

Die spezifische Lehrkraft-Selbstwirksamkeitserwartung bei Lehramtsstudierenden – Eine Integration zweier Messinstrumente

Lena Hannemann, Gesa Uhde, Barbara Thies

Zusammenfassung: Die Lehrkraft-Selbstwirksamkeitserwartung (SWE) im Studium kann die spätere Berufstätigkeit von Lehrkräften langfristig positiv beeinflussen. Im deutschen Sprachraum fehlen Instrumente, die die Lehrkraft-SWE mit spezifischen Subskalen zu einem frühen Zeitpunkt im Studium messen. Vorgestellt wird eine adaptierte Skala basierend auf der Fragestruktur der Skala Lehrerselbstwirksamkeit (WirkLehr; Schmitz & Schwarzer, 2000) sowie den Inhalten der Teachers' Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001) mit den Subskalen Engagement, Instruktionen und Classroom-Management. Die psychometrischen Kennwerte der adaptierten Skala wurden mittels konfirmatorischer Faktorenanalyse anhand von N = 618 Studierenden (drittes Semester) überprüft. Es zeigte sich ein besserer Modellfit für ein Drei-Faktoren- gegenüber einem Ein-Faktor-Modell. Bezüglich der drei Subskalen ergaben sich positive Korrelationen mit WirkLehr, negative Korrelationen mit dem Stresserleben (Perceived Stress Scale; Büssing, 2011) sowie positive Korrelationen mit einer zweiten Messung einen Monat später. Zusätzlich wurde die adaptierte Skala im Rahmen einer Classroom-Management-Trainingsevaluation eingesetzt. Hier zeigte sich eine höhere Lehrkraft-SWE einer Trainings- gegenüber einer Kontrollgruppe für die Bereiche Classroom-Management und Instruktionen. Aufgrund der zufriedenstellenden Validitäts- und Reliabilitätsbelege kann die adaptierte Skala zur Messung der spezifischen Lehrkraft-SWE bei Studierenden eingesetzt werden.

Schlüsselwörter: Classroom-Management, Lehramtsstudium, Lehrkraft-Selbstwirksamkeitserwartung, Skalenkonstruktion

Specific teacher self-efficacy in teacher students – an integration of two measurement instruments

Abstract: Students' self-efficacy can have a long-term beneficial influence on their later teaching profession. However, there is a lack of German instruments that consider specific aspects of teacher self-efficacy at an early stage of teacher training. An adapted teacher self-efficacy scale is presented based on the scale Lehrerselbstwirksamkeit (WirkLehr; Schmitz & Schwarzer, 2000) and on the Teachers' Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001) with the subscales engagement, instruction, and classroom management. Confirmatory factor analysis (N = 618 students, third semester) indicated a better model fit for a three-factor-model compared to a single-factor-model. Investigations revealed positive correlations of the three subscales with teacher self-

efficacy (WirkLehr), negative correlations with stress (Perceived Stress Scale, PSS; Büssing, 2011), and positive correlations of the subscales with their second measurements one month later. The adapted scale was also implemented in a classroom management training evaluation; the training group showed higher self-efficacy in classroom management and instructions than the control group. Given the evidence of satisfactory validity and reliability, the adapted scale can be used to assess specific teacher self-efficacy in teacher students.

Keywords: classroom management, scale construction, teacher self-efficacy, teacher training

1 Einleitung

Die Bedeutsamkeit der Lehrkraft-Selbstwirksamkeitserwartung (SWE), also der Erwartung, Herausforderungen des Lehrkraftberufs bewältigen zu können, wurde vielfach untersucht (Klassen, Tze, Betts & Gordon, 2011; Zee & Koomen, 2016). Die Lehrkraft-SWE steht in positivem Zusammenhang mit Berufszufriedenheit, Verbleibewahrscheinlichkeit im Beruf und Unterrichtsenthusiasmus (Coladarci, 1992; Tschannen-Moran & Woolfolk Hoy, 2001) sowie in negativem Zusammenhang mit Stress und Belastung der Lehrkräfte (Klassen & Chiu, 2010). Auf Seiten der Schülerinnen und Schüler geht die Lehrkraft-SWE mit positiven Verhaltensänderungen wie etwa einem höheren Anteil von On-Task-Verhalten (Robertson & Dunsmuir, 2013), höherer Motivation und besseren Schulleistungen einher (Caprara, Barbaranelli, Steca & Malone, 2006; Goddard, Hoy & Woolfolk Hoy, 2000). Bereits in der Lehramtsausbildung ist die Lehrkraft-SWE bedeutsam, da sie sich zu einem frühen Zeitpunkt entwickelt und langfristig die spätere Berufstätigkeit beeinflussen kann (Bandura, 1997; Woolfolk Hoy & Burke Spero, 2005). Die Entwicklung der Lehrkraft-SWE im Lehramtsstudium lässt sich dabei als ein zyklischer Prozess beschreiben und beeinflusst Ziele, Aufwand und Ausdauer der Studierenden in schwierigen Situationen (Tschannen-Moran & Woolfolk Hoy, 1998). Entsprechend ist eine günstige und ungünstige Entwicklung der Lehrkraft-SWE im Lehramtsstudium möglich (Seifert & Schaper, 2018). Dabei scheint die zu Studienbeginn oft hohe und globale Lehrkraft-SWE insbesondere nach den ersten schulpraktischen Phasen abzusinken (Schüle, Besa, Schriek & Arnold, 2016). Die Lehrkraft-SWE steht jedoch in positivem Zusammenhang mit Classroom-Management-Kompetenzen, die zentral für erfolgreiches Unterrichten sind (Seidel & Shavelson, 2007) und als eine der wichtigsten Ressourcen im Umgang mit Belastungen im Lehramtsberuf gelten (Klusmann, Kunter, Voss & Baumert, 2012). Eine Skala zur spezifischen Messung der Lehrkraft-SWE zu einem frühen Zeitpunkt im Studium ist für Untersuchungen im Rahmen der Lehramtsausbildung essenziell, beispielsweise zur Evaluation von Interventionen zur Förderung der Classroom-Management-Kompetenzen der Studierenden.

Die vorliegende Studie prüft die Validität einer adaptierten Skala zur Lehrkraft-SWE, die die Fragestruktur der etablierten, deutschsprachigen Skala Lehrerselbstwirksamkeit (WirkLehr; Schmitz & Schwarzer, 2000) mit den spezifischen Inhalten zu Instruktion, Engagement und Classroom-Management der amerikanischen Teachers' Sense of Efficacy Scale (TSES; Tschannen-Moran & Woolfolk Hoy, 2001) verbindet. Hierzu wird zunächst die Übersetzungs- und Adaptionsprozedur zur Item-Generierung beschrieben. Anschließend werden die Ergebnisse einer konfirmatorischen Faktorenanalyse zur Überprüfung der

Faktorenstruktur der adaptierten Skala präsentiert sowie Validitäts- und Reliabilitätsbelege dargelegt. Darüber hinaus werden Ergebnisse zum Einsatz der adaptierten Skala im Rahmen einer Trainingsevaluation berichtet.

2 Forschungsstand

2.1 Das Konstrukt der Lehrkraft-Selbstwirksamkeitserwartung

Zwei Forschungsansätze liegen dem Konstrukt der Lehrkraft-SWE zugrunde, und zwar die Locus-of-Control-Theorie (Rotter, 1966) sowie die sozial-kognitive Lerntheorie (Bandura, 1997). In der Locus-of-Control-Theorie wurde die Lehrkraft-SWE verstanden als das Ausmaß wahrgenommener Kontrolle der Lehrkräfte, ob Handlungsergebnisse im Sinne der Kontrollattribution durch eigenes Handeln oder durch Faktoren außerhalb des eigenen Handelns bestimmt werden. In der sozial-kognitiven Lerntheorie Banduras wird die Selbstwirksamkeitserwartung hingegen als die Erwartung formuliert, zukünftige Herausforderungen aufgrund eigener Kompetenz selbst bewältigen zu können. Auf Grundlage von Rotters Theorie wurden zahlreiche Instrumente entwickelt wie z.B. die „Teacher Efficacy Scale“ von Gibson und Dembo (1984). Messungen der Selbstwirksamkeitserwartung basierend auf dieser Theorie führten immer wieder zu Problemen, da Kontrollattribution und Selbstwirksamkeitserwartung eben nicht dasselbe messen (z.B. Klassen et al., 2011; Schmitz & Schwarzer, 2000), die Selbstwirksamkeitserwartung Verhalten jedoch besser vorhersagt als die Kontrollüberzeugung (Bandura, 1997). Folglich wurde der Bedarf an validen Instrumenten in Übereinstimmung mit Banduras sozial-kognitiver Lerntheorie (1997) konstatiert. Nach Bandura (1997) kann die Selbstwirksamkeitserwartung je nach Handlungsdomäne stark innerhalb einer Person variieren, weswegen die Domänenspezifität, d.h. die Spezifikation für den jeweiligen Tätigkeitsbereich, essentiell für die Messung der Selbstwirksamkeitserwartung ist (Bandura, 2006; Civitillo, Juang, Schachner & Börner, 2016). Die Lehrkraft-SWE als eine solche Spezifikation der Selbstwirksamkeitserwartung für das Unterrichten wird bei Tschannen-Moran und Woolfolk Hoy in Abgrenzung zu Schmitz und Schwarzer definiert als „a judgement of his or her capabilities to bring about desired outcomes of student engagement and learning, even among those students who may be difficult or unmotivated“ (2001, S. 783).

2.2 Instrumente zur Erfassung der Lehrkraft-Selbstwirksamkeitserwartung

Ein im deutschen Sprachraum häufig verwendetes Instrument zur Lehrkraft-SWE ist die Skala WirkLehr von Schmitz und Schwarzer (2000) mit zehn Items, wie z.B. „Ich weiß, dass ich es schaffe, selbst den problematischsten Schülern den prüfungsrelevanten Stoff zu vermitteln.“. Befragte geben auf einer vierstufigen Likert-Skala von 1 (stimmt nicht) bis 4 (stimmt genau) ihre jeweilige Zustimmung an. Die Item-Konstruktion ist eng an den Empfehlungen Banduras (2006) zur Konstruktion von SWE-Skalen orientiert und wird seinen Anforderungen bezüglich der Aufnahme eines Schwierigkeitsgrads gerecht, indem durchgängig Handlungsbarrieren eingebaut werden durch Formulierungen wie „selbst wenn ...“.

Im Zuge der Item-Generierung wurden für die vier Bereiche „berufliche Leistung“, „berufliche Weiterentwicklung“, „soziale Interaktion“ sowie „Umgang mit Berufsstress“ jeweils fünf bis neun Items entwickelt und mehrfach revidiert. Die Skala wurde schließlich aus einem Pool mit 27 empirisch erprobten Items erstellt. Ziel der Skalenkonstruktion war es, den Tätigkeitsbereich des Lehrkraftberufs möglichst sparsam und gleichzeitig heterogen abzubilden (Schmitz & Schwarzer, 2000). Aufgrund der schnellen Durchführbarkeit wurde die WirkLehr-Skala in der Forschung mit angehenden Lehrkräften vielfach eingesetzt. Durch die Kürze der Skala wird jedoch Spezifität hinsichtlich der Unterrichtsanforderungen eingeübt, laut Bandura (2006) ist aber die Domänenspezifität wesentlich zur validen Erfassung der Selbstwirksamkeitserwartung.

Das international am häufigsten verwendete Instrument zur Erfassung der Lehrkraft-SWE ist die TSES von Tschannen-Moran und Woolfolk Hoy (2001) (Klassen et al., 2011; Zee & Koomen, 2016). Die TSES fokussiert verschiedene Tätigkeiten einer Lehrkraft im Unterricht über die drei Subskalen Engagement, Instruktionen und Classroom-Management. Die Skala Engagement bezieht sich auf die Förderung des Interesses der Schülerinnen und Schüler, z.B. „How much can you do to get students to believe they can do well in school work?“. Die Skala Instruktionen umfasst Items zum Einsatz von Unterrichtsstrategien und dem Formulieren von Instruktionen, z.B. „To what extent can you provide an alternative explanation or example when students are confused?“. Die Skala Classroom-Management bezieht Strategien zur Prävention von und Intervention bei Unterrichtstörungen ein, z.B. „How much can you do to control disruptive behavior in the classroom?“. Neben einer Langform mit 24 Items kann eine Kurzform mit zwölf Items verwendet werden. Die Items wurden in einem umfangreichen Verfahren produziert, bei dem eine Gruppe aus zehn Expertinnen und Experten aus der Forschung beziehungsweise dem Lehrberuf zunächst 100 Items generierte, aus denen in wiederholten Revisionsrunden erst 52 Items und schließlich 24 Items entstanden (Tschannen-Moran & Woolfolk Hoy, 2001). Die gefundene Drei-Faktoren-Struktur für den Einsatz bei Lehrkräften wurde in zahlreichen Studien belegt (z.B. Fives & Buehl, 2010). Ergebnisse verschiedener Studien mit Lehramtsstudierenden zur Faktorenstruktur waren hingegen nicht eindeutig. Einige Autorinnen und Autoren wiesen eine Ein-Faktor-Struktur als bestes Modell für die Repräsentation ihrer Daten nach (z.B. Fives & Buehl, 2010; Tschannen-Moran & Woolfolk Hoy, 2001), andere zeigten eine bessere Passung bei einem Drei-Faktoren-Modell (Pfitzner-Eden, Thiel & Horsley, 2014; Poulou, 2007). Annahmen über einen möglichen Zusammenhang der gefundenen Faktorenstruktur mit den Studienerfahrungen der Teilnehmenden sind aufgrund der ungenauen (siehe Studien mit Drei-Faktoren-Struktur) bzw. der nicht vorhandenen Stichprobenbeschreibungen (siehe Studien mit Ein-Faktor-Struktur) der genannten Untersuchungen nicht möglich. Duffin, French und Patrick (2012) postulierten, dass Lehramtsstudierende zu Beginn des Studiums die Domänen nicht unterscheiden können, und deklarierten in diesem Zusammenhang den Forschungsbedarf an Studien, die die Faktorenstruktur des TSES zu verschiedenen Zeitpunkten im Lehramtsstudium systematisch untersuchen. Zu kritisieren ist insgesamt, dass die Autorinnen der TSES ein SWE-Instrument im Sinne Banduras entwickeln wollten, die Item-Formulierung jedoch mit dem Konstrukt der Kontrollüberzeugung konfundiert ist und streng genommen nicht den Empfehlungen Banduras (2006) zur Item-Formulierung entspricht. Denn nach dem Verständnis von Selbstwirk-

samkeitserwartung bei Bandura muss in den Items nach dem Überzeugungsgrad gefragt werden, verschiedene herausfordernde Tätigkeiten aufgrund der eigenen Fähigkeiten selbst ausführen zu können (z.B. „Ich kann ..., auch wenn ...“ auf einer Zustimmungsskala von „stimmt genau“ (1) bis „stimmt gar nicht“ (5)). Im TSES wird jedoch nach dem Grad des Einflusses auf einen bestimmten Aspekt des Unterrichts gefragt („How much can you do to ...?“ auf einer Antwortskala von „nothing“ (1) bis „a great deal“ (9)), d.h. es geht um den Einfluss, den die Person auf das Handlungsergebnis in einer bestimmten Situation hat – in Relation zu Einflüssen außerhalb der eigenen Kontrolle. Im deutschen Sprachraum wurden bereits adaptierte Übersetzungen der TSES verwendet (Dicke et al., 2014) und validiert (Pfitzner-Eden et al., 2014), die den Versuch einer Anpassung der Skala an Banduras Empfehlungen unternehmen. Allerdings wurden nur Kurzversionen der Ursprungsskala gebildet, die sich zudem von der Kurzversion der TSES-Autorinnen unterscheiden, was die Vergleichbarkeit mit der Originalskala erschwert und die jeweilige alleinige Verwendung der drei Subskalen einschränkt.

2.3 Ziele und Hypothesen

Trotz zahlreicher Studien im deutschen Sprachraum, die die Lehrkraft-SWE untersuchen (z.B. Dicke et al., 2014), und des Bedarfs an einer Skala zur Messung spezifischer Lehrkraft-SWE zu einem frühen Zeitpunkt in der Lehrkräfteausbildung (Duffin et al., 2012) liegen kaum validierte Instrumente zur Messung der Lehrkraft-SWE mit spezifischen Subskalen bei Lehramtsstudierenden vor. Ziel der vorliegenden Studie war daher, ein Instrument zur Lehrkraft-SWE für den deutschen Sprachraum zu übersetzen, zu adaptieren und hinsichtlich seiner Validität zu überprüfen, welches die Lehrkraft-SWE in Anlehnung an Banduras Definition domänenspezifisch mit Subskalen zum Engagement, zu Instruktionen sowie zum Classroom-Management differenziert. Durch die Skala soll die Evaluation von classroom-management-spezifischen Interventionen sowie die Forschung zur Entwicklung der Lehrkraft-SWE im Rahmen der Lehramtsausbildung im deutschen Sprachraum verbessert werden.

Angenommen wird, dass sich für die neu adaptierte Skala eine Drei-Faktoren-Struktur zeigen wird (Pfitzner-Eden et al., 2014; Poulou, 2007), da die untersuchten Lehramtsstudierenden in Reflexionsseminaren und Praktika erste eigene Erfahrungen gesammelt haben und so zwischen den drei Domänen differenzieren können sollten. Es wird erwartet, dass ein Drei-Faktoren-Modell des Instruments mit den Subskalen Engagement, Instruktionen und Classroom-Management einem Ein-Faktor-Modell überlegen ist. Da die Subskalen alle einen spezifischen Teil der Lehrkraft-SWE abbilden, wird dabei ein positiver Zusammenhang der Subskalen untereinander angenommen (= Hypothese 1).

Da, wie eingangs ausgeführt, eine höhere Lehrkraft-SWE mit geringen Stresswerten einhergeht (Klassen & Chiu, 2010), wird ein negativer Zusammenhang der Skala mit der Perceived Stress Scale (PSS; Büssing, 2011) als einem Maß für Stress als zusätzlicher Validitätsindikator angenommen. Außerdem wird eine positive Korrelation mit der im deutschen Sprachraum etablierten Skala WirkLehr (Schmitz & Schwarzer, 2000) erwartet (= Hypothese 2).

Zur Überprüfung der Reliabilität wurde untersucht, ob sich akzeptable Retest-Reliabilitätswerte für die Skala zeigen. Es wird erwartet, dass sich jeweils eine positive Korrelation der drei Subskalen mit einer Messung nach vier Wochen zeigt (= Hypothese 3).

Schließlich wurde die Skala im Rahmen der Evaluation eines Classroom-Management-Trainings eingesetzt. Es wird erwartet, dass Lehramtsstudierende, die ein Classroom-Management-Training absolvierten, im Vergleich mit einer Kontrollgruppe ohne Classroom-Management-Training einen signifikant stärkeren Zuwachs auf den drei Subskalen zeigen, insbesondere auf der Subskala Classroom-Management (= Hypothese 4).

3 Methode

3.1 Übersetzungsprozedur

Die Übersetzungsprozedur der TSES-Items ist an der „Kidscreen Translation and Validation Procedure“ (The KIDSCREEN group, 2004) orientiert. Zwei englische Muttersprachler (ein Lehrer sowie ein Soziologiestudent) übersetzten die TSES-Items jeweils ins Deutsche. Eine Gruppe aus Expertinnen und Experten in den Bereichen Pädagogischer Psychologie und/oder Classroom-Management diskutierte die Versionen und passte unverständliche oder unklare Items an. Die adaptierte Version wurde von einer angehenden Lehrerin englischer Muttersprache zurück ins Englische übersetzt. Vor- und Rückübersetzung wurden von den Autorinnen der Studie überprüft. Abschließend wurden die Items an das Frageformat der WirkLehr-Skala (Schmitz & Schwarzer, 2000) angepasst (z.B. „Ich traue mir zu, selbst zu den schwierigsten Schülerinnen und Schülern durchzudringen.“ auf einer Antwortskala von 1 – „stimmt gar nicht“ bis 5 – „stimmt genau“ anstelle von „How much can you do to get through to the most difficult students?“ auf einer Antwortskala von 1 – „Nothing“ bis 9 – „A great deal“; Tschannen-Moran & Woolfolk Hoy, 2001). Die Antwort-Skala von Schmitz und Schwarzer (2000) wurde von einer vierstufigen zu einer fünfstufigen verändert, da nach Banduras Empfehlungen (2006) Skalierungen mit wenigen Stufen vermieden werden sollten, weil diese weniger sensitiv und reliabel sind und die Möglichkeit einer mittleren Ausprägung gegeben sein sollte. Zudem erschien es für die Zustimmungsskala inhaltlich wenig sinnvoll und für Teilnehmende schwierig, zwischen neun Stufen zu differenzieren, wie dies im TSES vorgegeben war.

3.2 Stichprobe und Vorgehen

Es wurden Daten von $N = 637$ Lehramtsstudierenden im dritten Semester in zwei aufeinanderfolgenden Kohorten erhoben, was einer Quote von knapp 95% der initial angeschriebenen Personen entspricht. 19 Studierende nahmen nicht vollständig an der Erhebung teil und wurden aus zukünftigen Analysen ausgeschlossen, sodass Daten von $N = 618$ Studierenden einbezogen wurden (72,4% weiblich, Alter: $M = 22.05$ Jahre, $SD = 3.41$). Die Kohorten unterschieden sich nicht signifikant in Geschlecht oder Alter und repräsentierten verschiedene Fächer (z.B. Mathematik, Deutsch, Englisch, Geschichte) sowie Lehramtsformen (gymnasiales sowie Grund-, Haupt-, Realschullehramt). Vor Untersuchungsbeginn hatten die Studierenden bereits je ein semesterbegleitendes wöchentliches und ein dreiwöchiges Block-

praktikum in der Schule absolviert. Da, wie eingangs beschrieben, Schulpraktika zur Entwicklung der Lehrkraft-SWE führen, legitimierte dies den Einsatz der adaptierten Skala bereits zu diesem frühen Zeitpunkt im Lehramtsstudium.

Die Datenerhebung erfolgte im Rahmen der Evaluation eines Classroom-Management-Trainings. Hierbei nahmen alle Studierenden zur Praktikumsvorbereitung verpflichtend an einer Classroom-Management-Veranstaltung teil und wurden zufällig einem kognitiv-behavioralen Training, einem Online-Literaturkurs als Alternativ-Treatment oder einer Wartekontrollgruppe zugewiesen. Eine Beschreibung des Trainings ist nachzulesen bei Hannemann, Uhde und Thies (2019a). Teilnehmende bearbeiteten zu Wintersemesterbeginn einen Fragebogen. Die Bearbeitung erfolgte überwiegend online¹. Studierende der ersten Stichprobe, die nicht an einer Classroom-Management-Veranstaltung teilnahmen (randomisiert zugewiesene Kontrollgruppe; $n = 68$; 73,5% weiblich, Alter: $M = 22.03$ Jahre, $SD = 3.49$), wurden nach einem Monat gebeten, den Fragebogen zu einem zweiten Messzeitpunkt erneut auszufüllen. Außerdem wurde in der zweiten Kohorte zusätzlich die PSS eingesetzt, um ein weiteres Validierungsmaß neben der WirkLehr-Skala zu erhalten. Alle Teilnehmenden wurden vor Studienbeginn über Freiwilligkeit und Anonymität ihrer Teilnahme informiert und willigten ein.

3.3 Messinstrumente

Lehrkraft-SWE. Die neu adaptierte Skala misst die Lehrkraft-SWE analog zur TSES über die Subskalen Engagement, Instruktionen und Classroom-Management mit jeweils acht Items. Die Instruktion für die 24 Items lautete: „Wie schätzen Sie Ihre Möglichkeiten und Fähigkeiten ein, mit verschiedenen Schwierigkeiten Ihrer schulischen Tätigkeiten umzugehen? Bitte geben Sie an, inwieweit die folgenden Aussagen auf Sie zutreffen! Erinnern Sie sich dazu an Ihre Erfahrungen aus den bisherigen Schulpraktika.“ Teilnehmende gaben auf einer fünfstufigen Likert-Skala von 1 (stimmt gar nicht) bis 5 (stimmt genau) ihre Zustimmung zu dem jeweiligen Item an.

Instrumente zur Untersuchung der Validität. Bezüglich der WirkLehr-Skala (s. Abschnitt 2.2) gaben Teilnehmende auf einer fünfstufigen Likert-Skala von 1 (stimmt gar nicht) bis 5 (stimmt genau) ihre Zustimmung zu dem jeweiligen Item an. Die interne Konsistenz betrug Cronbach's $\alpha = .76$. Die PSS von Cohen (1994) wurde in der deutschen Übersetzung von Büssing (2011) eingesetzt. Die Skala misst die Stresswahrnehmung über zehn Items, z.B. „Wie oft hatten Sie im letzten Monat das Gefühl, dass sich die Probleme so aufgestaut haben, dass Sie diese nicht mehr bewältigen können?“. Die Teilnehmenden gaben auf einer fünfstufigen Likert-Skala von 1 (nie) bis 5 (sehr oft) an, wie häufig sie sich während des letzten Monats durch Stress belastet gefühlt haben. Die interne Konsistenz betrug Cronbach's $\alpha = .87$.

1 In Kohorte 2 wurden bei einem Teil der Stichprobe im Rahmen des Classroom-Management-Trainings Papierfragebögen eingesetzt, um die Teilnahmequote zu erhöhen. Auf der Subskala Classroom-Management zeigten sich etwas höhere Werte für das Onlineformat ($M = 3.67$, $SD = 0.48$) im Vergleich zum Papierfragebogen ($M = 3.52$, $SD = 0.49$) mit geringer Effektstärke ($t(251) = 2.32$, $p = .021$, $d = 0.3$).

3.4 Datenanalyse

Eine konfirmatorische Faktorenanalyse wurde mit den Daten der Gesamtstichprobe durchgeführt, um die Drei-Faktoren-Struktur mit der Ein-Faktor-Struktur zu vergleichen. Die konvergente Validität wurde mit der Skala WirkLehr (Schmitz & Schwarzer, 2000) untersucht. In der zweiten Stichprobe wurde die Kriteriumsvalidität durch die PSS (Cohen, 1994; Übersetzung von Büssing, 2011) überprüft. Außerdem wurde die Retest-Reliabilität anhand einer Teilstichprobe (unbehandelte Kontrollgruppe aus Kohorte 1) nach einem Monat untersucht. Schließlich wurden die drei Subskalen zur Evaluation eines Classroom-Management-Trainings angewandt und dazu eine Varianzanalyse mit Messwiederholung der Prä-Post-Messungen zwischen Trainings- und Kontrollgruppe der ersten Stichprobe durchgeführt. Für alle Analysen wurden die Programme SPSS in Version 23 und MPlus in Version 8 angewandt. Das Alpha-Niveau wurde jeweils auf $p < .05$ festgelegt (Cohen, 1992).

4 Ergebnisse

4.1 Faktorvalidität

Zur Überprüfung der Faktorenstruktur der Skala wurden zwei Modelle mittels konfirmatorischer Faktorenanalyse getestet, und zwar ein Ein-Faktor-Modell sowie ein Drei-Faktoren-Modell mit drei korrelierten Faktoren. Ein robuster Schätzer wurde verwendet (MLR), der auch bei kleineren Stichprobengrößen und nicht normalverteilten Daten zuverlässig misst (Finney & DiStefano, 2006). Bezüglich fehlender Werte wurde das Full-Information-Maximum-Likelihood-Verfahren eingesetzt (Collins, Schafer & Kam, 2001). Für die Evaluation des Modell-Fit wurde eine Multi-Index-Strategie verwendet (z.B. Hu & Bentler, 1999). Ein absoluter Fit-Index gibt an, zu welchem Grad das A-Priori-Modell durch die vorliegenden Daten repräsentiert wird. Folgende absolute Fit-Indices wurden verwendet: Der χ^2 -Test, der CFI und der TLI, der SRMR sowie der RMSEA. In Modell 1 repräsentierten alle 24 Items eine einzige Dimension der Lehrkraft-SWE. In Modell 2 umfassen die 24 Items drei Subskalen, die unabhängige, aber korrelierte Faktoren bilden: die Lehrkraft-SWE bezüglich des Engagements der Schülerinnen und Schüler, die Lehrkraft-SWE bezüglich Instruktionen sowie die Lehrkraft-SWE bezüglich Classroom-Management. Wie Tabelle 1 zeigt, ergaben sich akzeptable Modellfits, wobei das Drei-Faktoren-Modell die Datenstruktur besser abbildete als das Ein-Faktor-Modell. Die Chi-Quadrat-Werte wiesen darauf hin, dass beide Modelle signifikant von den erhobenen Daten abweichen. Nach Hu und Bentler (1999) indizieren der SRMR und RMSEA jedoch einen guten Modellfit für beide Modelle (Cutoff-Werte: guter Fit $< .06$; akzeptabler Fit $< .08$). Der CFI und TLI indizieren lediglich für das Modell 2 einen akzeptablen Modellfit (Cutoff-Werte: guter Fit $> .95$, akzeptabler Fit $> .90$), sodass dieses dem Modell 1 überlegen war.

Tabelle 1: Fit Indices der konfirmatorischen Faktorenanalyse für die drei Subskalen

	χ^2	<i>df</i>	<i>p</i>	CFI/TLI	SRMR	RMSEA	KI (90%)	<i>p</i> (RMSEA- A≤0.05)
Modell1	685.69	252	<.001	0.89/0.87	0.05	0.05	.05–.06	.15
Modell2	590.24	249	<.001	0.91/0.90	0.04	0.05	.04–.05	.82

Anmerkungen: Modell 1 = Ein-Faktor Modell; Modell 2 = Modell mit drei korrelierten Faktoren; CFI = Comparative Model Fit Index ; TLI = Tucker-Lewis Index; SRMR = Standardized Root Mean Square Residual; RMSEA = Root Mean-Square Error of Approximation; KI = Konfidenzintervall.

4.2 Deskriptive Analyse und interne Konsistenz

Es ergaben sich folgende Mittelwerte, Standardabweichungen und interne Konsistenzen der drei Subskalen: Engagement (*Mean* = 3.65, *SD* = 0.50, α = .83), Instruktionen (*Mean* = 3.71, *SD* = 0.48, α = .80), Classroom-Management (*Mean* = 3.64, *SD* = 0.49, α = .81). Die interne Konsistenz ist somit als hoch zu bewerten. Die standardisierten Faktorladungen der adaptierten Skala lagen im Bereich zwischen .50 und .70, mit der Ausnahme eines Items mit .39, und wurden somit insgesamt als zufriedenstellend eingestuft (Cut-off: .40; z.B. Civitillo et al., 2016).

4.3 Konstruktvalidität

Um die Konstruktvalidität zu überprüfen, wurden die drei neu entstandenen Subskalen mit der WirkLehr-Skala (Schmitz & Schwarzer, 2000) und der PSS (Büssing, 2011) korreliert (Tabelle 2). Unter der Verwendung der Gesamtstichprobe korrelieren die drei Subskalen positiv miteinander sowie mit der WirkLehr-Skala (r = .65 – r = .73). Auf der Datengrundlage der zweiten Stichprobe ergab sich eine mittlere negative Korrelation der präsentierten Subskalen mit der PSS (r = -.22 – r = -.28).

Tabelle 2: Manifeste Interkorrelationen der neu adaptierten Skala mit WirkLehr und PSS

	Engagement	Instruktionen	Classroom-Management
Engagement	1		
Instruktionen	0.70*	1	
Classroom-Management	0.73*	0.71*	1
WirkLehr	0.73*	0.65*	0.70*
PSS	-0.22*	-0.28*	-0.25*

Anmerkungen: n = 616 für die neu adaptierte Skala und n = 617 für WirkLehr; n = 337 für PSS; * p < .05.

4.4 Retest-Reliabilität

Einen Monat nach der Bearbeitung des ersten Fragebogens erhielten die Teilnehmenden der Kontrollgruppe der ersten Kohorte einen zweiten Fragebogen. Wie Tabelle 3 zeigt, war die Retest-Reliabilität zufriedenstellend mit Werten für die Skala Engagement von r = .72, für die Skala Instruktionen von r = .58 und für die Classroom-Management-Skala von r = .71 (nach Bagozzi und Yi (1988) werden Werte größer als .6 empfohlen).

Tabelle 3: Mittelwerte, Standardabweichungen und Retest-Reliabilität für die neu adaptierte Skala

Skalen	n_1	Mean ₁	SD ₁	n_2	Mean ₂	SD ₂	r
Engagement	68	3.68	0.49	55	3.74	.47	.72*
Instruktionen	68	3.69	0.44	55	3.71	.46	.58*
Classroom-Management	68	3.69	0.48	55	3.66	.49	.71*

Anmerkung: * $p < .05$.

4.5 Evaluation des Classroom-Management-Trainings

Die Skala wurde im Rahmen der Evaluation eines 15-stündigen Classroom-Management-Trainings eingesetzt. Das Training erwies sich als wirksam in Bezug auf Veränderungen in den Selbsteinschätzungen hinsichtlich Classroom-Management-Wissen und -Kompetenzen (Hannemann et al., 2019a; Hannemann, Uhde & Thies, 2019b) und wurde somit als Außenkriterium herangezogen, um weitere Hinweise auf die Kriteriumsvalidität der adaptierten Skala zu erhalten bzw. deren Veränderungssensitivität zu überprüfen. Es zeigten sich signifikante Unterschiede zwischen der Trainingsgruppe ($n = 96$) und der Kontrollgruppe ($n = 55$) im Zuwachs zwischen Prä- und Postmessung mit einem starken Effekt für die Classroom-Management-Skala sowie einem kleinen Effekt für die Skala Instruktionen (siehe Tabelle 4). Die Interaktion der Skala Engagement zwischen Gruppen und Messzeitpunkten war nicht signifikant.

Tabelle 4: Trainingsevaluation: Interaktionseffekte der Prä-Post-Messung

Skalen	Gruppe	N	$M_1 (SD_1)$	$M_2 (SD_2)$	F	p	η^2
Engagement	TG	96	3.70 (0.53)	3.91 (0.58)	2.86	.093	.02
	KG	55	3.66 (0.46)	3.74 (0.47)			
Instruktionen	TG	96	3.74 (0.47)	3.96 (0.50)	8.31	.005	.03
	KG	55	3.67 (0.44)	3.69 (0.45)			
Classroom-Management	TG	96	3.70 (0.48)	4.05 (0.57)	25.83	<.001	.15
	KG	55	3.69 (0.49)	3.66 (0.49)			

Anmerkungen: TG = Trainingsgruppe; KG = Kontrollgruppe.

5 Diskussion

Ergebnisse zur Validität. Erwartungskonform zeigten die Ergebnisse der konfirmatorischen Faktorenanalyse analog zu den Studien von Pfitzner-Eden et al. (2014) und Poulou (2007) einen leicht besseren Modellfit eines Drei-Faktoren-Modells der Skala gegenüber einem Ein-Faktor-Modell. Das heißt, anders als von Tschannen-Moran und Woolfolk Hoy (2001) für die Verwendung der TSES-Skala mit Lehramtsstudierenden postuliert, schienen die Studierenden in der Lage zu sein, ihre Fähigkeiten in den drei Domänen differenziert wahrzunehmen. Möglicherweise ist dies auf den hohen Anteil an Praxiselementen zu einem frühen Zeitpunkt im Studium zurückzuführen. Hinsichtlich der Validität zeigte sich erwartungskonform eine positive Korrelation der drei Subskalen mit der WirkLehr-Skala (Schmitz & Schwarzer, 2000) und eine mittlere negative Korrelation mit dem Stresserleben (PSS;

Büssing, 2011). Bezüglich der Retest-Reliabilität zeigte sich erwartungskonform eine positive Korrelation der Subskalen zu einem zweiten Messzeitpunkt nach einem Monat. Aufgrund der Ergebnisse kann die Skala zur Messung der Selbstwirksamkeitserwartung bei Lehramtsstudierenden für die Bereiche Engagement, Instruktionen und Classroom-Management eingesetzt werden.

Limitationen der Studie. Eine Einschränkung stellt die Datengrundlage einiger Teilergebnisse dar. Die Gesamtstichprobe setzte sich aus Erhebungen in zwei aufeinanderfolgenden Kohorten von Lehramtsstudierenden im dritten Semester zusammen. Die Kriteriumsvalidität bezüglich des Stresserlebens wurde nur mit der zweiten Stichprobe durchgeführt, da diese Daten in der ersten Stichprobe nicht erhoben wurden. Ebenso wurde die Retest-Reliabilität anhand eines Teils der ersten Stichprobe überprüft, da die anderen Lehramtsstudierenden zu verschiedenen Zeitpunkten nach der Datenerhebung an unterschiedlichen Classroom-Management-Veranstaltungen teilnahmen, weshalb unterschiedliche SWE-Entwicklungen zu erwarten waren. Daher sind Studien wünschenswert, die die Ergebnisse zur Kriteriumsvalidität und die Retest-Reliabilität überprüfen. Um Aussagen zu der adaptierten Skala verglichen mit der originalen TSES hinsichtlich ihrer Konfundierung mit dem Konstrukt der Kontrollüberzeugungen zu treffen, sind Studien unter Verwendung spezifischer Instrumente zur Lehrkraft-SWE auf Grundlage der Locus-of-Control-Theorie notwendig.

Daneben ist die Entwicklung einer Kurzversion wünschenswert, denn bezüglich des Modellfit zeigten sich zufriedenstellende Werte für die Drei-Faktoren-Lösung, die Daten schienen jedoch nicht so gut die angenommene Faktorenstruktur abzubilden, wie dies in Validierungsstudien zur TSES berichtet wird (Pfitzner-Eden et al., 2014). Als Grund für diesen Befund lässt sich der Einschluss aller Items, auch derjenigen mit niedrigen Faktorladungen, ausmachen. Dieses Vorgehen wurde jedoch gewählt, um eine inhaltliche Vergleichbarkeit mit anderen Studien zum TSES zu gewährleisten. Durch die Auswahl der Items mit den jeweils vier höchsten Faktorladungen ließe sich ein besserer Modellfit erzielen. Eine solche Kurzversion würde sich aber in der Item-Auswahl sowohl von der Kurzversion der TSES (Tschannen-Moran & Woolfolk Hoy, 2001) als auch von den zwölf Items der TSES-Übersetzung von Pfitzner-Eden et al. (2014) unterscheiden, obwohl in beiden Studien die vier Items mit den jeweils größten Faktorladungen pro Unterskala ausgewählt wurden. Die hohen Faktorladungen als einziges Auswahlkriterium heranzuziehen, scheint vor dem Hintergrund dreier unterschiedlicher Kurzversionen beliebig, weshalb bei der Item-Reduktion zusätzlich die inhaltliche Ebene berücksichtigt werden muss. So wies beispielsweise das letzte Item der Classroom-Management-Skala („Ich bin mir sicher, dass ich in meiner Klasse ein Classroom-Management-System etablieren kann.“) eine niedrige Faktorladung auf, und ein Ausschließen des Items liegt inhaltlich nahe, da es für die untersuchten Lehramtsstudierenden aufgrund geringer Bekanntheit des Begriffs zu dem frühen Zeitpunkt im Studium vermutlich wenig verständlich war.

Alle anderen Faktorladungen lagen zwar im zufriedenstellenden Bereich, dennoch fielen diese bei einzelnen Items geringer aus. So zeigte sich eine niedrigere Faktorladung für das letzte Item der Subskala Engagement („Ich kann Familien darin unterstützen, ihren Kindern zu helfen, in der Schule erfolgreich zu sein.“). Vermutlich hing hier die eingangs beschriebene diffusere Selbstwirksamkeitserwartung (Schüle et al., 2016) mit dem Fehlen

von Erfahrungen etwa im Bereich Elterngespräche zusammen, da diese nicht zu den Haupttätigkeiten der Studierenden im Praktikum zählten. Auch die letzten zwei Items der Subskala Instruktionen wiesen etwas niedrigere Faktorladungen auf („Ich traue mir zu, neue Lehr-Lern-Methoden in meinem Unterricht einzuführen.“; „Ich bin in der Lage, verschiedene Verfahren zur Leistungserfassung einzusetzen.“). Die Selbstwirksamkeitserwartung könnte hier mit dem pädagogisch-psychologischen Wissen z.B. über Lehr-Lern-Methoden bzw. Verfahren zur Leistungsbeurteilung konfundiert sein. Die niedrigeren Faktorladungen könnten insofern mit Unterschieden im Lehramtscurriculum zusammenhängen. Zwar sind Zeitpunkt und Dauer der Schulpraktika der Bachelorstudierenden für die gesamte Kohorte vereinheitlicht. Daneben ergeben sich jedoch zahlreiche Unterschiede bezüglich behandelter Inhalte z.B. durch verschiedene Fächerkombinationen oder Schwerpunktsetzungen in den einzelnen Lehrveranstaltungen.

Praktische Implikationen. Die Selbstwirksamkeitserwartung steht im positiven Zusammenhang mit Merkmalen erfolgreichen Unterrichts (Caprara et al., 2006; Klassen et al., 2011; Zee & Koomen, 2016). Dennoch werden im deutschen Sprachraum wenig Forschungsergebnisse bezüglich spezifischer Aspekte der Lehrkraft-SWE berichtet, die genauere Vorhersagen ermöglichen können und wichtige Ansatzpunkte zur Intervention bei ungünstigen Voraussetzungen bereits während der Lehramtsausbildung darstellen (Civitillo et al., 2016). Mit der präsentierten Skala wird ein Instrument für den deutschen Sprachraum vorgestellt, das die Vorteile zweier Instrumente zur Messung der Lehrkraft-SWE bei angehenden Lehrkräften verbindet: die inhaltliche Spezifität der TSES (Tschannen-Moran & Woolfolk Hoy, 2001) mit der Fragestruktur und dem Antwortformat der WirkLehr-Skala (Schmitz & Schwarzer, 2000), die eine Messung der Lehrkraft-SWE im Sinne der sozial-kognitiven Lerntheorie Banduras (1997) ermöglicht. Neben den Analysen zur Validierung ist die Skala im Rahmen einer Classroom-Management-Trainingsevaluation eingesetzt worden. Ziel des Classroom-Management-Trainings war u.a. die Steigerung der Lehrkraft-SWE. Die Varianzanalyse zwischen Trainings- und Kontrollgruppe von Prä- zu Post-Test zeigte einen großen Effekt für die Classroom-Management-Skala sowie einen kleinen Effekt für die Skala Instruktionen zugunsten der Trainingsgruppe. Die Ergebnisse bekräftigen den Nutzen der Skala, insbesondere hinsichtlich der Änderungssensitivität der drei Subskalen sowie deren Einsatz in der Interventionsforschung.

Ausblick. Zukünftige Studien sollten die Validität der Skala zu verschiedenen Messzeitpunkten innerhalb des Studiums sowie in der zweiten und dritten Phase der Lehramtsausbildung prüfen, um Aussagen bezüglich der Messinvarianz machen zu können. Die Ergebnisse weisen darauf hin, dass Lehramtsstudierende bereits zu einem frühen Zeitpunkt ihres Studiums zwischen verschiedenen SWE-Bereichen differenzieren. Die neu adaptierte Skala ermöglicht somit die Untersuchung spezifischer Lehrkraft-SWE-Entwicklung im Lehramtsstudium.

Literatur

- Bagozzi, R. P. & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, 16 (1), 74–94. <https://doi.org/10.1007/BF02723327>
- Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York: Freeman.
- Bandura, A. (2006). Guide for constructing self-efficacy scales. In T. Urdan & F. Pajares (Hrsg.), *Self-efficacy beliefs of adolescents* (S. 307–337). Greenwich, CT: Information Age Publishing.
- Büssing, A. (2011). *Translation of Cohen's 10 item PERCEIVED STRESS SCALE (PSS)*. Verfügbar unter: <https://www.cmu.edu/dietrich/psychology/stress-immunity-disease-lab/scales/index.html>
- Caprara, G. V., Barbaranelli, C., Steca, P. & 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. <https://doi.org/10.1016/j.jsp.2006.09.001>
- Civitillo, S., Juang, L., Schachner, M. & Börnert, M. (2016). Validierung einer deutschen Version der „Culturally Responsive Classroom Management Self-Efficacy Scale“. *Empirische Sonderpädagogik*, Heft 3/2016, 279–288.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112 (1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Cohen, S. (1994). *Perceived Stress Scale*. Verfügbar unter: <http://mindgarden.com/documents/PerceivedStressScale.pdf>
- Coladarci, T. (1992). Teachers' sense of efficacy and commitment to teaching. *Journal of Experimental Education*, 60 (4), 323–337. <https://doi.org/10.1080/00220973.1992.9943869>
- Collins, L. M., Schafer, J. L. & Kam, C.-M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6 (4), 330–351. <https://doi.org/10.1037/1082-989X.6.4.330>
- Dicke, T., Parker, P. D., Marsh, H. W., Kunter, M., Schmeck, A. & Leutner, D. (2014). Self-efficacy in classroom management, classroom disturbances, and emotional exhaustion: A moderated mediation analysis of teacher candidates. *Journal of Educational Psychology*, 106 (2), 569–583. <https://doi.org/10.1037/a0035504>
- 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. <https://doi.org/10.1016/j.tate.2012.03.004>
- Finney, S. J. & DiStefano, C. (2006). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Hrsg.), *Structural equation modelling: A second course* (S. 269–314). Greenwich, CT: Information Age Publishing.
- Fives, H. & Buehl, M. M. (2010). Examining the factor structure of the teachers' sense of efficacy scale. *Journal of Experimental Education*, 78 (1), 118–134. <https://doi.org/10.1080/00220970903224461>
- Gibson, S. & Dembo, M. H. (1984). Teacher efficacy. A construct validation. *Journal of Educational Psychology*, 76 (4), 569–582. <https://doi.org/10.1037/0022-0663.76.4.569>
- Goddard, R. D., Hoy, W. K. & Woolfolk Hoy, A. (2000). Collective teacher efficacy. Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, 37 (2), 479–507. <https://doi.org/10.3102/00028312037002479>
- Hannemann, L., Uhde, G. & Thies, B. (2019a). Training zur Förderung von Classroom-Management-Kompetenzen von Lehramtsstudierenden – 2. Evaluationsstudie. In G. Uhde & B. Thies (Hrsg.), *Kompetenzentwicklung im Lehramtsstudium durch professionelles Training* (S. 69–82). Technische Universität Braunschweig.
- Hannemann, L., Uhde, G. & Thies, B. (2019b). Evaluation eines Classroom-Management-Trainings für Lehramtsstudierende. *Zeitschrift für Bildungsforschung*. <http://doi.org/10.1007/s35834-019-00255-1>

- Hu, L.-t. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6 (1), 1–55. <https://doi.org/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. <https://doi.org/10.1037/a0019237>
- 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 (1), 21–43. <https://doi.org/10.1007/s10648-010-9141-8>
- Klusmann, U., Kunter, M., Voss, T. & Baumert, J. (2012). Berufliche Beanspruchung angehender Lehrkräfte: Die Effekte von Persönlichkeit, pädagogischer Vorerfahrung und professioneller Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 26 (4), 275–290. <https://doi.org/10.1024/1010-0652/a000078>
- Pfitzner-Eden, F., Thiel, F. & Horsley, J. (2014). An adapted measure of teacher self-efficacy for pre-service teachers. Exploring its validity across two countries. *Zeitschrift für Pädagogische Psychologie*, 28 (3), 83–92. <https://doi.org/10.1024/1010-0652/a000125>
- Poulou, M. (2007). Personal teaching efficacy and its sources. Student teachers' perceptions. *Educational Psychology*, 27 (2), 191–218. <https://doi.org/10.1080/01443410601066693>
- Robertson, C. & Dunsmuir, S. (2013). Teacher stress and pupil behaviour explored through a rational-emotive behaviour therapy framework. *Educational Psychology*, 33 (2), 215–232. <https://doi.org/10.1080/01443410.2012.730323>
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 80 (1), 1–28. <https://doi.org/10.1037/h0092976>
- Schmitz, G. S. & Schwarzer, R. (2000). Selbstwirksamkeitserwartung von Lehrern. Längsschnittbefunde mit einem neuen Instrument. *Zeitschrift für Pädagogische Psychologie*, 14 (1), 12–25. <https://doi.org/10.1024/1010-0652.14.1.12>
- Schüle, C., Besa, K.-S., Schriek, J. & Arnold, K.-H. (2016). Die Veränderung der Lehrerselbstwirksamkeitsüberzeugung in Schulpraktika. *Zeitschrift für Bildungsforschung*. <https://doi.org/10.1007/s35834-016-0177-9>
- Seidel, T. & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade. The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77 (4), 454–499. <https://doi.org/10.3102/0034654307310317>
- Seifert, A. & Schaper, N. (2018). Die Veränderung von Selbstwirksamkeitserwartungen und der Berufswahlsicherheit im Praxissemester. In J. König, M. Rothland & N. Schaper (Hrsg.), *Learning to practice, learning to reflect?* (S. 195–222). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-19536-6_7
- The KIDSCREEN group. (2004). *Translation & validation procedure. Guidelines and documentation form*. Verfügbar unter: <http://s2f1ad284f5ffc52e.jimcontent.com/>
- Tschannen-Moran, M., & Woolfolk Hoy, A. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68 (2), 202–248. <https://doi.org/10.3102/00346543068002202>
- Tschannen-Moran, M. & Woolfolk Hoy, A. (2001). Teacher efficacy. Capturing an elusive construct. *Teaching and Teacher Education*, 17 (7), 783–805. [https://doi.org/10.1016/S0742-051X\(01\)00036-1](https://doi.org/10.1016/S0742-051X(01)00036-1)
- 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. <https://doi.org/10.1016/j.tate.2005.01.007>

Zee, M. & Koomen, H. M. Y. (2016). Teacher self-efficacy and its effects on classroom processes, student academic adjustment, and teacher well-being. *Review of Educational Research*, 86 (4), 981–1015. <https://doi.org/10.3102/0034654315626801>

Anhang

Mittelwerte, Standardabweichungen, interne Konsistenz der neu adaptierten Skala sowie Items mit standardisierten Faktorladungen.

	Faktor-
	ladungen
<i>Faktor 1: Selbstwirksamkeit Engagement</i> (Mean = 3.65, SD = 0.50, α = .83)	
Ich bin in der Lage, Schülerinnen und Schüler davon zu überzeugen, dass sie schulische Aufgaben gut bewältigen können.	.70
Ich traue mir zu, selbst diejenigen Schülerinnen und Schüler zu motivieren, die wenig Interesse am Unterricht zeigen.	.69
Ich traue mir zu, selbst zu den schwierigsten Schülerinnen und Schülern durchzudringen.	.65
Ich kann meine Schülerinnen und Schüler dazu bringen, das Lernen wertzuschätzen.	.63
Selbst wenn eine Schülerin oder ein Schüler schlechte Leistungen in der Schule erbringt, schaffe ich es, das Verständnis dieser Schülerin oder dieses Schülers zu verbessern.	.61
Ich bin in der Lage, bei meinen Schülerinnen und Schülern kritisches Denken zu fördern.	.59
Ich bin in der Lage, die Kreativität meiner Schülerinnen und Schüler zu fördern.	.57
Ich kann Familien darin unterstützen, ihren Kindern zu helfen, in der Schule erfolgreich zu sein.	.50
<i>Faktor 2: Selbstwirksamkeit Instruktionen</i> (Mean = 3.71, SD = 0.48, α = .80)	
Ich bin mir sicher, dass ich auch besonders begabten Schülerinnen und Schülern angemessene Herausforderungen bieten kann.	.64
Mir gelingt es, auch auf schwierige Fragen von Schülerinnen und Schülern einzugehen.	.63
Ich kann gut einschätzen, was meine Schülerinnen und Schüler von meinem Unterricht verstanden haben.	.60
Wenn Schülerinnen oder Schüler verwirrt sind, bin ich mir sicher, dass ich ihnen mit alternativen Erklärungen oder Beispielen weiterhelfen kann.	.60
Ich traue mir zu, meinen Unterricht an die unterschiedlichen Schwierigkeitsniveaus aller Schülerinnen und Schüler anzupassen	.57
Ich bin in der Lage, gute Fragen für meine Schülerinnen und Schüler zu formulieren.	.54
Ich traue mir zu, neue Lehr-Lern-Methoden in meinem Unterricht einzuführen.	.52
Ich bin in der Lage, verschiedene Verfahren zur Leistungserfassung einzusetzen.	.51
<i>Faktor 3: Selbstwirksamkeit Classroom-Management</i> (Mean = 3.64, SD = 0.49, α = .81)	
Selbst wenn störendes Verhalten im Unterricht auftritt, bin ich mir sicher, dass ich dieses schnell unter Kontrolle bringen kann.	.69
Ich bin in der Lage, einige problematische Schülerinnen und Schüler davon abzuhalten, die gesamte Unterrichtsstunde zu ruinieren.	.67
Ich weiß, dass ich es schaffe, Routinen zu etablieren, damit der Unterricht reibungslos verläuft.	.64
Ich weiß, dass ich es schaffe, meine Erwartungen gegenüber meinen Schülerinnen und Schülern bezüglich deren Verhalten deutlich zu machen.	.63
Mir gelingt es, eine Schülerin oder einen Schüler zu beruhigen, der oder die stört oder laut ist.	.61
Mir gelingt es, dass meine Schülerinnen und Schüler die Klassenregeln einhalten.	.59
Ich bin in der Lage, selbst bei aufsässigen Schülerinnen oder Schülern gelassen zu reagieren.	.58
Ich bin mir sicher, dass ich in meiner Klasse ein Classroom-Management-System etablieren kann.	.39

Kontakt:

Lena Hannemann, Dr. Gesa Uhde, Prof. Dr. Barbara Thies
Technische Universität Braunschweig
Institut für Pädagogische Psychologie
Bienroder Weg 82
38106 Braunschweig
E-mail: l.hannemann@tu-bs.de
E-mail: g.uhde@tu-bs.de
E-mail: barbara.thies@tu-bs.de