

THORID RABE, CLAUDIA MEINHARDT, OLAF KREY

Entwicklung eines Instruments zur Erhebung von Selbstwirksamkeitserwartungen in physikdidaktischen Handlungsfeldern

Teachers' self efficacy beliefs for teaching physics: Development of a new instrument

ZUSAMMENFASSUNG

In aktuellen Forschungsprojekten zur Lehrerprofessionalisierung wird vorwiegend die fachdidaktische Wissenskomponente professioneller Handlungskompetenz von Physiklehrkräften untersucht. Ziel dieses Beitrags ist es erstens, auf Basis des aktuellen Forschungsstandes zu begründen, dass Forschungsprojekte zu domänenspezifischen, d.h. physikdidaktischen Selbstwirksamkeitserwartungen (SWE) die vorhandenen Projekte sinnvoll ergänzen können und sollten. Das Konstrukt der SWE wird dazu im Sinne einer Kompetenzerwartung vorgestellt und Ergebnisse empirischer Studien zu Lehrer-SWE werden referiert. Anhand einer Pilotstudie mit Physiklehramtsstudierenden wird zweitens gezeigt, dass es mit Bezug auf die theoretische Fundierung möglich ist, SWE in physikdidaktischen Handlungsfeldern so zu operationalisieren, dass ein valides Instrument zur Erhebung des Konstrukts für umfassendere Studien zur Verfügung steht. Diese sollen zunächst auf die langfristige Entwicklung von SWE im Rahmen der Ausbildung von (zukünftigen) Physiklehrkräften unter besonderer Berücksichtigung von Praxiserfahrungen fokussieren.

Schlüsselwörter: Physikdidaktische Selbstwirksamkeitserwartungen, Lehrerprofessionalisierung, motivationale Orientierungen, Erhebungsinstrument, Validierungsstudie

ABSTRACT

Present studies on science teachers' professional development focus on the (pedagogical content) knowledge component while widely neglecting the affective aspects of professional performance. Based on an overview on current research this paper (1) argues that research on domain specific self-efficacy should supplement current studies on physics teacher development. For this purpose the psychological construct of (teacher) self-efficacy is specified and results of empirical studies on teacher self-efficacy are reported. We conclude that so far self-efficacy in the domain of physics teaching is not addressed at an adequate level of specificity for the benefit of high predictivity and (2) offer first instruments to close this gap. In a pilot study self-efficacy concerning specific tasks in the field of physics teaching was operationalized based on theoretical assumptions. The constructed scales proved their potential to be valid psychometric instruments. After their revision research on the development of domain specific self-efficacy in teacher education can be approached.

Keywords: physics teachers' self efficacy, teachers' professional development, motivational orientation, psychometric instrument, validation study

1 Professionelle Handlungskompetenz von Lehrkräften: Einordnung und erste Begründung des Forschungsinteresses

In aktuellen Modellierungen professioneller Handlungskompetenz von Lehrkräften (vgl. z. B. die Projekte COACTIV und MT 21 oder Baumert & Kunter, 2006) spiegeln sich sowohl die kognitive wie auch die motivationale Komponente des Kompetenzbegriffs nach Weinert (2001) wider. Neben dem Professionswissen – das sich auf Wissen und Können bezieht – werden Überzeugungen und Werthaltungen, selbstregulative Fähigkeiten sowie motivationale Orientierungen genannt, wobei letztere die Bereiche der intrinsischen Motivation einerseits und der Kontrollüberzeugungen und Selbstwirksamkeitserwartungen (SWE) andererseits umfassen (Baumert & Kunter, 2006, 504). Das hier im Fokus stehende Konstrukt der SWE in physikdidaktischen Handlungsfeldern kann in der Schnittmenge zwischen dem Bereich des fachdidaktischen Professionswissens und den motivationalen Orientierungen verortet werden.

Dem Modell aus der COACTIV Studie liegt die Annahme zugrunde, dass das

Fach den eigentlichen Handlungsrahmen von Lehrkräften bildet, da das Fach letztlich „Oberflächen- und Tiefenstruktur“ von Unterricht (vgl. Oser & Baeriswyl, 2001) maßgeblich beeinflusst. Auf der Grundlage dieser Annahme ist es naheliegend, Studien zur Handlungskompetenz von (zukünftigen) Lehrkräften domänenspezifisch anzulegen.

Leitmotivisch für die theoretische Unterteilung des Professionswissens werden die Arbeiten Shulmans (1986) aufgegriffen, ohne dass deshalb davon auszugehen wäre, dass ein Konsens über die Inhalte der Wissensbereiche, die Abgrenzung der Wissensbereiche voneinander und über ihre Relation zueinander bestünde (vgl. z. B. Park & Oliver, 2008 oder Gramzow et al., 2012). Aus Sicht der vorliegenden Studie ist zunächst relevant, dass in fast allen (heuristischen) Modellierungen des Professionswissens überhaupt eine fachdidaktische Komponente postuliert wird (PCK nach Shulman oder fachspezifisch-pädagogisches Wissen nach Bromme, 1997, 197).

Bei dem Versuch, sich einen Überblick über die Forschung zu physikdidaktischer Handlungskompetenz im deutschsprachigen Raum zu verschaffen, entsteht

der Eindruck, dass inzwischen zwar intensiv im Bereich des Professionswissens und damit zu kognitiven Aspekten von Lehrerkompetenz geforscht wird (u. a. von Arbeitsgruppen in Paderborn, Duisburg-Essen bzw. Aachen), dass hingegen motivationale Aspekte eine eher untergeordnete bzw. randständige Rolle spielen. Die Relevanz, physikdidaktische und fachliche Kompetenzen auf kognitiver Ebene zu modellieren, zu operationalisieren und in ihrer Auswirkung auf den Physikunterricht und Lernprozesse im Bereich der Physik zu untersuchen, bleibt unbestritten. Nicht zuletzt, weil es noch nicht ausreichend belastbare Modelle oder Daten dazu gibt, welche Kompetenzbereiche überhaupt zu fachlichem und physikdidaktischem Wissen gezählt werden sollen und welche Komponenten sich als wirksam hinsichtlich eines guten, das heißt, eines lernförderlichen und motivierenden Physikunterrichts erweisen.

Allerdings sollten im Sinne des breit formulierten Kompetenzbegriffs nach Weinert (2001) und der darauf aufbauenden Modelle der Handlungskompetenz motivationale Aspekte nicht vernachlässigt werden. Hier setzen wir mit unserem Interesse für die Selbstwirksamkeitserwartungen an. Die domänenspezifischen SWE, so die Hypothese, die auf den folgenden Ausführungen zum Forschungsstand beruht, stellen einen Faktor dar, der sich potentiell dahingehend auswirkt, ob mehr oder weniger kompetente Handlungsversuche von (zukünftigen) Physiklehrkräften überhaupt unternommen werden. Insbesondere im Rahmen der Ausbildung von Physiklehrkräften können

SWE, vermittelt über die Handlungsversuche, sogar beeinflussen, welche (kognitiven) Kompetenzen auf welchem Niveau erworben werden. Das im Folgenden vorgestellte Projekt versteht sich damit als Beitrag zur Grundlagenforschung im Bereich der Professionalisierung (zukünftiger) Physiklehrkräfte.

Nachfolgend soll das bisher nur grob umrissene Forschungsinteresse theoretisch fundiert und vertiefend begründet werden. Damit steht im Zentrum des Artikels einerseits die Klärung der Frage, was unter SWE in physikdidaktischen Handlungsfeldern überhaupt zu verstehen ist bzw. verstanden werden soll. Dazu werden zunächst bisherige Forschungsergebnisse und theoretische Vorarbeiten referiert (Abschnitte 2 und 3). Andererseits werden auf Grundlage dieser Befunde geeignete Items / Skalen zur Messung physikdidaktischer SWE konstruiert und die Ergebnisse einer ersten Validierungsstudie vorgestellt (Abschnitt 4).

2 Selbstwirksamkeitserwartungen – eine Begriffsklärung

Das Konstrukt der Selbstwirksamkeitserwartung stammt aus der sozialkognitiven Theorie nach Bandura (2001), die auf der Vorstellung basiert, dass Merkmale einer Person, ihr Verhalten und die Rahmenbedingungen bzw. Umwelt interagieren und sich in Abhängigkeit voneinander konstituieren. Das Handeln und die Motivation einer Person hängen demnach stark davon ab, welche Wahrnehmung und Vorstellung diese Person von sich selbst hat.

2.1 Arbeitsdefinition und Relevanz von SWE

Bandura umschreibt SWE in erster Näherung wie folgt: „[p]erceived self-efficacy refers to beliefs in one`s capabilities to organize and execute the courses of action required to produce given attainments“ (Bandura, 1997, 3). Aus dieser Definition lassen sich zwei wichtige Eigenschaften des SWE-Konstruktes entnehmen: Erstens beziehen sich SWE immer auf Kompetenzen im Sinne einer Disposition, bestimmte Handlungen erfolgreich zu vollziehen, wobei diese Kompetenzen kontext- und domänenspezifisch sind. Zweitens handelt es sich um eine Selbsteinschätzung einer Person darüber, ob und wie stark eine bestimmte Kompetenz bei ihr ausgeprägt ist. Damit entspringen SWE subjektiven Wahrnehmungen und Einschätzungen und machen keine Aussage über tatsächlich vorhandene Fähigkeiten. Sie werden deshalb synonym auch als Kompetenzerwartungen (Schwarzer & Jerusalem, 2002) bezeichnet.

Für den deutschsprachigen Raum hat sich der Begriff der SWE in Anlehnung an Schmitz & Schwarzer (2000) dahingehend ausgeschärft, dass von Handlungen ausgegangen wird, die mit einer Schwierigkeit verbunden sind bzw. von Zielen, die die Überwindung einer Barriere verlangen: „Unter Selbstwirksamkeit versteht man die subjektive Gewißheit [sic], eine neue oder schwierige Aufgabe auch dann erfolgreich bearbeiten zu können, wenn sich Widerstände in den Weg stellen“ (Schmitz & Schwarzer, 2000, 13). Diese Definition liegt unseren Ausführungen und auch der

Skalenentwicklung zugrunde (vgl. Abschnitt 4.1).

Von Bedeutung sind Selbstwirksamkeitserwartungen, weil sie sich regulierend darauf auswirken, ob Handlungen, die mit einem bestimmten Schwierigkeitsgrad verbunden sind, überhaupt aufgenommen werden und wie viel Anstrengung investiert und Ausdauer an den Tag gelegt wird, um das Handlungsziel zu erreichen (vgl. Schmitz & Schwarzer, 2000, 13). Vermittelt über diese Variablen können – so die plausible theoretische Annahme – SWE schließlich auch den Handlungserfolg selbst beeinflussen. Voraussetzung dafür, von einem Gefühl der Selbstwirksamkeit zu sprechen, ist, dass der Erfolg der Handlung auf die eigenen Fähigkeiten und das realisierte Handeln zurückgeführt wird und nicht auf äußere Einflüsse oder den Zufall (vgl. Schmitz & Schwarzer, 2000, 13).

Überträgt man dieses Modell auf Physiklehrkräfte, so kommt man zu dem Schluss, dass die Realisierung von (kompetentem) Handeln eben nicht nur von dem bloßen Vorhandensein der Wissenskomponente der jeweiligen Kompetenz, sondern unter anderem auch von der damit verbundenen SWE abhängig ist.

2.2 Selbstwirksamkeitserwartungen, Handlungsergebniserwartungen und Selbstkonzept

Sinnvoll ist es, die SWE gegenüber zwei benachbarten motivationalen Konstrukten abzugrenzen, die mit ihnen häufig in einem Atemzuge genannt werden. Zu-

nächst soll eine Abgrenzung gegenüber dem Konstrukt der Handlungsergebniserwartungen vorgenommen werden. Handlungsergebniserwartungen (engl. „outcome expectation“, vgl. Bandura, 1977) beschreiben die Erwartung oder Vorstellung einer Person, welche Handlungen bestimmte Handlungsergebnisse herbeiführen können – und zwar unabhängig von der Einschätzung der eigenen Kompetenzen. Ein Beispiel für eine solche Handlungsergebniserwartung ist die Überzeugung, dass selbständiges Experimentieren von SuS nicht durch offene Experimentierangebote realisiert werden kann. Diese Überzeugung ist offensichtlich unabhängig von der Einschätzung der eigenen Fähigkeit, offenes Experimentieren begleiten zu können, also den SWE in diesem Bereich. Bei der Erfassung von Handlungsergebniserwartungen drückt sich der postulierte Zusammenhang zwischen einem Verhalten und dem Ziel meist in einer „Wenn-Dann“-Formulierung aus. Als Konstrukte sind die SWE und Handlungsergebniserwartungen auf theoretischer Ebene klar voneinander abzugrenzen, empirisch zeigt sich ein mittelstarker Zusammenhang (vgl. Schmitz & Schwarzer, 2000, 14).

Unter dem zweiten hier gegen die SWE abzugrenzenden Konstrukt, dem Selbstkonzept, versteht man eine mentale Repräsentation des Selbst, die die eigenen Erfahrungen mit Kohärenz und Sinn erfüllt, einschließlich der sozialen Beziehungen zu anderen Menschen (vgl. Pajares, 1996; Moschner, 2001; Simon & Trötschel, 2007). Damit handelt es sich bei dem Selbstkonzept um eine globalere Ein-

schätzung zu dem, wer man ist, was man kann und unter Umständen auch, wie man diese Eigenschaften bewertet. Das Selbstkonzept wird auf einem allgemeineren Niveau gemessen als die SWE. Es gilt dabei zwar auch als domänen-, aber nicht als aufgabenspezifisch: „Self-concept judgments can be domain specific but are not task specific. Compared to self efficacy judgments, they are more global and less context dependent.“ (Pajares, 1996, 561) Dieser Unterschied zwischen den genannten Konstrukten auf theoretischer Ebene führt zu entsprechenden Operationalisierungen und muss auch auf der Item-Ebene deutlich erkennbar bleiben (vgl. Abschnitt 4.1).

Von den allgemeinen SWE, die bei einer hohen Ausprägung „eine optimistische Einschätzung der generellen Lebensbewältigungskompetenz zum Ausdruck bringen“ (Schwarzer & Jerusalem, 2002, 40), können situationsspezifische und bereichs- bzw. domänenspezifische SWE unterschieden werden (Roberts, Henson, Tharp & Moreno, 2001). Dass SWE demnach nicht nur handlungsbezogen, sondern auch spezifisch und unter Umständen unterschiedlich für bestimmte (professionelle) Handlungsbereiche bzw. Domänen sind, hat Konsequenzen für die Messung des Konstrukts. Je nach Fragestellung muss ein angemessenes Spezifitätsniveau für die Erhebung von domänenspezifischen SWE ausgewählt werden und sich in dem Messinstrument niederschlagen: „Because judgements of self-efficacy are task and domain specific, global or inappropriately defined self-efficacy assessments weaken effects. Consequently, [...]“

to increase accuracy of prediction, [researchers] would be well advised to follow theoretical guidelines regarding *specificity* of self-efficacy assessment and *correspondence* with criterial tasks.” (Pajares, 1996, 547) Dementsprechend gibt es bereits zahlreiche Skalen zur allgemeinen Lehrer-selbstwirksamkeit und einzelne Skalen zur SWE bezogen auf naturwissenschaftliches Lehren (vgl. Abschnitt 3). Angesichts der Differenziertheit, mit der physikdidaktische Kompetenzen von (zukünftigen) Lehrkräften erhoben werden, ist es naheliegend, auch SWE mit einem äquivalenten Spezifitätsniveau zu untersuchen.

3 Lehrer-selbstwirksamkeitserwartung

Unter der Lehrer-Selbstwirksamkeitserwartung (Lehrer-SWE) versteht man “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, 233) Sie bezieht sich auf typische Handlungsfelder des Lehrerberufs, wie zum Beispiel die allgemeine berufliche Leistung, die berufsbezogenen sozialen Interaktionen, innovatives Handeln und den Umgang mit Berufsstress und Emotionen (vgl. Schmitz & Schwarzer, 2002, 194). Einen Überblick zu existierenden Instrumenten zur Erfassung von Lehrer-SWE findet man bei Tschannen-Moran, Hoy & Hoy (1998) sowie bei Schmitz & Schwarzer (2000) in Form eines historischen Rückblicks. Letztere haben das gängige deutschsprachige

Instrument entwickelt, mit dem sie zeigen konnten, dass ihre spezifische Skala zur Lehrer-SWE in einem engeren Zusammenhang mit anderen für den Lehrerberuf relevanten Konstrukten steht als die Allgemeine SWE und demnach prädiktiv überlegen ist (Schmitz & Schwarzer, 2000, 19). Dieses Ergebnis lässt darauf hoffen, dass z. B. physikdidaktische Skalen entwickelt werden können, die in bestimmten Untersuchungskontexten wiederum der Lehrer-SWE überlegen sind.

Als Quellen für die Ausprägung von Lehrer-SWE identifizieren Tschannen-Moran et al. (1998) in ihrem verschiedene theoretische Positionen und empirische Forschungsergebnisse integrierenden Modell der Lehrer-SWE mehrere Faktoren (vgl. Abbildung 1). Dazu gehören „verbale Beeinflussung“ (Feedback), „stellvertretende oder beobachtend gewonnene Erfahrungen“ (Vorbilder/peers), „physiologische und emotionale Reaktionen“ und „eigene prägende Handlungserfahrungen“ (Unterrichtsdurchführung), wobei letzteren der größte Einfluss zugesprochen wird (vgl. auch Bandura, 1997, 79ff.). Die Lehrer-SWE werden nach diesem Modell in einem zyklischen Prozess immer wieder der Belastung durch reale Lehrerfahrungen ausgesetzt, wobei diese Erfahrungen bereits durch die vorhandenen SWE vorgeprägt sind und entsprechend unterschiedlich angesteuert werden. Je nachdem, wie die neuen Erfahrungen ausfallen, bestätigen sie die bisherigen Lehrer-SWE oder werden zu neuen Quellen, aus denen die Lehrkraft veränderte Lehrer-SWE konstruieren kann. Deutlich wird dabei die Möglichkeit, dass sich SWE in diesem zyklischen

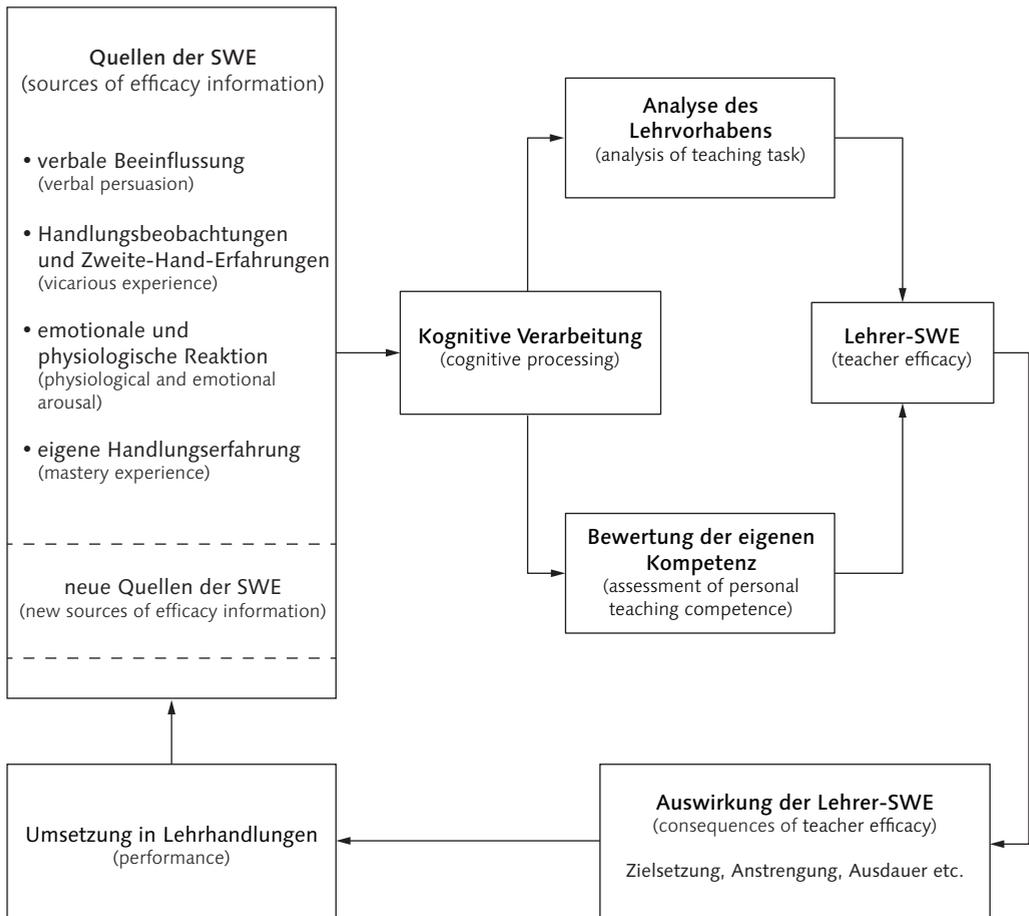


Abb. 1: Modell der zyklischen Entwicklung von Lehrer-SWE nach Tschannen-Moran et al. 1998, S. 228 ff.

Prozess im Sinne einer sich selbst erfüllenden Prophezeiung letztlich selbst bestätigen und verstärken. Für die Ausbildung zukünftiger Lehrkräfte kann es insofern sinnvoll sein, bereits die Studierenden für diesen Zusammenhang zu sensibilisieren, um ihnen Auswege aus dem Zirkel zu ermöglichen, falls es sich um negative oder unrealistische Selbstwirksamkeitserwartungen handelt.

3.1 Forschungsstand zur Lehrer-selbstwirksamkeitserwartung

In ihrem Review zur (allgemeinen) Lehrer-SWE verweisen Tschannen-Moran et al. (1998, 213ff.) auf Forschungsergebnisse, die zeigen, dass Lehrkräfte mit höheren Lehrer-SWE größeren Enthusiasmus für ihren Unterricht aufbringen, mit höherer Wahrscheinlichkeit im Beruf verbleiben

und sich stärker an die Unterrichtstätigkeit gebunden fühlen. Weiterhin beeinflussen die Lehrer-SWE in erwünschter Richtung die Unterrichtsvorbereitung und -durchführung, und hier insbesondere das konstruktive Unterstützungsverhalten. Schließlich werden auch Zusammenhänge mit dem Wissenserwerb von Schülerinnen und Schülern, deren Motivation und deren Selbstwirksamkeitserwartung berichtet.

Für den deutschsprachigen Raum können Schwarzer & Schmitz (2000) ausgeprägtere Lehrer-SWE als Resilienzfaktor bei der Bewältigung von Berufsstress identifizieren und positive Zusammenhänge zum beruflichen Engagement sowie zur Berufsverbleibswahrscheinlichkeit beobachten. Interessant sind auch die Annahmen zur Entwicklung von Lehrer-SWE im Laufe der Ausbildung, weil sie auf den ersten Blick widersprüchlich wirken. Einerseits ist aufgrund des zyklischen Entwicklungsmodells (vgl. Abbildung 1) zu erwarten, dass sich Lehrer-SWE mit zunehmender Berufserfahrung auf einem stabilen Niveau konsolidieren, weil angenommen werden kann, dass sich mit wachsender Erfahrung auch eine zunehmend realistische Sichtweise einstellt. Dies hat zur Folge, dass sich erst einmal ausgeprägte Lehrer-SWE, als änderungsresistent erweisen (Tschannen-Moran et al., 1998, 235). Auf der anderen Seite ist zu beobachten, dass Lehrer-SWE in Praxisphasen während der universitären Ausbildung, in denen also erste eigene Lehrerfahrungen gesammelt werden, in ihrer Ausprägung zurückgehen (Tschannen-Moran et al., 1998, 232). Die zukünftigen Lehrkräfte

erleiden in diesem Sinne einen „Realitätsschock“, indem sie mit ihren tatsächlichen Handlungskompetenzen unter Realbedingungen konfrontiert werden. Man könnte auch sagen, sie werden bei zu optimistischen Selbsteinschätzungen durch die Praxis „geerdet“. Vorteilhaft scheint es zu sein, in der Lehrerausbildung nicht nach dem Prinzip „sink or swim“ vorzugehen, sondern mit überschaubaren Unterrichtssituationen zu beginnen und die Komplexität langsam zu steigern (Tschannen-Moran et al., 1998, 236).

Insgesamt, so konstatieren Baumert & Kunter (2006, 503), ist über die Entwicklung von SWE in den verschiedenen Phasen der Lehreraus- und -weiterbildung noch wenig bekannt und die Forschungslage zum Zusammenhang von Selbstwirksamkeitsüberzeugen von Lehrkräften und ihrem Unterrichtsverhalten erscheint noch nicht zufriedenstellend.

3.2 Forschungsstand zu physikdidaktischen Selbstwirksamkeitserwartungen

Auf dem nächsten Spezifitätsniveau existieren zwei Instrumente für Lehrer-SWE im Bereich der Naturwissenschafts- bzw. Physikdidaktik, auf die im folgenden Abschnitt näher eingegangen wird. Die Skalen von Riggs & Enochs (1990) fokussieren auf den Bereich des naturwissenschaftlichen Lehrens an Grundschulen und sind damit nicht spezifisch auf Physik ausgelegt. Riese (2009) adaptiert diese Skalen für den Bereich der Physik.

Die Forschungsbefunde zu SWE als Be-

standteil physikdidaktischer Kompetenzen sind damit vergleichsweise dünn gesät. Aus Studien, die mit dem Instrument STEBI (Science Teaching Efficacy Belief Instrument; Riggs & Enochs, 1990) durchgeführt wurden, deutet sich an, dass höhere SWE im Bereich naturwissenschaftlicher Handlungsfelder mit der in die Unterrichtsplanung investierten Zeit korrelieren und auch die Qualität des Lehrens dahingehend positiv beeinflussen, dass innovative und herausfordernde Methoden eingesetzt werden (Tschanen-Moran et al., 1998, 216; Roberts et al., 2001). So zeigt sich beispielsweise im Rahmen eines Lehrerfortbildungsprogramms (Mathematics and Science Partnership, MSP), das über drei Jahre stattfindet, ein positiver Zusammenhang zwischen den Veränderungen der Lehrer-SWE der teilnehmenden Lehrkräfte und einer Orientierung ihrer Unterrichtspraxis hin zu forschend-entdeckendem Lernen (Lakshmanan, Heath, Perlmutter & Elder, 2011). In seiner Studie zur professionellen Handlungskompetenz von (angehenden) Physiklehrkräften findet Riese (Riese, 2009; Riese & Reinhold, 2010) keine Korrelation zwischen der allgemeinen SWE und dem physikalischen Fachwissen, dem fachdidaktischen oder dem pädagogischen Wissen, was die These erhärtet, dass relevante Kompetenzbereiche in den Blick genommen werden sollten. Allerdings korreliert auch die „SWE speziell bezogen auf das Physik-Lehren“, die in Anlehnung an die STEBI-B-Skala in der Fassung nach Bleicher (2004) erhoben wird, lediglich mit dem Fachwissen in Physik ($r = .35$, $p < .001$), nicht jedoch mit dem fachdi-

daktischen Wissen, wie zu erwarten wäre: „Offenbar stellt die Selbstwirksamkeitserwartung für die betrachtete Population doch eher ein fachbezogenes Konstrukt dar, wenn es um den Zusammenhang mit der gemessenen Leistung als Indiz für den Kompetenzerwerb innerhalb der Universität geht.“ (Riese, 2009, 157–158). Plausibler wird dieses Ergebnis, wenn man beachtet, dass in dieser Studie das fachdidaktische Wissen speziell bezogen auf das Experimentieren erhoben wurde. In der Skala zu den physikdidaktischen SWE gibt es hingegen nur zwei Items, die sich auf den Umgang mit Experimenten im Physikunterricht beziehen. Naheliegender ist also wiederum die Vermutung, dass die SWE nicht auf gleichem Spezifitätsniveau gemessen wurden wie das fachdidaktische Wissen.

3.3 Forschungsdesiderata

Man könnte angesichts der bisherigen Forschungsergebnisse behaupten, dass die SWE bereits eine Erfolgsgeschichte aufweisen – Schwarzer & Jerusalem (2002, 36) verweisen auf über 500 wissenschaftliche Veröffentlichungen zu diesem Konstrukt. Erfasst werden (Lehrer-) SWE typischerweise mit Fragebögen, die auf Selbsteinschätzungen beruhen. Eine Vielzahl an Instrumenten ist dabei entwickelt und erfolgreich eingesetzt worden. Dennoch ergeben sich aus den bisherigen Ausführungen einige Forschungsanlässe, die im Folgenden dargelegt werden sollen. Zunächst scheint es sinnvoll, der Frage der angemessenen Spezifität für die Messung

von Lehrer-SWE nachzugehen. „Teachers’ sense of instructional efficacy is not necessarily uniform across different subjects [...] Therefore teacher efficacy scales should be linked to the various knowledge domains.” (Bandura, 1997, 242) Somit ist es auch nicht überraschend, dass – wie bereits mehrmals angedeutet wurde – eine Erhebung von SWE auf einem globalen Niveau bei der Aufklärung von domänen- oder aufgabenspezifischem Handeln nicht oder nur begrenzt hilfreich ist (Schmitz & Schwarzer, 2000): „self-efficacy beliefs should be assessed at the optimal level of specificity that corresponds to the criterial task being assessed and the domain of functioning being analyzed.“ (Pajares, 1996, 547) Insofern ist es zugunsten einer höheren Prädiktivität sinnvoll, Skalen zu physikdidaktischen SWE zu entwickeln, die sich auf relevante und für den Physikunterricht typische Handlungsfelder bzw. die für eine Physiklehrkraft spezifischen Kompetenzen beziehen, denn „when self-efficacy beliefs closely correspond to the criterial task with which they are compared, prediction is enhanced.“ (Pajares, 1996, 555). Insofern hängt es von der jeweiligen Fragestellung ab, auf welchem Spezifitätsniveau SWE erfasst werden sollten, vorausgesetzt ein passendes Instrument liegt vor. Pajares (1996) formuliert dieses Problem im Sinne eines Abwägens zwischen „omnibus measures“ und einem „atomistic level“ bei der Erhebung. Den STEBI könnte man in diesem Zusammenhang als ein „domain-specific omnibus measure“ bezeichnen, denn es werden kaum fachdidaktische Kompetenzen im engeren Sinne angesprochen. Vielmehr

handelt es sich bei den in den Items beschriebenen Situationen eher um solche, die eine Beurteilung des eigenen (naturwissenschaftlichen) Fachwissens erfordern (z. B. „I understand science concepts well enough to be effective in teaching elementary science“ (Enochs & Riggs, 1990, 703). Selbst wenn man in den Items das Wort „science“ durch „Physik“ ersetzt wie in der Adaption von Riese (2009), wird aus der Skala noch kein Instrument, das auf (verschiedene) fachdidaktische Handlungsbereiche fokussiert.

Ein zweiter Problembereich liegt darin begründet, dass in den Itemformulierungen nur begrenzt Handlungsbarrieren sichtbar werden, die – schließt man sich der Definition von SWE nach Schmitz & Schwarzer (2000) an – konstitutives Element der jeweiligen SWE sind (vgl. Abschnitt 2 und 4.1)

Erst wenn ein valides Instrument zu SWE in physikdidaktischen Handlungsfeldern vorliegt, werden Untersuchungen möglich, die dem moderierenden Einfluss von SWE auf reale Handlungen nachgehen. Sobald also Untersuchungen darauf zielen, Zusammenhänge zwischen fachdidaktischem Wissen und Können und SWE zu beobachten oder SWE als Kontrollvariable genutzt werden sollen, wird ein Instrument benötigt, das die SWE passend zu dem fokussierten Handlungsbereich erhebt. Die Entwicklung von Skalen zu SWE in typischen physikdidaktischen Handlungsfeldern ist also ein erster wichtiger Schritt.

Als nicht vollständig geklärt gilt auch, wie sich Lehrer-SWE langfristig und unter dem Einfluss von Praxiserfahrungen entwickeln. Noch weniger weiß man ent-

sprechend über die Entwicklung von physikdidaktischen SWE der (zukünftigen) Physiklehrkräfte im Verlauf des Hochschulstudiums, des Referendariats und der Berufseingangsphase. Dem nachzugehen könnte – neben der Betrachtung der betreffenden Kompetenzentwicklung selbst – Hinweise auf die Wirksamkeit der (physikdidaktischen) (Aus-)Bildung geben. Ein weiterer, bisher nur am Rande erwähnter Aspekt weist in die Richtung, dass die Bewusstmachung und Auseinandersetzung der (zukünftigen) Lehrkräfte mit ihren eigenen SWE sich positiv auf ihre Lernprozesse und Handlungen auswirken kann.

4 Pilotstudie zur Entwicklung der SWE-Skalen zu physikdidaktischen Kompetenzbereichen

Im Folgenden soll eine Pilotstudie vorgestellt werden, deren Ziel vorrangig darin besteht, zu überprüfen, ob es langfristig möglich und sinnvoll erscheint, ein valides Instrument zu SWE in physikdidaktischen Kompetenzbereichen zu entwickeln. Am Ende dieser ersten Pilotierung werden also keine fertigen, ausreichend erprobten Skalen stehen, sondern die Entscheidung darüber, ob es sich aus Forschungsperspektive lohnt, einer solchen Konstruktion weiter nachzugehen.

Dazu wird zunächst offengelegt, nach welchem Konstruktionsmuster Items für ausgewählte physikdidaktische Kompetenzfelder erstellt und zu vorläufigen Skalen zusammengefügt werden, um dann in

einer Pilotierung mit Studierenden des Lehramts Physik eingesetzt zu werden. Berichtet werden die Ergebnisse von konfirmatorischen Faktorenanalysen als Voraussetzung für eine erste Überprüfung der Konstruktvalidität. Am Beispiel von zwei ausgewählten Skalen werden diese Ergebnisse und die Überarbeitung der Skalen im Detail erläutert. Schließlich werden erste Schritte zur Validierung der Skalen durch korrelations- und varianzanalytische Verfahren dargestellt.

4.1 Skalenkonstruktion

Die Konstruktion von Skalen zu physikdidaktischen SWE erfolgte in zwei Dimensionen. Nach Baumert & Kunter (2006) sind Vorbereitung, Inszenierung und Durchführung von Unterricht das Kerngeschäft von Lehrpersonen. Es wurde also in den ausgewählten physikdidaktischen Kompetenzbereichen jeweils unterschieden, ob sich die SWE auf die Planung oder die Durchführung von Physikunterricht beziehen. Diese Trennung ist sinnvoll, da sowohl zukünftige als auch bereits im Beruf tätige Lehrkräfte mit dem Instrumentarium befragt werden sollen, die jeweils unterschiedlich viel Erfahrung mit diesen Tätigkeiten besitzen. So ist es denkbar, dass Studierende bereits ausgeprägte SWE im Bereich der Planung haben, sich aber bei der Durchführung von Unterricht aufgrund der fehlenden Praxis weniger zutrauen.

Die Auswahl geeigneter physikdidaktischer Kompetenzbereiche gestaltet sich angesichts der heterogenen Forschungs-

landschaft schwierig, da es keinen Konsens darüber gibt, welche Fähigkeiten konstituierende Elemente physikdidaktischer Kompetenz sind und als notwendiger Bestandteil physikdidaktischer Ausbildung von Lehrkräften gelten. Auf der Grundlage aktueller Forschungsprojekte wie dem der Paderborner Gruppe (Riese, 2009), dem QUIP-Projekt (Olszewski et al., 2009), aber auch dem fachlich benachbarten COACTIV Projekt (Krauss et al., 2008) und vor allem der DPG-Quereinsteiger-Studie (Korneck et al., 2010) wurde eine erste Auswahl von Kompetenzbereichen getroffen. Letztere kommt wohl dem gesuchten Konsens von Experten aus der Physikdidaktik darüber, welche Kompetenzen in der universitären Ausbildung erworben werden sollen, recht nahe. Die Bereiche des Experimentierens, des Umgangs mit Schülervorstellungen und der Aufbereitung und Elementarisierung von physikalischen Inhalten (Stichwort „Elementarisierung“ im Sinne von Hopf et al., 2011, 72 ff) wurden zunächst für die Operationalisierung gewählt. Im Bereich des Experimentierens geht es dabei um die adressaten- und zielpassende Auswahl und Gestaltung eines Experimentes sowie seine lernförderliche Inszenierung in einem nach Möglichkeit differenzierenden Unterricht. Hinsichtlich des Umgangs mit Schülervorstellungen wurden Selbsteinschätzungen dazu erbeten, inwiefern die Befragten Schülervorstellungen diagnostizieren und ihren Schülerinnen und Schülern bewusst machen können und über geeignete Strategien verfügen, um sie zum Ausgangspunkt von Lernprozessen zu machen. Im Bereich Aufbereitung

und Elementarisierung schließlich geht es um das Identifizieren von Kernideen, die begründete Schwerpunktsetzung und Strukturierung von Inhalten auch quer zur „linear-thematischen Fachsystematik“ sowie die Auswahl geeigneter Vereinfachungen, Analogien oder Modelle.

Theoriegeleitet entworfen wurden folglich jeweils zwei Skalen in jedem Kompetenzbereich, wobei, wie gesagt, eine Skala auf „Planungskompetenzen“ und eine weitere Skala auf „Durchführungskompetenzen“, die beim Unterrichten von Physik benötigt werden, abzielt. Insgesamt wurden also sechs Skalen (drei physikdidaktische Kompetenzbereiche, zwei Dimensionen des Lehrerhandelns) entworfen.

Die Konstruktionsvorschrift bei der Formulierung der Items orientiert sich an der oben angeführten Definition der SWE (vgl. Abschnitt 2.1) und damit an der deutschsprachigen Forschung in diesem Bereich (Schmitz & Schwarzer, 2000). Die Vorgaben bestehen darin, dass

1. die Items in der 1. Person Singular zu verfassen sind, um sie gegenüber allgemeinen Handlungsergebniserwartungen abzugrenzen und um sicherzustellen, dass es sich um Selbsteinschätzungen handelt,
2. die gewählten Verbalphrasen Kompetenzen adressieren sollten (z. B. ich kann..., ich bin in der Lage...),
3. der Kompetenzbereich deutlich werden muss, wobei nur solche Handlungen auszuwählen sind, für die Anstrengung oder Ausdauer notwendig ist,
4. ein Schwierigkeitsgrad bzw. ein Widerstand in Form von Handlungsbarrieren

Tab. 1: Überblick über die konstruierten und zusätzlich verwendeten Skalen (Quellenangabe im Text). Items der Skalen konnten – sofern nicht anders vermerkt – mit „stimmt nicht“, „stimmt kaum“, „stimmt eher“ oder „stimmt genau“ bewertet werden.

	Skala	Abkürzung	Beispielitem
konstruierte Skalen	SWE bzgl. der Planung von Experimenten	SWE-Ex-P	Auch Zeitdruck während der Unterrichtsvorbereitung hindert mich nicht daran, ein zu meinen Unterrichtszielen passendes Experiment zu entwickeln. (<i>exp5</i>)
	SWE bzgl. der Durchführung von Experimenten im Unterricht	SWE-Ex-D	Immer wenn der Unterrichtsverlauf es sinnvoll erscheinen lässt, gelingt es mir, ein Experiment angemessen zu variieren, auch wenn ich das vorher nicht geplant habe. (<i>exd1</i>)
	SWE bzgl. der Elementarisierung unterrichtsrelevanter Inhalte	SWE-EI-P	Ich schaffe es, auch komplexe Themen der modernen Physik für eine Unterrichtseinheit so zu vereinfachen, dass meine Schülerinnen und Schüler sie verstehen. (<i>elp3</i>)
	SWE bzgl. der Umsetzung der Elementarisierung im Unterricht	SWE-EI-D	Es gelingt mir, die zentralen physikalischen Inhalte einer Physikstunde herauszustellen, auch wenn ich spontan für einen Kollegen einspringen muss. (<i>eld1</i>)
	SWE bzgl. der Berücksichtigung von Schülervorstellungen bei der Unterrichtsplanung	SWE-Sv-P	Ich kann die Vorstellungen meiner Schülerinnen und Schüler in meine Unterrichtsplanung einbeziehen, auch wenn sie sehr vielfältig sind. (<i>svp1</i>)
	SWE bzgl. des Umgangs mit Schülervorstellungen bei der Unterrichtsdurchführung	SWE-Sv-D	Ich kann geäußerte Schülervorstellungen noch in derselben Unterrichtsstunde für Lernprozesse nutzbar machen, auch wenn ich darauf nicht vorbereitet war. (<i>svd2</i>)
zusätzliche Skalen	Physikalisches Selbstkonzept	PS	Ich verstehe den Stoff der Physikvorlesungen (sehr gut/gut/mittel/schlecht/sehr schlecht). (<i>ps1</i>)
	Allgemeine SWE	A-SWE	Wenn sich Widerstände auftun, finde ich Mittel und Wege, mich durchzusetzen. (<i>aswe1</i>)
	Lehrer-SWE	L-SWE	Ich weiß, dass ich es schaffe, selbst den problematischsten Schülerinnen und Schülern den Stoff zu vermitteln. (<i>lswe1</i>)

oder (fehlenden) Ressourcen vorkommen muss (z. B. erkennbar an Formulierungen wie „auch wenn“, „obwohl“, „nur wenn“).

Beispiele für die so konstruierten Items sind Tabelle 1 zu entnehmen. Der erste Fragebogenentwurf zu den physikdidaktischen SWE umfasste die sechs Skalen (vgl. Tabelle 1) bestehend aus jeweils sieben Items, die auf einer vierstufigen Likertskala (stimmt nicht / kaum / eher / genau) zu bewerten waren.

4.2 Stichprobe und Gegenstand der Pilotierung

Die auf theoretischer Basis entworfenen Skalen wurden in einem ersten Durchlauf am Ende des Wintersemesters 10/11 an der Universität Potsdam pilotiert. Befragt wurden insgesamt 84, davon 29 weibliche, Studierende des Lehramts mit dem Fach Physik, die mindestens die erste Lehrveranstaltung im Bereich der Physikdidaktik durchlaufen hatten. Der überwiegende Anteil der Studierenden ist zum Zeitpunkt der Erhebung in das Lehramt für Gymnasien eingeschrieben, nur acht studieren das Lehramt für die Sekundarstufe I/ Primarstufe. Von den Studierenden streben 59 einen Bachelor, 22 einen Master und 3 das erste Staatsexamen als nächsten Abschluss an. Eine Mehrheit (52) studiert Physik in Kombination mit Mathematik, nur 24 aller Studierenden haben sich für Physik als erstes Fach entschieden. Der Gesamtumfang der physikdidaktischen Lehrveranstaltungen ist zu diesem Zeitpunkt in allen an der Universität Potsdam vorhandenen

Physik-Lehramtsstudiengängen gleich. Es hängt damit nur von dem individuellen Studienverlauf ab, welche Anteile bereits belegt wurden.

Neben Fragen zu den allgemeinen Angaben zur Person, zum Studium und zur Lehrerfahrung wurden weitere Skalen mit dem Ziel der Konstruktvalidierung eingesetzt (vgl. Tabelle 1), und zwar:

- Skala zur Lehrer-SWE (Schmitz & Schwarzer, 2000a) (L-SWE, 10 Items),
- Allgemeine SWE (nach Schwarzer & Jerusalem, 1999) (A-SWE, 10 Items),
- Fachspezifisches Selbstkonzept bzgl. Physik im Studium (adaptiert nach Hoffmann et al., 1998) (PS, 7 Items).

Erwartet wurden kleine bis mittlere Korrelationen zwischen den adaptierten Skalen und den Skalen im Bereich der physikdidaktischen SWE. Derartige Korrelationen wären ein Hinweis darauf, dass es sich bei den erhobenen Skalen um ähnliche, aber trotzdem unterschiedliche Konstrukte handelt.

4.3 Konfirmatorische Faktorenanalyse

Zur Prüfung der Skalen wurden konfirmatorische Faktorenanalysen (KFA) unter Verwendung des Maximum Likelihood-Verfahrens in AMOS 18.0 durchgeführt. Als globale Gütekriterien wurden neben dem χ^2 -Wert bei gegebenem Freiheitsgrad df des Modells und dem zugehörigen p -Wert insbesondere der Quotient χ^2/df , der Comparative Fit Index (CFI), der

Tucker-Lewis-Index (TLI) und der Root Mean Square Error of Approximation (RMSEA) verwendet. Folgende Bedingungen für die globalen Fit-Indizes wurden einer positiven Beurteilung zugrunde gelegt: $\chi^2/df < 2.0$, $CFI > 0.95$, $TLI > 0.95$, $RMSEA < 0.08$ (vgl. Bühner, 2011, 254 ff.). Zusätzlich wurde Cronbachs Alpha (α_c) als Reliabilitätsmaß verwendet. Für den Reliabilitätskoeffizienten wurde in Anlehnung an die Argumentation von Schermelleh-Engel und Werner (2008, 129) ein relativ niedriger Wert ($\alpha_c > 0.6$) als akzeptabel festgelegt, weil es sich erstens nicht um Leistungsvariablen handelt, die generell einfacher präzise zu messen sind als affektive Konstrukte und zweitens bei fehlender Alternative ein Instrument mit niedrigerem α_c sinnvoller ist als der vollständige Verzicht auf ein passendes Instrument. Faktorladungen über 0.6 werden bei der Erstpilotierung als gut betrachtet, Faktorladungen zwischen 0.4 und 0.6 als akzeptabel, Werte darunter gelten als zu niedrig. In Fällen, in denen die Indizes eine Ablehnung des Messmodells nahelegen, wurde

über die KFA insofern hinausgegangen, als dann anhand der Modifikationsindizes nach Verbesserungsmöglichkeiten der Skalen durch Ausschluss einzelner Items oder durch Parceling (bei ausreichend hoher Korrelation ($r > 0.65$) homogener Items (vgl. Bühner, 2006, 265)) gesucht wurde. Neben den statistischen Kennwerten wurden bei der Entscheidung über den Umgang mit einzelnen Items immer auch inhaltliche Aspekte ausschlaggebend in die Erwägungen einbezogen.

Fehlende Werte wurden mit der Methode „Zeitreihen-Mittelwert“ in SPSS 19 geschätzt, sofern maximal drei fehlende Werte in einer Skala vorlagen (weniger als 20 % der Fälle). Gab es mehr als drei fehlende Werte wurde der Fall von der jeweiligen Untersuchung ausgeschlossen. (Dies war genau einmal der Fall.) Die Ergebnisse der KFA für die auf theoretischer Basis entworfenen und ggf. angepassten Skalen (revidiertes Modell) sind in Tabelle 2 im Überblick dargestellt.

Interessant ist zunächst, dass von den übernommenen bzw. adaptierten Skalen nur

Tab. 2: Güte der verwendeten Skalen (ggf. revidiertes Modell); Cut-Off-Werte im Text.

Modell	Anzahl Items	χ^2	df	χ^2/df	p	CFI	TLI	RMSEA	α_c
PS	6	7.139	9	0.739	0.623	1.000	1.021	0.000	0.813
A-SWE	10	24.682	35	0.705	0.903	1.000	1.098	0.000	0.794
L-SWE	9	22.618	35	0.838	0.705	1.000	1.078	0.000	0.693
SWE-Ex-P	5	5.563	5	1.113	0.351	0.990	0.980	0.037	0.699
SWE-Ex-D	6	11.458	9	1.273	0.246	0.954	0.923	0.057	0.652
SWE-EI-P	6	11.053	9	1.228	0.272	0.971	0.952	0.052	0.707
SWE-EI-D	7	11.768	14	0.841	0.625	1.000	1.027	0.000	0.791
SWE-Sv-P	6	11.237	9	1.249	0.260	0.969	0.949	0.055	0.724
SWE-Sv-D	7	16.963	14	1.212	0.258	0.975	0.963	0.051	0.772

die Allgemeine SWE-Skala ohne Überarbeitung angenommen werden kann (wenn auch mit Faktorladungen, die zum Teil unter .50 liegen), während für das Physikalische Selbstkonzept und die Lehrer-SWE einzelne Items ausgeschlossen oder zusammengefasst werden müssen.

Die KFA führte im Ergebnis dazu, dass zwei der neu konstruierten Skalen unverändert angenommen wurden (El-D, Sv-D), bei denen der Modellfit schon beim ersten Durchlauf angemessen war. In den anderen Fällen (Ex-P, Ex-D, El-P, Sv-P) wurden einzelne Items aus der Skala ausgeschlossen oder zwei hoch korrelierende Items durch „parceling“ zusammengefasst, sofern inhaltlich plausible Gründe für das jeweilige Vorgehen vorlagen. In einigen Fällen führten zu geringe Faktorladungen dazu, nach Möglichkeiten zur Überarbeitung der Items in ihrer inhaltlichen Ausrichtung oder sprachlichen Formulierung zu suchen.

4.4 Korrelative Validierung der Skalen

Im Anschluss an die Überprüfung der Skalen mittels KFA als Voraussetzung für die Konstruktvalidierung können nun Zusammenhänge zwischen den Skalen im Sinne von Korrelationen analysiert werden. Datengrundlage sind die Skalenmittelwerte der 84 Probanden, die im Anschluss an die KFA für die angenommenen bzw. revidierten Skalen für jede Person berechnet wurden (vierstufige Skala; Minimum 0, Maximum 3). Da nicht für alle Skalen Normalverteilungen ange-

nommen werden können (dies wurde mit dem Shapiro-Wilk-Test für kleine Stichproben sowie dem Kolmogorov-Smirnov-Test überprüft), wurden neben dem Korrelationskoeffizienten nach Pearson auch Kendalls Tau B (τ) und Spearmans Rho (ρ) berechnet. Da sich die Ergebnisse im Wesentlichen decken, sind in Tabelle 3 nur die Ergebnisse für Kendalls Tau B dargestellt.

Ein Blick auf die Korrelationen zwischen den Skalen zeigt, dass keine oder nur sehr geringe Korrelationen zwischen dem erhobenen physikalischen Selbstkonzept und fast allen eingesetzten SWE-Skalen zu beobachten sind. Dieses Ergebnis unterstützt die theoretisch begründete Trennung der Konstrukte. Eine Ausnahme bilden die Skalen zur Elementarisierung, bei denen eine schwache, aber signifikante Korrelation zum physikalischen Selbstkonzept vorliegt. Inhaltlich lässt sich dies damit begründen, dass es in beiden Fällen um den Umgang mit physikalischen Inhalten bzw. um Fachwissen geht, so dass es erwartbar ist, dass ein hohes physikalisches Selbstkonzept mit einer positiven Kompetenzüberzeugung bzgl. der didaktisch-methodischen Aufbereitung der physikalischen Inhalte einhergeht.

Zwischen der Allgemeinen SWE und den pilotierten Skalen zur physikdidaktischen SWE sind schwache signifikante Korrelationen zu beobachten (vgl. Tabelle 3), die dafür sprechen, dass sich die erhobenen Konstrukte zwar ähneln, aber nicht das gleiche Konstrukt gemessen wurde. Die Zusammenhänge mit der Lehrer-SWE sind uneinheitlich und nicht einleuchtend interpretierbar, was damit zusammenhän-

Tab. 3: Korrelationskoeffizienten berechnet nach Kendalls Tau B, ** hoch signifikant mit $\alpha=0.01$. Werte zwischen 0.2 und 0.4/ 0.4 und 0.6/ 0.6 und 0.8 sprechen für eine schwache/ mittlere/ starke Korrelation.

Kendalls Tau B	PS	L-SWE	A-SWE	SWE-Ex-P	SWE-Ex-D	SWE-EI-P	SWE-EI-D	SWE-Sv-P	SWE-Sv-D
PS	1								
L-SWE	-0.016	1							
A-SWE	0.054	0.519**	1						
SWE-Ex-P	0.149	0.139	0.230**	1					
SWE-Ex-D	0.141	0.221**	0.255**	0.404**	1				
SWE-EI-P	0.244**	0.226**	0.277**	0.240**	0.372**	1			
SWE-EI-D	0.209**	0.142	0.240**	0.329**	0.376**	0.437**	1		
SWE-Sv-P	0.061	0.117	0.225**	0.431**	0.356**	0.270**	0.362**	1	
SWE-Sv-D	0.087	0.360**	0.397**	0.344**	0.479**	0.411**	0.433**	0.452**	1

gen kann, dass die verwendete Skala trotz ihrer Revision, wie oben erwähnt, in der Pilotierungsstichprobe nur unbefriedigende Skalenskennwerte zeigt.

Die zusammengehörigen Skalen zu Planungs- und Durchführungsaspekten in den einzelnen Kompetenzbereichen korrelieren in mittlerem Maße miteinander (vgl. Tabelle 3). Diese Ergebnisse unterstützen damit die theoretische Annahme, dass zwischen SWE, die sich auf Planungs- bzw. Durchführungshandlungen beziehen, zu unterscheiden ist.

Auffällig ist, dass die Allgemeine SWE deutlich stärker mit der Lehrer-SWE korreliert als mit den physikdidaktischen SWE-Skalen. Dies ist zu erwarten, da das Spezifikationsniveau zwischen Allgemeiner SWE und Lehrer-SWE ähnlicher ist als zwischen den physikdidaktischen SWE und der Allgemeinen SWE (vgl. die Ausführungen am Ende von Abschnitt 2.2).

Insgesamt ergibt sich im Sinne einer konvergenten bzw. diskriminanten Validitäts-

prüfung ein ermutigendes Bild, das die Validität der hier operationalisierten Konstrukte stützt.

4.5 Validierung durch erste Mittelwertvergleiche

Im Folgenden werden die Ergebnisse erster Mittelwertvergleiche referiert, die der weiteren Prüfung der Skalen auf äußere Kriteriumsvalidität dienen und explizit nicht der Beschreibung der Stichprobe. Da in der Literatur von einem Rückgang der SWE bei ersten Unterrichtserfahrungen berichtet wird (Hoy & Spero, 2005; Hoy & Woolfolk, 1990), wäre bei einer Eignung der Instrumente zu erwarten, dass sich dieser Rückgang in den Mittelwerten der physikdidaktischen SWE-Skalen abbildet. Für die Pilotierungsstichprobe bietet es sich an, die Studierenden danach zu unterscheiden, ob sie bereits die obligatorische Lehrveranstaltung „Schulpraktische

Tab. 4: Ergebnisse des Vergleichs der Ausprägungen von physikdidaktischen SWE Studierenden mit und ohne Praxiserfahrung. Skalenminimum: 0, Skalenmaximum: 3. $d = 0.2 / d = 0.5 / d = 0.8$ indiziert einen kleinen / mittleren / starken Effekt.

Konstrukt	Gruppe	Mittelwert	SD	p	Effektstärke d
SWE-Ex-P	gesamt	1.496	0.485		
	mit Praxis ohne Praxis	1.452 1.536	0.487 0.485	0.429	0.174
SWE-Ex-D	gesamt	1.721	0.391		
	mit Praxis ohne Praxis	1.600 1.832	0.369 0.381	0.006**	0.617
SWE-EI-P	gesamt	1.757	0.413		
	mit Praxis ohne Praxis	1.711 1.799	0.401 0.425	0.332	0.213
SWE-EI-D	gesamt	1.920	0.424		
	mit Praxis ohne Praxis	1.846 1.987	0.399 0.439	0.129	0.335
SWE-Sv-P	gesamt	1.605	0.433		
	mit Praxis ohne Praxis	1.471 1.727	0.335 0.478	0.006**	0.624
SWE-Sv-D	gesamt	1.679	0.418		
	mit Praxis ohne Praxis	1.556 1.792	0.372 0.430	0.009**	0.586

Übungen“ im Rahmen des Bachelorstudiums im Fach Physik belegt haben, in der sie erste, begleitete Lehrerfahrten im Physikunterricht sammeln. Da es sich bei den vorliegenden Daten um eine Querschnittserhebung handelt, sollte allerdings bei der Interpretation der Daten keine Entwicklung unterstellt werden.

Die Ergebnisse eines Mittelwertvergleichs für diese beiden Untergruppen – „mit Praxiserfahrung“ (40 Studierende), „ohne Praxiserfahrung“ (44 Studierende) – sind in Tabelle 4 wiedergegeben. Berechnet wurden t-Tests, deren Ergebnisse wegen der fehlenden Normalverteilung mit dem

Mann-Whitney-U-Test als nicht-parametrischem Verfahren bestätigt wurden.

Im Ergebnis zeigt sich, dass sich für Studierende mit Unterrichtserfahrung im Mittel niedrigere SWE auf den physikdidaktischen Skalen ergeben, wobei der Unterschied bei drei Skalen signifikant wird (SWE-Ex-D, SWE-Sv-P, SWE-Sv-D: $p < 0.006$, $d \geq 0.586$) und mittlere Effektstärken aufweist. Insofern lässt sich der in der Literatur berichtete „Praxischock“ von zukünftigen Lehrkräften mit Hilfe eines Teils der konstruierten Skalen replizieren, womit die Validität der eingesetzten Skalen weiter gestützt wird.

In einem weiteren Vergleich wurden die Mittelwerte auf der Planungs- und der Durchführungsskala eines physikdidaktischen Kompetenzbereichs verglichen (vgl. Tabelle 5)

Es ergeben sich durchgängig signifikante Unterschiede bei der Durchführung paarweiser Vergleiche, wobei die SWE in Bezug auf die Durchführung von Unterricht interessanterweise höher eingeschätzt werden (t -Test $p < 0.001$, $d \geq 0.201$). Eine vorläufige und nur mit Vorsicht zu behandelnde Interpretation könnte davon ausgehen, dass den Studierenden die Durchführung von Unterricht im Sinne einer stellvertretenden / passiven Erfahrung durch die eigenen Physiklehrkräfte vertraut ist, während die meisten Studierenden wahrscheinlich lediglich geringfügige Erfahrungen oder Einblicke in Planungs-handeln erhalten haben.

Diskussion und Ausblick

Der Ertrag der vorgestellten Studie setzt sich aus einer theoretischen Begründung

eines Forschungsinteresses im Bereich physikdidaktischer SWE einerseits und einem „Belastungstest“ einer ersten Skalenkonstruktion andererseits zusammen. Aus der Durchsicht und Zusammenschau von Literatur zu SWE sowie empirischen Forschungsprojekten zur Erhebung von SWE konnte die Relevanz des Konstrukts für die Erklärung von (Lehrer-)Handeln herausgearbeitet werden. Sowohl theoretische Annahmen als auch die Ergebnisse von Studien zu (Lehrer-)SWE weisen darauf hin, dass eine Beobachtung von SWE auf einem angemessenen Spezifitätsniveau entscheidend für das Maß an Prädiktivität ist. Das angemessene Maß zu finden ist eine erste fortbestehende Forschungsaufgabe: „perhaps the greatest challenge has to do with finding the appropriate level of specificity for measurement.“ (Tschannen-Moran et al., 1998, 219). Insofern kann ein Forschungsprojekt zu SWE bzgl. physikdidaktischer Handlungsbereiche theoretisch begründet und legitimiert werden.

Ein erstes Ergebnis der Pilotierungsstudie zu SWE im Bereich physikdidaktischer Handlungsfelder besteht darin, dass es

Tab. 5: Ergebnisse des Vergleichs der Ausprägungen von physikdidaktischen SWE Studierender auf den Konstruktebenen Planung bzw. Durchführung. Skalenminimum: 0, Skalenmaximum: 3. $d = 0.2$ / $d = 0.5$ / $d = 0.8$ indiziert einen kleinen / mittleren / starken Effekt.

Gruppe	Mittelwert	SD	p	Effektstärke d
SWE-Ex-P SWE-Ex-D	1.496 1.721	0.485 0.391	<0.001**	0.518
SWE-EI-P SWE-EI-D	1.757 1.920	0.413 0.424	<0.001**	0.452
SWE-Sv-P SWE-Sv-D	1.605 1.679	0.433 0.418	<0.001**	0.201

sich als möglich erweist, Skalen mit Hilfe der theoretisch fundierten Konstruktionsvorschriften zu entwickeln, die statistisch akzeptabel sind. Die Konstruktionsvorschriften haben sich damit als tragfähig für die Entwicklung weiterer Skalen erwiesen. Neben den hier vorgestellten Skalen sind deshalb weitere bereits in der Entwicklung begriffen, um in zusätzlichen physikdidaktischen Handlungsfeldern wie zum Beispiel dem Einsatz von Aufgaben oder dem Umgang mit der Mathematik im Physikunterricht die SWE erheben zu können.

Weiterhin deutet sich an, dass physikdidaktische SWE valide erfasst werden können, da die Ergebnisse der Pilotstudie dem theoretischen wie auch empirischen Forschungsstand zu Lehrer-SWE entsprechen. Während sich die SWE einerseits gegenüber einem benachbarten Konstrukt wie dem Selbstkonzept abgrenzen lassen, erweisen sich die Zusammenhänge zwischen den einzelnen SWE-Skalen als stark genug, um von einem gemeinsamen zugrundeliegenden Konstrukt ausgehen zu können. In diesem Sinne kann die Pilotstudie als Ausgangspunkt für die Konstruktion von SWE-Skalen in weiteren Kompetenzbereichen gewertet werden. Die identifizierten Skalen können jetzt in weiteren Pilot-Studien eingesetzt werden, die zum Beispiel andere Hochschulstandorte und damit andere Studienverläufe einbeziehen. Von Interesse ist auch, ob sich die Skalen in allen (Aus-)Bildungsphasen für das Lehramt Physik und bei im Beruf stehenden Lehrkräften bewähren können. Neben der rein quantitativ-statistischen Erprobung der Skalen soll

die Validierung in Zukunft auch qualitativ über Interviews mit Experten und Probanden vorgenommen werden.

Langfristig geplant sind Querschnittserhebungen an der Universität Potsdam, sowie weiteren Hochschulen in verschiedenen Phasen des Physiklehramtsstudiums, die zu einer Längsschnittstudie ausgeweitet werden. Eine solche Längsschnittstudie sollte die Entwicklung von physikdidaktischen SWE in der ersten sowie zweiten Ausbildungsphase (also Hochschulstudium und Vorbereitungsdienst) sowie die Berufseinstiegsphase in den Blick nehmen. Erst durch eine Längsschnittbetrachtung können dann Verläufe von SWE beschrieben und in Zusammenhang zum Beispiel zum (physikdidaktischen) Wissenserwerb gesetzt werden. Ein besonderer Fokus wird dabei auf die Einflüsse von Praxiserfahrungen auf SWE zu legen sein. Langfristig sind Zusammenhänge zum Handeln der Lehrerinnen und Lehrer und die Auswirkungen auf die Kognitionen und Emotionen der Schülerinnen und Schüler von Interesse. Insofern stellt das vorgestellte Projekt auf lange Sicht eine sinnvolle und grundlegende Ergänzung der Forschungsbemühungen zur Lehrerprofessionalisierung von Physiklehrkräften dar.

Literatur

- Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, 84(2), 191–215.
- Bandura, A. (1997). *Self-Efficacy. The Exercise of Control*. New York: W.H. Freeman and Company.
- Bandura, A. (2001). Social Cognitive Theory: An Agentic Perspective. *Annual Review of Psychology*, 52, 1–26.
- Baumert, J. & Kunter, M. (2006). Stichwort: Professionelle Kompetenz von Lehrkräften. *Zeitschrift für Erziehungswissenschaft*, 9(4), 469–520.
- Bleicher, R. E. (2004). Revisiting the STEBI-B : Measuring Self-Efficacy in Preservice Elementary Teachers. *School Science and Mathematics*, 104(8), 383–391.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. In F. E. Weinert (Ed.), *Psychologie des Unterrichts und der Schule* (pp. 177–212). Göttingen [u. a.]: Hogrefe.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Enochs, L. G. & Riggs, I. M. (1990). Further Development of an Elementary Science Teaching Efficacy Belief Instrument: A Preservice Elementary Scale. *School Science and Mathematics*, 90(8), 694–706.
- Gramzow, Y., Riese, J. & Reinhold, P. (2012). Fachdidaktisches Wissen von Lehramtsstudierenden und Lehrenden an der Hochschule im Fach Physik. In S. Bernholt (Ed.), *Konzepte fachdidaktischer Strukturierung für den Unterricht. Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Oldenburg 2011* (pp. 619–621). Berlin: LIT Verlag.
- Hoffmann, L., Häußler, P. & Lehrke, M. (1998). *Die IPN-Interessenstudie Physik*. Kiel: Institut für die Pädagogik der Naturwissenschaften.
- Hoy, A. W. & Spero, R. B. (2005). Changes in teacher efficacy during the early years of teaching: A comparison of four measures. *Teaching and Teacher Education*, 21(4), 343–356.
- Hoy, W. K. & Woolfolk, A. E. (1990). Socialization of Student Teachers. *American Educational Research Journal*, 27(2), 279–300.
- Korneck, F. et al. (2010). Quereinsteiger in das Lehramt Physik. Lage und Perspektiven der Physiklehrausbildung in Deutschland. Bad Honnef: DPG
- Krauss, S., Neubrand, M., Blum, W., Baumert, J., Brunner, M., Kunter, M. & Jordan, A. (2008). Die Untersuchung des professionellen Wissens deutscher Mathematik-Lehrerinnen und -Lehrer im Rahmen der COACTIV-Studie. *JMD*, 29(3/4), 223–258.
- Lakshmanan, A., Heath, B. P., Perlmutter, A. & Elder, M. (2011). The Impact of Science Content and Professional Learning Communities on Science Teaching Efficacy and Standards-Based Instruction. *Journal of Research in Science Teaching*, 48(5), 534–551.
- Schermelleh-Engel, K. & Werner, C (2008). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger, & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 113–133). Berlin, Heidelberg: Springer.
- Moschner, B. (2001). Selbstkonzept. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie (Vol. 2)* (pp. 629–635). Weinheim: Beltz.
- Oser, F. K. & Baeriswyl, F. J. (2001). Choreographies of Teaching: Bridging Instruction to Learning. In E. Richardson (Ed.), *Handbook of Research on Teaching. Fourth Edition* (pp. 1031–1065). Washington: American Educational Research Association.
- Pajares, F. (1996). Self-Efficacy Beliefs in Academic Settings. *Review of Educational Research* 66, 543–578.
- Park, S. & Oliver, J.S. (2008). Revisiting the Conceptualization of Pedagogical Content Knowledge (PCK): PCK as a Conceptual Tool to Understand Teachers as Professionals. *Research in Science Education*, 38, 261–284.
- Riese, J. (2009). *Professionelles Wissen und professionelle Handlungskompetenz von (angehenden) Physiklehrkräften*. Berlin: Logos Verlag.
- Riese, J. & Reinhold, P. (2010). Empirische Erkenntnisse zur Struktur professioneller Handlungskompetenz von angehenden Physiklehrkräften. *ZfDN* 16, 167–187

- Riggs, I. M. & Enochs, L. G. (1990). Towards the Development of an Elementary Teacher's Science Teaching Efficacy Belief Instrument. *Science Education*, 74(6), 625–637.
- Roberts, K. J., Henson, R. K., Tharp, B. Z. & Moreno, N. P. (2001). An Examination of Change in Teacher Self-efficacy Beliefs in Science Education Based on the Duration of Inservice Activities. *Journal of Science Teacher Education*, 12(3), 199–213.
- Schmitz, G. S. & Schwarzer, R. (2002). Individuelle und kollektive Selbstwirksamkeitserwartung von Lehrern. *Zeitschrift für Pädagogik, Beiheft 44*, 192–214.
- Schmitz, G. S. & Schwarzer, R. (2000). Selbstwirksamkeitserwartung von Lehrern: Längsschnittbefunde mit einem neuen Instrument. *Zeitschrift für Pädagogische Psychologie 14*(1), 12–25.
- Schwarzer, R. & Jerusalem, M. (1999). Skalen zur Erfassung von Lehrer- und Schülermerkmalen. Dokumentation der psychometrischen Verfahren im Rahmen der Wissenschaftlichen Begleitung des Modellversuchs Selbstwirksame Schulen. Berlin. (http://userpage.fu-berlin.de/~health/self/skalendoku_selbstwirksame_schulen.pdf, zuletzt abgerufen am 06.03.2012)
- Schwarzer, R. & Jerusalem, M. (2002). Das Konzept der Selbstwirksamkeit. *Zeitschrift für Pädagogik, Beiheft 44*, 28–53.
- Sekretariat der ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2008). Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung. Beschluss der Kultusministerkonferenz vom 16.10.2008 i.d.F. vom 16.09.2010
- Shulman, L.S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4–14.
- Simon, B. & Trötschel, R. (2007). Das Selbst und die soziale Identität. In K. Jonas, W. Stroebe, & M. Hewstone (Eds.), *Sozialpsychologie* (pp. 147–185). Heidelberg: Springer.
- Tschannen-Moran, M., Woolfolk-Hoy A. & Hoy, W. K. (1998). Teacher Efficacy: Its Meaning and Measure. *Review of Educational Research*, 68, 202–248.
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Ed.), *Leistungsmessungen in Schulen* (pp. 17–31). Weinheim und Basel: Beltz Verlag.

KONTAKT

Prof. Dr. Thorid Rabe
Universität Potsdam
Institut für Physik und Astronomie
Arbeitsgruppe Didaktik der Physik
Karl-Liebknecht-Str. 24/25
14476 Potsdam-Golm
rabeth@uni-potsdam.de

AUTORENINFORMATION

Prof. Dr. Thorid Rabe ist Juniorprofessorin und leitet derzeit den Lehrstuhl für Didaktik der Physik an der Universität Potsdam. Ihre derzeitigen Arbeitsschwerpunkte liegen im Bereich der physikdidaktischen Lehr-Lern-Forschung, der Lehrerprofessionalisierung und dem Einsatz von E-Learning-Szenarien.

Claudia Meinhardt ist Doktorandin am Lehrstuhl für Didaktik der Physik an der Universität Potsdam mit dem Arbeitsschwerpunkt Selbstwirksamkeitserwartungen in physikdidaktischen Handlungsfeldern.

Dr. Olaf Krey ist wissenschaftlicher Mitarbeiter am Lehrstuhl Didaktik der Physik an der Universität Potsdam. Im Rahmen seiner Promotion beschäftigte er sich mit der Rolle der Mathematik in der Physik.

