept, and attitudes toward the implementation of instructional innovation. *Teaching and Teacher Education, 4*(1), 63–69. doi:10.1016/0742-051X(88)90025-X
- Guskey, T. R., & Passaro, P. D. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal, 31*(3), 627–643. Retrieved from <http://aer.sagepub.com/content/31/3/627>
- Hattie, J. (2009). Visible learning: A synthesis of over 800 meta-analyses relating to achievement. London, New York: Routledge.
- Henson, R. K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist, 37*(3), 137–150. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/S15326985EP3703_1#.UnfabKMwfcs
- Henson, R. K., Kogan, L. R., & Vacha-Haase, T. (2001). A Reliability Generalization Study of the Teacher Efficacy Scale and Related Instruments. *Educational and Psychological Measurement, 61*(3), 404–420. doi:10.1177/00131640121971284

- Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology, 105*(3), 774–786.
- Hoy, W. K., & Woolfolk, A. E. (1990). Socialization of student teachers. *American Educational Research Journal, 27*(2), 279–300.
- Jamil, F. M., Downer, J. T., & Pianta, R. C. (2012). Association of pre-service teachers' performance, personality, and beliefs with teacher self-efficacy at program completion. *Teacher Education Quarterly, 39*(4), 119–138.
- Klassen, R. M., Bong, M., Usher, E. L., Chong, W. H., Huan, V. S., Wongd, I. Y. F., & Georgiou, T. (2009). Exploring the validity of a teachers' self-efficacy scale in five countries. *Contemporary Educational Psychology, 34*, 67–76.
- Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology, 102*(3), 741–756.
- Klassen, R. M., & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching context. *Contemporary Educational Psychology, 36*(2), 114–129.
doi:10.1016/j.cedpsych.2011.01.002
- Klassen, R. M., & Durksen, T. L. (2014). Weekly self-efficacy and work stress during the teaching practicum: A mixed methods study. *Learning and Instruction, 33*, 158–169.
doi:10.1016/j.learninstruc.2014.05.003
- Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review, 12*, 59–76.
doi:10.1016/j.edurev.2014.06.001
- Klassen, R. M., Tze, V. M. C., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise? *Educational Psychology Review, 23*, 21–43.
- Kleinginna, P. R., & Kleinginna, A. M. (1981). A categorized list of motivation definitions, with a suggestion for a consensual definition. *Motivation and Emotion, 5*(3), 263–291.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, P., & Prenzel, M. (2003). *Zur Entwicklung nationaler Bildungsstandards: Eine Expertise*. Berlin: Bundesministerium für Bildung und Forschung (BMBF).

- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, *52*, 876–903.
- KMK. (2004). *Standards für die Lehrerbildung: Bildungswissenschaften*. Beschluss der Kultusministerkonferenz vom 16.12.2004. Retrieved from http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2004/2004_12_16-Standards-Lehrerbildung.pdf
- Knoblauch, D., & Woolfolk Hoy, A. (2008). “Maybe I can teach those kids.” The influence of contextual factors on student teachers’ efficacy beliefs. *Teaching and Teacher Education*, *24*, 166–179.
- Kunter, M. (2014). Forschung zur Lehrermotivation. In E. Terhart, H. Brennewitz, & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (2nd ed., pp. 698–711). München, New York: Waxmann.
- Liaw, E.-C. (2009). Teacher efficacy of pre-service teachers in Taiwan: The influence of classroom teaching and group discussions. *Teaching and Teacher Education*, *25*(1), 176–180. doi:10.1016/j.tate.2008.08.005
- Lin, H.-L., & Gorrell, J. (2001). Exploratory analysis of pre-service teacher efficacy in Taiwan. *Teaching and Teacher Education*, *17*, 623–635.
- Lindqvist, P., Nordänger, U. K., & Carlsson, R. (2014). Teacher attrition the first five years – A multifaceted image. *Teaching and Teacher Education*, *40*, 94–103. doi:10.1016/j.tate.2014.02.005
- Marsh, H. W., & O'Mara, A. J. (2008). Self-concept is as multidisciplinary as it is multidimensional: A review of theory, measurement, and practice in self-concept research. In D. M. McInerney, H. W. Marsh, & R. Craven (Eds.), *Self-processes, learning, and enabling human potential. Dynamic new approaches* (pp. 87–115). Charlotte, NC: Information Age Pub.
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, *60*(1), 577–605. doi:10.1146/annurev.psych.60.110707.163612
- Meade, A. W., & Bauer, D. J. (2007). Power and precision in confirmatory factor analytic tests of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 611–635. doi:10.1080/10705510701575461

- Mergler, A. G., & Tangen, D. (2010). Using microteaching to enhance teacher efficacy in pre-service teachers. *Teaching Education, 21*(2), 199–210. Retrieved from <http://dx.doi.org/10.1080/10476210902998466>
- Midgley, C., Feldlaufer, H., & Eccles, J. S. (1989). Change in teacher efficacy and student self- and task-related beliefs in mathematics during the transition to junior high school. *Journal of Educational Psychology, 81*, 247–258.
- Millsap, R. E., & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology, 73*(3), 574–584.
- Moulding, L. R., Stewart, P. W., & Dunmeyer, M. L. (2014). Pre-service teachers' sense of efficacy: Relationship to academic ability, student teaching placement characteristics, and mentor support. *Teaching and Teacher Education, 41*, 60–66.
doi:10.1016/j.tate.2014.03.007
- National Board for Professional Teaching Standards. (2002). *What teachers should know and be able to do*. Retrieved from http://www.nbpts.org/sites/default/files/what_teachers_should_know.pdf
- OECD. (2005). *Teachers matter: Attracting, developing and retaining effective teachers* (Education and training policy). Paris.
- OECD. (2014). *TALIS 2013 results: An international perspective on teaching and learning*. Paris. Retrieved from <http://dx.doi.org/10.1787/9789264196261-en>
- Ophardt, D., & Thiel, F. (2013). *Klassenmanagement: Ein Handbuch für Studium und Praxis*. Schulpädagogik. Stuttgart: Kohlhammer.
- Oser, F. (1997). Standards in der Lehrerbildung: Teil 1 : Berufliche Kompetenzen, die hohen Qualitätsmerkmalen entsprechen. *Beiträge zur Lehrerbildung, 15*(1), 26–37.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543–578. Retrieved from http://www.jstor.org/stable/1170653?seq=1#page_scan_tab_contents
- Pajares, F. (1997). Current directions in self-efficacy research. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement* (Vol. 10, pp. 1–49). Greenwich, Conn.: Jai Press.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307–332.
doi:10.3102/00346543062003307

- Pendergast, D., Garvis, S., & Keogh, J. (2011). Pre-service student-teacher self-efficacy beliefs: An insight into the making of teachers. *Australian Journal of Teacher Education*, 36(12), 46–58. Retrieved from <http://ro.ecu.edu.au/ajte/vol36/iss12/4/>
- Piwowar, V. (2013). Konzeptualisierung, Erfassung und Entwicklung von Kompetenzen im Klassenmanagement (doctoral dissertation). Freie Universität Berlin.
- Poulou, M. (2007). Personal teaching efficacy and its sources: Student teachers' perceptions. *Educational Psychology*, 27(2), 191–218. doi:10.1080/01443410601066693
- Riggs, I. M., & Enochs, L. G. (1990). Toward the development of an elementary teacher's science teaching efficacy belief instrument. *Science Education*, 74(6), 625–637.
- Ross, J. A. (1998). The antecedents and consequences of teacher efficacy. *Advances in Research on Teaching*, 7, 49–74.
- Ross, J. A., Hogaboam-Gray, A., & Hannay, L. (2001). Effects of teacher efficacy on computer skills and computer cognitions of Canadian students in grades K-3. *The Elementary School Journal*, 102(2), 141–156. Retrieved from http://www.jstor.org/stable/1002205?seq=1#page_scan_tab_contents
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 80(1), 1–28.
- Ruble, L. A., Usher, E. L., & McGrew, J. H. (2011). Preliminary investigation of the sources of self-efficacy among teachers of students with autism. *Focus on Autism and Other Developmental Disabilities*, 26(2), 67–74. doi:10.1177/1088357610397345
- Schleicher, A. (2012). Preparing teachers and developing school leaders for the 21st century: Lessons from around the world. Paris: OECD Publishing.
- Schulte, K., Bögeholz, S., & Watermann, R. (2008). Selbstwirksamkeitserwartungen und Pädagogisches Professionswissen im Verlauf des Lehramtsstudiums. *Zeitschrift für Erziehungswissenschaft*, 11(2), 268–287.
- Schunk, D. H., & Zimmermann, B. J. (2006). Competence and control beliefs: Distinguishing the means and ends. In P. A. Alexander (Ed.), *Handbook of educational psychology* (2nd ed., pp. 349–367). New York, NY: Routledge.
- Schwarzer, R., & Hallum, S. (2008). Perceived teacher self-efficacy as a predictor of job stress and burnout: Mediation analyses. *Applied Psychology: An International Review*, 57, 152–171.
- Schwarzer, R., Schmitz, G. S., & Daytner, G. T. (1999). *Teacher Self-Efficacy*. Retrieved from http://userpage.fu-berlin.de/~health/teacher_se.htm

- Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46(3), 407–441.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford, New York: Oxford University Press.
- Siwatu, K. O., & Chesnut, S. R. (2015). The career development of preservice and inservice teachers: Why teachers' self-efficacy beliefs matter. In H. Fives & M. G. Gill (Eds.), *Educational psychology handbook series. International handbook of research on teachers' beliefs* (pp. 212–229). New York, NY: Routledge.
- Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology*, 99(3), 611–625. doi:10.1037/0022-0663.99.3.611
- Skaalvik, E. M., & Skaalvik, S. (2008). Teacher self-efficacy: Conceptual analysis and relations with teacher burnout and perceived school context. In D. M. McInerney, H. W. Marsh, & R. Craven (Eds.), *Self-processes, learning, and enabling human potential. Dynamic new approaches* (pp. 223–247). Charlotte, NC: Information Age Pub.
- Skaalvik, E. M., & Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: A study of relations. *Teaching and Teacher Education*, 26(4), 1059–1069. doi:10.1016/j.tate.2009.11.001
- Skinner, E. A. (1996). A guide to constructs of control. *Journal of Personality and Social Psychology*, 71(3), 549–570.
- Somech, A., & Drach-Zahavy, A. (2000). Understanding extra-role behavior in schools: The relationships between job satisfaction, sense of efficacy, and teachers' extra-role behavior. *Teaching and Teacher Education*, 16(5-6), 649–659. doi:10.1016/S0742-051X(00)00012-3
- Soodak, L. C., & Podell, D. M. (1996). Teacher efficacy: Toward the understanding of a multi-faceted construct. *Teaching and Teacher Education*, 12(4), 401–411.
- Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research - Online*, 2(1), 21–33.

- Swan, B. G., Wolf, K. J., & Cano, J. (2011). Changes in teacher self-efficacy from the student teaching experience through the third year of teaching. *Journal of Agricultural Education, 52*(2), 128–139.
- Tatto, M. T., Schwille, J., Senk, S., Ingvarson, L., Peck, R., & Rowley, G. (2008). Teacher education and development study in mathematics (TEDS-M): Policy, practice, and readiness to teach primary and secondary mathematics. Conceptual framework. East Lansing, MI: Teacher Education and Development International Study Center, College of Education, Michigan State University.
- Terhart, E. (2005). Standards für die Lehrerbildung - ein Kommentar. *Zeitschrift für Pädagogik, 51*(2), 275–279.
- Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805.
- Tschannen-Moran, M., & Woolfolk Hoy, A. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education, 23*(6), 944–956. doi:10.1016/j.tate.2006.05.003
- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*(2), 202–248. Retrieved from <http://www.jstor.org/stable/1170754?origin=JSTOR-pdf&>
- Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research, 54*(2), 143–178. Retrieved from <http://rer.sagepub.com/content/54/2/143>
- Vieluf, S., Kunter, M., & van de Vijver, F. J. R. (2013). Teacher self-efficacy in cross-national perspective. *Teaching and Teacher Education, 35*, 92–103. doi:10.1016/j.tate.2013.05.006
- Wang, H., Hall, N. C., & Rahimi, S. (2015). Self-efficacy and causal attributions in teachers: Effects on burnout, job satisfaction, illness, and quitting intentions. *Teaching and Teacher Education, 47*, 120–130. doi:10.1016/j.tate.2014.12.005
- Weinert, F. E. (2001). *Leistungsmessungen in Schulen*. Weinheim: Beltz.
- Wheatley, K. F. (2002). The potential benefits of teacher efficacy doubts for educational reform. *Teaching and Teacher Education, 18*(5-22).
- Wheatley, K. F. (2005). The case for reconceptualizing teacher efficacy research. *Teaching and Teacher Education, 21*(7), 747–766. doi:10.1016/j.tate.2005.05.009
- Woolfolk, A. E., & Hoy, W. K. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology, 82*(1), 81–91.

- Woolfolk Hoy, A., & Burke Spero, R. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, 21(4), 343–356. doi:10.1016/j.tate.2005.01.007
- Woolfolk Hoy, A., Davis, H., & Pape, S. J. (2006). Teacher knowledge and beliefs. In P. A. Alexander (Ed.), *Handbook of educational psychology* (2nd ed., pp. 715–737). New York, NY: Routledge.
- Wyatt, M. (2013). Towards a re-conceptualization of teachers' self-efficacy beliefs: Tackling enduring problems with the quantitative research and moving on. *International Journal of Research & Method in Education*, 37(2), 166–189. doi:10.1080/1743727X.2012.742050

2

Study 1: An adapted measure of teacher self-efficacy for preservice teachers: Exploring its validity across two countries

Author's note: This article does not exactly replicate the final version published in the journal "Zeitschrift für Pädagogische Psychologie". It is not a copy of the original published article and is not suitable for citation.

Pfitzner-Eden, F., Thiel, F., & Horsley, J. (2014). An adapted measure of teacher self-efficacy for preservice teachers: Exploring its validity across two countries. *Zeitschrift für Pädagogische Psychologie*, 28(3), 83-92. DOI-link: <http://dx.doi.org/10.1024/1010-0652/a000125>

© 2014 by Hans Huber

2 An adapted measure of teacher self-efficacy for preservice teachers: Exploring its validity across two countries

Abstract

Teacher self-efficacy (TSE) is an important construct in the prediction of positive student and teacher outcomes. However, problems with its measurement have persisted, often through confounding TSE with other constructs. This research introduces an adapted TSE instrument for preservice teachers, which is closely aligned with self-efficacy experts' recommendations for measuring self-efficacy, and based on a widely used measure of TSE. Further, we provide first evidence of construct validity for this instrument. Participants were 851 preservice teachers in three samples from Germany and New Zealand. Results of the multiple-group confirmatory factor analyses showed a uniform 3-factor solution for all samples, metric measurement invariance, and a consistent and moderate correlation between TSE and a measure of general self-efficacy across all samples. Despite limitations to this study, there is some first evidence that this measure allows for a valid 3-dimensional assessment of TSE in preservice teachers.

Key words:

teacher self-efficacy; preservice teachers; measurement invariance; construct validity

Zusammenfassung

LehrerSelbstwirksamkeit (LSW) ist ein zentrales Konstrukt zur Vorhersage von positiven Outcomes von Schülerinnen und Schülern sowie Lehrpersonen. Aufgrund der Konfundierung von Selbstwirksamkeit mit anderen Konstrukten, werden Probleme der Erfassung von LSW seit längerer Zeit diskutiert. Diese Studie stellt eine adaptierte Version eines häufig verwendeten englischsprachigen Instruments zur Messung der LSW für Lehramtsstudierende vor. Neben einer zielgruppenspezifischen Anpassung, erfolgte eine Anpassung des ursprünglichen Instruments im Hinblick auf Banduras Richtlinien zur trennscharfen Erfassung von Selbstwirksamkeit. Wir berichten erste Belege für die Konstruktvalidität des Instruments. Untersucht wurden 851 Lehramtsstudierende in drei Stichproben aus Deutschland und Neuseeland. Die Ergebnisse der konfirmatorischen Mehrgruppen-Faktorenanalysen zeigten eine uniforme 3-Faktoren Struktur für alle Gruppen, metrische Messinvarianz, sowie eine konsistente, moderate Korrelation zu genereller Selbstwirksamkeit über alle drei Stichproben hinweg. Trotz einiger Limitationen der Studie, gibt es erste Belege dafür, dass das vorgestellte Instrument eine valide drei-dimensionale Messung der LSW für Lehramtsstudierende zulässt.

Schlüsselwörter:

LehrerSelbstwirksamkeit; Lehramtsstudierende; Messinvarianz; Konstruktvalidität

2.1 Introduction

Over the last three decades, research attention has increasingly focused on teacher self-efficacy (TSE) (Klassen, Tze, Betts, & Gordon, 2011). While the effects of TSE on student and teacher outcomes have been extensively researched, the operationalization of TSE has been troubled by issues with construct validity and invalid measurement (Denzine, Cooney, & McKenzie, 2005; Henson, 2002; Klassen et al., 2011). In 2001, Tschannen-Moran and Woolfolk Hoy introduced a new measure of TSE, the Teacher's Sense of Efficacy Scale (TSES). By 2012, it had advanced to become the predominant measure of TSE (Duffin, French, & Patrick, 2012). In this paper, we propose an adaptation of the TSES for preservice teachers, intended to capture the underlying construct of self-efficacy more in line with Bandura's (2006) recommendations, which were published after the debut of the TSES.

The construct of teacher self-efficacy

The relevance of TSE for research in educational settings is high, given its predictive power for a range of positive outcomes for students and teachers. At the student level, research has shown that TSE is positively related to students' achievement and motivation (e.g. Klassen et al., 2011; Skaalvik & Skaalvik, 2007; Woolfolk Hoy & Burke Spero, 2005). At the teacher level, TSE has been found to positively influence the effort teachers invest into their work, the degree of aspiration teachers show, the goals they set for themselves (Woolfolk Hoy & Burke Spero, 2005), and the level of job satisfaction they experience (Klassen et al., 2009). Moreover, researchers have linked TSE and burnout (e.g. Schwarzer & Hallum, 2008), even in preservice teachers (e.g. Fives, Hamman, & Olivarez, 2007). Underlying this decade-spanning body of research, are two separate strands of theory. We briefly review the basic understanding of each as it pertains to TSE.

Representing an integral part of Bandura's social cognitive theory (1977, 1986), the construct of self-efficacy constitutes the basis for TSE. Bandura (1997) defines self-efficacy as „...beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments.” (p.3). These beliefs are formed when people interpret information about their capabilities from four different sources: mastery experiences, vicarious experiences, social persuasion, and physiological and affective states (Bandura, 1997). According to Bandura (1997), self-efficacy beliefs can vary with the domain of action. Based on this, TSE can be understood as self-efficacy beliefs specific to the domain of teaching rather than merely self-efficacy beliefs held by teachers. Domain-specificity is essential to the measurement of self-efficacy (e.g. Pajares, 1996; Pajares & Schunk, 2001). Items need to be

tailored to the domain of interest, requiring a comprehensive analysis of that domain, and should be formulated at an intermediate level of specificity (Bandura, 2006).

Especially the early research on and measures of TSE have been dominated by a second conceptual approach to TSE grounded in Rotter's social learning theory. Specifically, the first research to note that teachers' beliefs of efficacy influence student outcomes was conducted by the RAND corporation in 1976 (Armor et al.). They used two items reflecting Rotter's locus-of-control beliefs, applying his concept to teachers' beliefs regarding the effect they could or could not have on student outcomes. Gibson and Dembo (1984) developed these items further into the *Teacher Efficacy Scale*. This scale comprises two factors, representing the original RAND items. Guskey and Passaro (1994) showed that the distinction between those factors corresponded closely to an external internal locus-of-control distinction. By the late 1990s, TSE research had "suffered an adolescent identity crisis" (Henson, 2002, p. 139) due to ongoing problems with poor measurement and construct validity, largely supported by the fact that the Gibson and Dembo Scale was used as a standard measure of TSE, despite being based on locus-of-control theory. Review articles (e.g. Henson, 2002; Klassen et al., 2011; Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998) have again and again highlighted the need for valid measurement and sound conceptualization of TSE in line with Bandura's theory of self-efficacy.

The Teacher's Sense of Efficacy Scale

The TSES (Tschannen-Moran & Woolfolk Hoy, 2001) comprises three subscales: instructional strategies (IS), classroom management (CM) and student engagement (SE). It is available in a long (24-items) and a short (12-items) version. The factor structure of the TSES has posed an ongoing source of debate. When used with preservice teachers, results regarding the factor structure have been ambivalent. While several researchers (e.g. Fives & Buehl, 2009; Tschannen-Moran & Woolfolk Hoy, 2001) found a 1-factor model to best represent their data, Poulou (2007) found a 3-factor model to represent their data from fourth year preservice teachers better. Duffin et al. (2012) explained those contradictory results by suggesting that beginning preservice teachers were unable to distinguish between the three factors, while preservice teachers who were advanced in their studies had gained that ability. However, an instrument that distinguishes the three TSE factors for preservice teachers at any stage would be of great advantage, as it would allow research into the development of those factors early on. Since researchers (e.g. Veenman, 1984) suggest that TSE in CM is of particular concern to preservice teachers, this would be a worthy endeavour.

When introducing the TSES, Tschannen-Moran and Woolfolk Hoy (2001) stated that their objective was to develop a valid and reliable measure of TSE. Close inspection of the wording of the TSES-items, response scale and introduction, shows that they do not closely reflect Bandura's (2006) recommendations for measuring self-efficacy. According to Bandura's Guide for constructing self-efficacy scales (2006), respondents are to rate their degree of *confidence* that they *can do* different activities. The TSES though asks respondents to rate their degree of *influence* on teaching and teacher activities with most items phrased "How much can you do to...?" (Tschannen-Moran & Woolfolk Hoy, 2001), and a response scale ranging from "nothing" (1) to "a great deal" (9). This phrasing of the items and response scale confounds the locus-of-control concept with self-efficacy. For example, asking participants to rate how much influence they can exert over e.g. "the motivation of students with low interest in schoolwork", answering with "a great deal" would indicate high internal locus-of-control and "nothing" would indicate high external locus-of-control. Wheatley (2005) also noted that most teacher efficacy scales, including the TSES, blur the distinction between self-efficacy and locus-of-control. This is detrimental to predictive validity because self-efficacy predicts behavior better than locus-of-control beliefs (Bandura, 1997). Additionally, evidence regarding the construct validity of the TSES presented by Tschannen-Moran and Woolfolk Hoy (2001) was based on correlations with instruments designed to assess TSE that confound locus-of-control beliefs (an adaptation of the Gibson and Dembo Scale and the Rand items). A recent Master thesis (Boettcher, 2013) examining a sample of 239 beginning preservice teachers and 465 advanced preservice teachers, found the correlation between the TSES short form and general self-efficacy (GSE) as measured by the *General Perceived Self-Efficacy Scale* (Schwarzer & Jerusalem, 1999) to be: $r = .34$ for IS; $r = .21$ for CM; $r = .23$ for SE; and $r = .33$ for IS; $r = .33$ for CM; $r = .26$ for SE, for the beginning and advanced preservice teachers respectively. These correlations can be judged as low compared to correlations between GSE and TSE of $r = .53$ for preservice teachers (Schulte, Bögeholz, & Watermann, 2008), and $r = .71$ for inservice teachers reported by Schwarzer and Hallum (2008), who still interpreted TSE as a separate meaningful construct. These comparisons lend empirical support to the notion that the TSES does not measure the underlying construct of self-efficacy closely.

The item contents on the other hand are a real strength of the TSES. In generating the items of the TSES, Tschannen-Moran and Woolfolk Hoy (2001) followed a rigorous process in order to ensure relevance to the teaching context, to assess a comprehensive range of teacher capabilities, and to establish an appropriate level of domain-specificity.

The current study

The objective of the current study was twofold: first, to revise an existing TSE instrument to measure TSE more in line with Bandura's conceptualization; second, to create an instrument that assesses the three factors of TSE for preservice teachers. Thus, we revised the original TSES items in line with Bandura's guidelines on measuring self-efficacy, while keeping item contents referring to the teaching domain identical. First evidence for construct validity of this adapted version was then gathered across three diverse samples of preservice teachers. All samples came from a non-North American setting (two from Germany and one from New Zealand), addressing a call by Klassen et al. (2011) for more TSE research in a non-North American context. First, we selected items for the adapted measure that represented the three factors of TSE well. Then, we examined the dimensionality (1- vs. 3-dimensional) of the adapted measure. Second, we conducted tests of measurement and construct level invariance. Third, to examine convergent validity, we inspected the correlation between TSE and GSE, as well as between TSE and study satisfaction, by testing construct level invariance. Showing that a measure of TSE operates similarly for those groups of preservice teachers would be a powerful argument for good construct validity of the adapted measure, because self-efficacy theory states the universality of the construct across contexts and nations (e.g. Bandura, 1997). Based on results by Schulte et al. (2008), we expected a moderate to large correlation with GSE, indicative of an assessment of the underlying construct of self-efficacy. Study satisfaction was chosen as a convergent validity criterion as an equivalent or proxy of job satisfaction for preservice teachers, which has repeatedly been found to show a moderate positive relationship to TSE (e.g. r between .17 and .45 in Klassen et al., 2009; Klassen & Chiu, 2010). Also, Boettcher (2013) found the correlation between the TSES subscales and study satisfaction to range between $r = .12$ and $.21$, thus providing a benchmark for our examination of study satisfaction. We expect a positive association with study satisfaction that should be greater than the one observed by Boettcher (2013).

2.2 Method

Participants and procedure

Participants were 851 preservice teachers in three samples. Two samples were drawn from a large German university. Its teacher preparation program encompasses a 3-year Bachelor degree followed either by a 1-year Master degree for elementary/middle school teachers, or a 2-year Master degree for secondary school teachers. Participants in the first sample (Sample 1) were 438 undergraduate preservice teachers (19% elementary/middle school) at the

beginning of their studies, 94% were in their first month of Bachelor study (63% female; age: $M = 22.81$ years, $SD = 4.93$). Participants in the second sample (Sample 2) were 282 preservice teachers (31% elementary/middle school) at an advanced stage of their studies (at least third year), of whom 80% were in their Master study (68% female; age: $M = 25.93$ years, $SD = 4.88$). All participants of Sample 2 had previously completed a one month practicum at school. Participants in both samples represented a wide range of academic subjects (e.g. mathematics, German, biology, physics, history, English). Participation was voluntary and data were collected in a paper-pencil format in seminars. Response rates were high, averaging 75%. Participants in the third sample (Sample 3) were 131 preservice teachers (82% female) undertaking a 1-year graduate diploma course at a large New Zealand university that qualifies them to teach in either primary or secondary schools. In line with ethics committee standards of this university, participants' age was collected in ranges (59% 20-30 years, 29% 31-40 years, and 18% over 41 years). Requirements for eligibility to this course include an undergraduate degree of three or four years. Participants in this sample also represented a wide range of academic subjects in their undergraduate degree (e.g. sports, English, Te Reo Māori, biology, physics, history). All participants had previously completed a seven-week teaching practicum at school. Participants received an emailed information sheet that described the study and invited them to click on a link connecting them to the online survey. The survey was anonymous, with the front page containing ethical considerations relating to giving informed consent. The response rate was 33%.

Measures

Adaptation of the TSES

Our goal was to develop a short, economic instrument equivalent to the 12-item short form of the TSES. Starting with the 24 TSES-items, we applied the following process. We revised the wording of the TSES items and response scale in accordance with Bandura's 2006 guidelines. Participants were instructed to rate their degree of *confidence* that they "...can successfully perform the following every-day tasks of teachers". The introductory wording of the items was changed to: "*How certain are you that you can...?*" with a 9-point response scale ranging from "*not at all certain can do*" to "*absolutely certain can do*".

In order to establish not only linguistic but also functional and cultural equivalence, we followed guidelines for translation of instruments for cross-cultural research suggested by Peña (2007). First, two bilingual translators, translated the 24 revised TSES-items into German and back-translated them, ensuring linguistic equivalence. Then, two teacher education experts judged how applicable each item was in a non-North American context,

thus ensuring functional and cultural equivalence. Based on an unequivocal result of the translation back-translation process and the judgments of the teacher education experts, six items from each TSES subscale were selected. The German version of these 18 items was then administered to the German samples, and the English version to the New Zealand sample. Preliminary analysis of the data applying an ant colony optimization algorithm (Marcoulides & Drezner, 2003), was then used to select four items that best represented each of the three subscales in all three samples. This procedure can be characterized as a partly a-priori, partly data-driven approach to the selection of the final 12 items for the adapted instrument, the *Scale for Teacher Self-Efficacy* (STSE)²¹. A-priori, because all items were assigned to their respective factor before going into the analysis. Data-driven, because the algorithm selected the four items out of six that best represented each factor in all samples.

Further measures

We used a well-established measure of GSE the *General Perceived Self-Efficacy Scale* by Schwarzer and Jerusalem in its German (1999) and English (1995) version²² for the respective samples (see Scholz, Gutiérrez Doña, Sud, & Schwarzer, 2002 for a description of cross-cultural psychometric properties and validity of the scale). This measure uses a 4-point response scale, ranging from “*not at all true*” (1) to “*exactly true*” (4). Internal consistency coefficients ($\alpha = .78$, $\alpha = .83$, $\alpha = .87$, for Sample 1, Sample 2, and Sample 3, respectively) were good. Study satisfaction was assessed using a 3-item measure, developed and tested as part of the study program evaluation (Thiel & Blüthmann, 2009) at the German university where data were collected. The items were: *In general, I am satisfied with my studies/study program. I would choose a teaching degree again. I would choose the same academic field/major again.* This measure uses a 9-point response scale, ranging from “*not at all true*” (1) to “*exactly true*” (9). Internal consistency coefficients ($\alpha = .70$, $\alpha = .59$, for Sample 1 and Sample 2, respectively) were low.

Analysis

A series of multiple-group confirmatory factor analyses (CFA) were conducted using Mplus (Version 6). First, we evaluated whether the adapted measure was better specified as a 1-factor or a 3-factor model. We then tested measurement invariance of the chosen model across the three groups. This approach is ideally suited to assess cross-group comparability of measures (Stein, Lee, & Jones, 2006; Vandenberg & Lance, 2000), an important aspect of

²¹ The German version of the STSE items can be obtained from the first author.

²² Both versions were retrieved July 21, 2011 from <http://userpage.fu-berlin.de/~health/selfscal.htm>

construct validity (Yin & Fan, 2003). Following recommendations by Brown (2006), tests of invariance were conducted in the following order. First, the CFA model was tested separately in each group. Second, equal form of the CFA model was tested in the combined groups (configural invariance). Third, the equality of factor loadings across groups was tested (metric invariance). These tests assessed measurement invariance, meaning the same constructs are measured in the same manner across groups (Riordan & Vandenberg, 1994). There are more levels of measurement invariance, but for the purpose of research on construct validity Byrne and Stewart (2006) recommend testing measurement invariance up to metric level invariance. Then we examined population heterogeneity by assessing two aspects of construct level invariance. On the basis of metric invariance, first the equality of factor variances, and then factor covariances were tested (Brown, 2006).

We used a robust estimator (MLR) recommended for both small samples and non-normal data (Finney & DiStefano, 2006; Satorra & Bentler, 2001). Model fit was evaluated using a combination of fit indices as recommended by various methodologists (e.g. Beauducel & Wittmann, 2005; Hu & Bentler, 1999): the comparative fit index (CFI) with a cut-off value close to .95, the root mean square error of approximation (RMSEA) with a cut-off value close to .06; and the χ^2/df values for which values between 2 and 3 indicate a good to acceptable fit (Schermelleh-Engel, Moosbrugger, & Müller, 2003). Following recommendations by Cheung and Rensvold (2002), we assessed invariance using the more practical difference in CFI whereby a value of up to -.01 indicates invariance, rather than the traditional χ^2 difference test²³ for nested models which is nevertheless reported. Further, we calculated construct reliabilities (Fornell & Larcker, 1981) for the latent factors in the chosen model as a local fit index, which in addition to global fit indices provide a closer look at internal aspects of the model. Construct reliabilities indicate how well each latent construct is measured in terms of internal consistency, whereby values greater than .60 are desirable (Bagozzi & Yi, 1988).

Finally, we examined the correlations between the STSE subscales and GSE across all groups, and the STSE subscales and study satisfaction in the German samples. The correlation coefficient r served as an indicator of effect size, whereby according to Cohen (1992) the effect size r can be judged small at .10, medium at .30 and large at .50. In order to test whether the correlations with GSE and study satisfaction varied across groups we used invariance testing once more.

²³ Due to the robust estimator used, Satorra-Bentler scaling-corrected Chi square differences were computed as described in the Mplus User Guide (1998-2010).

The alpha-level was set at $p < .05$, following recommendations by Cohen (1992), with regard to power for the relatively small Sample 3.

2.3 Results

Fit indices for the 1-factor, 3-factor, and constrained models are shown in Table 1.

Table 1. Model fit for the competing models in the combined groups, three-factor model in each group, and constrained models for the combined groups

Model	χ^2	<i>df</i>	<i>p</i>	χ^2/df	CFI	RMSEA	SCF	$\Delta\chi^2_{SBscaled}$
<i>1-factor model</i>								
Combined groups (<i>n</i> = 851)	1204.06	162	<.001	7.43	.701	.151	1.419	-
<i>3-factor model</i>								
Combined groups (<i>n</i> = 851)	229.64	153	<.001	1.50	.978	.042	1.396	-
<i>3-factor model in each group</i>								
Germany beginning (<i>n</i> = 438)	63.71	51	.109	1.25	.992	.024	1.424	-
Germany advanced (<i>n</i> = 282)	93.58	51	<.001	1.83	.964	.054	1.340	-
New Zealand (<i>n</i> = 131)	73.35	51	.022	1.44	.968	.058	1.425	-
<i>3-group CFA (n = 851)</i>								
Factor structure equal (unconstrained)	229.64	153	<.001	1.50	.978	.042	1.396	-
Factor loadings equal	264.90	171	<.001	1.55	.973	.044	1.393	35.41(18)*
Factor variances (and loadings) equal	273.31	177	<.001	1.54	.972	.044	1.384	8.09(6) <i>ns</i>
Factor co-variances (and variances and loadings) equal	291.77	183	<.001	1.59	.969	.046	1.386	17.48(6)*

Note: CFI = comparative fit index; RMSEA = root mean square error of approximation; SCF = scaling correction factor; * $p < .05$

One-factor vs. three-factor model

The 1-factor model showed an unacceptable fit in the combined groups, with a CFI well below its cut-off value, the χ^2/df ratio and the RMSEA well above their respective cut-offs. The 3-factor model showed a very good fit in the combined groups, as indicated by all fit indices. On the basis of all indicators of model fit, we rejected the 1-factor model and proceeded with the invariance tests for the 3-factor model.

Invariance of the three-factor model

The 3-factor model also showed a very good fit in each group, with all χ^2/df ratios below 2, all CFIs above .96, and all RMSEAs below .06 (see Table 1). The 3-group unconstrained baseline model (factor structure equal) showed a very good fit, supporting the assumption of configural invariance across all groups. Constraining the factor loadings to be equal, resulted in a significant $\Delta\chi^2_{\text{SBscaled}}$, with an acceptable change in CFI of .005. Further constraining the factor variances across groups, resulted in a non-significant $\Delta\chi^2_{\text{SBscaled}}$, with a negligible change in CFI of .001. In a last step, constraining the covariances among the three latent factors across all groups, yielded a significant $\Delta\chi^2_{\text{SBscaled}}$, with a small change in CFI of .003. According to the practical evaluation perspective, measurement and construct level invariance could be established across all three groups.

Evaluating the three-factor model further

Standardized factor loading coefficients for all 12 items of the STSE, construct reliability for the three latent constructs, latent interfactor correlations, as well as manifest means and standard deviations are presented in Table 2. According to guidelines by Guadagnoli and Velicer (1988), with as few as four items per factor, an accurate representation of the population can still be assumed for the estimated solution when factor loadings are small (around .40) but sample size is large ($N > 300$), which is the case for the two German samples. If sample size is small ($N < 150$), as in the case of Sample 3, factor loadings above .60 indicate an accurate representation. All items fulfilled these recommendations, with the exception of the first item on the IS factor, which only approached those recommendations in Sample 1 and 3. Construct reliability can be judged as good to very good for all latent factors. Latent interfactor correlations were moderate ranging from .45 to .69, except for the relatively high .82 correlation between IS and SE in Sample 3. The STSE subscales showed adequate internal consistency. Respondents mostly used the maximum range of response categories (minimum 1, maximum 9 in both German samples; minimum 2, maximum 9 in the New Zealand sample).

Table 2. Standardized factor loadings, construct reliability (CR), internal consistency (α), latent interfactor correlations and manifest means and standard deviations for the Scale for Teacher Self-Efficacy

How certain are you that you can...	Germany (beginning, $n = 438$)	Germany (advanced, $n = 282$)	New Zealand (advanced, $n = 131$)
Standardized factor loadings			
<i>Instructional Strategies</i> (CR = .71/.74/.82)			
...provide an alternative explanation or example when students are confused?	.28	.43	.54
...adjust lessons to the proper level for individual students?	.71	.75	.86
...gauge students' comprehension of what has been taught?	.81	.76	.79
...provide appropriate challenges for very capable students?	.58	.63	.73
<i>Classroom Management</i> (CR = .91/.93/.94)			
...control disruptive behaviour in the classroom?	.89	.90	.92
...get students to follow classroom rules?	.88	.92	.90
...calm a student who is disruptive and noisy?	.87	.92	.87
...keep a few problem students from ruining an entire lesson?	.72	.78	.86
<i>Student Engagement</i> (CR = .82/.78/.86)			
...help students value learning?	.69	.72	.80
...motivate students who show low interest in schoolwork?	.84	.85	.85
...improve the understanding of a student who is failing?	.88	.71	.79
...help students to think critically?	.45	.43	.69
Latent interfactor correlations			
Instructional Strategies with Classroom Management	.53	.45	.69
Instructional Strategies with Student Engagement	.60	.66	.82
Classroom Management with Student Engagement	.45	.54	.62
Manifest Means (Standard Deviations)			
Instructional Strategies ($\alpha = .68/.73/.81$)	6.78 (0.99)	6.44 (1.11)	7.12 (1.03)
Classroom Management ($\alpha = .91/.93/.94$)	6.41 (1.46)	5.98 (1.62)	6.79 (1.41)
Student Engagement ($\alpha = .80/.76/.86$)	6.85 (1.13)	6.47 (1.19)	7.30 (1.02)

Note: CRs and α s for each sample are reported in the same order in which the samples appear in the right columns of the table. All coefficients were statistically significant.

Convergent validity

Table 3 shows the latent and manifest correlations between the STSE subscales, GSE and study satisfaction. All latent and manifest correlations between the STSE subscales and GSE were statistically significant (all $ps < .001$) with medium to large effect sizes. Except for CM in Sample 1 ($p = .78$), all correlations between the STSE subscales and study satisfaction were significant (all $ps < .001$), with small-to-medium effects in Sample 1, and large effects in Sample 2.

Table 3. Latent (manifest) correlations of the STSE subscales with general self-efficacy (GSE) and study satisfaction (StSa)

Scale	Germany (beginning, $n = 438$)		Germany (advanced, $n = 282$)		New Zealand (advanced, $n = 131$)
	GSE	StSa	GSE	StSa	GSE
IS	.51* (.45*)	.24* (.21*)	.49* (.42*)	.42* (.32*)	.49* (.44*)
CM	.45* (.40*)	.01 (.01)	.46* (.43*)	.36* (.24*)	.45* (.41*)
SE	.46* (.39*)	.23* (.19*)	.36* (.35*)	.51* (.35*)	.50* (.45*)
GSE	-	.20* (.19*)	-	.31* (.20*)	-

Note: Study satisfaction data were not collected in New Zealand; IS = instructional strategies; CM = classroom management; SE = student engagement; * $p < .05$

In order to examine group differences in the way both validity criteria related to the STSE subscales, we first tested metric invariance for the 3-group 3-factor TSE model correlated with GSE. Table 4 shows that constraining the factor loadings to be equal, resulted in a significant $\Delta\chi^2_{\text{SBscaled}}$, with an acceptable change in CFI of .004. Further constraining the factor variances across groups, resulted in a non-significant $\Delta\chi^2_{\text{SBscaled}}$, with a negligible change in CFI of .001. In a last step, constraining the covariances between the three latent factors and GSE across all groups, yielded a non-significant $\Delta\chi^2_{\text{SBscaled}}$, with no change in CFI. According to the practical evaluation perspective, the factor variances as well as the correlations between the TSE factors and GSE are invariant across the three groups.

Second, we tested metric invariance for the 2-group 3-factor TSE model correlated with study satisfaction. Constraining the factor loadings to be equal, resulted in a significant $\Delta\chi^2_{\text{SBscaled}}$, with an acceptable change in CFI of .006. Further constraining the factor variances across groups, resulted in a non-significant $\Delta\chi^2_{\text{SBscaled}}$, with a negligible change in CFI of .001. In a last step, constraining the covariances between the three latent factors and study satisfaction across both groups, yielded a significant $\Delta\chi^2_{\text{SBscaled}}$, with a small change in CFI of .002. According to the practical evaluation perspective, the factor variances and the

correlations between the TSE factors and study satisfaction are invariant across the two German groups.

Table 4. Model fit for the 3-factor model and constrained models with general self-efficacy (GSE) in the 3 combined groups, and study satisfaction (StSat) in the two German groups combined

Model	χ^2	<i>df</i>	<i>p</i>	χ^2/df	CFI	RMSEA	SCF	$\Delta\chi^2_{SBscaled}$
<i>3-group CFA with GSE (n = 851)</i>								
Factor structure equal (unconstrained)	505.63	339	<.001	1.49	.967	.042	1.226	-
Factor loadings equal	554.75	365	<.001	1.52	.963	.043	1.233	48.41(26)*
Factor variances (and loadings) equal	566.07	373	<.001	1.52	.962	.043	1.232	11.28(8) <i>ns</i>
Factor co-variances (and variances and loadings) equal	571.48	379	<.001	1.51	.962	.042	1.231	3.20(6) <i>ns</i>
<i>2-group CFA with StSat (n = 720)</i>								
Factor structure equal (unconstrained)	242.76	168	<.001	1.45	.977	.035	1.280	-
Factor loadings equal	276.83	179	<.001	1.55	.971	.039	1.289	31.26(11)*
Factor variances (and loadings) equal	283.72	183	<.001	1.55	.970	.039	1.287	6.96(4) <i>ns</i>
Factor co-variances (and variances and loadings) equal	292.23	186	<.001	1.57	.968	.040	1.282	9.65 (3)*

Note: CFI = comparative fit index; RMSEA = root mean square error of approximation; SCF = scaling correction factor; * $p < .05$

2.4 Discussion

This article presents a 3-dimensional TSE instrument for preservice teachers, which is closely aligned with Bandura's (2006) recommendations for measuring self-efficacy. Despite several notable limitations to our study, we found first evidence in favor of good construct validity of this measure.

First, we found a consistent 3-factor structure for all groups of preservice teachers, meaning all groups conceptualized TSE in the same way. Both global and local fit indices indicated a very good model fit. Contrary to previous findings derived using the TSES, beginning preservice teachers in this study distinguished between the three TSE factors using the adapted measure. This result can be due to (a) the differences between the original instrument and the STSE in terms of item selection and (b) a changed assessment of self-efficacy (construct). Since we deliberately selected four items per scale which represented the three factors well, this item selection would have at least partly produced this result. In terms

of construct, the main difference is that the TSES assesses the degree of *influence* that preservice teachers believe they could exert over tasks relating to IS, CM and SE, whereas the adapted instrument assesses the degree of *confidence* with which they believe they could successfully perform those tasks. So beginning preservice teachers might feel that they have the same degree of *influence* over all three factors, but might differentiate between being *confident* in successfully performing classroom management tasks, in applying instructional strategies, and in engaging students.

Second, measurement invariance could be established for all three groups of preservice teachers up to a critically important level for judging construct validity: metric level invariance. This is promising with regard to generalizability and external validity of our findings, because metric invariance could be established despite the numerous differences of the samples (e.g. teacher preparation programs, languages, gender distribution, administration modes, response rates).

Third, invariance of factor variances and covariances across the three groups was found, indicating that there were no meaningful cross-group differences in the degree to which responses varied on each of the three latent factors, and the degree to which the latent factors related to each other. However, a higher interfactor correlation in Sample 3 was reflected in the significant χ^2 -difference test result, which in light of the small change in CFI seems of little practical relevance.

Fourth, according to both the practical and statistical evaluation perspective, the relationship between GSE and the STSE subscales was comparable between the three groups. In line with our hypothesis, there was a moderate to high level of association between the two constructs of self-efficacy, providing first evidence of convergent validity. Manifest correlations ranged from $r = .35$ to $.45$, comparable to results by Schulte et al. (2008), and higher than the correlations between the TSES subscales and GSE ($r = .21$ to $.34$) observed by Boettcher (2013). Taken together, these results add support to the claim that the adapted measure reflects the underlying construct of self-efficacy more closely.

Fifth, the correlation between the STSE subscales and study satisfaction, another indicator of convergent validity, showed the expected positive direction. However, it also revealed some unexpected results, namely that the relationship, particularly with CM, was weaker for the beginning preservice teachers than for the advanced preservice teachers, as indicated by the latent and manifest correlation coefficients. However, this was not picked up by the practical evaluation perspective, according to which latent correlations between the STSE subscales and study satisfaction were invariant, but was reflected in the significant χ^2 -

difference test result. The weaker association between the STSE subscales and study satisfaction may not be surprising considering that the beginning preservice teachers had only started studying one month prior to data collection. Thus in hindsight, study satisfaction might not have been a good choice of a convergent validity criterion for this group, because TSE might not have been related to study satisfaction yet, since the ratings of study satisfaction, due to little experience, might not carry much meaning this early. Conversely for advanced preservice teachers, study satisfaction should be based on experiences with the study program, and should thus be related in a meaningful way to TSE. The results of this group showed a greater association with the STSE subscales. This association ($r = .24$ to $.35$) was comparable to correlations between TSE and job satisfaction that have been reported in previous studies of inservice teachers ($r = .17$ to $.45$, in Klassen et al., 2009), and were greater than the correlation between the TSES subscales and study satisfaction reported by Boettcher (2008; $r = .12$ to $.21$), whose beginning preservice teachers were at the end of their first year.

Limitations and implications

Several limitations to our study need to be considered. Study satisfaction data was only assessed for the German samples, yet our study would have greatly benefited from an inspection of the relationship between study satisfaction and TSE in another sample of advanced preservice teachers. Further, our study offered only limited empirical proof of validity. The argument that the adapted instrument measures TSE less confounded with locus-of-control beliefs than the original instrument, has mostly been made theoretically. Lacking is a concurrent examination of the original and adapted measure, for example with regard to their respective relationships with locus-of-control beliefs. However, because of the similarity in item content of the two measures, we did not administer the TSES and STSE to our samples at the same time, as we expected that would confuse participants. Nonetheless, this could have been avoided by administering each instrument at different times, for example a week apart. Empirical proof of validity was limited to examining convergent validity with GSE and study satisfaction compared to a benchmark provided by an unpublished Master thesis. This comparison would be more convincing if conducted with the same sample. Furthermore, considering that the correlation between CM and study satisfaction was significantly different from zero in the advanced group and non-significant in the beginning group, it seems questionable to what degree the practical evaluation perspective (i.e. change in global CFI) is able to pick up on between-group differences of a single parameter within the invariance model. The unequal sample sizes in our study might have played an exacerbating role in this issue (see Brown, 2006, p. 279). The partly data-driven approach to

the STSE item selection also needs to be pointed out as a limiting factor, because it could be criticized as optimizing the 3-factor model for this study only. But, the diversity of the three samples suggests that the 3-factor structure of the STSE will be replicated in other samples of preservice teachers, yet this remains to be proven. Furthermore, future studies could explore if the STSE is also applicable to samples of inservice teachers. One instrument for measuring TSE reliably and validly across those reasonably distinct groups (i.e. beginning and advanced preservice teachers, inservice teachers) would be preferable to separate instruments, as results would be directly comparable across groups.

This study presented first evidence of validity and reliability of an adapted 3-factor measure of TSE for preservice teachers, which has meaningful implications for educational research as well as for practitioners in teacher education. A measure of TSE that is closely aligned with self-efficacy theory should produce results with greater predictive validity. This would allow for more accurate insights into the significant relationships between TSE and desired outcomes. The consistent 3-factor structure of the STSE allows studying the development of each TSE dimension, starting at the beginning of teacher preparation. This could be a valuable endeavor, especially considering previous research (e.g. Veenman, 1984) suggesting that classroom management is of particular concern to preservice teachers. More practically, a valid measure of TSE could for example be used as a screening instrument regarding the development of burnout syndrome in preservice teachers, which has been found to be an issue (e.g. Fives et al., 2007). If low levels of TSE were found, interventions targeted at increasing TSE could be recommended specifically for those at-risk. Further, since practical experiences are thought to influence TSE most, by using a measure of TSE that is closely aligned with self-efficacy theory, teacher educators could attempt to design a practicum experience (based on the sources of self-efficacy) that enhances the development of TSE.

2.5 References

- Armor, D. J., Conry-Oseguera, P., Cox, M., King, N., McDonnell, L., Pascal, A., Pauly, E., & Zellman, G. (1976). *Analysis of the school preferred reading program in selected Los Angeles minority schools, REPORT NO. R-2007-LAUSD*. Santa Monica, CA: Rand Corporation.
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 16*, 74-94.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191–215.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. M. Pajares, & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich: Information Age Publishing.
- Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling, 12*, 41–75.
- Boettcher, F. (2013). *Beitrag zur Validierung eines Instrumentes zur Erfassung der Lehrerinnen- und Lehrer-Selbstwirksamkeit*. Unpublished Master Thesis, Freie Universität Berlin.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Byrne, B. M., & Stewart, S. M. (2006). Teacher's corner: The MACS approach to testing for multigroup invariance of a second-order structure: A walk through the process. *Structural Equation Modeling, 13*, 287–321.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255.
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159.
- Denzine, G. M., Cooney, J. B., & McKenzie, R. (2005). Confirmatory Factor Analysis of the Teacher Efficacy Scale for Prospective Teachers. *British Journal of Educational Psychology, 75*, 689–708.
- Duffin, L. C., French, B. F., & Patrick, H. (2012). The teachers' sense of efficacy scale: Confirming the factor structure with beginning pre-service teachers. *Teaching and Teacher Education, 28*, 1–8.

- Finney, S. J., & Distefano, C. (2006). Nonnormal and categorical data in structural equation modelling. In G. R. Hancock, & R. O. Mueller (Eds.), *Structural equation modelling: A second course* (pp. 269-314). Greenwich, CT: Information Age Publishing.
- Fives, H., & Buehl, M. M. (2009). Examining the factor structure of the teachers' sense of efficacy scale. *The Journal of Experimental Education*, 78, 118–134.
- Fives, H., Hamman, D., & Olivarez, A. (2007). Does burnout begin with student-teaching? Analyzing efficacy, burnout, and support during the student-teaching semester. *Teaching and Teacher Education*, 23, 916–934.
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18, 39-50.
- Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology*, 76, 669–682.
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103, 265-275.
- Guskey, T. R., & Passaro, P. D. (1994). Teacher efficacy: A study of construct dimensions. *American Educational Research Journal*, 31, 627–643.
- Henson, R. K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist*, 37, 137–150.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Klassen, R. M., Bong, M., Usher, E. L., Har Chong, W., Huan, V. S., Wong, I. F. Y., & Georgiou, T. (2009). Exploring the validity of a teachers' self-efficacy scale in five countries. *Contemporary Educational Psychology*, 34, 67–76.
- Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology*, 102, 741–756.
- Klassen, R. M., Tze, V. M. C., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise?. *Educational Psychology Review*, 23, 21–43.
- Marcoulides, G. A., & Drezner, Z. (2003). Model specification searches using ant colony optimization algorithms. *Structural Equation Modeling*, 10, 154–164.

- Muthén, L.K., & Muthén, B.O. (1998-2010). *Mplus user's guide (6th ed.)*. Los Angeles, CA: Muthén & Muthén.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*, 543–578.
- Pajares, F., & Schunk, D. H. (2001). Self-beliefs and school success: Self-efficacy, self-concept, and school achievement. In R. Riding & S. Rayner (Eds.), *Self-perception* (pp. 239-266). London: Ablex Publishing.
- Peña, E. D. (2007). Lost in translation: Methodological considerations in cross-cultural research. *Child Development, 78*, 1255–1264.
- Poulou, M. (2007). Personal teaching efficacy and its sources: Student teachers' perceptions. *Educational Psychology, 27*, 191–218.
- Riordan, C. M., & Vandenberg, R. J. (1994). A central question in cross-cultural research: Do employees of different cultures interpret work-related measures in an equivalent manner? *Journal of Management, 20*, 643–671.
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*, 507–514.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8*, 23–74.
- Scholz, U., Gutiérrez Doña, B., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment, 18*, 242-251.
- Schulte, K., Bögeholz, S., & Watermann, R. (2008). Selbstwirksamkeitserwartungen und pädagogisches Professionswissen im Verlauf des Lehramtsstudiums. *Zeitschrift für Erziehungswissenschaft, 11*, 268–287.
- Schwarzer, R., & Hallum, S. (2008). Perceived teacher self-efficacy as a predictor of job stress and burnout: Mediation analyses. *Applied Psychology, 57*, 152–171.
- Schwarzer, R. & Jerusalem, M. (1995). Generalized self-efficacy scale. In J. Weinman, S. Wright & M. Johnston, *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35-37). Windsor, England: NFER-NELSON.
- Schwarzer, R. & Jerusalem, M. (Eds.) (1999). *Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen*. Berlin: Freie Universität Berlin.

- Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology, 99*, 611–625.
- Stein, J. A., Lee, J. W., & Jones, P. S. (2006). Assessing cross-cultural differences through use of multiple-group invariance analyses. *Journal of Personality Assessment, 87*, 249–258.
- Thiel, F., & Blüthmann, I. (2009). *Ergebnisse der Evaluation der lehrerbildenden Studiengänge an der Freien Universität Berlin: Sommersemester 2009*. Retrieved January 11, 2013 from http://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/schulentwicklungsforschung/downloads/Lehramtsmasterbefragung_2009.pdf?1310986750.
- Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805.
- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research, 68*, 202–248.
- Vandenberg, R. J., & Lance, C. E. (2000). A Review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70.
- Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research, 54*, 143–178.
- Wheatley, K. F. (2005). The case for reconceptualizing teacher efficacy research, *Teaching and Teacher Education, 21*, 747–766.
- Woolfolk Hoy, A., & Burke Spero, R. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education, 21*, 343–356.
- Yin, P., & Fan, X. (2003). Assessing the factor structure invariance of self-concept measurement across ethnic and gender groups: Findings from a national sample. *Educational and Psychological Measurement, 63*, 296–318.

2.6 Acknowledgements

The authors gratefully acknowledge funding support to the first author from the Center for Teacher Education at Freie Universität Berlin. We would also like to extend our sincere thanks to Martin Schultze for the ant optimization algorithm he programmed for us, and to the reviewers and the editor for providing excellent comments that greatly improved the quality of this work.

3

Study 2: I feel less confident so I quit? True changes in teacher self-efficacy predict changes in preservice teachers' intention to quit their teaching degree

Author's note: This is the author's version of a manuscript that was submitted to the journal *Teaching and Teacher Education*:

Pfitzner-Eden, F. (under review). I feel less confident so I quit? True changes in teacher self-efficacy predict changes in preservice teachers' intention to quit their teaching degree.

3 I feel less confident so I quit? True changes in teacher self-efficacy predict changes in preservice teachers' intention to quit their teaching degree

Abstract

This study explored how teacher self-efficacy (TSE) of two cohorts of preservice teachers (advanced and beginning) changes during coursework at university and during a practicum at school. Further, it determined if changes in TSE were related to changes in preservice teachers' intention to quit their degree. Changes in TSE differed between TSE dimensions, the two aspects of teacher preparation, and the two cohorts. Generally, TSE increases during the practicum were associated with decreases in preservice teachers' intention to quit. Results are discussed with regard to future directions in research and practical implications for teacher attrition.

Key words:

teacher self-efficacy, longitudinal study, preservice teachers, latent-true-change

3.1 Introduction

Research on teacher self-efficacy (TSE) has steadily increased in volume over the last thirty odd years (e.g., Klassen, Tze, Betts, & Gordon, 2011; Kleinsasser, 2014). However, previous research has mainly focused on the outcomes and correlates of TSE, employing cross-sectional designs. This research has produced a large body of data in favor of the notion that TSE is of significant importance in predicting various desirable outcomes at the student as well as at the teacher level (e.g., Klassen et al., 2011; Tschannen-Moran & Woolfolk Hoy, 2001; Vieluf, Kunter, & van de Vijver, 2013; Woolfolk Hoy & Burke Spero, 2005). Since TSE is such a fruitful construct, it seems worthwhile to study its origins. When and how does this belief form in teachers? What factors influence the development of TSE? How does the development of TSE impact on other outcomes? So far, very little research has addressed these and similar questions, and even less research has employed longitudinal designs, which are a necessity for studying the development of TSE over time. The present study offers a first longitudinal insight into the development of TSE, starting as early as at the beginning of teacher education. Latent changes in TSE are documented for two cohorts at different stages of their formal teacher education program (beginning and advanced), and in relation to different aspects of the program (coursework at university vs. practicum at school). Further, it was explored whether the changes in TSE affected changes in preservice teachers' intention to quit (ITQ) the teacher education program. Since attrition of novice teachers is of great concern to teacher educators (e.g., Ingersoll & Strong, 2011; Veenman, 1984), exploring the relationship between a change in preservice teachers' ITQ and changes in TSE could make a meaningful contribution to addressing this issue early on.

Theoretical foundations of teacher self-efficacy

The construct of teacher self-efficacy is grounded in Bandura's (1977; 1986; 1997) social cognitive theory. However, its measurement has also been influenced by Rotter's (1966) locus of control theory. In 2002, Henson characterized the first decades of research on TSE as "hampered by both construct validity and measurement problems" (p. 144). He reasons that this was mostly due to TSE being researched with instruments that were based on locus of control theory (e.g., the Gibson and Dembo scale, 1984). While locus of control theory has traditionally played a key role in TSE research (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998), for years influential researchers conducting reviews in the field of TSE (e.g., Henson, 2002; Klassen et al., 2011; Tschannen-Moran et al., 1998) have demanded a sound conceptualization and valid measurement of TSE, which is closely aligned with Bandura's construct of self-efficacy. Bandura (1977) introduced the construct of self-efficacy as the

“conviction that one can successfully execute the behavior required to produce” (p.193) a given outcome. This conviction varies with different domains of action. So in Bandura’s terms, TSE can be understood as the belief that one holds about one’s capability with regard to the domain of teaching. Self-efficacy beliefs are a powerful predictor of behavior. Self-efficacy influences behavior for example, by affecting what challenges and goals people set for themselves, what effort they invest in pursuing those, and to what degree they persevere when faced with obstacles (e.g., Bandura, 2006).

The development of teacher self-efficacy

According to Bandura (1997), self-efficacy beliefs are formed when people interpret information about their capabilities from four sources: mastery experiences, vicarious experiences, social persuasion, and physiological and affective states, whereby mastery experiences are hypothesized to have the strongest effect on the formation of self-efficacy beliefs. For preservice teachers, mastery experiences in an actual classroom would generate a strong TSE belief, because they provide authentic evidence of whether they are able to accomplish a given outcome (i.e., teaching successfully). Modeled attainments (vicarious experiences) of other teachers or preservice teachers in a practicum at school can affect preservice teachers’ TSE both by providing a social point of reference for judging their own attainments and by creating an opportunity for model learning. Being persuaded by others that one has great capability in the domain of teaching (social persuasion) would influence the formation of self-efficacy beliefs of preservice teachers most when they are being given feedback regarding their performance by a person who they regard as being competent in the field concerned (e.g., their mentor during a practicum at school). Physiological and affective states provide preservice teachers with a last source of information on their teaching capabilities, for example, through negative (e.g., stress/anxiety) or positive (e.g., enjoyment) emotions while student teaching.

In order to study the development of TSE, these general theoretical guidelines for the development of self-efficacy beliefs need to be explored empirically for TSE of preservice teachers specifically. To date, the development of TSE is an under researched area. In 2002, Henson noted that longitudinal research on the development of TSE was effectively non-existent but much needed. A recent review of teacher efficacy research of the last decade (Klassen et al., 2011) found that only six out of 218 teacher efficacy studies examined the sources of TSE, five of which used qualitative approaches. These studies provide some evidence that the proposed sources do affect the development of TSE, although the strength of the influence of each source seems to vary with a multitude of factors (e.g. professional

experience in Gabriele & Joram, 2007) and does not seem to be equally effective for each dimension of TSE. In response to the low number of studies on the sources of TSE, Klassen et al. (2011) strongly recommended focusing research efforts on longitudinal studies that explore how TSE changes over time. However, their reviewed studies on the sources were limited to inservice teachers, thus providing no insights into how the sources of TSE impact on its development while teachers are still in training. Examining changes in TSE at an early stage is of significance, because TSE is understood to be most malleable in preservice teachers (Henson, 2002; Woolfolk & Hoy, 1990), and rather resistant to change once firmly established (Bandura, 1997). There is some research regarding changes in TSE of preservice teachers, which is reviewed in the next section.

Previous research examining changes in teacher self-efficacy of preservice teachers

Few studies have explored changes in the TSE of preservice teachers. Some of these used qualitative approaches (e.g., Mulholland & Wallace, 2001), offering valuable insights into factors that might impact on the development of TSE. However, for the purpose of this study, a systematic documentation of changes in TSE for preservice teachers is of interest. Hence, only quantitative research that focused on such changes is reviewed in this section.

Following Bandura's (1997) notion that mastery experiences are the most powerful source influencing self-efficacy beliefs, most previous research exploring the development of TSE in preservice teachers has centered on practical experiences during teacher preparation (i.e., a teaching practicum at school). Using pre-post designs, these studies have reported increases in TSE during the teaching practicum (Hoy & Woolfolk, 1990; Knoblauch & Woolfolk Hoy, 2008), but also inconclusive differences (Liaw, 2009). Some studies (Garvis, Pendergast, & Keogh, 2012; Lin & Gorrell, 2001; Pendergast, Garvis, & Keogh, 2011) reported decreases in TSE, but used designs which do not allow for disentangling the effects of the practical experiences. Lin and Gorrell (2001) employed a cross-sectional design to examine two groups of preservice teachers at different stages of their teacher education program, a beginning group (at the end of their first year) and an advanced group (at the end of their third year). While their results showed that the beginning group scored higher on TSE than the advanced group, this does not carry much meaning in relation to a development of TSE, which cannot be inferred from a cross-sectional design. Pendergast et al. (2011) assessed TSE at the beginning of the first semester of a one-year graduate diploma course and again after a practicum at school²⁴ at the beginning of the second semester. Their results showed

²⁴ The authors did not state if the practicum involved student teaching.

that TSE was high at the beginning of the first semester, and was lower after the practicum. Garvis et al. (2012) assessed TSE at the beginning of the first year and the beginning of the second year of a group of preservice teachers studying a 4-year bachelor degree in primary education. Preservice teachers attended one week of observations at school prior to the second measurement. Again, the group means were high at the first measurement and lower at the second measurement. However, both studies do not allow for separating the effects of the university year from the practicum, as that would have required another measurement point before the practicum. Another study by Fives, Hamman, and Olivarez (2007), which found increases in TSE during a teaching practicum, also employed a design which could not adequately capture the TSE change during a practicum, as their two measurement occasions took place during the practicum (between week 4-6 and week 9-11 of a 12-week practicum) rather than before and after. They assessed TSE using the Teacher's Sense of Efficacy Scale (TSES; Tschannen-Moran and Woolfolk Hoy, 2001) which distinguishes three dimensions of TSE: instructional strategies, classroom management, and student engagement. The increase was most pronounced for the instructional strategies dimension of TSE. However, the three-dimensional structure of the TSES has been discussed controversially when used with preservice teachers (e.g., Duffin, French, & Patrick, 2012), and when introducing the TSES in 2001, Tschannen-Moran and Woolfolk Hoy recommended to treat their instrument as one-dimensional when used with preservice teachers.

The only study that did not examine changes in TSE that occurred during some form of practical experience at school, but rather during coursework at university, was conducted by Mergler and Tangen (2010). They examined how the TSE of preservice teachers changed during a microteaching assessment task in the first semester of a one-year graduate diploma course. Using a pre-post design, they found an increase in TSE.

To date, there are three studies which surpassed the pre-post design, measuring TSE at more than two points in time. Woolfolk Hoy and Burke Spero (2005) assessed TSE at the beginning of a postgraduate teacher education program, at the end of the student teaching practicum, and again after the first year of teaching. Overall, TSE increased during the student teaching practicum and declined during the first year of teaching. Swan, Wolf, and Cano (2011) assessed TSE at the end of a student teaching practicum, and at the end of the first, second, and third year of teaching. Overall, TSE was highest at the end of the student teaching experience and lowest at the end of the first year of teaching. Most recently, Klassen and Durksen (2014) examined the development of TSE during a teaching practicum of preservice teachers in their final year. TSE was assessed eight times at weekly intervals. Analyzing their

longitudinal data with latent growth curve models, they found a linear increase of TSE over the course of the practicum. The authors pointed out that their results were limited by their use of a global 3-item measure of TSE and that it would be advantageous if future studies inspected changes in TSE linked to particular tasks (i.e., the three TSE dimensions) with validated full versions of scales.

In sum, previous research on the changes in the TSE of preservice teachers indicates that TSE tends to increase with the student teaching experience. Yet, most studies focused on changes that occur during student teaching and have thereby been limited to samples of preservice teachers that are at an advanced stage of their teacher education. Only one study (Garvis et al., 2012) administered TSE measures early on. Results from this study indicate that there might be different patterns of changes at play, depending on the stage of teacher education. Furthermore, with the exception of one study (by Mergler & Tangen, 2010), all studies that employed pre-post designs focused on changes occurring during a practical part of teacher education at school (i.e., student teaching or observation). The study by Mergler and Tangen (2010) demonstrated that changes in TSE can also occur during a theoretical part of teacher education at university. Thus, a systematic examination of changes in TSE during the coursework at university as well as during the practicum at school seems to be warranted.

However, evidence for these findings stems from research that is generally restricted by very small sample sizes (Fives et al., 2007; Liaw, 2009; Swan et al., 2011; Woolfolk Hoy & Burke Spero, 2005), by simple pre-post-test designs, and by using different instruments to assess TSE that (a) do not closely capture the construct of self-efficacy as recommended by Bandura (2006) i.e., the Gibson and Dembo scale; and (b) do not show a clear factor structure when used with preservice teachers²⁵, so that a valid examination of the three proposed dimensions of TSE is lacking. Another significant limitation in previous research is the lack of data analysis at the intra-individual (within-subject) level in the majority of studies (Garvis et al., 2012; Hoy & Woolfolk, 1990). This makes it impossible to judge if actual changes within people occurred during the study (e.g., see McArdle, 2009 for a discussion of differences and changes).

Furthermore, with the exception of Klassen and Durksen (2014), all of the reviewed studies have determined changes based on the analysis of manifest data that are prone to measurement error to the extent that changes are unreliable (e.g., Cronbach & Furby, 1970).

While this review focused on non-subject specific TSE, there are a number of studies that investigated changes in subject-specific teaching efficacy, primarily elementary science

²⁵ For a discussion of this issue see Author et al. (2014).

teaching efficacy. Results from this body of research have shown increases during a teaching practicum (e.g., Charalambous, Philippou, & Kyriakides, 2008; Swars, Bart, Smith, & Tolar, 2007) and during university courses (Palmer, 2006; Richardson & Liang, 2008), but also no change during a teaching practicum (Plourde, 2002). However, the above-mentioned limitations also apply to this line of research. Moreover, it is unclear to what degree these different constructs (TSE vs. elementary science/mathematics teaching efficacy) are related and can thus be considered to inform this study.

Teacher self-efficacy and retention

Generally, retaining teachers in the profession is of great concern. While attrition figures worldwide differ in magnitude (e.g., between 5% and 30%, reported in Macdonald, 1999), it is undisputed that a great proportion of teachers leave the profession prematurely and voluntarily, a phenomenon termed the “revolving door” by Ingersoll (e.g., 2001; 2002). While a large body of research has examined factors that contribute to the retention or attrition rates of inservice teachers (e.g., Borman & Dowling, 2008; Macdonald, 1999), studies exploring those rates and factors for preservice teachers are less common. It is particularly unclear what proportion of preservice teachers actually complete their degree and subsequently enter teaching, and what factors affect their decision to do so. With regard to attrition, a study conducted by Swan et al. (2011), found preservice teachers with higher TSE to be more likely to enter the teaching profession than preservice teachers that showed lower TSE after their student teaching experience. Moreover, only 50% of their sample entered the teaching profession after their student teaching. A sobering result in light of the attrition rates of inservice teachers, highlighting the issue of retaining teachers even at this early stage. Most recently, a meta-analysis by Chesnut and Burley (2015) shed some light on the role that TSE might play with regard to commitment of preservice and inservice teachers. The authors reported an effect size of $r = .30$ for the association between TSE and commitment to the teaching profession for preservice teachers, and $r = .32$ for inservice teachers. However, their inclusion criteria for what constituted commitment were quite broad and included variables such as burnout. In total they identified 14 studies with preservice teachers, of which none reported a correlation between TSE and specifically ITQ. Interestingly, Chesnut and Burley (2015) also found that studies, which used TSE measures deemed conceptually accurate showed greater associations between commitment and TSE than studies that used measures deemed conceptually inaccurate (e.g., the Gibson and Dembo Scale). This adds further support to the above-mentioned limitations to the assessment of TSE in previous research on the development of TSE. However, all studies reviewed in this meta-analysis used a cross-

sectional approach and thus the reported associations are based on correlations between the static levels of both variables. The same is true for the study by Swan et al. (2011), even though they did use a longitudinal design, they only compared static TSE levels across groups. Thus, to date no previous study has demonstrated a link between changes in TSE and changes in ITQ. Finding this ‘change-change association’ can be viewed as a key prerequisite to any efforts targeted at changing the predictor (i.e., TSE) in order to produce a behavior change (in ITQ) (Reuter et al., 2010). Nonetheless, targeting interventions at TSE has routinely been recommended in previous research which only documented ‘level-level associations’ (e.g., Klassen & Chiu, 2011), despite the fact that these associations cannot be equated to a change-change association. For example, Klassen and Chiu (2011) found an association between TSE and stress levels in preservice teachers, while Klassen and Durksen (2014) found no association between TSE changes during a teaching practicum and changes in perceived stress during the practicum.

The current study

This study is designed to address several gaps identified in the preceding review of research in this area. Subsequently, the purpose of the current study is twofold: (1) to document the direction and magnitude of latent true changes for three dimensions of TSE, occurring at two different stages of teacher education (at the beginning and at an advanced stage) and during different aspects of the teacher education program (coursework at university vs. practicum at school); (2) to assess if there is a relationship between changes in TSE and changes in preservice teachers’ ITQ their teaching degree. For these purposes, data was collected, using a longitudinal design, at three points in time. Latent true change (LTC) modeling (e.g., Steyer, Eid, & Schwenkmezger, 1997) was applied to analyze latent changes at an intra-individual level over time. There are two main advantages of a LTC approach over a manifest approach. First, latent changes represent true change in the construct that is free of measurement error. Second, the latent approach allows for a direct examination of longitudinal measurement invariance (for a detailed description see e.g., Vandenberg & Lance, 2000), which is a key prerequisite for interpreting changes over time so as not to compare “apples and pears”. Consequently, latent changes are only analyzed if the scale is found to assess the underlying constructs in exactly the same manner over all measurement occasions. This is of particular concern when a construct is assessed repeatedly because change can be attributed either to a true change of the construct under examination, or to changes in participants’ perceptions of either the target construct or the scale used to measure it, thus rendering an interpretation of changes invalid (Meade & Lautenschlager, 2004; Millsap & Hartog, 1988).

TSE was assessed using an adapted version of the TSES, which allows for a reliable and valid 3-dimensional measurement of TSE for beginning and advanced preservice teachers (Author et al., 2014), and is closely aligned with Bandura's (2006) recommendations. Since there is some indication from past research (Fives et al., 2007) that the three dimensions of TSE change differently over time, and researchers (e.g., Veenman, 1984) suggest that preservice teachers are particularly apprehensive about classroom management, it seems valuable to explore changes in TSE separately for the three dimensions.

The following research questions were explored:

1. Are there any changes in the three dimensions of TSE during the coursework at university (between T1 and T2) and during the practicum at school (between T2 and T3)?
2. Are changes in TSE and ITQ related to their respective initial levels (e.g., Do participants with higher or lower levels of TSE at the beginning of the study experience less or more change over time)?
3. Are TSE and ITQ changes similar across individuals?
4. What is the relationship between changes in TSE and changes in preservice teachers' ITQ?

3.2 Method

Participants and procedure

Data in this study stem from a longitudinal project on the development of TSE in preservice teachers at a large German university. Participants were preservice teachers in two cohorts, which were enrolled in a teacher education program at a large university in Germany, representing a broad range of academic subjects (e.g., mathematics, biology, physics, German, English, history, chemistry). This program encompasses a three-year Bachelor degree followed either by a one-year Master degree (elementary/middle school teachers), or a two-year Master degree (secondary school teachers).

The first cohort (Sample 1) comprised 559 beginning preservice teachers in total. The second cohort (Sample 2) totaled 352 preservice teachers at an advanced stage of their studies. Data for Sample 1 were collected in the first month of participants' first year of undergraduate study in October (Time 1; $N = 438$), at the end of their first year of undergraduate study before undertaking their first practicum at school in July the following year (Time 2; $N = 359$), and at the beginning of their second year of undergraduate study in October, after they had completed a one-month practicum at school (Time 3; $N = 226$). The main purpose of this

first practicum is for preservice teachers to observe the classroom of experienced teachers in school. Data for Sample 2 were collected at an advanced stage of their studies (at least third year with 80% in their Master study) at the beginning of the semester in October simultaneously to data collection in Sample 1 (Time 1; $N = 274$), at the end of that semester before undertaking their teaching practicum at school in February the following year (Time 2; $N = 255$), and at the beginning of their next semester in April after they had completed a one-month teaching practicum (Time 3; $N = 143$). The main rationale of this teaching practicum is for advanced preservice teachers to prepare lessons and teach classes independently with an experienced teacher observing them and giving them feedback. Prior to the commencement of the study, all subjects in Sample 2 had completed their observation practicum during their first year of undergraduate study. Measures were administered in a paper-pencil format in seminars at Time 1 and 2, and online at Time 3.

Sample attrition

Based on enrollment data, at T1 the response rate was 84% for Sample 1 and 83% for Sample 2. Nonetheless, self-selection bias at T1 can be assumed to be negligible due to students missing at random during data collection in seminars, and only five preservice teachers in total refusing to participate in the study. The response rates are comparable to other studies investigating preservice teachers (e.g., 76% for beginning preservice teachers reported in Garvis et al., 2012; 77% for advanced preservice teachers reported in Klassen & Durksen, 2014). Retention rates between T1 and T2 were 60% for Sample 1 and 72% for Sample 2. Retention rates between T2 and T3 were 50% for Sample 1 and 45% for Sample 2. Retention rates across all three occasions of measurement were 32% for Sample 1 and 36% for Sample 2. While retention in general was disappointing, the average retention rates in this study compare favorably to those reported in similar studies (55% in Sample 1 and 59% in Sample 2 as compared to 47% in Malmberg, 2008; 46% in Klassen & Durksen, 2014).²⁶

Table 1 shows descriptive statistics for the longitudinal samples. To test for attrition bias, T1 characteristics were compared between the participants that completed measures at T1 and T2 and those that dropped out between T1 and T2. No significant differences were found for Sample 1 (all $ps > .05$) for age and TSE variables. Males and participants with a

²⁶ However, it is difficult to judge what constitutes a similar study as there are a great number of variables that can differ between studies (e.g., type of participants, number of measurement occasions, interval between measurement occasions, assessment format). For example, in Klassen and Durksen (2014) 36% of the initial sample participated across all eight measurement occasions but only over seven weeks. In Malmberg (2008) only 8% of the initial sample participated across all five measurement occasions but over a period of four years. Furthermore hindering a comparison is the fact that most of the studies reviewed earlier did not report retention or attrition rates.

higher ITQ mean at the first measurement were more likely to drop out. No significant differences were found for Sample 2 (all $ps > .05$) for gender, age, TSE variables and ITQ. In a second step, T2 characteristics were compared between the participants that completed measures at T2 and T3 and those that dropped out between T1 and T2. No significant differences were found for Sample 1 (all $ps > .05$) for all TSE variables. Males, older participants and participants with a higher ITQ mean were more likely to drop out. Again, no significant differences were found for Sample 2 (all $ps > .05$) for gender, age, TSE variables and ITQ.

Measures

The German version of the Scale for Teacher Self-Efficacy (STSE) was used to assess TSE in this study (see Author et al., 2014, for items, a description of psychometric properties and construct validity of the scale). This scale is an adaptation of the Teacher's Sense of Efficacy Scale (TSES) by Tschannen-Moran and Woolfolk Hoy (2001). It comprises 12 items, and shows a stable three-factor structure for preservice teachers with 4 items each measuring TSE for instructional strategies, classroom management, and student engagement. Participants rate their confidence in their capabilities on a 9-point response scale ranging from 1 (*not at all certain can do*) to 9 (*absolutely certain can do*). Internal consistency was satisfactory to good (see Table 2).

Table 1. Longitudinal sample description

Variable at T1/T2	T1	T1 & T2	T1 not T2	T2 & T3	T2 not T3
<i>Beginning cohort</i>	<i>n</i> = 438	<i>n</i> = 263	<i>n</i> = 175	<i>n</i> = 178	<i>n</i> = 181
Female	62.6%	66.9%	56.0%	70.1%	59.5%
Mean Age (SD)	22.82 (4.94)	22.46 (4.86)	23.35 (5.00)	22.74 (4.75)	24.70 (5.36)
Mean TSE Instruction (SD)	6.78 (0.99)	6.81 (0.98)	6.75 (1.01)	6.79 (1.08)	6.69 (1.06)
Mean TSE Classroom Management (SD)	6.41 (1.46)	6.38 (1.50)	6.43 (1.40)	6.22 (1.44)	6.02 (1.47)
Mean TSE Student Engagement (SD)	6.85 (1.13)	6.76 (1.20)	6.98 (1.02)	6.87 (1.14)	6.75 (1.26)
Mean ITQ (SD)	2.54 (1.49)	2.39 (1.51)	2.85 (1.65)	2.31 (1.58)	2.79 (1.91)
<i>Advanced cohort</i>	<i>n</i> = 274	<i>n</i> = 198	<i>n</i> = 76	<i>n</i> = 115	<i>n</i> = 140
Female	68.6%	68.8%	67.1%	73.9%	62.9%
Mean Age (SD)	25.93 (4.82)	25.60 (4.32)	26.66 (5.85)	25.76 (4.65)	26.35 (4.42)
Mean TSE Instruction (SD)	6.46 (1.11)	6.42 (1.20)	6.55 (0.92)	6.63 (1.06)	6.58 (1.09)
Mean TSE Classroom Management (SD)	6.00 (1.60)	5.95 (1.60)	6.10 (1.63)	6.26 (1.35)	6.16 (1.44)
Mean TSE Student Engagement (SD)	6.47 (1.20)	6.47 (1.20)	6.47 (1.19)	6.70 (1.01)	6.60 (1.23)
Mean ITQ (SD)	2.54 (1.49)	2.58 (1.47)	2.45 (1.53)	2.60 (1.54)	2.69 (1.57)

A 4-item measure was used to assess the intention to prematurely quit (ITQ) the teacher education program. This measure was developed and tested as part of the study program evaluation (Thiel & Blüthmann) at the German university where data were collected. Participants rated their ITQ on a 9-point response scale ranging from 1 (*not at all true*) to 9 (*exactly true*). The items were: *If I had a good alternative, I would quit this program. If I had known what was awaiting me, I would have studied something else. During the course of my studies, I mostly lost interest in it. I am seriously considering quitting this program.* Internal consistency coefficients for the ITQ scale were acceptable for such a brief measure that covers quite diverse aspects of ITQ (see Table 2).

Table 2. Internal consistency coefficients (α) for the employed measures

Scale	T1	T2	T3
<i>Beginning cohort</i>			
TSE Instruction	.68	.75	.79
TSE Classroom Management	.91	.91	.94
TSE Student Engagement	.80	.82	.88
Intention to Quit	.76	.82	.82
<i>Advanced cohort</i>			
TSE Instruction	.74	.76	.74
TSE Classroom Management	.93	.92	.93
TSE Student Engagement	.78	.83	.81
Intention to Quit	.72	.74	.67

Note: TSE = Teacher Self-Efficacy

Analysis

The longitudinal data were analyzed using latent true change (LTC) modeling as introduced by Steyer et al. in 1997. Applying a structural equation modeling (SEM) approach, LTC models allow for the examination of intra-individual changes in the true scores of repeatedly measured constructs. Changes can be analyzed in relation to the baseline measure (i.e., T3-T1), or in relation to measurements taken at neighboring occasions of measurement (i.e., T2-T1 and T3-T2) (see Steyer, Partchev, & Shanahan, 2000 for a detailed description). For the purpose of this study, LTC was examined using the neighbor model, in order to separately determine changes that occurred during the coursework at university (between T1 and T2) and changes that occurred during the practicum at school (between T2 and T3). The latent change

scores in the model represent a latent (not measured) construct that is defined as the part of the score of the neighboring measurement (e.g., at T2) that is not part of the score at the preceding measurement (e.g., at T1) (e.g., McArdle, 2009). The latent change score can then be treated as any other latent construct in SEM, thus allowing for an examination of its mean, variance and covariance with initial scores. Measured change scores on the other hand have long been recognized as leading to “fallacious conclusions” (Cronbach & Furby, 1970) because they are systematically related to measurement error.

An imperative prerequisite for the modeling of LTC is the invariance of factor loadings and intercepts (scalar invariance) across all occasions of measurement (e.g., Steyer et al., 1997). Before proceeding to the LTC model, a series of invariance tests was run to establish separate measurement models²⁷ for each dimension of TSE and for ITQ that fulfilled the requirement of scalar invariance in each cohort. A change in model fit was evaluated using the traditional χ^2 -difference test for nested models. The requirement of measurement invariance over time was fulfilled for all latent constructs in both samples. Model fit of the invariance models is identical to the fit of the latent change models reported in the results section.

The LTC model was computed using *Mplus* Version 7 (Muthén & Muthén, 1998-2012) and the estimation method used was Full Information Maximum Likelihood (FIML; see Wothke, 2000 for more information) recommended by Steyer et al. (2000) as the method of choice for longitudinal studies with a large amount of missing data. Measurement errors were allowed to correlate across time (as mentioned in Steyer et al., 2000), but only on neighboring occasions and not between the first reference indicators whose factor loadings were fixed.

Model fit was assessed using a multi index strategy as recommended by prominent methodologists (e.g., Hu & Bentler, 1999): the Comparative Fit Index (CFI) representing an incremental fit index, less affected by sample size, with a cut-off value close to .95, and the Standardized Root Mean Square Residual (SRMR) representing an absolute fit index with a cut-off value close to .08 were used in combination as they were recommended to be particularly suited to small sample sizes (see Hu & Bentler, 1999, p. 28); further the χ^2 test-statistic of exact fit is reported but not used to examine model fit, as it has been found to be a very sensitive indicator of model fit, whose assumptions are not often met in practical applications (e.g., Schermelleh-Engel, Moosbrugger, & Müller, 2003).

²⁷ Indicators of the measurement models were item parcels, whereby two items were parceled to one indicator, resulting in two indicators for each latent construct. Items were parceled considering item-to-construct balance (see Little, Cunningham, Shahar, and Widaman (2002) for a discussion of merits and pitfalls as well as techniques of item parceling).

In order to judge the practical relevance of the latent changes, Cohen's (1992) d is reported as an indicator of effect size. As there are no previous effect sizes derived using the same design (longitudinal), outcome (3 TSE dimensions), and method (latent change) as this study, it is unclear what magnitude of d constitutes a meaningful effect. Consequently, the effect sizes derived here can be regarded as a first benchmark in this area. Nonetheless, results from a recent meta-analysis in the area of health psychology (Ashford, Edmunds, & French, 2010) could serve as a very distant point of reference. Ashford et al. (2010) found a mean d of .16 for intervention-induced changes in self-efficacy. For the purpose of this study, I consider an equal or larger sized effect as substantial. The significance level is set at $p < .05$ for the whole study.

3.3 Results

Goodness of Fit

Goodness of fit indices for the separate LTC models for each TSE dimension, for ITQ, and for the LTC models predicting ITQ changes from TSE changes in each dimension are presented in Table 3.

Beginning preservice teachers: the LTC model for instruction and classroom management showed a very good fit, and a good fit for student engagement. The LTC model for ITQ showed a very good fit. The LTC model, whereby the latent changes in the three factors of TSE predict latent changes in ITQ, showed a good fit for all three TSE dimensions.

Advanced preservice teachers: the LTC model for instruction and student engagement showed a good fit, and a very good fit for classroom management. The LTC model for ITQ showed an acceptable fit. The LTC model, whereby the latent changes in the three factors of TSE predict latent changes in ITQ, showed an acceptable fit for instruction, and a good fit for classroom management and for student engagement.

Table 3. Model fit indices for the latent true change models

Latent Change Model	χ^2	df	N	<i>p</i>	CFI	SRMR
<i>Beginning cohort</i>						
Instruction TSE	10.5	8	559	>.05	.996	.031
Classroom Management TSE	17.9	8	559	<.05	.993	.036
Student Engagement TSE	28.4	8	559	<.05	.979	.055
ITQ	14.9	8	559	<.05	.991	.039
Instruction TSE changes predict changes in ITQ	71.6	47	559	<.05	.983	.031
Classroom Management TSE changes predict changes in ITQ	76.4	47	559	<.05	.987	.046
Student Engagement TSE changes predict changes	84.4	47	559	<.05	.979	.052
<i>Advanced cohort</i>						
Instruction TSE	22.9	8	352	<.05	.963	.046
Classroom Management TSE	14.7	8	352	>.05	.993	.026
Student Engagement TSE	22.6	8	352	<.05	.973	.051
ITQ	31.2	8	352	<.05	.959	.058
Instruction TSE changes predict changes in ITQ	94.7	47	352	<.05	.953	.065
Classroom Management TSE changes predict changes in ITQ	81.9	47	352	<.05	.978	.050
Student Engagement TSE changes predict changes	92.5	47	352	<.05	.960	.060

Note: ITQ = Intention to quit; CFI = comparative fit index; SRMR = standardized root mean square residual

Changes over time

The unstandardized means, variances and correlations of the LTC model for each cohort are presented in Table 4. In the cohort of the beginning preservice teachers the average latent change during the university year (T2-T1) is negligible for two TSE dimensions: instruction ($p = .549$; $d = -.04$) and student engagement ($p = .775$; $d = -.02$). TSE for classroom management however shows a significant ($p = .000$; $d = -.22$) decrease during the university year. In the cohort of the advanced preservice teachers the average latent change in TSE during the term (T2-T1) is significantly positive for two dimensions: classroom management ($p = .016$; $d = .18$), and instruction ($p = .047$; $d = .17$). TSE for student engagement however shows a non-significant ($p = .068$; $d = .14$) increase during the term. In the cohort of the beginning preservice teachers TSE increases on average significantly during the practicum (T3-T2) for two dimensions: classroom management ($p = .007$; $d = .20$), and instruction ($p =$

.005; $d = .24$). The TSE increase in student engagement is not significant ($p = .222$; $d = .09$). Regarding the change during the practicum (T3-T2) in the advanced cohort, there is a significant average increase in classroom management TSE only ($p = .002$; $d = .28$), while the increases in instruction and student engagement TSE are non-significant ($p = .082$; $d = .19$ and $p = .726$; $d = .03$, respectively). On average, ITQ does not change significantly for both cohorts either during the university year/term (Sample 1: $p = .65$; $d = .03$; Sample 2: $p = .80$; $d = .02$), or during the practicum (Sample 1: $p = .18$; $d = -.10$; Sample 2: $p = .74$; $d = -.03$).

Table 4. Unstandardized means, variances and correlations of the latent variables for the neighbor model (Full Information Maximum Likelihood Estimates)

Latent Variable		η_1	$\eta_2 - \eta_1$	$\eta_3 - \eta_2$
<i>Beginning cohort</i>				
Instruction	Means	6.741	-0.037	0.205*
	Variances	0.770*	0.841*	0.747*
	η_1	1.000		
	$\eta_2 - \eta_1$	-0.416*	1.000	
	$\eta_3 - \eta_2$.158	-.473*	
Classroom Management	Means	6.412	-0.295*	0.279*
	Variances	2.047*	1.816*	2.011*
	η_1	1.000		
	$\eta_2 - \eta_1$	-.467*	1.000	
	$\eta_3 - \eta_2$.156	-.491*	1.000
Student Engagement	Means	6.805	-0.021	0.107
	Variances	1.391*	1.646*	1.512*
	η_1	1.000		
	$\eta_2 - \eta_1$	-.512*	1.000	
	$\eta_3 - \eta_2$	-.012	-.374*	1.000
Intention to Quit	Means	2.919	0.051	-0.177
	Variances	2.971*	3.048*	2.889*
	η_1	1.000		
	$\eta_2 - \eta_1$	-.301*	1.000	
	$\eta_3 - \eta_2$	-.160	-.341*	1.000

Latent Variable		η_1	$\eta_2 - \eta_1$	$\eta_3 - \eta_2$
<i>Advanced cohort</i>				
Instruction	Means	6.495	0.145*	0.147
	Variances	0.893*	0.773*	0.584*
	η_1	1.000		
	$\eta_2 - \eta_1$	-.491*	1.000	
	$\eta_3 - \eta_2$	-.176	-.387*	1.000
Classroom Management	Means	6.008	0.213*	0.330*
	Variances	2.445*	1.430*	1.415*
	η_1	1.000		
	$\eta_2 - \eta_1$	-.547*	1.000	
	$\eta_3 - \eta_2$	-.151	-.412*	1.000
Student Engagement	Means	6.442	0.164	0.031
	Variances	1.533*	1.450*	0.941*
	η_1	1.000		
	$\eta_2 - \eta_1$	-.506*	1.000	
	$\eta_3 - \eta_2$	-.143	-.452*	1.000
Intention to Quit	Means	2.771	0.028	-0.045
	Variances	3.132*	1.606*	2.135*
	η_1	1.000		
	$\eta_2 - \eta_1$	-.302*	1.000	
	$\eta_3 - \eta_2$	-.002	-.608*	1.000

Note: * $p < .05$.

An inspection of the correlations among the latent variables reveals that in both cohorts and across all three factors of TSE the change between T1 and T2 is significantly (all $ps = .000$) negatively correlated with the initial level of TSE at T1. Consequently, for the beginning cohort, which shows an overall decrease in TSE during the university year, this decrease is greater, when the initial level of TSE was high at the beginning of their studies. For the advanced cohort, which shows an overall increase in TSE during the university term, this increase is greater, the lower the initial level of TSE was at the beginning of that term (all $ps = .000$). Across all three dimensions of TSE the change that occurs during the practicum is not significantly (all $ps > .05$) correlated with the initial level of TSE at T1 in either cohort. Thus, the overall increase in TSE, which takes place during the practicum, does not depend on the

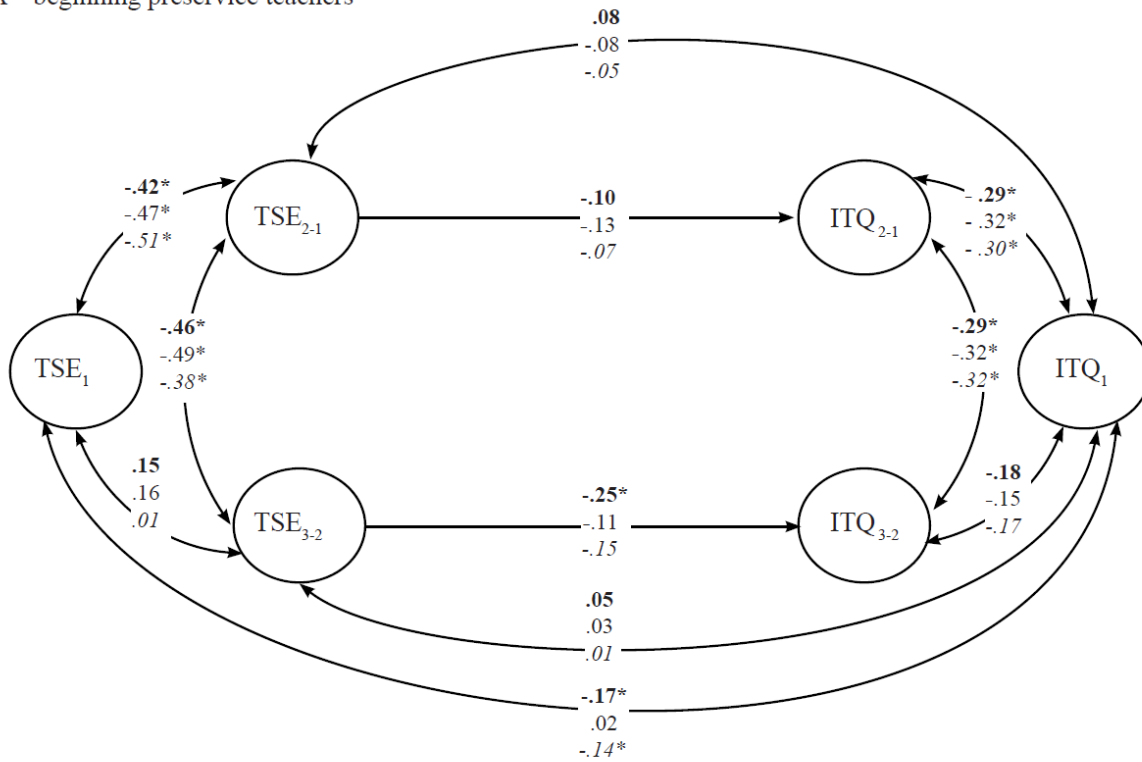
initial level of TSE. Regarding ITQ, a similar pattern of correlations holds. In both cohorts, the change in ITQ during the term/university year is significantly (beginning preservice teachers: $p = .002$; advanced preservice teachers: $p = .017$) negatively correlated with the initial level of ITQ, while the change during the practicum is not ($p > .05$). Thus, between T1 and T2, ITQ increased more for preservice teachers who rated their ITQ lower at the beginning of the study.

All variances of the latent changes were significant (all $ps = .000$). Thus, there is considerable inter-individual variability in the latent changes of TSE and ITQ in each cohort.

Changes in teacher self-efficacy predicting changes in intention to quit

The structural model of the latent change model, whereby the latent changes in TSE predict the latent changes in ITQ, is depicted in Figure 1. Latent changes during the term/university year in either dimension of TSE did not significantly predict changes in ITQ that occurred between T1 and T2 for both cohorts (all R^2 's $< .02$). In contrast, latent changes in TSE that occurred during the practicum (between T2 and T3) predicted changes in ITQ during the practicum negatively, but did so significantly only for TSE for instruction in the beginning group ($R^2 = .064$), and TSE for instruction and student engagement in the advanced group ($R^2 = .059$; $R^2 = .064$; respectively). The TSE change in classroom management and student engagement did not significantly predict changes in the ITQ of the beginning preservice teachers ($R^2 = .012$; $R^2 = .022$; respectively), and changes in TSE for classroom management did not significantly predict changes in the ITQ of the advanced preservice teachers ($R^2 = .028$).

A – beginning preservice teachers



B – advanced preservice teachers

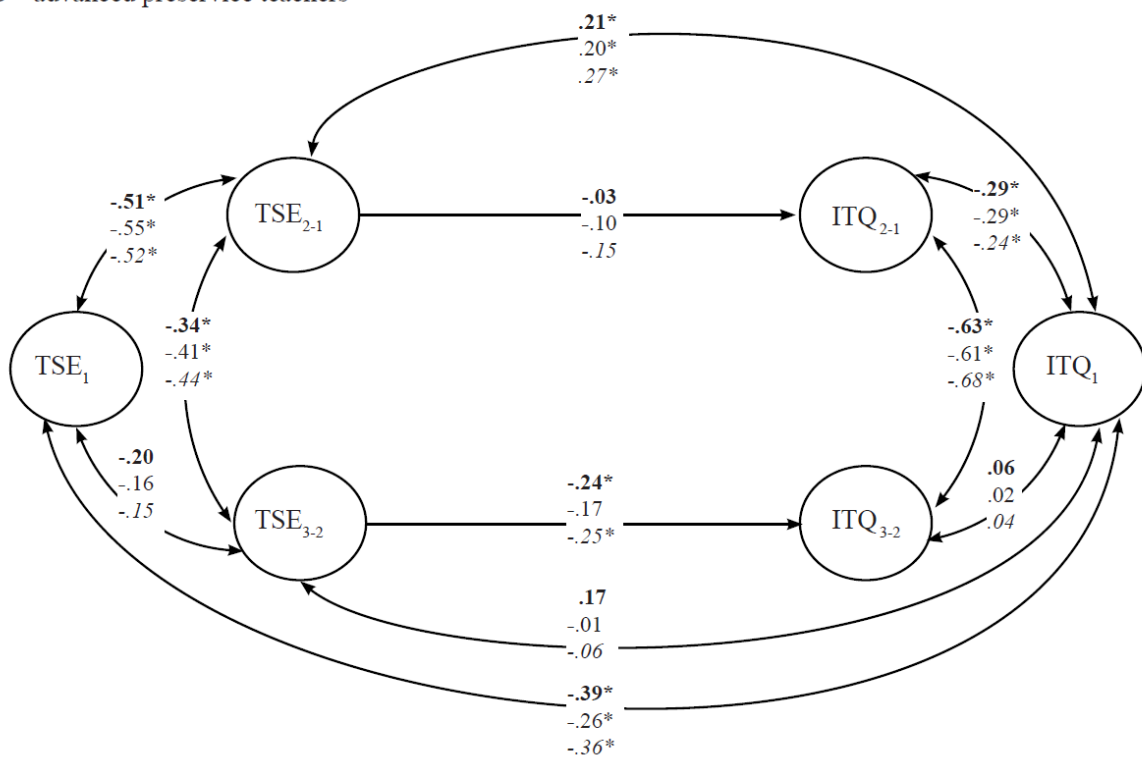


Figure 1. Structural latent true change model with standardized path coefficients for each group of pre-service teachers

Note: Bold printed coefficients refer to instructional strategies, normal print refers to classroom management, and italic print refers to the student engagement dimension of teacher self-efficacy. TSE = teacher self-efficacy. ITQ = intention to quit. * p < .05.

3.4 Discussion

This study has offered a first systematic documentation of changes in three dimensions of TSE, during two different program aspects and at two stages of teacher education. In the following, the results are discussed and placed in context.

There were some significant changes in TSE during the coursework at university in both cohorts. For the beginning preservice teachers, however, this change only occurred for one dimension of TSE: classroom management, which declined, while the other two dimensions remained unchanged. For the advanced preservice teachers TSE increased for classroom management and instruction, while student engagement remained unchanged. So why do TSE changes of opposite directionality occur during the coursework at university? What is the difference between the two cohorts? While there might be many differences between the two cohorts, the one that matters most in the context of TSE is their difference in professional experience. At the start of the study, the TSE beliefs of the advanced cohort would have been influenced by at least three years of teacher education, conveying both theoretical and practical experience. In comparison, the beginning cohort would have had no professional experience at the beginning of the study. They might have based their initial TSE beliefs, for example, on experiences they gathered from a student perspective during their own schooling, a phenomenon termed *apprenticeship of observation* by (Lortie, 1975). But why would TSE decline only for classroom management during the first university year? It might be that the beginning preservice teachers had either no or a very limited concept of classroom management and thus overestimated their TSE for this dimension at the start of the study. In their limited experience of teaching, most likely solely derived from a student perspective in a classroom, they might have had ample opportunity to observe teachers' instruction and student engagement, but might not have recognized teachers' classroom management skills. By the end of their first year of teacher education (T2) they might have had a much greater understanding of what (difficulties) classroom management actually encompasses, and have thus lowered their TSE beliefs for this dimension. This phenomenon termed *response shift* (e.g., Howard & Dailey, 1979) is well documented in the evaluation of treatment or training interventions. More support for explaining this phenomenon might be provided by Weinstein (1988), who documented *unrealistic optimism* of preservice teachers prior to any teaching experience. This unrealistic optimism was particularly pronounced for classroom management. Preservice teachers could be prone to unrealistic optimism at the beginning of the teacher preparation program resulting in inflated TSE for classroom management, which subsequently decreased during the university year. A 'miscalibration' of preservice teachers'

TSE, possibly due to incomplete knowledge, has also been described by Klassen and Chiu (2011). However, these are only speculations that would need to be further explored in future studies. In sum, what seems clear is that there are meaningful differences in the way the three dimensions of TSE change during the coursework at university at different stages of teacher education.

There were some significant changes in TSE during the practicum at school in both cohorts. For the beginning preservice teachers, this change was positive and most pronounced for TSE for classroom management and instruction, while TSE for student engagement did not change significantly. For the advanced preservice teachers, TSE increased significantly for classroom management. TSE also increased substantially (but non-significantly) for instruction, and did not change for student engagement. Taken together, these results show that TSE increased similarly in both cohorts during their practicum experience, despite the fact that the practicum experience had a different focus in each cohort (i.e., observation vs. teaching). So it seems that even mere vicarious experiences can elicit a similar change in TSE for the beginning preservice teachers, as mastery experiences do for the advanced preservice teachers. This result is in line with what would be predicted by Bandura's (e.g., 1997) theory of the sources. While Bandura stated that "vicarious experiences are generally weaker than direct ones" (1997, p.88), he also proposed that vicarious experiences can have a similar effect on self-efficacy beliefs under conditions of little prior direct experience regarding one's own capability, which would have been the case for the beginning preservice teachers. Again, the TSE changes that occurred during the practicum differed across TSE dimensions, with student engagement showing no change over time in both cohorts. While the current data does not allow for an explanation of this pattern, it documents it. Regarding the engagement of students in class, the practicum experience simply might not have offered an opportunity to evaluate one's own capabilities, either by vicarious or direct experience. Capabilities in the classroom management and instruction domain might be more salient to observers in the case of the beginning cohort and more salient to oneself when teaching in the case of the advanced cohort. It may be that for a change in student engagement TSE to occur, a longer time-frame is needed. Possibly one month of practicum is not enough time to effect change on this dimension.

Changes in all dimensions of TSE but also in ITQ during the practicum did not depend on the initial level of the respective construct at the first measurement. However, changes during the university year/term did depend on the initial level of each respective construct. This might indicate that while changes in TSE and ITQ during the coursework at university

are still very much influenced by their initial levels, there are different factors at play during the practicum at school that exert an influence over and above initial levels of TSE and ITQ.

Furthermore, the results show that changes occurring during the coursework at university as well as during the practicum vary considerably between individuals. This result provides a solid basis for targeting further efforts at identifying factors that can predict such variability.

With regard to the ‘change-change’ association, the results show that TSE and ITQ changes during the university course were not related in either cohort, but changes that occurred during the practicum were, albeit not for all TSE dimensions. All observed effects followed the expected negative direction, whereby an increase in TSE meant a decrease in ITQ. The results also show, somewhat unexpectedly, that while the greatest changes during the practicum occurred for classroom management TSE, these changes did not affect changes in preservice teachers’ ITQ. This indicates that a change in TSE for instruction during the practicum seems to be more important with regard to preservice teachers’ ITQ than a change in TSE for classroom management. A change in TSE for student engagement during the practicum, on the other hand, seems to be only of relevance to the advanced preservice teachers’ ITQ. This might be because the beginning preservice teachers had no opportunity to examine their own student engagement, since they only observed classes during their practicum, and maybe for this particular dimension of TSE the effect on ITQ cannot be transported vicariously.

Taken together, this study has provided five important findings contributing new insights to the field of TSE research. First, changes in TSE of preservice teachers occur during both aspects of a teacher education program: the coursework at university and the practicum at school. Second, these changes differ by TSE dimension, with classroom management generally showing the greatest changes. Third, the patterns of TSE change are different at different stages of teacher education (i.e., beginning vs. advanced). Fourth, TSE and ITQ changes during the coursework at university, but not during the practicum, depend on their respective initial levels. Fifth, changes in TSE during the practicum experience, but not during the coursework at university, are associated negatively with changes in preservice teachers’ intention to quit their teaching degree.

Limitations and future directions

While this study addressed many shortcomings of previous research in this area (e.g., longitudinal design, latent-change-analysis approach, greater sample size, 3-dimensional TSE instrument for preservice teachers), there are some limitations that need to be pointed out.

Sample mortality is usually a problem in longitudinal designs. This study is no exception. In general, sample mortality effectively reduced the sample size over time. This has repercussions for the statistical power of the study, making it harder to detect small and medium sized effects. Moreover, the longitudinal samples of the beginning cohort showed some attrition bias with regard to gender, ITQ and age. The gender bias should not have affected the results, as Klassen and Durksen (2014) found no relationship between gender and TSE changes. Similarly, the age difference between the sample that completed T2 and T3 and those that did not should not have impacted on the results, as according to Bandura (e.g., 1997) TSE should not be related to age but rather to (professional) experience. A greater tendency of beginning preservice teachers with higher ITQ scores to drop out might have resulted in underestimated latent change mean scores for ITQ and an underestimation of the relationship between TSE and ITQ changes in this cohort.

Unfortunately, a causal relationship between the changes in TSE and ITQ could not be established in the longitudinal design used in this study, as the changes happened concurrently between T2 and T3. Thus, a basic prerequisite for causal inferences (e.g., Cohen & Cohen, 1983) was not fulfilled. To overcome this issue, future studies could employ a design similar to the one in the Klassen and Durksen (2014) study, whereby measurements are taken at weekly intervals. Such an approach might reveal if TSE changed prior to ITQ.

Furthermore, the generalizability of the results from this study is limited. The examined samples represented only one teacher education program in one country. However, Klassen et al. (2009) demonstrated good intercultural generalizability of the TSE construct for inservice teachers and Author et al. (2014) for preservice teachers, which seems promising in terms of generalizability.

The longitudinal perspective of this study was limited to only one year for the beginning cohort and to only six months for the advanced cohort. A longer timeframe would allow for an even more systematic mapping of the changes that occur from the beginning of preservice teacher preparation through to the first years of service. Nonetheless, multiple year-long longitudinal studies are very resource intensive and thus longitudinal designs that use more than one cohort at different stages of teacher education might be a feasible alternative, keeping in mind that cohort effects will limit the interpretability of the results.

This study can be regarded as taking a first step toward a thorough examination of changes in TSE and their relationship to changes in preservice teachers' ITQ. At this stage, the examination was limited to a description of the changes that occur early on and later during university teacher preparation. In addition it was demonstrated that such changes are

meaningfully associated with preservice teachers' changes in their intention to quit their study program, providing a necessary basis for further exploration of predictors for such a relationship. Based on the present results, future studies on the relationship between TSE and ITQ could, for example, focus on the period during the practicum and even on certain dimensions of TSE (e.g., instruction for beginning preservice teachers). Since the present data showed large inter-individual variability in changes, future studies could also focus on identifying predictors and mediators that can explain this variability. To date, the literature (e.g., Bandura, 1997; Woolfolk Hoy & Burke Spero, 2005) offers a large pool of such factors. The challenge will be to determine the significance of each factor and their interplay in relation to the different stages of teacher education.

Future studies could also explore some avenues opened up in the interpretation of the present findings. For example, in a similar design, knowledge in classroom management, instruction and student engagement could be assessed in addition to TSE in order to determine if differential patterns of change among the three TSE dimensions are related to differences in knowledge conveyed through the coursework. Previous research has mostly focused on changes associated with the practicum experience, yet this study and the study by Mergler and Tangen (2010) have demonstrated that TSE is also malleable by university course content. More research could be directed at examining under which conditions coursework impacts positively or negatively on TSE.

Implications

Lastly, I would like to highlight some meaningful implications both for theory development and practice associated with the present results. In TSE research, there is some debate about the ability of beginning preservice teachers to distinguish the three dimensions of TSE (e.g., see Duffin et al., 2012). The differential patterns of change exhibited by the three dimensions of TSE in the present study provide a strong rationale for examining these dimensions separately in future studies, even as early as in beginning preservice teachers. Interestingly, if the TSE of the beginning preservice teachers had only been measured at the beginning and the end of the university year, or been analyzed using the baseline LTC model, then virtually no change would have been detected. This fact demonstrates a need for choosing meaningful measurement occasions when assessing changes in TSE, otherwise no changes might be observed. Choosing such measurement occasions should be guided by theory, for example, regarding the sources of self-efficacy.

This study found a link between a TSE change during the practicum at school and preservice teachers' ITQ their teaching degree. Intentions are a very good predictor of actual

behavior (e.g., Ajzen, Czasch, & Flood, 2009). Considering problematic teacher attrition rates, and unknown preservice teacher attrition rates, such a link is highly relevant. In light of this result, supporting preservice teachers during their practicum experience in teacher education seems of as much importance as supporting them in their first years of teaching, where efforts have mostly been guided so far. While the present results could not prove that an increase in TSE would cause a decrease in preservice teachers' ITQ, for the first time it demonstrated a 'change-change' link. Furthermore, keeping self-efficacy theory in mind, this assumed direction of influence is more plausible than preservice teachers' ITQ changing first during the practicum, and subsequently causing a decrease in TSE. However, before teacher educators can design a practicum experience that will foster TSE, research will need to pinpoint the factors that influence TSE. The results from this study can contribute an imperative for doing so.

3.5 References

- Author et al. (2014) [details removed for peer review]
- Ajzen, I., Czasch, C., & Flood, M. G. (2009). From intentions to behavior: Implementation intention, commitment, and conscientiousness. *Journal of Applied Social Psychology, 39*(6), 1356–1372. doi:10.1111/j.1559-1816.2009.00485.x
- Ashford, S., Edmunds, J., & French, D. P. (2010). What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis. *British Journal of Health Psychology, 15*(2), 265–288. doi:10.1348/135910709X461752
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191–215.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Bandura, A. (2006). Guide for creating self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich: Information Age Publishing.
- Borman, G. D., & Dowling, N. M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research, 78*(3), 367–409. doi:10.3102/0034654308321455
- Charalambous, C. Y., Philippou, G. N., & Kyriakides, L. (2008). Tracing the development of preservice teachers' efficacy beliefs in teaching mathematics during fieldwork. *Educational Studies in Mathematics, 67*(2), 125–142. doi:10.1007/s10649-007-9084-2
- Chesnut, S. R., & Burley, H. (2015). Self-efficacy as a predictor of commitment to the teaching profession: A meta-analysis. *Educational Research Review, 15*, 1–16. doi:10.1016/j.edurev.2015.02.001
- Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155–159.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.
- Cronbach, L. J., & Furby, L. (1970). How should we measure "change" - or should we? *Psychological Bulletin, 74*(1), 68–80.
- Duffin, L. C., French, B. F., & Patrick, H. (2012). The teachers' sense of efficacy scale: Confirming the factor structure with beginning pre-service teachers. *Teaching and Teacher Education, 28*(6), 827–834.

- Fives, H., Hamman, D., & Olivarez, A. (2007). Does burnout begin with student-teaching? Analyzing efficacy, burnout, and support during the student-teaching semester. *Teaching and Teacher Education, 23*, 916–934.
- Gabriele, A. J., & Joram, E. (2007). Teachers' reflections on their reform-based teaching in mathematics: Implications for the development of teacher self-efficacy. *Action in Teacher Education, 29*(3), 60–74. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/01626620.2007.10463461#.UnkBL6Mwfcs>
- Garvis, S., Pendergast, D., & Keogh, J. (2012). Changes in teacher self-efficacy in the first year of primary school teacher education study. *The Journal of the World Universities Forum, 5*(1).
- Gibson, S., & Dembo, M. H. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*(4), 669–682.
- Henson, R. K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist, 37*(3), 137–150. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/S15326985EP3703_1#.UnfabKMwfcs
- Howard, G. S., & Dailey, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology, 64*, 144–150.
- Hoy, W. K., & Woolfolk, A. E. (1990). Socialization of student teachers. *American Educational Research Journal, 27*(2), 279–300.
- Hu, L., & 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
- Ingersoll, R. M. (2002). The teacher shortage: A case of wrong diagnosis and wrong prescription. *NASSP Bulletin, 86*, 16–31. doi:10.1177/019263650208663103
- Ingersoll, R. M., & Strong, M. (2011). The impact of induction and mentoring programs for beginning teachers: A critical review of the research. *Review of Educational Research, 81*(2), 201–233. Retrieved from <http://rer.sagepub.com/content/81/2/201>
- Klassen, R. M., Bong, M., Usher, E. L., Chong, W. H., Huan, V. S., Wong, I. Y. F., & Georgiou, T. (2009). Exploring the validity of a teachers' self-efficacy scale in five countries. *Contemporary Educational Psychology, 34*, 67–76.
- Klassen, R. M., & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching

- context. *Contemporary Educational Psychology*, 36(2), 114–129.
doi:10.1016/j.cedpsych.2011.01.002
- Klassen, R. M., & Durksen, T. L. (2014). Weekly self-efficacy and work stress during the teaching practicum: A mixed methods study. *Learning and Instruction*, 33, 158–169.
doi:10.1016/j.learninstruc.2014.05.003
- Klassen, R. M., Tze, V. M. C., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise? *Educational Psychology Review*, 23, 21–43.
- Kleinsasser, R. C. (2014). Teacher efficacy in teaching and teacher education. *Teaching and Teacher Education*, 44, 168–179. doi:10.1016/j.tate.2014.07.007
- Knoblauch, D., & Woolfolk Hoy, A. (2008). “Maybe I can teach those kids.” The influence of contextual factors on student teachers’ efficacy beliefs. *Teaching and Teacher Education*, 24, 166–179.
- Liaw, E.-C. (2009). Teacher efficacy of pre-service teachers in Taiwan: The influence of classroom teaching and group discussions. *Teaching and Teacher Education*, 25(1), 176–180. doi:10.1016/j.tate.2008.08.005
- Lin, H.-L., & Gorrell, J. (2001). Exploratory analysis of pre-service teacher efficacy in Taiwan. *Teaching and Teacher Education*, 17, 623–635.
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 151–173. doi:10.1207/S15328007SEM0902_1
- Lortie, D. C. (1975). *Schoolteacher: A sociological study*. London: University of Chicago Press.
- Macdonald, D. (1999). Teacher attrition: A review of literature. *Teaching and Teacher Education*, 15(8), 835–848. doi:10.1016/S0742-051X(99)00031-1
- Malmberg, L.-E. (2008). Student teachers' achievement goal orientations during teacher studies: Antecedents, correlates and outcomes. *Learning and Instruction*, 18(5), 438–452. doi:10.1016/j.learninstruc.2008.06.003
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60(1), 577–605.
doi:10.1146/annurev.psych.60.110707.163612
- Meade, A. W., & Lautenschlager, G. J. (2004). A Monte-Carlo Study of Confirmatory Factor Analytic Tests of Measurement Equivalence/Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(1), 60–72. doi:10.1207/S15328007SEM1101_5

- Mergler, A. G., & Tangen, D. (2010). Using microteaching to enhance teacher efficacy in pre-service teachers. *Teaching Education, 21*(2), 199–210. Retrieved from <http://dx.doi.org/10.1080/10476210902998466>
- Millsap, R. E., & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology, 73*(3), 574–584.
- Mulholland, J., & Wallace, J. (2001). Teacher induction and elementary science teaching: enhancing self-efficacy. *Teaching and Teacher Education, 17*, 243–261.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Palmer, D. (2006). Durability of changes in self-efficacy of preservice primary teachers. *International Journal of Science Education, 28*(6), 655–671. doi:10.1080/09500690500404599
- Pendergast, D., Garvis, S., & Keogh, J. (2011). Pre-service student-teacher self-efficacy beliefs: An insight into the making of teachers. *Australian Journal of Teacher Education, 36*(12), 46–58. Retrieved from <http://ro.ecu.edu.au/ajte/vol36/iss12/4/>
- Plourde, L. A. (2002). The influence of student teaching on preservice elementary teachers' science self-efficacy and outcome expectancy beliefs. *Journal of Instructional Psychology, 29*(4), 245–253.
- Reuter, T., Ziegelmann, J. P., Wiedemann, A. U., Geiser, C., Lippke, S., Schüz, B., & Schwarzer, R. (2010). Changes in intentions, planning, and self-efficacy predict changes in behaviors: An application of latent true change modeling. *Journal of Health Psychology, 15*(6), 935–947. doi:10.1177/1359105309360071
- Richardson, G. M., & Liang, L. L. (2008). The use of inquiry in the development of preservice teacher efficacy in mathematics and science. *Journal of Elementary Science Education, 20*(1), 1–16.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*(1), 1–28.
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research - Online, 8*(2), 23–74.
- Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research - Online, 2*(1), 21–33.

- Steyer, R., Partchev, I., & Shanahan, M. J. (2000). Modeling true intraindividual change in structural equation models: The case of poverty and children's psychosocial adjustment. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data* (pp. 109–126). Mahwah, NJ: Lawrence Erlbaum Associates.
- Swan, B. G., Wolf, K. J., & Cano, J. (2011). Changes in teacher self-efficacy from the student teaching experience through the third year of teaching. *Journal of Agricultural Education*, 52(2), 128–139.
- Swars, S., Bart, L. C., Smith, S. Z., & Tolar, T. (2007). A longitudinal study of elementary pre-service teachers' mathematics beliefs and content knowledge. *School Science and Mathematics*, 107(8), 325–335.
- Thiel, F. & Blüthmann, I. *Ergebnisse der Evaluation der lehrerbildenden Studiengänge an der Freien Universität Berlin: Sommersemester 2009* [Results of the evaluation of the teacher preparation programs at the Freie Universität Berlin]. Berlin. Retrieved from http://www.ewi-psy.fu-berlin.de/einrichtungen/arbeitsbereiche/schulentwicklungsforschung/downloads/Lehramtmasterbefragung_2009.pdf?1310986750
- Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, 17, 783–805.
- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68(2), 202–248. Retrieved from <http://www.jstor.org/stable/1170754?origin=JSTOR-pdf&>
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70. doi:10.1177/109442810031002
- Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research*, 54(2), 143–178. Retrieved from <http://rer.sagepub.com/content/54/2/143>
- Vieluf, S., Kunter, M., & van de Vijver, F. J. R. (2013). Teacher self-efficacy in cross-national perspective. *Teaching and Teacher Education*, 35, 92–103. doi:10.1016/j.tate.2013.05.006
- Weinstein, C. S. (1988). Preservice teachers' expectations about the first year of teaching. *Teaching and Teacher Education*, 4(1), 31–40. doi:10.1016/0742-051X(88)90022-4
- Woolfolk, A. E., & Hoy, W. K. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology*, 82(1), 81–91.

- Woolfolk Hoy, A., & Burke Spero, R. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, 21(4), 343–356. doi:10.1016/j.tate.2005.01.007
- Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data* (pp. 219–240). Mahwah, NJ: Lawrence Erlbaum Associates.

3.6 Acknowledgments

I would like to express my sincere thanks to Martin Schultze from the Freie Universität Berlin for his tremendous help with the trickier aspects of latent true change modeling.

4

Study 3: Why do I feel more confident? Bandura's sources predict preservice teachers' latent changes in teacher self-efficacy

Author's note: This is the author's version of a manuscript that was submitted to the journal *Contemporary Educational Psychology*:

Pfitzner-Eden, F. (under review). Why do I feel more confident? Bandura's sources predict preservice teachers' latent changes in teacher self-efficacy.

4 Why do I feel more confident? Bandura's sources predict preservice teachers' latent changes in teacher self-efficacy

Abstract

Teacher self-efficacy (TSE) is associated with a multitude of positive outcomes for teachers and students. However, the development of TSE is an under-researched area. Bandura (e.g., 1997) proposed four sources of self-efficacy: mastery experiences, vicarious experiences, verbal persuasion, and physiological and affective states. Currently there is no quantitative measure of Bandura's four sources for TSE. This study introduces a measure of the four sources for TSE. Further, gathering first evidence of validity, the contribution that each source makes to the development of TSE during a practicum at school is explored for two samples of German preservice teachers. The first sample ($N = 359$) were beginning preservice teachers who completed an observation practicum at school. The second sample ($N = 395$) were advanced preservice teachers who completed a teaching practicum. TSE was assessed before and after the practicum in both groups. The sources of TSE were assessed retrospectively after the practicum. The source measure showed good psychometric properties. Latent true change modeling was applied to explore how the sources predict changes in TSE. Three different models were compared. As expected, results showed that TSE changes in both groups were significantly predicted by mastery experiences, with a stronger relationship in the advanced group. Further, the results indicated that mastery experiences are largely informed by the other three sources to varying degrees depending on the type of practicum. Implications for the practice of teacher education are discussed in light of the results.

Key words:

teacher self-efficacy; sources; preservice teachers; practicum; latent-true-change

4.1 Introduction

Research into teacher self-efficacy (TSE) has proven to be a fruitful endeavor in the field of educational psychology. There are a multitude of studies that have identified meaningful associations between TSE and a range of sought after outcomes for inservice teachers (e.g., resilience - Beltman, Mansfield, & Price, 2011; instructional quality - Holzberger, Philipp, & Kunter, 2013; occupational commitment - Klassen & Chiu, 2011; job satisfaction - Klassen & Chiu, 2010; teaching performance - Klassen & Tze, 2014; burnout - Schwarzer & Hallum, 2008), preservice teachers (e.g., burnout - Fives, Hamman, & Olivarez, 2007; occupational commitment - Klassen & Chiu, 2011; commitment to finishing a teaching degree – Author, submitted), and students (e.g., academic achievement - Caprara, Steca, Barbaranelli, & Malone, 2006). However, studies focusing on the development of TSE and factors that can predict such development were essentially non-existent in the first decades of TSE research (Henson, 2002) and are still rare now (e.g., Klassen, Tze, Betts, & Gordon, 2011). The present study addresses this gap by providing an assessment of the four main theorized sources of TSE, and by exploring how they might predict latent changes in TSE that occurred during a practicum at school. Data are assessed in two groups of preservice teachers at different stages of a teacher preparation program (beginning vs. advanced), which allowed for a comparison of two different types of practicum format (observation vs. teaching).

Teacher self-efficacy and its development

TSE, as understood here, is based on Bandura's self-efficacy construct, which is a central feature in his social cognitive theory (1977; 1986; 2001). Bandura defines self-efficacy as the "...beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments." (1997, p. 3). TSE can then be understood as the beliefs that inservice and preservice teachers hold about their capabilities in the teaching domain. Tschannen-Moran and Woolfolk Hoy (2001) operationalized TSE as encompassing three dimensions: instruction, classroom management, and student engagement, as reflected in their Teachers' Sense of Efficacy Scale (TSES).

Once firmly established, self-efficacy beliefs are reasonably stable and fairly resistant to change (Bandura, 1997). Thus, TSE beliefs are understood to be most malleable early on in a teacher's career, during teacher education (Henson, 2002; Woolfolk & Hoy, 1990). Bandura (1997) proposes that self-efficacy beliefs develop based on people's cognitive interpretation of information regarding their own capabilities. He hypothesizes that this information stems from four principal sources: mastery experiences, vicarious experiences, verbal persuasion, and physiological and affective states (PAS). Mastery experiences provide information about

one's successes, but also failures. Successful experiences increase self-efficacy beliefs, while experiences of failure lower them. Bandura (1997) states that mastery experiences have the strongest effect on self-efficacy formation because they are the most authentic indicators of one's capabilities. For preservice teachers, mastery experiences or successes generated in an actual classroom during a practicum should thus have a strong effect on TSE development, because they provide genuine evidence of whether or not they can accomplish a given outcome (i.e., giving a class successfully). Vicarious experiences provide information about modeled attainments of others, which influence one's self-efficacy beliefs by demonstrating and transferring competencies (model learning), and by providing a point of reference for social comparison. According to Bandura (1997), vicarious experiences exert a greater influence when people have little prior experience in the domain to be evaluated. Observing classes of other teachers during a practicum at school provides preservice teachers with an opportunity for model learning. This will be particularly beneficial for TSE development if several competent models, but also models who overcome difficult situations, can be observed. Verbal persuasion by "significant others" (Bandura, 1997; p. 101) can convince people of their capabilities, especially if this persuasion comes from a credible source. A credible source would be somebody who is regarded as an expert in the respective field, and who is also experienced in judging the accomplishments of different aspirants. During teacher education, a mentor teacher assigned to preservice teachers in their school practicum fits this profile perfectly. PAS provide information about physiological and affective arousal during situations in which capability in the domain in question is demonstrated. Bandura (1997) elaborates that in stressful situations people tend to read this somatic information as an indicator of dysfunction, thus impacting negatively on self-efficacy beliefs. However, as with the other sources, such negative impact depends on the cognitive processing of this information. For example, physical arousal (e.g., sweating, increased heart rate) in a teaching situation can be interpreted as anxiety in the face of the teaching task and attributed to one's own inadequacy, or it can be interpreted as a common reaction to the teaching task that does not necessarily convey any negative information about one's capability. Regarding the relationship between the sources, Bandura (1997) provides no specific guidelines. He states that "the weights assigned to different types of efficacy information may vary across different domains of functioning" (p. 114), further that the sources "vary in their informativeness and degree of interrelatedness" (p. 114), and that the sources "vary in the complexity of their relations to" (p. 114) self-efficacy judgements.

Following Bandura's (e.g., 1997) description of the sources, research on the development of TSE has typically focused on the practical phases during teacher education, since they provide the most authentic opportunity for mastery experiences to occur, and also for the other three sources to effect TSE development.

Previous research on the sources of teacher self-efficacy

Previous research on the sources of TSE has produced some results regarding the four sources proposed by Bandura, but has also aimed to determine any other factors that correlate with or predict TSE (e.g., university training in Cheung, 2008; cognitive pedagogical mastery in Palmer, 2006). In line with the purpose of this study, this review focuses only on research related to Bandura's (e.g., 1997) proposed four sources.

So far, only one quantitative measure of the sources of TSE has been published in a peer-reviewed journal, the Teaching Efficacy Sources Inventory (TESI) by Poulou (2007). Starting with an inductive qualitative approach, Poulou (2007) identified statements by preservice teachers regarding factors that would influence and promote their sense of teaching efficacy. Among other factors (e.g., motivation), the TESI also reflects the four sources, whereby mastery experiences and verbal persuasion, however, were combined into one factor. Among the sources, this combined factor was found to be the only one to predict the TSE level for all three TSE dimensions. O'Neill and Stephenson (2011) adapted Poulou's (2007) TESI for their study of Australian preservice teachers, using 19 of the original 30 statements and focusing on TSE for classroom management. They found that among the assessed sources only PAS (negatively) predicted the level of TSE for classroom management.

In the absence of valid measures of the four sources of TSE, some researchers used proxies for assessing the sources. Moulding, Stewart, and Dunmeyer (2014) documented a relationship between preservice teachers' ratings of support from their mentor during student teaching (verbal persuasion) and their TSE level. Woolfolk Hoy and Burke Spero (2005) participants' ratings of how satisfied they were with their professional performance in their induction year (mastery experiences), as well as their ratings of various kinds of support received during that year (verbal persuasion) correlated positively with TSE change scores during the same period. Ruble, Usher, and McGrew (2011) used proxy indicators for three sources: teaching experience in years (mastery experiences), perceived support by school leaders (verbal persuasion), and self-reported levels of burnout (PAS). They found a significant correlation only between burnout and the level of TSE, which is a well-established relationship (e.g., Brouwers & Tomic, 2000; Schwarzer & Hallum, 2008). Tschannen-Moran and Woolfolk Hoy (2007) also used proxy indicators for mastery experiences (rated

satisfaction with own professional performance in the last year) and verbal persuasion (rated quality of interpersonal support by school administration and colleagues, and rated quality of parental and community support for classrooms). For novice teachers, mastery experiences predicted their TSE level positively, while verbal persuasion predicted TSE level negatively. For experienced teachers, only mastery experiences predicted TSE significantly. Most recently, Klassen and Durksen's (2014) analysis of qualitative data on preservice teachers' TSE changes during a teaching practicum emphasize the importance of verbal persuasion by the mentor teacher as an influential source. Their verbatim examples provide some indication that preservice teachers draw on feedback (i.e., verbal persuasion) by their mentor teacher to inform their judgment of their mastery experiences. For example, a participant in Klassen and Durksen's (2014) study said "I have been doing a lot of marking this week and it's building my confidence" (p.165) – a mastery experience. Another participant stated that he/she was "asking my mentor teacher for help to know I am marking correctly" (p. 165) – this participant engaged in the same form of mastery experience as the former participant. However, this participant did not rely on their own appraisal of their performance, most likely because he/she was attempting this task for the first time. In this case, the participant used the mentor feedback to appraise their own performance (i.e., mastery experience).

There is some evidence from case studies with inservice teachers highlighting the importance of mastery experiences (e.g., Milner & Woolfolk Hoy, 2003), particularly during preservice teaching practicums (Mulholland & Wallace, 2001). Milner and Woolfolk Hoy (2003) identified mastery experiences as past successes but also enthusiasm shown by students, which would rather be indicative of verbal persuasion. In Mulholland and Wallace's (2001) case study, mastery experiences and verbal persuasion were likewise closely linked. There was also some indication that verbal persuasion by students was used to inform mastery experiences during preservice teaching experiences (p. 249). In addition (the lack of) vicarious experiences and PAS were identified as impacting negatively on TSE development. Morris and Usher (2011) interviewed research professors who identified mastery experiences and verbal persuasion as the most influential sources, whereby both sources were again thought to be closely related.

Since there is little quantitative research on the sources specifically of TSE, it might be useful to look into research on the sources of self-efficacy in another more researched domain. Usher and Pajares (2008) provide a comprehensive review of research on the sources of self-efficacy in school, in which they conclude that mastery experiences consistently predict self-efficacy of students, but that evidence for the other three sources was less consistent. They

point out that so far the quantitative assessment of the sources can only be regarded as preliminary, as it suffers from low reliabilities, aggregate scores for more than one source that mask each source's contribution, and inconsistencies with Bandura's (1997) theoretical guidelines, according to which an evaluative component is key to the assessment of the sources.

In sum, there is some evidence from qualitative findings regarding the importance of mastery experiences and verbal persuasion for TSE. Regarding verbal persuasion, there is little information on which significant "others" give preservice teachers feedback. Results from quantitative research so far have mostly been derived using proxy indicators. These studies focused on mastery experiences and verbal persuasion, generally also showing their association with TSE. The assessment of vicarious experiences has been entirely neglected in these studies and PAS featured in only one study (Ruble et al., 2011), whereby it is questionable to what degree burnout can be considered an indicator of PAS in line with Bandura's description of this source. Among the studies using proxy indicators, Ruble et al.'s (2011) study was also the only study not to detect a relationship for mastery experiences and verbal persuasion with TSE. Taking Usher and Pajares' (2008) analysis of the shortcomings in quantitative assessment of the sources into account, this might have been due to a lack of an evaluative component in the proxies used for mastery experiences and verbal persuasion. Among the quantitative research, two studies used the only instrument (TESI) to date assessing the sources, producing conflicting results regarding which sources do or do not predict TSE. However, since the TESI features several methodological weaknesses (i.e., a combined factor of mastery experiences and verbal persuasion, a lack of an evaluative component in the majority of items, and little alignment with Bandura's specification of the sources), it is unclear what weight these results carry. Taken together, previous research on the four sources has (a) shown a strong relationship between mastery experiences and verbal persuasion, (b) produced few insights regarding vicarious experiences and PAS, (c) underscored the significance of practical phases during teacher education, (d) often applied inadequate quantitative measures (i.e., proxies), and (e) focused on predicting levels (i.e., state), rather than changes (i.e., development), of TSE. With regard to the last point, it is essential for research on the sources to predict changes in TSE because changes, not states, signify development, which is exactly what the sources are theorized to predict. TSE changes were only predicted in one study (Woolfolk Hoy & Burke Spero, 2005), albeit using manifest change scores that are associated with serious methodological drawbacks (Cronbach & Furby, 1970). For this reason, the current study uses a latent approach.

The current study

The purpose of the current study is to present a first quantitative measure of all four sources of TSE, which is closely aligned with Bandura's description of the sources. Thus, a deductive approach was used in the development of this measure. A second aim of this study was to explore how Bandura's (e.g., 1997) four sources predict the development of TSE during a practical phase of teacher education: the school practicum. This exploration should also provide first validity information for the new source measure. For this purpose, TSE changes that occurred in two different cohorts (beginning and advanced preservice teachers) during two different practicum formats at school (observation vs. teaching) are predicted. The practicum format for the beginning preservice teachers focuses on observing teaching, while the format for the advanced preservice teachers focuses on teaching independently. Both formats should differ with respect to the impact the sources exert over the development of TSE. Specifically, one could expect vicarious experiences to have a greater impact during the observation practicum in the beginning group, and mastery experiences to have a greater impact during the teaching practicum in the advanced group. Another indicator of convergent validity would be if mastery experiences, in comparison to the other sources, show the strongest association with TSE changes in each sample, as Bandura (1997) described this as the most influential source. Further, vicarious experiences and verbal persuasion should be positively, and PAS should be negatively related to TSE changes. Since Bandura did not provide any specific guidance with regard to the nature of the relationship between the sources, three different plausible models are tested in this study.

In the first and most simple model, the relationship is modelled as correlations between all sources (i.e., the *simple model*). This model requires (and tests) no assumptions about the relationships between the sources. In the second model, verbal persuasion informs mastery experiences (i.e., the *mediation model 1*). Since Usher and Pajares (2008) reported problems with multicollinearity among the sources, and there is some indication from previous qualitative research that preservice teachers might use verbal persuasion to appraise their mastery experiences, this is another plausible model. In the third model, mastery experiences are informed, not only by verbal persuasion, but also by vicarious experiences and PAS (i.e., the *mediation model 2*). This model is plausible for preservice teachers, because they have had no prior opportunity to gain actual mastery experiences. After all, how would they know if their experiences during a practicum at school were successful or not? They could know through feedback from significant others during their practicum (verbal persuasion), through observing other teachers in their practicum (vicarious experiences), and through their

physiological and affective responses (PAS) to their own teaching, or to teaching related tasks in the case of beginning preservice teachers. Thus, preservice teachers might use all other three sources to appraise the experiences they gain during the practicum. So far, there is little empirical evidence in favor of such a hypothesis, since, as outlined earlier, particularly vicarious experiences and PAS have not really featured in previous research. This idea is not entirely new though, as previously Tschannen-Moran, Woolfolk Hoy, and Hoy (1998) already conceptualized PAS as influencing TSE via mastery experiences. In order to test the three plausible models, specifically the following models are compared:

- (a) all four sources predict TSE changes directly = simple model
- (b) mastery experiences, vicarious experiences and PAS predict TSE changes directly, while verbal persuasion predicts mastery experiences = mediation model 1a
- (c) mastery experiences, vicarious experiences and PAS predict TSE changes directly, while verbal persuasion predicts mastery experiences and TSE directly = mediation model 1b
- (d) only mastery experiences predict TSE changes directly, while vicarious experiences, verbal persuasion, and PAS predict mastery experiences = mediation model 2a
- (e) mastery experiences predict TSE changes directly, while vicarious experiences, verbal persuasion, and PAS predict mastery experiences and TSE directly = mediation model 2a.

In addition, the current study also explores qualitatively who, apart from the mentor teacher at school, provides relevant feedback to preservice teachers during their practicum.

4.2 Method

Participants and procedure

Participants were two groups of German preservice teachers at two different stages of a teacher education program. This program consists of a three-year Bachelor degree, which is followed either by a one-year or two-year Master degree for elementary/middle school teachers or secondary school teachers, respectively. Participants represented a broad range of academic subjects (e.g., mathematics, French, physics, English, geography, chemistry).

Participants in Sample 1 were beginning preservice teachers at an early stage of their Bachelor degree (19% elementary/middle school). Data for Sample 1 were collected at the end of their first year (T1: $N = 359$) and three months later at the beginning of their second

year of Bachelor study (T2: $N = 226$). Between T1 and T2, participants in Sample 1 completed a one-month practicum at school, which was mainly aimed at preservice teachers observing the classroom and teaching of experienced teachers. Participants in Sample 2 were advanced preservice teachers, either enrolled in the Master degree (80%), or in their last year of Bachelor study (20%) (31% elementary/middle school). Data for T1 were collected at the end of the winter semester ($n_1 = 255$) and the end of the following summer semester ($n_2 = 140$) to increase numbers in the advanced sample (T1: $N = 395$). For both groups in the advanced sample, data for T2 were collected three months later at the beginning of the next semester respectively ($n_1 = 143$; $n_2 = 80$; T2: $N = 223$). Between T1 and T2, participants in Sample 2 undertook a one-month teaching practicum at school, which focused on preservice teachers preparing lessons and teaching independently. Lessons were observed by an experienced mentor teacher who gave the preservice teachers feedback afterwards. In this practicum format, advanced preservice teachers do also observe lessons of other teachers, but this task is secondary. Data for each sample were collected in the same university year so that no participants overlapped between Sample 1 and Sample 2. Measures at T1 were administered in a paper and pencil format in seminars that prepared participants for their practicum at school. At T2, measures had to be administered via an online survey, since the same preservice teachers did not attend another seminar together after the practicum.

In compliance with privacy guidelines of the university, participation was voluntary and data were collected anonymously. In order to guarantee anonymity, data had to be matched using a personal code across both times of measurement. The personal code was an anonymous identifier (e.g., first and last letter of the mother's first name) that respondents completed each time. Self-selection bias in entering the study was minimal, as participation at T1 was declined by only three preservice teachers in Sample 1 and five preservice teachers in Sample 2.

Attrition

Descriptive statistics for the longitudinal samples are shown in Table 1. Retention rates between T1 and T2 were 50% for Sample 1 and 42% for Sample 2. While these rates are low, they are comparable to those reported in similar studies with preservice teachers (47% in Malmberg, 2008; 46% in Klassen & Durksen, 2014), or conducted online (44% in Reuter et al., 2010). In order to test attrition bias, T1 characteristics were compared between those that participated at T1 and T2 and those that dropped out between T1 and T2. In Sample 1, there were no significant differences (all $ps > .05$) for all TSE variables. However, males and older

participants were more likely to drop out. In Sample 2, there were no significant differences (all $ps > .05$) for all TSE variables, as well as gender and age.

Table 1. Longitudinal sample description

Variable at T1	T1	T1 & T2	T1 not T2
<i>Beginning group</i>	<i>n</i> = 359	<i>n</i> = 178	<i>n</i> = 181
Female	64.4%	70.1%	59.5%
Mean Age (SD)	23.80 (5.05)	22.74 (4.75)	24.70 (5.36)
Mean TSE Instruction (SD)	6.74 (1.07)	6.79 (1.08)	6.69 (1.06)
Mean TSE Classroom Management (SD)	6.12 (1.46)	6.22 (1.44)	6.02 (1.47)
Mean TSE Student Engagement (SD)	6.80 (1.20)	6.87 (1.14)	6.75 (1.26)
<i>Advanced group</i>	<i>n</i> = 395	<i>n</i> = 167	<i>n</i> = 228
Female	67.2%	71.2%	64.3 %
Mean Age (SD)	26.13 (4.55)	25.86 (4.60)	26.32 (4.51)
Mean TSE Instruction (SD)	6.54 (1.10)	6.63 (1.04)	6.48 (1.14)
Mean TSE Classroom Management (SD)	6.14 (1.42)	6.26 (1.34)	6.05 (1.47)
Mean TSE Student Engagement (SD)	6.57 (1.16)	6.65 (1.02)	6.53 (1.25)

Note: TSE = teacher self-efficacy; SD = standard deviation.

Measures

To assess TSE, the German version of the Scale for Teacher Self-Efficacy (STSE) was used (see Author et al., 2014, for reliability and validity estimates of the scale). The STSE is an adapted version of the TSES (Tschannen-Moran and Woolfolk Hoy, 2001), which provides a stable three-dimensional assessment of TSE for preservice teachers at different stages of teacher preparation. Four items each assess TSE for instructional strategies (e.g., “*How certain are you that you can adjust lessons to the proper level for individual students?*”), classroom management (e.g., “*How certain are you that you can control disruptive behavior in the classroom?*”), and student engagement (e.g., “*How certain are you that you can motivate students who show low interest in schoolwork?*”). Participants rate how confident they are regarding their capabilities in the three dimensions using a 9-point response scale ranging from 1 (*not at all certain can do*) to 9 (*absolutely certain can do*).

The sources of TSE were assessed using self-developed items (see Table 2 for wording). Item development followed a deductive process in order to achieve maximum alignment with Bandura’s theorized sources. Two experts in self-efficacy theory rated the items for their

congruence with Bandura's (1997) description of the sources. The self-developed items were pilot tested with a sample of German advanced preservice teachers (N = 282) and a sample of New Zealand advanced preservice teachers (N = 131). The mastery experiences scale comprises four items, whereby two items were not identical across the two samples. In order to accommodate the different nature of the practicum, these items focused on teaching in the advanced group, but were phrased without reference to teaching in the beginning group. The vicarious experiences scale also comprises four items. These items focus on the model learning aspect of this source. Originally four items were constructed for the verbal persuasion scale, but preliminary confirmatory factor analyses with four source factors and a low internal consistency among the verbal persuasion items indicated that the verbal persuasion items were better conceptualized as two factors reflecting the source of the persuasion: by the mentor teacher and by others. There was an open response format next to this item where respondents could indicate who these "others" were. The PAS scale comprises three items, which all refer to negative PAS. Participants rate their agreement using a 9-point response scale ranging from 1 (*not at all true*) to 9 (*exactly true*).

Table 2. Standardized factor loadings, construct reliabilities (CR), manifest means and standard deviations, and internal consistencies (α) of the source items/scales for each group

Source Items/Scales	Beginning Cohort (<i>N</i> = 226)	Advanced Cohort (<i>N</i> = 223)
Standardized factor loadings		
<i>Mastery Experiences</i> (CR = .87/.93)		
I was very satisfied with my own achievements (<i>in teaching</i>) during my practicum.	.95	.88
During my practicum, I had many successful experiences (<i>during my own teaching</i>).	.79	.89
During my practicum, I noted that I could be a very good teacher.	.63	.90
I successfully mastered the requirements of the practicum.	.75	.82
<i>Vicarious Experiences</i> (CR = .93/.95)		
I could observe teachers from whom I learned how to be a good teacher.	.89	.95
I observed teachers that managed difficult classroom situations successfully.	.78	.81
I observed teachers that conducted very good lessons.	.92	.94
I could observe teachers from whom I learned a lot.	.92	.92
<i>Verbal Persuasion by the Mentor</i> (CR = .89/.85)		
My mentor at school told me that I would be a good teacher.	.91	.75
My mentor at school gave me positive feedback.	.89	.96
<i>Verbal Persuasion by Others</i> (CR = .86/.80)		
Other people in my practicum told me that I would be a good teacher.	.85	.82
Other people in my practicum gave me positive feedback.	.88	.81
<i>Physiological and Affective States</i> (CR = .83/.85)		
During my practicum, I often felt anxious.	.69	.75
During my practicum, I mostly felt uncomfortable.	.76	.85
During my practicum, I often felt quite down.	.90	.83

Source Items/Scales	Beginning Cohort (<i>N</i> = 226)	Advanced Cohort (<i>N</i> = 223)
Manifest Means (Standard Deviations)		
Mastery Experiences ($\alpha = .85/.93$)	7.23 (1.33)	6.99 (1.28)
Vicarious Experiences ($\alpha = .93/.95$)	7.51 (1.50)	6.60 (1.83)
Verbal Persuasion by the Mentor ($\alpha = .89/.83$)	6.14 (3.16)	7.19 (2.15)
Verbal Persuasion by Others ($\alpha = .86/.79$)	6.18 (2.89)	6.97 (2.14)
Physiological and Affective States ($\alpha = .82/.84$)	1.99 (1.47)	2.67 (1.67)

Note: All standardized factor loadings were significant at the $p < .001$ level. Italic print in brackets shows the wording for the advanced preservice teachers. The first α coefficient refers to the beginning cohort, the second to the advanced cohort. Response categories range from 1 (*not at all certain can do*) to 9 (*absolutely certain can do*).

Analysis

First, a confirmatory factor analysis (CFA) was conducted to examine the construct validity and reliability of the newly developed source measures. Several descriptive statistics were also computed for this purpose. In a second step, latent true change (LTC) modelling (Steyer, Eid, & Schwenkmezger, 1997) was applied to analyze TSE changes that occurred during the practicum (between T1 and T2). LTC models are structural equation models (SEM) that allow for determining changes in the latent true scores (free of measurement error) of repeatedly measured constructs at the intra-individual level. In order to do so, the LTC model uses a latent (i.e., not measured) change score, defined as the part of a score (i.e., TSE at T2) that is not part of that score at the previous measurement (i.e., TSE at T1) (e.g., McArdle, 2009). Once latent change scores are specified, they can be treated like a regular latent construct in SEM, allowing for the introduction of predictor variables (i.e., the sources) that can explain variance in the latent change scores. Since the TSE measure provides a three-dimensional assessment, a general TSE factor was modeled as a second-order factor, whereby the three TSE dimensions served as first-order factors. In order to be able to model LTC, scalar measurement invariance (i.e., equal factor structure, factor loadings, and intercepts) over time needs to be demonstrated for all repeatedly measured constructs (e.g., Steyer et al., 1997). In order to test this prerequisite, invariance tests were conducted for the entire hierarchical CFA model of TSE in each sample in the following order recommended by Brown (2006). First, it was tested whether the factorial structure was equal at both assessment points (configural invariance); second, whether the factor loadings were equal (metric invariance); and third, whether the intercepts were equal (scalar invariance) across both times. In order to evaluate invariance, changes in model fit were tested with the χ^2 -difference test for nested models (e.g., Chan, 1998; Palmer, 2001). If the fit does not decrease significantly at each step, invariance can be assumed. After establishing invariant LTC models, model fit for the competing models was compared as follows. The simple model was compared to the mediation models 1a and 2a, in order to judge whether mastery experiences acted as a mediator. Then, the mediation models 1a and 2a were compared, in order to judge whether mastery experiences acted as a mediator for all other sources. The mediation model 1b and 2b were compared to models 1a and 2a, respectively, in order to test whether including a direct path to TSE was necessary (i.e., improved the fit) in the mediation models, or if the entire influence on TSE changes was mediated by mastery experiences. Model fit was compared using the χ^2 -difference test for nested models. Models were nested, as each of the compared models could be derived by constraining parameters of the less restrictive models they are compared with. If a significant

difference is found, the less restrictive model should be favored, if there is no significant difference the more restrictive model should be favored (e.g., Schermelleh-Engel, Moosbrugger, & Müller, 2003).

A multi index strategy (e.g., Hu & Bentler, 1999) was used to evaluate model fit of the LTC models. As an incremental fit index, the Comparative Fit Index (CFI) was inspected, for which Hu and Bentler (1999) recommend a cut-off value of .95. Further, as an absolute fit index the Root Mean Square Error of Approximation (RMSEA) was inspected, for which Hu and Bentler (1999) recommend a cut-off value of .06. The χ^2 test-statistic of exact fit is also reported.

In order to reduce the amount of estimated parameters in the model in relation to the fairly small sample size, item parcels were used as indicators of the measurement models for the three latent TSE factors that were measured by four items respectively (for a discussion of pitfalls and merits of item parceling see Little, Cunningham, Shahar, & Widaman, 2002). Taking item-to-construct balance into account, two items were parceled to one indicator so that each of the three latent TSE factors featured two indicators. Measurement errors of the second indicators were allowed to correlate between T1 and T2 (Steyer, Partchev, & Shanahan, 2000).

Mplus Version 7 (Muthén & Muthén, 1998-2012) was used to compute all models. The Full Information Maximum Likelihood estimation method (FIML; see Wothke, 2000 for a detailed discussion) was used, as recommended by Steyer et al. (2000) for longitudinal data with a large number of missings. The FIML method takes into account all information available from the observed data when estimating model parameters, which results in a less biased estimate than other common practices (e.g., listwise deletion, mean-imputation) of dealing with missing data (Wothke, 2000). The significance level was set at $p < .05$ for all analyses. As an indicator of effect size, standardized path coefficients can be regarded as the effect size r (Durlak, 2009), for which Cohen (1992) suggested to interpret values of .10, .30 and .50 as small, medium and large effects, respectively.

Simple counts are reported for the open question regarding who (in addition to their mentor teacher) gave preservice teachers feedback.

4.3 Results

Source items

CFA results for the source measure are presented in Table 2. Model fit was good in both groups: beginning group – CFI = .974; RMSEA = .055; $\chi^2(80, N = 226) = 134.88, p < .001$; advanced group – CFI = .972; RMSEA = .063; $\chi^2(80, N = 223) = 148.98, p < .001$. All factor loadings exceeded the .60 mark for accurate representation recommended by Guadagnoli and Velicer (1988). All construct reliabilities of the latent source factors were well above the .60 value recommended by Bagozzi and Yi (1988), thus indicating an excellent internal consistency of each latent source construct. Internal consistencies, manifest means, and standard deviations for the source scales are also presented in Table 2. Internal consistencies were reasonably high, indicating a good reliability for such brief scales. Factor intercorrelations between source constructs (see Table 3) did not exceed .62. This is an indication that the source factors should indeed be treated as distinct factors (Brown, 2006). Bivariate correlations between the latent source factors and the latent change in TSE are presented in Table 3. As expected, the bivariate correlations are highest between mastery experiences and TSE changes, and all other sources show the expected direction of relationship. Unexpectedly, the bivariate correlation between vicarious experiences and TSE changes is equivalent in both groups. The intercorrelations between the latent factors of the CFA are also presented in Table 3. Intercorrelations between the sources were mostly of medium to large effect size, with few small effects (e.g., between VPo and PAS). Controlling for initial levels of TSE, intercorrelations between each source and TSE change drop as compared to the bivariate correlations, reflecting common variance shared among the sources, and between the sources and initial TSE levels. When all sources and initial TSE levels are taken into account, TSE changes in the advanced group show a stronger association to mastery experiences than in the beginning group.

Table 3. Latent intercorrelations between all constructs and bivariate correlations between the source variables and TSE₂₋₁

Source Variable	ME	VE	VPm	VPo	PAS	TSE1	TSE ₂₋₁
<i>Beginning cohort</i>							
Mastery Experiences (ME)	–						.48*
Vicarious Experiences (VE)	.45*	–					.33*
Verbal Persuasion Mentor (VPm)	.44*	.32*	–				.32*
Verbal Persuasion Others (VPo)	.39*	.25*	.62*	–			.30*
Physiological and Affective States (PAS)	-.50*	-.43*	-.29*	-.23*	–		-.35*
TSE1	.35*	.25*	.15	.19*	-.13	–	–
TSE ₂₋₁	.25*	.17	.26*	.20*	-.28*	-.31*	–
<i>Advanced cohort</i>							
Mastery Experiences (ME)	–						.51*
Vicarious Experiences (VE)	.48*	–					.33*
Verbal Persuasion Mentor (VPm)	.61*	.56*	–				.40*
Verbal Persuasion Others (VPo)	.48*	.38*	.55*	–			.37*
Physiological and Affective States (PAS)	-.47*	-.26*	-.31*	-.15	–		-.35*
TSE1	.25*	.19*	.13	.10	-.36*	–	–
TSE ₂₋₁	.36*	.19*	.26*	.32*	-.09	-.57*	–

Note: TSE₂₋₁ = latent change variable; bivariate correlations are reported in the right-hand column. * $p < .05$.

Measurement invariance over time

The model fit for the configural invariance model was very good for the beginning preservice teachers: CFI = .984; RMSEA = .046; $\chi^2(42, N = 407) = 78.0, p < .05$; and very good for the advanced preservice teachers: CFI = .978.; RMSEA = .051; $\chi^2(42, N = 448) = 91.2, p < .05$. Model fit did not decrease significantly between the configural invariance and the metric invariance model for either sample: beginning preservice teachers: $\Delta\chi^2 = 8.5; \Delta df = 5; p = .13$; advanced preservice teachers: $\Delta\chi^2 = 4.1; \Delta df = 5; p = .53$. Further constraining the intercepts to be equal over time resulted in a non-significant change in model fit: beginning preservice teachers: $\Delta\chi^2 = 5.1; \Delta df = 5; p = .40$; advanced preservice teachers: $\Delta\chi^2 = 10.2; \Delta df = 5; p = .07$. Accordingly, measurement invariance over time was demonstrated for all repeatedly measured constructs in the latent change analysis. Model fit of the final scalar invariance model is identical to the model fit of the LTC model reported in the next section.

Predicting latent changes in teacher self-efficacy

Due to the substantial intercorrelations between the sources of TSE (see Table 3), all sources were allowed to correlate with each other in the latent change model. Since latent changes are usually related to their initial level, this correlation was also allowed. The LTC model showed a good fit in the beginning group: CFI = .983; RMSEA = .043; $\chi^2(52, N = 407) = 91.7, p < .05$; and in the advanced group: CFI = .976.; RMSEA = .048; $\chi^2(52, N = 448) = 105.5, p < .05$. The standardized group means of the latent changes were .232 and .165 for the beginning and advanced group, respectively.

In a next step, the sources were introduced to the LTC model as predictors of the TSE change according to the different models outlined in the current study section. In each of the competing models, correlations between the initial level of TSE and each source were allowed, since they were meaningfully related (see Table 3 – all effect sizes were at least small). This could be expected, as for instance, a preservice teacher who started the practicum with a relatively high level of TSE, might not experience the same amount of anxiety (PAS) during teaching as if he/she had started their practicum with a lower TSE belief.

As indicated by all model fit indices in both groups (see Table 4), all three models showed an identical good fit to the data. The results of the χ^2 -difference test for nested models (see also Table 4) showed no significant difference in model fit between the simple model and the more restrictive mediation model 1, indicating that the mediation model 1 should be favored over the simple model. There is also no significant difference in model fit between the

Table 4. Model fit for the competing models

Model	χ^2	<i>df</i>	<i>p</i>	CFI	RMSEA	$\Delta\chi$	Δdf	compared to
<i>Beginning cohort</i>								
Direct Model	466.31	302	<.001	.963	.037	-	-	-
Mediation Model 1a	469.29	304	<.001	.963	.037	2.98 <i>ns</i>	2	direct model
Mediation Model 1b ²⁸ (with direct effect)	466.31	302	<.001	.963	.037	2.98 <i>ns</i>	2	1a
Mediation Model 2a	472.39	306	<.001	.963	.037	6.09 <i>ns</i>	4	direct model
						3.11 <i>ns</i>	2	1a
Mediation Model 2b (with directs effects)	466.31	302	<.001	.963	.037	6.09 <i>ns</i>	4	2a
<i>Advanced cohort</i>								
Direct Model	491.75	302	<.001	.961	.037	-	-	-
Mediation 1a	493.53	304	<.001	.961	.037	1.79 <i>ns</i>	2	direct model
Mediation Model 1b (with direct effect)	491.75	302	<.001	.961	.037	1.79 <i>ns</i>	2	1a
Mediation Model 2a	494.49	306	<.001	.961	.037	2.74 <i>ns</i>	4	direct model
						0.96 <i>ns</i>	2	1a
Mediation Model 2b (with directs effects)	491.75	302	<.001	.961	.037	2.74 <i>ns</i>	4	2a

Note: CFI = comparative fit index; RMSEA = root mean square error of approximation; *ns* = non-significant; * $p < .05$.

²⁸ All mediation models with direct effects are mathematically identical to the direct model.

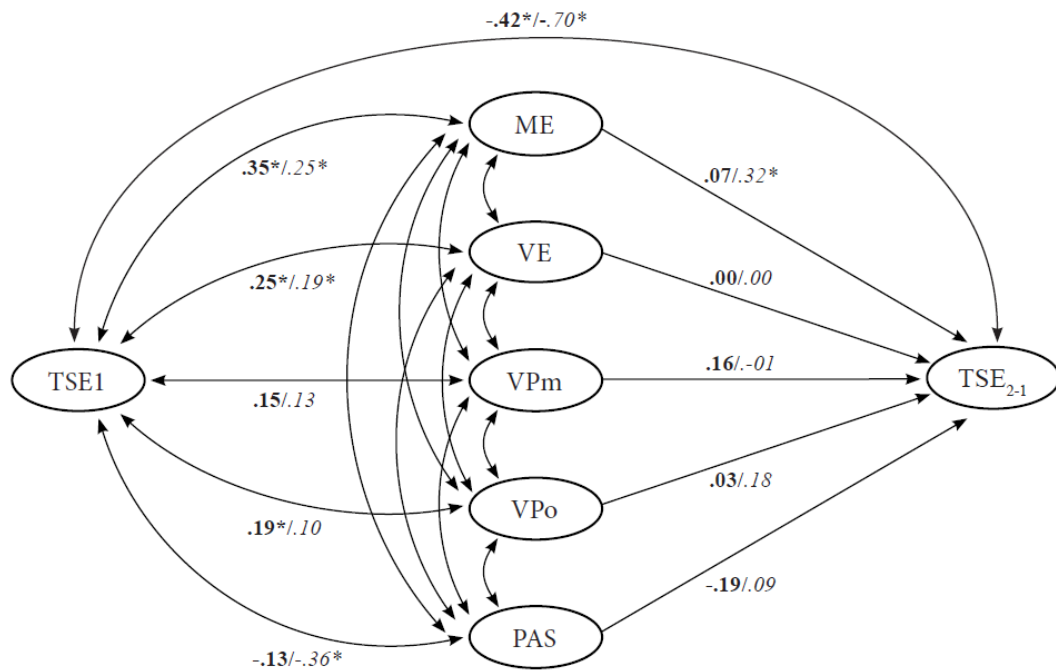
simple and the more restrictive mediation model 2, indicating that the mediation model 2 should be favored over the simple model. Comparing the mediation models also resulted in no significant difference, indicating that the mediation model 2 should be favored over the simple model. Further, including direct paths in each of the mediation models did in no case result in an improved model fit as indicated by the non-significant χ^2 -difference test for nested models. Since in each case the inclusion of direct paths leads to a less restrictive model with equal fit, the more restrictive model (1a and 2a) should be favored in both cases.

To further evaluate the competing models their respective structural models showing the standardized path coefficients are depicted in Figure 1.

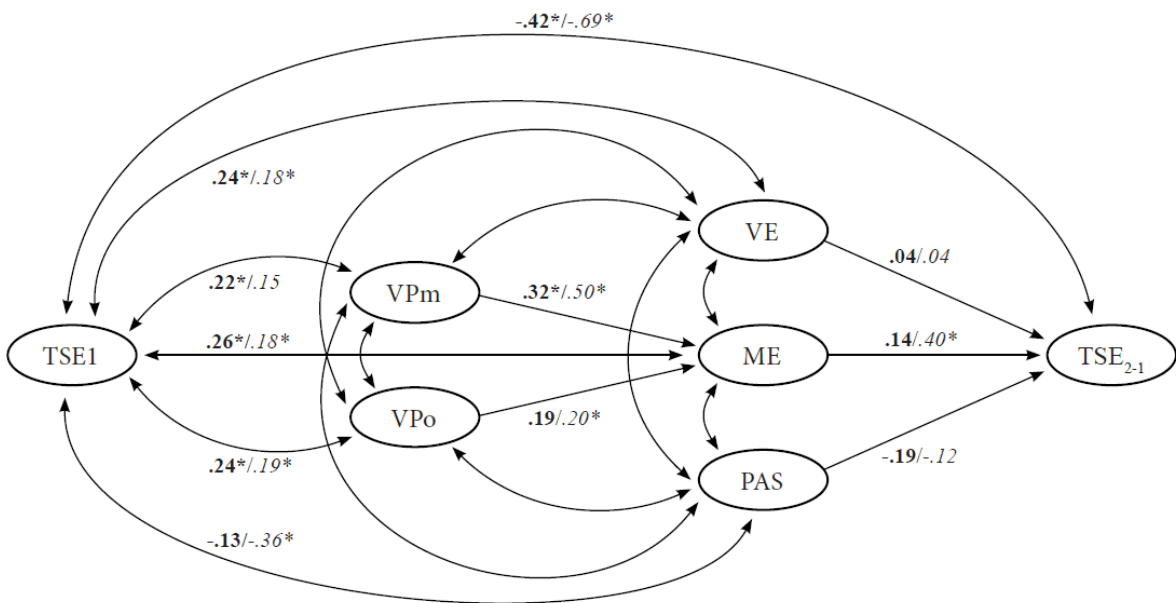
Simple model

When all five sources predict TSE changes simultaneously, all path coefficients are non-significant in the beginning group, and only mastery experiences shows a significant effect on TSE changes in the advanced group. Together, the sources explain 11.7% of the variance in TSE changes in the beginning group, and 16.4% of TSE changes in the advanced group. Since the low weights are somewhat perplexing, but most likely a reflection of the variance shared among the sources, a hierarchical analysis was conducted to further examine this phenomenon. By introducing one source at a time, the unique contribution of each predictor to the variance in TSE change can be determined (e.g., Cohen & Cohen, 1983). The results are reported in Table 5. Including mastery experiences as the first predictor is in line with Bandura's (1997) theory of the sources, whereby mastery experiences are assumed to exert the greatest influence over changes in self-efficacy. Indeed, in both groups, mastery experience is the one single predictor that explains the most variance in TSE changes. In the beginning group, the effect of mastery experiences is attenuated with each source added. Particularly adding verbal persuasion by the mentor teacher and PAS reduces the weight of mastery experiences and improves the prediction of TSE change (ΔR^2). Beta weights are equivalent to partial correlations (e.g., Kline, 2011) and the partial correlation of a predictor x signifies the proportion of the criterion variance that is not associated with the other predictors but is associated with the predictor x (e.g., Cohen & Cohen, 1983). Thus, in this group, while mastery experiences explain a large proportion of the variance in TSE changes, this particular proportion of variance is nearly entirely shared with all other sources (i.e., the beta weight drops to .07). In the advanced group, the effect of mastery experiences is much less attenuated by the inclusion of any other source. In this group, particularly verbal persuasion by others makes a unique contribution to the prediction of TSE changes.

Simple Model



Mediation Model 1



Mediation Model 2

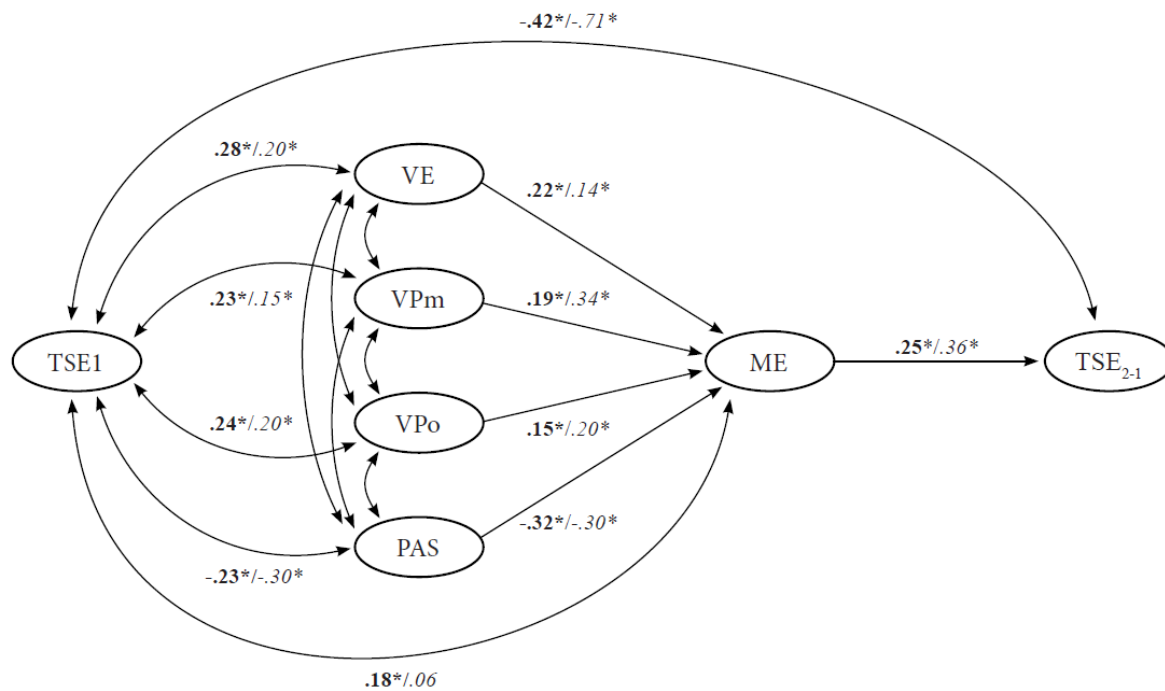


Figure 1. Structural latent true change model with standardized path coefficients of the competing models for each group of pre-service teachers.

Note: The first coefficients in bold print refer to the beginning preservice teachers, and the second coefficients in italic print refer to the advanced preservice teachers. TSE₂₋₁ = latent change in teacher self-efficacy; ME = mastery experiences; VE = vicarious experiences; VPm = verbal persuasion by the mentor teacher; VPo = verbal persuasion by “others”; PAS = physiological and affective states. Correlation coefficients between the predictor sources have been omitted from presentation to improve readability. * p < .05.

Table 5. Hierarchical analysis of the direct effects of the sources on TSE change

Predictor	Standardized path coefficient	R ²	ΔR ²
<i>Beginning group</i>			
ME	.24*	5.8%	-
ME/VE	.21*/.07	6.3%	0.5%
ME/VE/VPm	.15/.04/.18	8.6%	2.3%
ME/VE/VPm/VPo	.14/.04/.16/.04	8.8%	0.2%
ME/VE/VPm/VPo/PAS	.07/.00/.16/.03/-.19	11.7%	2.9%
<i>Advanced group</i>			
ME	.36*	12.8%	-
ME/VE	.34*/.04	13.0%	0.2%
ME/VE/VPm	.30*/.01/.09	13.2%	0.2%
ME/VE/VPm/VPo	.28*/-.01/.00/.19	15.7%	2.5%
ME/VE/VPm/VPo/PAS	.32*/.00/-.01/.18/.09	16.4%	0.7%

Note: Beta weights are reported in the same order as the corresponding predictors.

Mediation model 1

In this model the impact of the verbal persuasion sources on TSE changes is fully mediated by mastery experiences. Subsequently this model explains less variance in TSE changes, specifically 9.7% in the beginning group, and 14.3% in the advanced group. Changes in TSE were significantly predicted by preservice teachers' rating of their mastery experiences during the practicum in the advanced group only, although there was still a small effect in the beginning group. In sum, both verbal persuasion factors explained 21.4% and 40.6% of variance in the mastery experiences source in the beginning and the advanced group, respectively. Verbal persuasion by the mentor teacher significantly predicted mastery experiences in each sample, and did so to a greater degree for the advanced group. Verbal persuasion by others predicted mastery experiences in both groups to virtually the same degree, but the significance level was just missed in the beginning group. Similarly, to the direct model, vicarious experiences did not exert a direct effect on TSE changes in either group. PAS exerted a small but insignificant direct effect on TSE changes in both groups of preservice teachers.

Mediation model 2

In this model the impact of all other sources on TSE changes is fully mediated by mastery experiences. Consequently this model explains less variance in TSE changes, specifically 6.2% in the beginning group, and 13.1% in the advanced group. In both samples, changes in TSE were significantly predicted by preservice teachers' rating of their mastery experiences during the practicum. As expected, this effect was smaller in the beginning group (small to medium effect size) than in the advanced group (medium effect size). In sum, the vicarious experiences, verbal persuasion, and PAS sources predicted 39.8% and 50.5% of variance in the mastery experiences source in the beginning and the advanced group, respectively. Each source significantly predicted mastery experiences in each sample. However, the weights associated with each source predictor were not identical for the two samples. Specifically, vicarious experiences exerted a greater effect on mastery experience in the beginning than in the advanced group. This is in line with the expectation that vicarious experiences should have a greater impact during the observation practicum. Verbal persuasion by the mentor teacher had a smaller effect on mastery experiences in the beginning group, and the largest (medium-sized) effect of all predictors on mastery experiences in the advanced group. Verbal persuasion by others also carried a slightly greater weight in the advanced group. PAS exerted a strong (medium-sized) effect on mastery experiences in both groups of preservice teachers.

Verbal persuasion by “others”

The response rate to this open question was 52% in Sample 1 and 59% in Sample 2. Answers of participants who responded are presented in Table 6. Students were most likely to be a source of verbal persuasion in both samples, however, even more so in the advanced group. The next most common source of verbal persuasion in both samples were other teachers, meaning teachers that were not the mentor teacher. The next two most common sources of verbal persuasion named by the beginning group were school principals and other school staff (e.g., teacher aides, social workers), while the next two most common sources of verbal persuasion for the advanced preservice teachers were their peers and university supervisor. Only the advanced group had a supervisor at university, who taught the preparatory seminar, observed a lesson taught by the preservice teachers during their practicum, and then provided feedback afterwards. Peers were other preservice teachers undertaking their practicum at the same time. Verbal persuasion from peers played a lesser role in the beginning group. A small proportion of beginning preservice teachers even reported feedback from parents, whereby this source was virtually non-existent in the advanced group.

Table 6. List of “others” that acted as a source of verbal persuasion for preservice teachers during the practicum

Others	Beginning group ($n = 118$)		Advanced group ($n = 132$)	
	Count	Percent	Count	Percent
Students	76	64.4%	94	71.2%
Other teachers	39	33.1%	38	28.8%
School principal	25	21.2%	11	8.3%
Other school staff	15	12.7%	3	2.3%
University supervisor	–	–	22	16.7%
Peers	6	5.1%	23	17.4%
Parents	4	3.4%	1	0.8%

Note: Participants could list multiple “others”, thus percentages do not add up to 100.

4.4 Discussion

Results and implications

The current study presented a new measure of Bandura’s (e.g., 1997) proposed four sources of self-efficacy for TSE. In addition, it was explored how the four sources could predict TSE changes during two different practicum formats of preservice teachers at different stages of teacher education. Taken together, the results are a first indication of the good construct validity, convergent validity, and reliability of the self-developed source scales. All three competing models showed the same good fit to the data. The model comparisons provided first evidence in support of the mediation model 2. This model showed an equivalent fit to both the less constrained simple model and the mediation model 1, and should thus be favored (e.g., Schermelleh-Engel et al., 2003). Furthermore, including direct paths between the sources predicting mastery experiences and TSE change did not improve model fit. Due to the shared variance among the sources, the simple model (all four sources directly predict TSE changes) showed that the unique contributions of each source to TSE changes were small (with the exception of mastery experiences in the advanced group). Similarly, the mediation model 1 also showed that the unique direct contributions of the sources were small (with the exception of mastery experiences in the advanced group). Taken together, the results from the model comparison provide first evidence in favor of the hypothesis that preservice teachers’ changes in TSE are directly influenced by the mastery experiences they gather during a

practicum at school. Their mastery experiences, in turn, are largely informed by the other three sources, whereby the contribution of each source differs for the two groups (i.e., practicum formats). Whereas PAS, vicarious experiences, and verbal persuasion by the mentor teacher had the largest influence on beginning preservice teachers' ratings of mastery experiences, verbal persuasion by the mentor teacher, PAS, and verbal persuasion by others had the largest effect on advanced preservice teachers' ratings of mastery experiences. In total, the three sources explained about half of the variation in the mastery experience ratings in the advanced group and 40% in the beginning group. With respect to Bandura's (e.g., 1997) theory of the sources, this does not mean that there is no influence of the verbal persuasion, vicarious experiences, and PAS sources on the change in TSE, but rather that their effects on the TSE development of preservice teachers could be entirely mediated by mastery experiences.

Furthermore, as expected, the mediation model 2 showed stronger associations between mastery experiences and TSE changes in the advanced group than in the beginning group, while associations between vicarious experiences and mastery experiences were stronger in the beginning group as compared to the advanced group. Both results are in line with Bandura's (e.g., 1997) description of the sources, adding evidence in favor of the construct validity of the source measures. According to Bandura's description, the mastery experiences in the advanced group should carry a greater weight with regard to TSE development because they are based on actual teaching experiences, which provide more authentic evidence of one's teaching capabilities when compared to the beginning group. Vicarious experiences, on the other hand, are expected to play a greater role in the beginning group, which had more opportunity to observe other teachers.

Perhaps most interesting to practitioners and teacher educators, this study showed that positive feedback from the mentor teacher has a significant positive influence on the development of preservice teachers' TSE, possibly via the perception of mastery experiences. Again, this result is in line with Bandura's (1997) description of verbal persuasion as being particularly influential when it comes from somebody who has expert knowledge in the evaluated field: the mentor teacher. Most likely the influence was stronger during the teaching practicum than the observation practicum because mentors could here base their feedback on preservice teachers' actual teaching performance, which would make the feedback more credible. According to Bandura (1997), credibility is another important aspect that increases the effect that verbal persuasion has on self-efficacy. This finding underscores the significant role mentor teachers at school play in the development of preservice teachers' TSE,

particularly during teaching experiences. It further adds to the qualitative evidence put forward by Klassen and Durksen (2014) regarding the significance of verbal persuasion by the mentor teacher on preservice teachers' TSE changes during a teaching practicum. Consequently, one recommendation to teacher educators, interested in supporting preservice teachers' TSE development, would be to prepare mentor teachers at school with regard to the process of giving feedback to preservice teachers. Drawing on Bandura's (1997) description of the framing of performance feedback, mentor teachers should phrase their feedback in terms of achieved progress towards a certain standard, in order to support TSE development. In comparison, feedback phrased with regard to shortfalls from a certain standard is likely to have detrimental effects on TSE development. In addition, with respect to attributional feedback, emphasizing preservice teachers' ability rather than their invested effort, should also enhance TSE. This is particularly true early on during skill development. Telling preservice teachers that their performance progress shows they have ability for teaching, increases their TSE more than feedback that only and repeatedly focuses on effort. The latter approach conveys that they can only ever meet a standard when they expend consistent high effort, which produces a lower sense of self-efficacy than ability attribution.

The qualitative findings of the present study offer a first systematic insight into who, in addition to the mentor teacher, influences preservice teachers' TSE beliefs via verbal persuasion. Here, especially students seem to play a noteworthy role. It would then make sense to take this fact into account, for example, when assigning a particular class to preservice teachers during their practicum. Or it might even be useful to brief students in advance, regarding the impact that their feedback could have on preservice teachers.

Especially interesting in light of the few previous empirical findings regarding PAS, the current study shows that negative PAS contributed strongly across both groups to reduced mastery experiences and consequently to a decrease in TSE. Moulding et al. (2014) state that traditional teacher preparation programs generally involve all sources of TSE except PAS. The current results provide a convincing reason for addressing this gap in teacher preparation. Since Bandura (1997) highlights that PAS (and the other sources) do not influence self-efficacy beliefs directly but via cognitive processing, it seems worthwhile to formally integrate, for instance, practicing emotional and physiological self-regulation strategies into the practicum experience. This could possibly be implemented as part of reflective processes guided by the mentor teacher in pre- and post-observation conferences before and after first teaching attempts (for a review on mentoring of beginning teachers see Hobson, Ashby, Malderez, & Tomlinson, 2009).

Since there are no prior studies quantifying the contributions of each source to TSE changes, it is difficult to judge the practical relevance of the predictive magnitudes seen in the current study. Nonetheless, the large amount of unexplained variance clearly indicates that there are other factors at play in TSE development. As of yet, some researchers have explored the impact of other factors, usually under the umbrella term of context factors (Knoblauch & Woolfolk Hoy, 2008; Moulding et al., 2014; Tschannen-Moran & Woolfolk Hoy, 2007). However, this research has not related those factors to actual *changes* in TSE, but rather to states, which cannot offer any indication of TSE development. Consequently, a comparison of the influence exerted by the sources vs. other factors is not yet possible. Nevertheless, in terms of theory development, the current magnitudes provide a benchmark. This benchmark will allow for a comparison of results produced by other factors, as well as the contribution of the four sources produced by other teacher education programs or differently designed practicum experiences.

Limitations and future directions

The design used in this study is correlational and thus no causality can be inferred between the sources and changes in TSE. A strictly uni-directional causal relationship between the sources and TSE development is unrealistic as predicted by theory (e.g., Tschannen-Moran et al., 1998), and as the correlations between the sources and the level of TSE at T1 in this study demonstrate. However, the timing of the study (i.e., examining TSE development in preservice teachers) might have allowed for minimal effects of TSE on the sources, and maximal effects for the sources on TSE, since TSE is hypothesized to be most malleable by the sources early on (e.g., Woolfolk Hoy & Burke Spero, 2005). While one model of how the sources interact in predicting TSE changes emerged as the most parsimonious model with an identical fit to the other models (and should thus be favored), there could be numerous other relationships between the sources that might explain the data equally well. Future research employing experimental designs would be needed to demonstrate the assumed causal nature among the sources. Since teacher education programs differ considerably across the world (OECD, 2005), examining to what degree results are similar would contribute to validating the present findings or, in any case, to enhancing our understanding of how the sources interact.

Furthermore, there was a small attrition bias with regard to gender and age in the beginning group. This might impact the generalizability of the results from the observation practicum to male and older preservice teachers. While there was no attrition bias with regard

to the dependent variables (TSE), it is still possible that changes in TSE that occurred during the practicum were related to retention.

The present assessment of the sources is not independent of context and is thus limited in its general application. The developed source scales are designed specifically to be used for preservice teachers after a practicum experience at school. There are potentially many other aspects of teacher education (e.g., university course, induction year) for which this assessment of the sources would be unsuitable. However, the context-specific design might have made this measure successful. At the same time, this context-specificity also carries another limitation. The wording in two of the four mastery experiences items is not identical in the two samples, thus limiting the comparability of the mastery experiences source between the two groups.

In order to match the same level of generality of the source measure, TSE changes were examined for the higher-order TSE factor, rather than for the three TSE factors²⁹ (i.e., instructional strategies, classroom management, student engagement). Researchers might be interested in detecting more dimension-specific effects, requiring the development of dimension-specific source items. For example, mastery experience items could specifically refer to successes in using a certain method of instruction, managing difficult students, or engaging under-performing students. Exploring this avenue could be useful, for example, when evaluating or developing a teacher education program that has a certain profile focusing on one dimension, for example classroom management, which is often reported to be challenging for preservice teachers (e.g., Veenman, 1984). Dimension-specific source items could then be used in determining if this emphasis is reflected in preservice teachers' development of TSE in classroom management over the course of the program.

Building on the qualitative findings this study produced regarding verbal persuasion, future studies could be directed at determining what influence feedback of the different 'persuaders' has on TSE development. Are there interactions, for example, between feedback given by the mentor teacher at school and feedback given by the university supervisor? Whose feedback has a stronger influence: that of the students or the university supervisor?

Future research could also focus on how the contribution of each source, or even the interrelationships between the sources, change with career stage. A central assumption in this study is that due to a lack of previous mastery experiences, preservice teachers rely on the other three sources to inform their judgment of mastery experiences. This should not hold true

²⁹ In fact, the analysis was also conducted using the three dimensions of TSE as dependent variables. This showed that the effects of the sources were equivalent across all three dimensions.

for more experienced inservice teachers, who have accumulated an abundance of mastery experiences. Consequently, the simple model may well be valid for inservice teachers. Regarding the contribution of each source, PAS, for instance, might have a great impact early on in TSE development, but its influence might fade later on when teachers get used to being in front of a class. Determining what weight is associated with each source at different career stages could inform the design of professional development programs aimed at increasing TSE for inservice teachers, or teacher education programs for teacher candidates.

References

- Author et al. (2014) [details removed for peer review]
- Bagozzi, R. P., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *16*(1), 74–94. doi:10.1007/BF02723327
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52*, 1–26.
- Beltman, S., Mansfield, C., & Price, A. (2011). Thriving not just surviving: A review of research on teacher resilience. *Educational Research Review*, *6*(3), 185–207. doi:10.1016/j.edurev.2011.09.001
- Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education*, *16*, 239–253.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.
- Caprara, G. V., Steca, P., Barbaranelli, C., & Malone, P. S. (2006). Teachers' self-efficacy beliefs as determinants of job satisfaction and students' academic achievement: A study at the school level. *Journal of School Psychology*, *44*(6), 473–490. doi:10.1016/j.jsp.2006.09.001
- Chan, D. (1998). The conceptualization and analysis of change over time: An integrative approach incorporating Longitudinal Mean and Covariance Structures analysis (LMACS) and Multiple Indicator Latent Growth Modeling (MLGM). *Organizational Research Methods*, *1*(4), 421–483. doi:10.1177/109442819814004
- Cheung, H. Y. (2008). Teacher efficacy: A comparative study of Hong Kong and Shanghai primary in-service teachers. *The Australian Educational Researcher*, *35*(1), 103–123.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cronbach, L. J., & Furby, L. (1970). How should we measure "change" - or should we? *Psychological Bulletin*, *74*(1), 68–80.

- Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34*(9), 917–928. doi:10.1093/jpepsy/jsp004
- Fives, H., Hamman, D., & Olivarez, A. (2007). Does burnout begin with student-teaching? Analyzing efficacy, burnout, and support during the student-teaching semester. *Teaching and Teacher Education, 23*, 916–934.
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological bulletin, 103*(2), 265–275.
- Henson, R. K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist, 37*(3), 137–150. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/S15326985EP3703_1#.UnfabKMwfcs
- Hobson, A. J., Ashby, P., Malderez, A., & Tomlinson, P. D. (2009). Mentoring beginning teachers: What we know and what we don't. *Teaching and Teacher Education, 25*(1), 207–216. doi:10.1016/j.tate.2008.09.001
- Holzberger, D., Philipp, A., & Kunter, M. (2013). How teachers' self-efficacy is related to instructional quality: A longitudinal analysis. *Journal of Educational Psychology, 105*(3), 774–786.
- Hu, L., & 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
- Klassen, R. M., & Chiu, M. M. (2010). Effects on teachers' self-efficacy and job satisfaction: Teacher gender, years of experience, and job stress. *Journal of Educational Psychology, 102*(3), 741–756.
- Klassen, R. M., & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching context. *Contemporary Educational Psychology, 36*(2), 114–129. doi:10.1016/j.cedpsych.2011.01.002
- Klassen, R. M., & Durksen, T. L. (2014). Weekly self-efficacy and work stress during the teaching practicum: A mixed methods study. *Learning and Instruction, 33*, 158–169. doi:10.1016/j.learninstruc.2014.05.003
- Klassen, R. M., & Tze, V. M. (2014). Teachers' self-efficacy, personality, and teaching effectiveness: A meta-analysis. *Educational Research Review, 12*, 59–76. doi:10.1016/j.edurev.2014.06.001

- Klassen, R. M., Tze, V. M. C., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise? *Educational Psychology Review, 23*, 21–43.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). *Methodology in the social sciences*. New York, NY: Guilford Press.
- Knoblauch, D., & Woolfolk Hoy, A. (2008). “Maybe I can teach those kids.” The influence of contextual factors on student teachers’ efficacy beliefs. *Teaching and Teacher Education, 24*, 166–179. doi:10.1016/j.tate.2007.05.005
- Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 151–173. doi:10.1207/S15328007SEM0902_1
- Malmberg, L.-E. (2008). Student teachers' achievement goal orientations during teacher studies: Antecedents, correlates and outcomes. *Learning and Instruction, 18*(5), 438–452. doi:10.1016/j.learninstruc.2008.06.003
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology, 60*(1), 577–605. doi:10.1146/annurev.psych.60.110707.163612
- Milner, R. H., & Woolfolk Hoy, A. (2003). A case study of an African American teacher's self-efficacy, stereotype threat, and persistence. *Teaching and Teacher Education, 19*(2), 263–276. doi:10.1016/S0742-051X(02)00099-9
- Morris, D. B., & Usher, E. L. (2011). Developing teaching self-efficacy in research institutions: A study of award-winning professors. *Contemporary Educational Psychology, 36*(3), 232–245. doi:10.1016/j.cedpsych.2010.10.005
- Moulding, L. R., Stewart, P. W., & Dunmeyer, M. L. (2014). Pre-service teachers' sense of efficacy: Relationship to academic ability, student teaching placement characteristics, and mentor support. *Teaching and Teacher Education, 41*, 60–66. doi:10.1016/j.tate.2014.03.007
- Mulholland, J., & Wallace, J. (2001). Teacher induction and elementary science teaching: enhancing self-efficacy. *Teaching and Teacher Education, 17*, 243–261.
- Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- OECD. (2005). *Teachers matter: Attracting, developing and retaining effective teachers* (Education and training policy). Paris.

- O'Neill, S. C., & Stephenson, J. (2011). The measurement of classroom management self-efficacy: A review of measurement instrument development and influences. *Educational Psychology Review*, *31*(3), 261–299. doi:10.1080/01443410.2010.545344
- Palmer, D. H. (2001). Factors contributing to attitude exchange amongst preservice elementary teachers. *Science Teacher Education*, *86*(1), 122–138. doi:10.1002/sce.10007
- Palmer, D. H. (2006). Sources of self-efficacy in a science methods course for primary teacher education students. *Research in Science Education*, *36*(4), 337–353. Retrieved from <http://link.springer.com/article/10.1007%2Fs11165-005-9007-0>
- Poulou, M. (2007). Personal teaching efficacy and its sources: Student teachers' perceptions. *Educational Psychology*, *27*(2), 191–218. doi:10.1080/01443410601066693
- Reuter, T., Ziegelmann, J. P., Wiedemann, A. U., Geiser, C., Lippke, S., Schüz, B., & Schwarzer, R. (2010). Changes in intentions, planning, and self-efficacy predict changes in behaviors: An application of latent true change modeling. *Journal of Health Psychology*, *15*(6), 935–947. doi:10.1177/1359105309360071
- Ruble, L. A., Usher, E. L., & McGrew, J. H. (2011). Preliminary investigation of the sources of self-efficacy among teachers of students with autism. *Focus on Autism and Other Developmental Disabilities*, *26*(2), 67–74. doi:10.1177/1088357610397345
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research - Online*, *8*(2), 23–74.
- Schwarzer, R., & Hallum, S. (2008). Perceived teacher self-efficacy as a predictor of job stress and burnout: Mediation analyses. *Applied Psychology: An International Review*, *57*, 152–171.
- Steyer, R., Eid, M., & Schwenkmezger, P. (1997). Modeling true intraindividual change: True change as a latent variable. *Methods of Psychological Research - Online*, *2*(1), 21–33.
- Steyer, R., Partchev, I., & Shanahan, M. J. (2000). Modeling true intraindividual change in structural equation models: The case of poverty and children's psychosocial adjustment. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data* (pp. 109–126). Mahwah, NJ: Lawrence Erlbaum Associates.
- Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education*, *17*, 783–805.

- Tschannen-Moran, M., & Woolfolk Hoy, A. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education*, 23(6), 944–956. doi:10.1016/j.tate.2006.05.003
- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68(2), 202–248. Retrieved from <http://www.jstor.org/stable/1170754?origin=JSTOR-pdf&>
- Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research*, 78(4), 751–796. Retrieved from <http://rer.sagepub.com/content/78/4/751>
- Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research*, 54(2), 143–178. Retrieved from <http://rer.sagepub.com/content/54/2/143>
- Woolfolk, A. E., & Hoy, W. K. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology*, 82(1), 81–91.
- Woolfolk Hoy, A., & Burke Spero, R. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, 21(4), 343–356. doi:10.1016/j.tate.2005.01.007
- Wothke, W. (2000). Longitudinal and multigroup modeling with missing data. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data* (pp. 219–240). Mahwah, NJ: Lawrence Erlbaum Associates.

4.5 Acknowledgments

I would like to express my sincere thanks to Martin Schultze from the Freie Universität Berlin for his help with the trickier aspects of latent true change modeling.

5

General Discussion

5 General Discussion

Since the mid-1970s, teacher self-efficacy (TSE), in one form or another, has attracted a large amount of interest in the field of educational psychology in general (e.g., Klassen, Tze, Betts, & Gordon, 2011; Woolfolk Hoy, Davis, & Pape, 2006), and research on teaching and teacher education in particular (e.g., Kleinsasser, 2014). Despite the many documented positive outcomes associated with TSE (summarized in section 1.4 of Chapter 1), quantitative research on the development of TSE beliefs has been scarce (e.g., Klassen et al., 2011). Furthermore, longitudinal approaches, which are a necessary prerequisite to any study of development (Singer & Willett, 2003), have essentially been absent from such endeavors (as detailed in section 1.3 of Chapter 3). The overarching purpose of this thesis was to address this gap in TSE research. In order to do so, the studies in this thesis were based on a thorough theoretical analysis of the TSE construct and the sources of self-efficacy proposed by Bandura (e.g., 1997), as outlined in the introductory Chapter 1. Following this analysis, the instrument development and first validation of the TSE measure used in this thesis is described (Chapter 2). Then, changes in preservice teachers' TSE during two different aspects and at two different stages of university teacher preparation are documented (Chapter 3). Further, it is explored how these changes in TSE relate to preservice teachers' intention to quit (ITQ) their teaching degree (Chapter 3). The last empirical chapter presents a newly developed instrument for the assessment of the four sources of TSE and explores different hypotheses of how the sources interact in predicting TSE changes during a practicum at school (Chapter 4).

In the following, first the results of the three studies presented in the preceding empirical chapters are summarized and discussed with regard to the four research questions posed at the beginning of this thesis. Second, the limitations and strengths of the thesis are outlined. Third, implications for theory and practice are highlighted. And fourth, future directions for further research are proposed, building on the results of this thesis.

5.1 Summary and discussion of results

A succinct overview of the aims, main findings and conclusions of the three studies in this thesis is presented in Table 1.

It was the first aim of this thesis to develop an instrument for measuring TSE that fulfilled a range of criteria. Following an analysis of the requirements for the purpose of this thesis on the one hand, and the available measures on the other hand, the Teachers' Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001), currently the predominant TSE instrument (Duffin, French, & Patrick, 2012), was adapted with regard to two aims: (1) to provide a valid TSE assessment on the conceptual basis of Bandura's (e.g., 1997; 2006) theory of self-efficacy; (2) to provide a reliable three-dimensional assessment of TSE not only for advanced but also for beginning preservice teachers. After a pre-selection of items, and an adjustment of the wording of the instruction to participants, the introductory phrases of the items, and the response categories of the TSES, this adapted measure was administered to three samples: beginning and advanced preservice teachers in Germany, and advanced preservice teachers in New Zealand. Using a data-driven process, four items that best represented each of the three TSE dimensions were selected for the adapted instrument. In order to evaluate this instrument, a range of psychometric properties were inspected, several tests of invariance were conducted, and the convergent validity with general self-efficacy and study satisfaction was examined. Results from this first study indicated good psychometric properties for the adapted instrument (i.e., the Scale for Teacher Self-Efficacy = STSE), an excellent fit and metric level invariance of the three-factor solution for all three samples.

Table 1. Summary of thesis aims, findings, and conclusions

	Aims	Findings	Conclusions
Chapter 2	<ul style="list-style-type: none"> to adapt the TSES with regard to a valid and reliable three-dimensional assessment of TSE in beginning and advanced preservice teachers 	<ul style="list-style-type: none"> acceptable to excellent internal consistencies and construct reliability of all scales metric level invariance of the three TSE scales for all samples of preservice teachers medium to large-sized correlations between all TSE scales and GSE, which were invariant across all samples, and larger than those found using the TSES correlation between study satisfaction and TSE in Sample 2 comparably related to job satisfaction and TSE found in other studies 	<ul style="list-style-type: none"> new measure presents a reliable three-dimensional assessment of TSE for beginning and advanced preservice teachers first indicators of good construct and convergent validity in two cultures (i.e., German and English speaking) improvement over the TSES in terms of assessing underlying self-efficacy needs to be validated for use with inservice teachers in the future
Chapter 3	<ul style="list-style-type: none"> to investigate intra-individual changes in TSE of two cohorts of preservice teachers (beginning and advanced) during the coursework at university and during the practicum at school to determine if TSE changes influenced changes in preservice teachers' intention to quit the program 	<ul style="list-style-type: none"> latent TSE changes occurred both during coursework at university and the practicum at school in each cohort TSE for classroom management changed the most average TSE changes differed by cohort (i.e., stage of the teacher preparation program) and across individuals TSE changes during a practicum but not during the coursework were associated with preservice teachers' intention to quit (ITQ) the teaching degree in each cohort 	<ul style="list-style-type: none"> TSE development in preservice teachers occurs not only during the practicum, where most research efforts have been directed so far, but also during the coursework at university TSE changes only influence preservice teachers' ITQ, when they occur during the practicum, however the causal link still needs to be demonstrated using a different design
Chapter 4	<ul style="list-style-type: none"> to develop a valid and reliable measure of Bandura's four sources for TSE to explore how the four sources of TSE interact in predicting changes in TSE 	<ul style="list-style-type: none"> good to excellent internal consistencies and construct reliability of all source scales, good CFA fit with 5 factors three models of different relationships among the sources showed identical fit, one was the most parsimonious mastery experiences (ME) emerged as the strongest predictor in each group and were largely informed by the other three sources of TSE verbal persuasion by mentor teachers and negative physiological and affective states also played an important role for TSE development, mediated by ME 	<ul style="list-style-type: none"> new measure is suited to assess Bandura's four sources for TSE with regard to both an observation and a teaching practicum model comparisons provided first evidence in favor of the hypothesis that ME mediate the influence of the other sources on TSE, since preservice teachers are lacking an internal frame of reference for judging ME future research needs to provide further evidence for this causal assumption between the sources

Note: TSE = teacher self-efficacy; GSE = general self-efficacy; TSES = teachers' sense of efficacy scale; CFA = confirmatory factor analysis.

Overall, the magnitude of the correlations between the STSE subscales and general self-efficacy as well as study satisfaction were in line with the hypotheses and thus provided first evidence of convergent validity. In addition, the magnitudes were all larger in comparison to the correlations between the same variables derived using the TSES in a master thesis (Böttcher, 2013) using comparable samples. Construct invariance for the associations found between STSE subscales and general self-efficacy across the three samples also provided evidence of convergent validity, as did construct invariance found for the associations between study satisfaction and TSE subscales across the two German samples where study satisfaction was assessed. Further evidence of a good convergent validity using another criterion stems from Study 2. Here, the results showed a negative association between initial TSE and ITQ levels at T1 (see Figure 1 in Chapter 3). This corresponds with what is known about the relationship between TSE and commitment from meta-analytic data (Chesnut & Burley, 2015). However, the comparison to meta-analytic data using commitment as a criterion might be somewhat too distal. A more specific comparison is possible to the effects reported by Klassen and Chiu (2011). In their study on the commitment *and* ITQ (which they distinguished as two separate outcomes) of practicing and preservice teachers, Klassen and Chiu (2011) report correlations between ITQ and the three TSE dimensions, assessed with the TSES (Tschannen-Moran & Woolfolk Hoy, 2001), for a sample of preservice teachers. Their sample is comparable, with regard to the stage of teacher education, to the advanced group in Study 2. The correlations found by Klassen and Chiu (2011) are also negative but of smaller magnitude and very similar across the three dimensions: $-.17$ for TSE in instructional strategies, $-.19$ for TSE in classroom management, and $-.18$ for TSE in student engagement. In the advanced group of Study 2 on the other hand, the negative association with ITQ was of medium³⁰ size for all TSE dimensions, but effect sizes did vary by dimension: $-.39$ for TSE in instructional strategies, $-.26$ for TSE in classroom management, and $-.36$ for TSE in student engagement. This could be indicative of two improvements of the STSE over the TSES. One, the adapted scale does indeed capture the underlying construct of self-efficacy better, because the relationship to the outcome is stronger. This is in line with Bandura's (1997; 2006) claim of greater predictive validity of measures closely aligned with self-efficacy theory, which has received first empirical support in the area of TSE from

³⁰ According to Cohen (1992) effect sizes of r can be judged as small = $.10$, medium = $.30$, and large = $.50$.

Chesnut and Burley (2015). Two, the three-dimensionality is clearer with preservice teachers, as the magnitudes of the effects found with the STSE differed by TSE dimension, whereas the effects found with the TSES did not. Again, as in the first study, the associations between TSE and a potential outcome of TSE (i.e., study satisfaction and ITQ) in the beginning group were of smaller effect size than in the advanced group, and of no effect for TSE in classroom management. This might add further evidence to the hypothesis discussed in the second study, according to which beginning preservice teachers at the very start of their university studies miscalibrate their TSE in classroom management. Since there are no studies with beginning preservice teachers that relate TSE to an outcome, it is hard to judge what the smaller effect sizes for the relationships observed in Study 1 and Study 2 mean. They may well be an accurate reflection of the association between TSE and outcome measures one could expect for preservice teachers at such an early stage of teacher education.

In addition, the good fit and longitudinal scalar measurement invariance (i.e., same factor structure, factor loadings, and intercepts in a sample over time) found for Sample 1 and 2 in Study 2, and Sample 2 and 4 combined in Study 3, provide further evidence of the good construct validity and measurement properties of the three-dimensional STSE.

What have the results of the second and third study shown with regard to the four central research questions of this thesis?

Does TSE develop in preservice teachers during their university preparation?

Study 2 (in Chapter 3) was designed to address this question. For this purpose, intra-individual changes in TSE were analyzed for two cohorts of preservice teachers at two different stages of teacher education and for two different aspects of a teacher preparation program: coursework at university and a practicum at school. This approach was chosen to extend previous research, which has mostly focused on TSE changes during a teaching practicum in advanced preservice teachers. In order to determine changes in each dimension of TSE, latent true change (LTC) analysis was applied, because it represents an improvement over previous studies using manifest approaches for the analysis of change, as such results can be unreliable (e.g., Millsap & Hartog, 1988).

Results from Study 2 showed that, on average, TSE increased during the practicum in both cohorts, but did so to a different extent for the three TSE dimensions.

TSE in classroom management increased the most and TSE in student engagement increased the least in absolute terms in both cohorts. Effect sizes were substantial³¹ for increases in instructional strategies and classroom management TSE in both cohorts, and showed no effects for TSE in student engagement. This rather similar result across two differently designed practicum experiences is interesting in light of Bandura's (e.g., 1997) theory of the sources. It shows that even though mastery experiences could only have occurred vicariously rather than enactively in the beginning cohort, the TSE changes that occurred in this group were nonetheless of similar size as compared to the advanced group.

Changes that occurred during the coursework at university, on the other hand, occurred in different directions for the two groups. While TSE increased substantially for each TSE dimension in the advanced group, it decreased substantially for TSE in classroom management in the beginning group, and did not change on average in the other two dimensions of TSE. In comparison to the only other study documenting TSE changes during coursework (by Mergler & Tangen, 2010), which found increases in TSE, this shows that the stage of teacher education does seem to matter for TSE development. Both Mergler and Tangen's (2010) sample of preservice teachers and Sample 2 in this second study were advanced in teacher education, and both showed increased TSE, while the beginning preservice teachers in Sample 1 experienced decreased TSE, albeit only in classroom management. One explanation offered in the discussion of Study 2 for this unexpected result, was that this downward change in TSE for classroom management of beginning preservice teachers might be due to some sort of miscalibration of TSE at the beginning of teacher education. This miscalibration could be unique to classroom management, due to either limited knowledge (e.g., Klassen & Chiu, 2011) particularly about this aspect of teaching, or unrealistic optimism (e.g., Weinstein, 1988), or both, whereby unrealistic optimism might result when little is known about a particular competence area (Dunning, Heath, & Suls, 2004). However, out of the three TSE dimensions, notably TSE for classroom management was rated lowest by all samples at all times of assessment (see Table 2 in Chapter 2, Table 2 in Chapter 3, and Table 1 in Chapter 4). Yet, this lower appraisal of TSE in classroom management, at the beginning of teacher education, might still not have been realistic for the beginning preservice teachers. The non-existent association

³¹ The effect size Cohen's (e.g., 1992) *d* of a latent change was deemed "substantial" if it was similar to the average effect of $d = .16$ found in the meta-analysis by Ashford, Edmunds, and French (2010) for intervention-induced changes in self-efficacy.

between TSE in classroom management and study satisfaction in Study 1, and also TSE in classroom management and ITQ in Study 2 could be further indicators that particularly TSE in classroom management is somehow miscalibrated in preservice teachers at the start of teacher education. At the same time, the correlation between TSE in classroom management and general self-efficacy reported in Study 1 (see Table 3 in Chapter 2) was equivalent to the other two TSE dimensions in the beginning group and to all other TSE dimensions in the advanced group. This shows that the underlying construct of self-efficacy was, in any case, represented as well in this dimension in the beginning group as it was in the other two dimensions in this group and in all dimensions in the advanced group.

So in response to the first research question, it can be stated that yes, TSE develops during university teacher preparation, but it does so differently for each TSE dimension as well as for different aspects of the teacher preparation program (i.e., coursework vs. practicum), and depends on the stage of teacher education (i.e., beginning vs. advanced).

Do changes in TSE have any impact on preservice teachers' changes in their intention to quit the teaching degree?

In order to answer this question, preservice teachers' ITQ the teaching degree they were currently enrolled in, was assessed at each time in Study 2. Previous research (e.g., Klassen & Chiu, 2011) had indicated that there is a negative relationship between the static levels of TSE and ITQ of preservice teachers, meaning that greater average values of TSE are associated with a lower ITQ. However, *changes* in each construct had not been related to each other previously, but a negative association could be expected. The results of Study 2 showed that overall this was the case for TSE changes, which occurred during the practicum, but not for TSE changes during the coursework at university. Again, these associations differed by TSE dimension. In the beginning group, the only significant effect was found for TSE in instructional strategies, with a medium-sized effect. Non-significant but small-sized effects did occur for the other TSE dimensions. In the advanced group, changes in TSE for instruction and classroom management exerted medium-sized significant effects on ITQ changes. In this group, the small effect of TSE in classroom management was non-significant. While none of the effects between TSE and ITQ changes during the coursework at university were significant, nearly all effects were of small size. As there are no previous studies that

could be used as prior benchmarks, however, it is simply not known yet whether this small effect has any meaningful implications for the attrition or retention of preservice teachers during their studies.

Taken together, the second question can be answered as follows: yes, changes in TSE impact on preservice teachers' ITQ, but do so to a greater extent during the practicum and only to a small, possibly negligible, extent during the coursework at university. Furthermore, the association between TSE and ITQ is most pronounced for TSE in instruction in the beginning group, and TSE in instruction and student engagement in the advanced group. However, while it is both tempting and in accordance with self-efficacy theory to assume a causal effect from TSE on ITQ, this causation could not be demonstrated due to the correlational design used in Study 2 (e.g., Shadish, Cook, & Campbell, 2002).

There was a significant amount of inter-individual variation in the changes in TSE, which brings us to the next research question.

Can changes in TSE during the practical phases of a teacher preparation program be predicted by the four sources of information assumed to be influential in forming TSE beliefs?

This question was addressed in Study 3, which was set up to attempt to explain some of the inter-individual variation in TSE development during the practicum. In order to do so, Bandura's (e.g., 1997) theory of the sources of self-efficacy was utilized for the development of a source measure of TSE. This development was necessary, as the only quantitative measure of the sources of TSE by Poulou (2007) showed little alignment with Bandura's (e.g., 1997) theory, for example combining the mastery experiences and verbal persuasion source. This practice has been criticized in a review by Usher and Pajares (2009) on the sources of self-efficacy in school. In addition, this instrument seems to have had little success in the scientific community, as so far only one other study (O'Neill & Stephenson, 2011) has been published using this instrument. Moreover, results from both studies did not converge with regard to their insights on the effects of the sources on TSE development.

As a first step in Study 3, the newly developed instrument was subjected to several tests of reliability and construct validity. Results showed good psychometric properties of the source measures. A first indication of construct validity was provided using confirmatory factor analysis (CFA). While a five-factor solution was needed to represent the sources well, this was deemed acceptable with regard to theory, as all

items fit their respective constructs as hypothesized in advance. Only the verbal persuasion source items were better represented by two factors, each reflecting a different source of the persuasion. Inspecting the bivariate correlations of the source constructs with TSE changes during the practicum provided further evidence of construct validity. As expected, all sources were significantly related to TSE changes with moderate to large effect sizes in the hypothesized directions. Also in line with expectations, among all sources, mastery experiences consistently showed the strongest association with TSE changes in both groups. Somewhat unexpectedly however, vicarious experiences were equally related to TSE changes in both groups. This association was expected to be larger in the beginning group, since here the focus of the practicum had been on the observation of classes. Also in line with predictions, the sources were substantially related to each other. In comparison to the instrument presented by Poulou (2007), it can be noted that the latent interfactor correlations between mastery experiences and verbal persuasion found in Study 3 did not warrant a common factor, as they ranged between .39 and .61, which is well below the threshold of .80 recommended for distinct factors by Brown (2006). Furthermore, mastery experiences showed similarly high intercorrelations with other source factors (e.g., -.50 with PAS).

Due to the substantial intercorrelations between the sources, when all sources were included and initial TSE levels were taken into account the individual weights of the sources in the prediction of TSE changes were quite different in comparison to the bivariate correlations. In response to the research question though, the results showed that together all four sources predicted about 12% of the variance in TSE changes in the beginning group, and about 16% of the variance in TSE changes in the advanced group. However, with the exception of mastery experiences in the advanced group, none of the weights of each source associated with this prediction were significantly different from zero. Further investigation of this rather perplexing phenomenon revealed that while indeed mastery experiences explained the most variance in TSE changes in each group (5.8% in the beginning group and 12.8% in the advanced group), the weights of the sources dropped with the introduction of each new source, because, due to their intercorrelations, the prediction that the sources made were not unique. So, how could these intercorrelations be modeled?

In which way do the four sources interact in predicting changes in TSE?

This question was explored in Study 3 by evaluating three different plausible models of how the sources could interact in the prediction of TSE changes during the practicum. Since there is no specific guidance by Bandura on this issue and no previous research has offered quantitative results in response to this question, these models are somewhat arbitrary and there could be many other plausible models that could have been tested. However, the three chosen models were all plausible with regard to theoretical considerations. In the first and least parsimonious model, no assumptions about the relationships between the sources were made. All sources were simply correlated and directly predicting TSE changes. The second and third model placed mastery experiences into a mediating role between the other sources and TSE changes. In the second model, only verbal persuasion was mediated, whereby in the third and most parsimonious model all other sources were mediated. It was hypothesized that a mediating role of mastery experiences may be plausible for preservice teachers, because they are likely to lack a frame of reference for judging what would constitute mastery. The three models were then evaluated with regard to their fit to the data. Since model fit was equally good in all three models, according to common principle (e.g., Schermelleh-Engel, Moosbrugger, & Müller, 2003) the most parsimonious model should be chosen over the less parsimonious models. In addition, including direct paths between the sources that were mediated by mastery experiences and TSE changes did not improve the model fit in either of the mediation models. This is further evidence that the direct influence of those sources on TSE changes is of little effect and does not add substantially to the prediction of variance that mastery experiences already account for. Nonetheless, including direct paths did increase the total amount of variance that could be explained, particularly in the beginning group.

Since the most parsimonious model also assumes causal relationships between the sources, which could not be proven in the correlational design of Study 3, this model can only be viewed as a preliminary result regarding the way that the sources interact in predicting TSE changes. Interestingly, the weights assigned to the sources in this model corresponded closely to what was hypothesized on the basis of Bandura's (e.g., 1997) theory of the sources. In this model, vicarious experiences had a greater weight in the beginning group than in the advanced group (i.e., .22 vs. .14, respectively). This effect, however, was through informing mastery experiences, rather than TSE changes directly. So, while initially the bivariate correlations for this source seemed to contradict

expectations formed on the basis of theory, and thus would have cast doubt on the validity of the source measure, taking a possible mediating role of mastery experiences into account, theory-corresponding effects are found. Similarly, in line with expectations, the effect of mastery experiences was stronger in the advanced than in the beginning group (i.e., .36 vs. .25, respectively). Together the sources accounted for half of the variation in mastery experiences in the advanced group, and about 40% in the beginning group. Physiological and affective states (PAS) exerted a medium-sized negative influence on mastery experiences in both groups. This source has been seriously neglected in previous research and the strong associations found in Study 3 provide a convincing rationale for paying more attention to preservice teachers' PAS in future research and teacher education practice. The results in Study 3 furthermore highlighted the importance of the role of the mentor teacher, which had previously been a focus of research in this area (e.g., Klassen & Durksen, 2014; Moulding, Stewart, & Dunmeyer, 2014; Mulholland & Wallace, 2001; Tschannen-Moran & Woolfolk Hoy, 2007; Woolfolk Hoy & Burke Spero, 2005).

The results from Study 3 provide some support for the hypothesis, in the discussion of Study 2, that vicarious experiences gathered during an observation practicum seem to have a similar effect on TSE development as the mastery experiences elicited during a teaching practicum. Even though the preservice teachers did not teach during the observation practicum, they still gathered mastery experiences which exerted a medium-sized effect on TSE change. The results from Study 3 show that these mastery experiences were fuelled by vicarious experiences, yet the results also show that the other sources, particularly PAS, likewise played a significant informing role.

Figure 1 presents a high-level overview of the results found in Study 2 and 3 with regard to the four research questions.

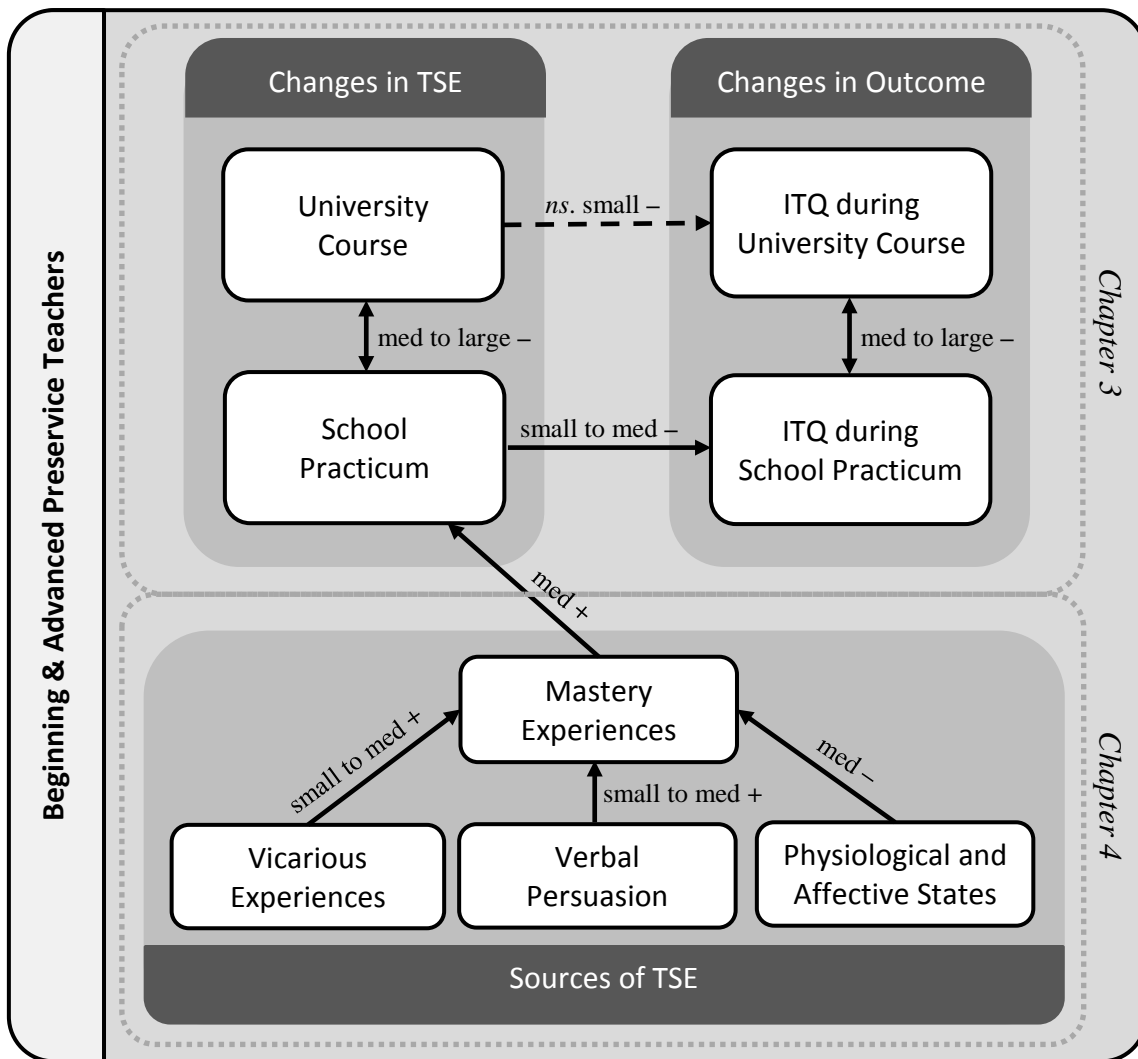


Figure 8. Size and direction of effects found between study variables
 Note: ITQ = Intention to Quit the teaching degree.

While numerous implications, limitations, strengths, and future directions of research have already been outlined in each of the three preceding empirical chapters, the following sections provide a structured integrated account of the main points, as well as some new insights.

5.2 Limitations and strengths

Limitations

Measures

The instrument for assessing TSE (the STSE) developed as part of this thesis is limited in its scope with regard to two factors: (a) it only focuses on the teaching domain, and (b) it has only been validated for use with preservice teachers. Arguably, teaching is not the only task that teachers fulfil (e.g., Skaalvik & Skaalvik, 2008). It can thus be

contended that a measurement of TSE should comprise a broader task focus, especially if it is supposed to be applicable to inservice teachers, which is a definite future aim for the STSE. Other researchers put forward TSE instruments that have taken a broader focus (e.g., Schwarzer, Schmitz, & Daytner, 1999; Skaalvik & Skaalvik, 2007). However, these in turn are not applicable to preservice teachers. While it would be advantageous if one instrument was available for both preservice and inservice teachers, this would have to come at the cost of limiting the scope of this measure to the common denominator of tasks that both preservice and inservice teachers would face.

The interfactor correlation between instructional strategies and student engagement just exceeded the threshold of .80 recommended by Brown (2006) for distinct factors. Since the correlation was close to the threshold (i.e., .82), the interfactor correlations were found to be invariant for all three groups, and the three-factor fit was good in the New Zealand sample, it was decided that the three-factorial structure was adequate for this sample. However, replicating the results with another English-speaking sample would provide better evidence for this claim.

Because of the data-driven process in selecting the items that best represented the three dimensions of TSE, there is the possibility of an “overfit” (e.g., Preacher, 2006) to the sample data in Study 1. This means it might be possible that the three-dimensional structure found with the three samples cannot be reproduced with other samples. Considering the diversity of the samples fitted in the first study, however, and the reproduced good fit with the beginning and advanced samples at T2 and T3, there is only a small chance of “overfit”.

While intentions are good predictors of actual behavior (e.g., Ajzen, 1985) and as such have been utilized previously as outcomes of TSE (e.g., Klassen & Chiu, 2011), an objective measure of retention of preservice teachers in Sample 2 would have allowed for a conclusion with regard to the effects of TSE on the actual quitting behavior of preservice teachers. However, actual quitting behavior might have only manifested itself if the study had had a longer time-frame.

The measure of the sources is limited in its scope to an assessment during a practicum experience. The practicum was chosen for the development of a new source measure for TSE and its validation, because Bandura (e.g., 1997) and TSE researchers (e.g., Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998) stress the importance of enactive mastery experiences for the development of (teacher) self-efficacy. Nonetheless, results from Study 2 showed that substantial development in TSE, which was coupled with a

large inter-individual variability, also occurred during the university coursework. In hindsight, developing a source measure for that aspect of teacher education would have also proven useful in order to account for these changes.

The assessment of the mastery experiences source was not identical for the two samples of preservice teachers. Instead, the items were tailored to some degree to accurately reflect the difference in the format of the practicum. This impacts adversely on the direct comparability of the effect of this source on the two groups.

The assessment of the verbal persuasion source with four items unexpectedly needed to be split into two factors, which then only featured two items each. At least three items would be needed to identify a CFA model if one of these sources was to be examined separately. There are several ways to deal with this issue. A higher-order factor could be introduced that captures the common aspects of both sources when the verbal persuasion source is modeled separately. However, the results showed quite different effects for each source of verbal persuasion, which would be masked in this way. Alternatively, each verbal persuasion source should not be modeled separately. Despite the small number of items (i.e., two), each verbal persuasion scale did show good internal consistency and the latent factors showed very good construct reliabilities, so there is little concern with regard to the reliability of both scales. Thus, this two-item limitation might be acceptable. However, the current measure could be extended to include more than two items in the future. Based on the open responses reported in Study 3, it would make sense to focus the “other” source of verbal persuasion on verbal persuasion by students.

The assessment of the vicarious experience source of TSE is solely focused on the model learning mechanism. Missing from the assessment of this source is the social comparison mechanism by which vicarious experiences also influence self-efficacy beliefs (see section 1.5 in Chapter 1). In fact, items for this facet of vicarious experiences were generated, but elicited a great amount of reactance in the pilot study, and were consequently dropped. Another attempt at creating such an assessment should add to the variance in TSE changes that can be accounted for by the four sources. Bouffard-Bouchard (1990), for example, demonstrated the strong effect of bogus peer norms on the level of perceived self-efficacy in students. Extrapolating from these results, it can be expected that the performance of peers would have a significant impact on TSE changes via the social comparison mechanism of vicarious experiences. This would be the case if preservice teachers completed their practicum experience at school

together with at least one other preservice teacher, which is often the case in the teacher preparation program examined in this thesis. Research on the big-fish-little-pond-effect (e.g., Seaton, Graven, & Marsh, 2008) has shown powerful reference group effects regarding self-concept. It would be interesting to examine these effects on TSE development.

Validity testing

A direct comparison of the TSES and the STSE using the same sample is lacking. This could be realized a few weeks apart and would present more convincing evidence than comparing the magnitudes of correlations derived from different samples.

Sample attrition

Attrition rates were high, and in some cases biased. This adversely impacts the generalizability of the results and conclusions that can be drawn on the basis of the data. It further reduces the sample size and thus impacts on significance levels. In addition, since the data had to be collected anonymously due to privacy guidelines of the university, data had to be matched across the different times of assessment using sometimes illegible hand-written personal codes (i.e., first letter of mother's first name etc.), rather than a unique identifier (e.g., student id). This procedure of code-matching contributed to an artificially inflated attrition, which could not be controlled.

Effect sizes vs. statistical significance

I encountered some difficulty in judging whether effects occurred or not, and should consequently be discussed and interpreted or not. While the APA³² Task Force on Statistical Inference (Wilkinson, 1999) stresses that effect sizes should be reported in any results section, their recommendations are less clear on whether sizes of effects which are non-significant should be discussed. A reviewer of Study 2 recommended a sole focus on discussing significant effects. This approach to the interpretation of results entails different conclusions to acknowledging small-sized effects. Was there no association between TSE and ITQ changes during the coursework at university, or was there a small association, which could be meaningful in this context? The answer to this question remains elusive until there are more results in this area which will provide further reference for interpretation.

³² American Psychological Association

A related issue regards judging how meaningful the average TSE changes observed in Study 2 are. This is difficult as there are no benchmarks, and the only benchmark used here (Ashford, Edmunds, & French, 2010) stems from a different area of self-efficacy research (i.e., health-related).

Design

Fundamentally, the longitudinal study features an observational non-randomized design. This correlational design does not allow for verifying the causal assumptions, which were specified in the structural equation models in Study 2 and 3 (e.g., Cohen & Cohen, 1983; Kline, 2011; Shadish et al., 2002).

The scope of Study 2 was restricted to examining the development of TSE in preservice teachers, and only over the period of one year for the beginning and half a year for the advanced group. Adopting a longer time-frame, ideally over the entire teacher preparation program (i.e., four to five years), would have allowed for a more comprehensive study of the development of TSE in preservice teachers. Realistically, such a study would be faced with multiple issues. For instance, sample mortality over such a long period of time would be a real challenge. If the attrition rate was high, it would make the interpretation of results difficult or in the worst case render them meaningless, and consequently these results might never be published. This is indeed a bleak outlook, which could dampen the motivation of any researcher to conduct longitudinal studies. As the example of the current thesis shows, attrition rates over such a short time span were already high, even though great care was taken to maximize retention in the longitudinal study (e.g., incentivized participation).

On a similar note, the two-cohort design, intended to shorten the timeframe of the study, does not allow for a direct comparison of the effects found at different stages of teacher education. The interpretation of the differences in changes between the beginning and advanced group observed in Study 2 would be more convincing if such differences were observed in the same cohort. Strictly speaking, this design also did not allow for TSE changes that occurred during the university term/year to be attributed to the input of university coursework, as no sources or other factors of that aspect of teacher preparation were assessed and linked to the respective TSE changes.

Missing moderators/mediators

Cognitive processing plays a central role in the way the sources affect self-efficacy formation (e.g., Bandura, 1997). The factors that affect the relationship between the sources and self-efficacy formation were detailed in section 1.5 of Chapter 1. Despite

this detailed analysis, the current work is lacking the inclusion of these factors entirely. Instead, the scope of this thesis was limited to assessing the four sources and determining the contribution of each. Nevertheless, without including any cognitive processing factors, the information gathered in this thesis can only be viewed as a first step towards an understanding of how the sources shape the development of TSE in preservice teachers.

Strengths

Despite the many shortcomings mentioned above, the overall strength of this work stems from its reasonably advanced methods of analysis as well as the longitudinal nature of data collection, which represents an improvement over previous research on this topic. In addition, the research questions addressed in this thesis are highly relevant with regard to teacher preparation, teacher retention, and teacher health (implications are further explored in the next section). Furthermore, examining TSE development during teacher education and trying to determine the role Bandura's (e.g., 1997) sources play in this development addresses a gap in TSE research that has become a sore spot. In addition, this first attempt at assessing the four sources for TSE was based on a deductive approach to maximize congruence with Bandura's conceptual framework. This sound theoretical base might account for the first promising results with regard to this measure presented in Chapter 4 of this thesis.

5.3 Implications

Despite several limitations to this thesis, there are a range of meaningful implications both for theory and practice that the results of this thesis entail.

Theory

The greater associations found between different validity criteria throughout this thesis (i.e., general self-efficacy, study satisfaction, ITQ) and TSE, as measured by the STSE in comparison to the original TSES, provide more evidence for Bandura's (e.g., 1997) claim of a greater predictive validity of measures in close alignment with self-efficacy theory.

There was some indication of a "miscalibration" of TSE in beginning preservice teachers, which has previously been hypothesized to occur (e.g., Klassen & Chiu, 2011). This phenomenon was not explored any further in this thesis, but warrants some further examination and theory building. For example, it could be tested if this miscalibration was indeed dependent on (a lack of) knowledge.

The results from Study 2 have demonstrated the value of separating the three dimensions of TSE. The dimensions changed differently over the course of the study, depending on the stage of teacher education (i.e., sample) and the different aspects of the teacher preparation program. Moreover, changes in the different dimensions were not equally related to the outcome measure (i.e., ITQ). So far, a separation of the three dimensions has not been possible for beginning preservice teachers with the TSES (Duffin et al., 2012). Thus, the developed STSE scale provides a sound instrument for exploring the three dimensions of TSE further with preservice teachers at any stage of teacher education.

As mentioned in the limitations, judging how meaningful changes that occurred in Study 2 were is difficult due to a lack of prior benchmarks. However, for the purpose of theory development determining average changes might be of little value. More importantly, the average changes showed large inter-individual variability and the results from Study 3 provide a first estimate of the variance that can be explained by the sources. Notwithstanding the fact that the source measures could be improved (especially with regard to including the social comparison aspect), there is a large amount of variance that could not be accounted for by the sources. This is where research can add value to further theory development by identifying other factors which also play a role in TSE development, possibly in interaction with the career stage of (future) teachers. Some researchers have already taken up this endeavor (Knoblauch & Woolfolk Hoy, 2008; Moulding et al., 2014; Tschannen-Moran & Woolfolk Hoy, 2007).

Additionally, the relationships between the sources explored in Study 3 provide some new impetus to theory development with regard to how the sources interact. Specifically, exploring the possible mediating role of mastery experiences reliant on career stage seems promising.

Lastly, I would like to highlight the contribution that a formal teacher education program made to the development of TSE. While it could not be reliably proven that the university course effected the change in TSE during the term/year, it is highly likely. It could, however, be demonstrated that some part of the change during the practicum was related to different factors of the practicum experience (i.e., mentor teacher). There is some debate about the effectiveness of teacher education (e.g., Cochran-Smith & Fries, 2005; Keller, 2002; Klieme et al., 2003), which has at times being accused of being entirely ineffective (Darling-Hammond, 2000). Considering Hattie's (2009) much cited meta-analysis only attests a "negligible" (p. 111) effect of $d = .11$ to teacher preparation

programs, this debate is still topical. As is the *personality approach* (Hascher, 2014) to teacher education, whereby key teacher attributes are assumed to be inherent to a person. Naturally, this assumption severely limits the influence teacher education is thought to have on preservice teachers (Hascher, 2014). In contrast to this assumption, this thesis could show that TSE beliefs can be influenced by a teacher preparation program.

Practice

The STSE instrument developed as part of this thesis is the first TSE measure that allows the separate assessment of the three dimensions of TSE introduced by Tschannen-Moran and Woolfolk Hoy (2001) for preservice teachers at different stages of teacher education. Since these dimensions are essential features of teaching and thus teacher preparation (e.g., Baumert & Kunter, 2006; Ophardt & Thiel, 2013), it is ideally suited for evaluating teacher education programs in general or the implementation of policy changes with regard to teacher education in particular. For example, the working group led by Mareike Kunter is planning to use this instrument in the evaluation of the “*Praxissemester*” (i.e., practical term at school during university teacher preparation), which is going to be implemented this year in the federal state of Hesse. In addition, there have been several other requests by German researchers to use the STSE. So this instrument seems to indeed fill a gap. Researchers interested in the development of TSE would face similar difficulties with the available measures that I encountered in planning this study, and would have similar requirements as outlined in section 1.3 of the first chapter. To enable easy access, the STSE will shortly be made available to the scientific community via the electronic test archive of the German Leibniz Centre for Psychological Information and Documentation (i.e., *Leibniz-Zentrum für Psychologische Information und Dokumentation*) and its inclusion in the PSYINDEX Tests database.

Considering that early failures are especially detrimental to self-efficacy development (Bandura, 1997), it seems of utmost importance to design the practical phases of teacher preparation in a way that fosters TSE development in preservice teachers. This view is also supported by the results from Study 2, which showed a link between TSE changes particularly during the practicum and changes in preservice teachers’ ITQ their degree. This is especially relevant with regard to the attrition, turnover, and teacher shortage rates outlined in the introduction. Klassen and Durksen (2014) provide some qualitative insights into the detrimental effects that mentor

teachers can exert on the TSE of preservice teachers during a practicum. The results from Study 3 in this thesis again highlight the important role that mentor teachers play, and thus provide an impetus to develop mentor training programs, which take into account insights provided by self-efficacy theory (e.g., Bandura, 1997). However, the prevailing view that greater TSE has only positive and beneficial outcomes has been somewhat challenged by Wheatley (2002; 2005). The potential detrimental effects of high TSE are not well examined (or at least published), especially with regard to developmental trajectories in TSE. While Bandura (1997) provides some guidance with regard to trajectories as summarized in section 1.5 of the introduction, it is entirely unclear how, for example, an easy first increase in TSE affects outcomes of preservice teachers.

Since PAS were also found to be influential with regard to TSE development, one more recommendation to teacher educators is to take this into account in teacher preparation. So far, this aspect has been entirely neglected in teacher education (Moulding et al., 2014), but interventions targeted at teaching preservice teachers to manage their physiological and affective responses could prove useful with regard to TSE development.

5.4 Future directions

Based on the current results and the state of the TSE research as summarized in the introductory chapter, I can see two major directions where future research in this area can add value.

- (1) A set of experimental studies ultimately aimed at developing interventions targeting:
 - (a) mentor teachers' feedback-giving process
 - (b) preservice teachers' interpretation of their physiological and affective states.

With respect to (a), this could, for example, be done by experimentally manipulating the feedback that mentor teachers give to preservice teachers during their practicum. Study 3 provided some specific guidelines on the framing of performance feedback, which are based on self-efficacy theory (e.g., Bandura, 1997). These are but a few examples that can be used to guide the design of such experimental studies. The ultimate aim of these studies would be to develop a comprehensive training program for mentors of preservice teachers at school, which focusses on establishing “resilient” (Bandura, 1997, p. 80) self-efficacy beliefs.

With regard to (b), since cognitive processing plays a vital role in the interpretation of PAS (e.g., Bandura, 1997), experimental studies could vary, for instance, by different self-regulation strategies and different stress levels in teaching situations to gauge the effectiveness of various strategies. Eventually, results from these experiments could be used to train preservice teachers to manage their PAS as part of formal teacher preparation. In light of the ample evidence showing that teachers suffer from high stress levels as part of their work (Schwarzer & Hallum, 2008), and the mediating role TSE can play in the development of burnout syndrome (e.g., Brisse, von Hoover-Dempsey, & Bassler, 1988; Brouwers & Tomic, 2000; Evers, Brouwers, & Tomic, 2002), such training might also be of benefit to inservice teachers.

(2) A research program examining the outcomes of TSE for preservice teachers, and the role that outcome expectancies play with regard to the relationship between TSE and outcomes. As outlined in Chapter 1, there is little research on the outcomes of TSE for preservice teachers. This thesis has shown some benefits of TSE development with regard to the ITQ of preservice teachers. However, due to the general lack of research in this area, it is unclear in what way TSE is beneficial or possibly even detrimental (see Wheatley, 2005 for a critique) to other outcomes at the teacher preparation stage. Furthermore, the role teachers' outcome expectancies play in the prediction of behavior warrants greater examination. Despite Bandura's (e.g., 1986) claim that outcome expectancies do not add much incremental value in predicting behavior in addition to self-efficacy, this claim needs more research to support or refute it. Due to the long tradition of the confused conceptual nature of TSE, previous research has assessed outcome expectancies to some degree as well as self-efficacy beliefs. Consequently, much research in the past has shown outcome expectancies to have some predictive value. However, because of the confusion between the constructs, neither one of them has been assessed in line with Bandura's (e.g., 1997) theoretical specifications. Assessing both constructs accordingly would enable researchers to study their effects on behavior both separately and in concert. The relationship between the two constructs is unclear. Early TSE research (using the Gibson and Dembo scale) indicates that at least what has been termed *personal efficacy* and *general efficacy* are reasonably independent (e.g., Woolfolk & Hoy, 1990). Bandura's (e.g., 1986) argument that people do not act because they are convinced of positive outcomes but because they believe they can successfully produce the necessary behavior required for such outcomes, seems to argue for independence or a low correlation. In contrast, Bandura (e.g., 1986) also argues that

often outcome expectations will be influenced by self-efficacy beliefs. Schunk and Zimmermann (2006), for example, clearly state that both self-efficacy and outcome expectancies are necessary for optimal motivation. The question is, how interlinked are the two constructs? I would expect that while TSE is a necessary prerequisite for engaging in an action, it is not sufficient. People would only engage if they believe the action to produce a desired outcome. This is turning Bandura's statement on its head. There are many examples from research on conceptual change, whereby teachers are highly resistant to change (Gregoire, 2003). This might be an indicator for the influence of outcome expectancies, because if (preservice) teachers do not believe that, for instance, a new teaching method will lead to success, they will not attempt it. The very same argument has been made (e.g., by Bandura, 1986) for the role of self-efficacy beliefs. But which belief needs to be changed first? The belief that the new teaching method will lead to better results for students or the belief that one can successfully apply this new method? An answer to this question would be highly relevant to teacher education.

5.5 Conclusion

This thesis contributes to the advancement of TSE research in three important ways. First, it provides an instrument that allows for a three-dimensional measurement of TSE in beginning *and* advanced preservice teachers. Second, it documents a link between changes in TSE during a practicum and changes in preservice teachers' intention to quit their teaching degree. Third, it provides a first instrument for the assessment of Bandura's four sources for TSE and offers some insight into their interaction. Despite many limitations to this thesis, these results have meaningful practical implications for the design and evaluation of teacher education programs.

5.6 References

- Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control. From cognition to behavior* (pp. 11–39). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Ashford, S., Edmunds, J., & French, D. P. (2010). What is the best way to change self-efficacy to promote lifestyle and recreational physical activity? A systematic review with meta-analysis. *British Journal of Health Psychology, 15*(2), 265–288.
doi:10.1348/135910709X461752
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Bandura, A. (2006). Guide for creating self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (pp. 307–337). Greenwich: Information Age Publishing.
- Baumert, J., & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft, 9*(4), 469–520.
- Böttcher, F. (2013). *Beitrag zur Validierung eines Instrumentes zur Erfassung der Lehrerinnen- und Lehrerselbstwirksamkeit* [Contribution to the validation of an instrument for the assessment of teacher self-efficacy] (Unpublished master's thesis). Freie Universität Berlin, Berlin.
- Bouffard-Bouchard, T. (1990). Influence of self-efficacy on performance in a cognitive task. *The Journal of Social Psychology, 130*(3), 353–363.
doi:10.1080/00224545.1990.9924591
- Brisse, J. S., von Hoover-Dempsey, K., & Bassler, O. C. (1988). Individual, situational contributors to teacher burnout. *Journal of Educational Research, 82*(2), 106–112.
- Brouwers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education, 16*, 239–253.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guilford Press.
- Chesnut, S. R., & Burley, H. (2015). Self-efficacy as a predictor of commitment to the teaching profession: A meta-analysis. *Educational Research Review, 15*, 1–16.
doi:10.1016/j.edurev.2015.02.001

- Cochran-Smith, M., & Fries, K. (2005). Researching teacher education in changing times: Politics and paradigms. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education. The report of the AERA panel on research and teacher education* (pp. 69–109). Mahwah, NJ: Lawrence Erlbaum Associates.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Darling-Hammond, L. (2000). How teacher education matters. *Journal of Teacher Education*, *51*(3), 166–173.
- Duffin, L. C., French, B. F., & Patrick, H. (2012). The teachers' sense of efficacy scale: Confirming the factor structure with beginning pre-service teachers. *Teaching and Teacher Education*, *28*(6), 827–834.
- Dunning, D., Heath, C., & Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, *5*(3), 69–106.
- Evers, W. J. G., Brouwers, A., & Tomic, W. (2002). Burnout and self-efficacy: A study on teachers' beliefs when implementing an innovative educational system in the Netherlands. *British Journal of Educational Psychology*, *72*(2), 227–243.
doi:10.1348/000709902158865
- Gregoire, M. (2003). Is it a challenge or a threat? A dual-process model of teachers' cognition and appraisal processes during conceptual change. *Educational Psychology Review*, *15*(2), 147–179.
- Hascher, T. (2014). Forschung zur Wirksamkeit der Lehrerbildung. In E. Terhart, H. Brennewitz, & M. Rothland (Eds.), *Handbuch der Forschung zum Lehrerberuf* (2nd ed., pp. 542–571). München, New York: Waxmann.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, New York: Routledge.
- Keller, H.-J. (2002). Standards in der Lehrerinnen- und Lehrerbildung: Blick in die USA und Ausblick auf die deutschsprachige Lehrpersonenausbildung. *Journal für Lehrer- und Lehrerinnenbildung*, *2*(1), 20–28.
- Klassen, R. M., & Chiu, M. M. (2011). The occupational commitment and intention to quit of practicing and pre-service teachers: Influence of self-efficacy, job stress, and teaching context. *Contemporary Educational Psychology*, *36*(2), 114–129.
doi:10.1016/j.cedpsych.2011.01.002

- Klassen, R. M., & Durksen, T. L. (2014). Weekly self-efficacy and work stress during the teaching practicum: A mixed methods study. *Learning and Instruction, 33*, 158–169. doi:10.1016/j.learninstruc.2014.05.003
- Klassen, R. M., Tze, V. M. C., Betts, S. M., & Gordon, K. A. (2011). Teacher efficacy research 1998–2009: Signs of progress or unfulfilled promise? *Educational Psychology Review, 23*, 21–43.
- Kleinsasser, R. C. (2014). Teacher efficacy in teaching and teacher education. *Teaching and Teacher Education, 44*, 168–179. doi:10.1016/j.tate.2014.07.007
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, P., & Prenzel, M. (2003). *Zur Entwicklung nationaler Bildungsstandards: Eine Expertise*. Berlin: Bundesministerium für Bildung und Forschung (BMBF).
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). *Methodology in the social sciences*. New York, NY: Guilford Press.
- Knoblauch, D., & Woolfolk Hoy, A. (2008). “Maybe I can teach those kids.” The influence of contextual factors on student teachers’ efficacy beliefs. *Teaching and Teacher Education, 24*, 166–179.
- Mergler, A. G., & Tangen, D. (2010). Using microteaching to enhance teacher efficacy in pre-service teachers. *Teaching Education, 21*(2), 199–210. Retrieved from <http://dx.doi.org/10.1080/10476210902998466>
- Millsap, R. E., & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology, 73*(3), 574–584.
- Moulding, L. R., Stewart, P. W., & Dunmeyer, M. L. (2014). Pre-service teachers' sense of efficacy: Relationship to academic ability, student teaching placement characteristics, and mentor support. *Teaching and Teacher Education, 41*, 60–66. doi:10.1016/j.tate.2014.03.007
- Mulholland, J., & Wallace, J. (2001). Teacher induction and elementary science teaching: Enhancing self-efficacy. *Teaching and Teacher Education, 17*, 243–261.
- O'Neill, S. C., & Stephenson, J. (2011). The measurement of classroom management self-efficacy: A review of measurement instrument development and influences. *Educational Psychology Review, 31*(3), 261–299. doi:10.1080/01443410.2010.545344
- Ophardt, D., & Thiel, F. (2013). *Klassenmanagement: Ein Handbuch für Studium und Praxis. Schulpädagogik*. Stuttgart: Kohlhammer.

- Poulou, M. (2007). Personal teaching efficacy and its sources: Student teachers' perceptions. *Educational Psychology, 27*(2), 191–218. doi:10.1080/01443410601066693
- Preacher, K. J. (2006). Quantifying parsimony in structural equation modeling. *Multivariate Behavioral Research, 41*(3), 227–259. doi:10.1207/s15327906mbr4103_1
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research - Online, 8*(2), 23–74.
- Schunk, D. H., & Zimmermann, B. J. (2006). Competence and control beliefs: Distinguishing the means and ends. In P. A. Alexander (Ed.), *Handbook of educational psychology* (2nd ed., pp. 349–367). New York, NY: Routledge.
- Schwarzer, R., & Hallum, S. (2008). Perceived teacher self-efficacy as a predictor of job stress and burnout: Mediation analyses. *Applied Psychology: An International Review, 57*, 152–171.
- Schwarzer, R., Schmitz, G. S., & Daytner, G. T. (1999). *Teacher Self-Efficacy*. Retrieved from http://userpage.fu-berlin.de/~health/teacher_se.htm
- Seaton, M., Graven, R. G., & Marsh, H. W. (2008). East meets west: An examination of the big-fish-little-pond effect in western and non-western countries. In D. M. McInerney, H. W. Marsh, & R. Craven (Eds.), *Self-processes, learning, and enabling human potential. Dynamic new approaches* (pp. 353–371). Charlotte, NC: Information Age Pub.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford, New York: Oxford University Press.
- Skaalvik, E. M., & Skaalvik, S. (2007). Dimensions of teacher self-efficacy and relations with strain factors, perceived collective teacher efficacy, and teacher burnout. *Journal of Educational Psychology, 99*(3), 611–625. doi:10.1037/0022-0663.99.3.611
- Skaalvik, E. M., & Skaalvik, S. (2008). Teacher self-efficacy: Conceptual analysis and relations with teacher burnout and perceived school context. In D. M. McInerney, H. W. Marsh, & R. Craven (Eds.), *Self-processes, learning, and enabling human potential. Dynamic new approaches* (pp. 223–247). Charlotte, NC: Information Age Pub.
- Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805.

- Tschannen-Moran, M., & Woolfolk Hoy, A. (2007). The differential antecedents of self-efficacy beliefs of novice and experienced teachers. *Teaching and Teacher Education*, 23(6), 944–956. doi:10.1016/j.tate.2006.05.003
- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68(2), 202–248. Retrieved from <http://www.jstor.org/stable/1170754?origin=JSTOR-pdf&>
- Usher, E. L., & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology*, 34(1), 89–101. doi:10.1016/j.cedpsych.2008.09.002
- Weinstein, C. S. (1988). Preservice teachers' expectations about the first year of teaching. *Teaching and Teacher Education*, 4(1), 31–40. doi:10.1016/0742-051X(88)90022-4
- Wheatley, K. F. (2002). The potential benefits of teacher efficacy doubts for educational reform. *Teaching and Teacher Education*, 18(5-22).
- Wheatley, K. F. (2005). The case for reconceptualizing teacher efficacy research. *Teaching and Teacher Education*, 21(7), 747–766. doi:10.1016/j.tate.2005.05.009
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594–604. doi:10.1037/0003-066X.54.8.594
- Woolfolk, A. E., & Hoy, W. K. (1990). Prospective teachers' sense of efficacy and beliefs about control. *Journal of Educational Psychology*, 82(1), 81–91.
- Woolfolk Hoy, A., & Burke Spero, R. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, 21(4), 343–356. doi:10.1016/j.tate.2005.01.007
- Woolfolk Hoy, A., Davis, H., & Pape, S. J. (2006). Teacher knowledge and beliefs. In P. A. Alexander (Ed.), *Handbook of educational psychology* (2nd ed., pp. 715–737). New York, NY: Routledge.

Danksagung

Viele Personen haben zu dieser Arbeit beigetragen und an dieser Stelle möchte ich mich dafür persönlich herzlich bedanken.

Mein besonderer Dank gilt meinen Gutachtern Felicitas Thiel und Ralf Schwarzer. Felicitas Thiel hat mich herzlich in ihrem Arbeitsbereich aufgenommen und mir von Anfang an vollkommene wissenschaftliche Freiheit, was die Wahl meines Dissertationsthemas betraf, gelassen. Auch wenn ich mir dann ein Thema suchte, was (noch) außerhalb der Schwerpunkte des ABs lag, hat sie mich ohne zu Zögern unterstützt und mir stets konstruktive Rückmeldungen gegeben. Außerdem hat sie diese konstruktive Art an ihrem AB konsequent kultiviert und damit konnte ich in allen Phasen dieser Arbeit von den exzellenten Rückmeldungen und Anregungen meiner Kolleginnen profitieren.

Ralf Schwarzer danke ich für sein Interesse an dieser Arbeit, die er von Anfang an begleitet hat. Damit hat er vor nunmehr vier Jahren schon in der Planungsphase dieser Arbeit maßgeblich zu deren Gelingen beigetragen. Diese Unterstützung setzte er während der gesamten Entstehungszeit dieser Arbeit fort, indem er sich die Zeit nahm, mir zu den Manuskripten eine Rückmeldung zu geben.

Mein ausdrücklicher Dank gilt meinen Kolleginnen, die mich in wechselnder Besetzung in den letzten vier Jahren am AB für Schulpädagogik und Schulentwicklungsforschung begleitet haben. Allen voran Valentina Piwowar, die mich sowohl fachlich als emotional während der gesamten Zeit unterstützt hat und deren Einsichten, Erfahrungen, und Anregungen ich sehr schätze. Diemut Ophardt, dafür, dass sie immer ein offenes Ohr für mich hatte und diese Arbeit auch finanziell unterstützt hat. Susanne Czaja, dafür, dass sie als meine Büropartnerin meine täglichen Erfolge und Misserfolge geteilt hat und mich mit der motivierenden Wirkung von sozialen Vergleichsnormen vertraut gemacht hat. Victoria Barth danke ich für ihre stets positive Einstellung und gute Laune, mit der sich mich an vielen Tagen zum Lachen gebracht hat. Katharina Krysmanski für ihre Expertise in statistischen Fragen und ihr Praxiswissen. Max Schlotthauer möchte ich danken für seine aufwändige Arbeit im Erstellen der Literaturdatei. Martin Schultze möchte ich an dieser Stelle noch einmal ganz herzlich für seine wertvolle statistische Beratungsarbeit danken, ohne die die Arbeit tatsächlich so nicht hätte entstehen können.

Bei meiner Familie möchte ich mich für ihre bedingungslose Unterstützung und Motivation bedanken. I am especially grateful to my husband who had to deal with the brunt of negative

impacts associated with the writing of a PhD thesis. Thank you so much for supporting me always, but particularly for your invaluable help at the end of this journey.

Erklärung

Hiermit versichere ich, die Dissertationsarbeit „Evaluation of a teacher preparation program using the development of teacher self-efficacy as an outcome - a longitudinal study“ selbstständig angefertigt zu haben. Sämtliche Hilfsmittel, die ich verwendet habe, sind angegeben. Sie ist in keinem früheren Promotionsverfahren angenommen oder abgelehnt worden.

Franziska Pfitzner-Eden

Berlin, im Mai 2015

Curriculum Vitae

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.