

Sprachbezogenes metakognitives Strategiewissen im Studienalltag internationaler Studierender: Der szenariobasierte Fragebogen ScenEx

Katrin Wisniewski,¹ Maria Parker,² Wolfgang Lenhard³
und Jennifer Seeger⁴

This article describes the development of a scenario-based strategy questionnaire for assessing metacognitive strategy knowledge (ScenEx: Scenario-Based Executive Metacognition in Studying in a Foreign Language). This instrument is currently used in the longitudinal research project SpraStu (Language and academic success of international students at German universities). A brief methodological overview of strategy assessment techniques is followed by a detailed description of the ScenEx rationale, contents, and psychometric quality. ScenEx assesses knowledge of strategy adequacy in contextualized and empirically validated study-related situations (scenarios) with reference to an expert-based qualitative standard. The piloting revealed good psychometric qualities of the instrument. First external validation resulted in robust correlations with measures of language competence, but not with other measures of self-regulation, pointing at the language-specificity of ScenEx. Potential uses of the questionnaire are discussed, as are constraints and limitations.

1. Einleitung

In einem empirischen Längsschnittprojekt untersucht das Projekt "Sprache und Studienerfolg bei Bildungsausländer/innen" (SpraStu, 2017-2020) den Zusammenhang sprachlicher und anderer Faktoren mit dem Studienerfolg von Bildungsausländer/innen (vgl. Wisniewski 2018; <http://home.uni-leipzig.de/sprastu/>). Dabei handelt es sich um Studierende mit nicht deutscher Staatsangehörigkeit und Hochschulzugangsberechtigung. Aus Studienabbruchsuntersuchungen ist bekannt, dass diese Gruppe deutlich häufiger ihr Studium abbricht als deutsche Studierende. Zuletzt lag die Abbruchquote von Bildungsausländer/innen im BA-Studium bei 41% versus 28% der deutschen Studierenden (Heublein, Ebert, Hutzsch, Isleib, König, Richter & Woisch 2017: 263, 270), wobei sich die Zahlen je nach Herkunftsregion, Hochschulart und

1 Korrespondenzadresse: Dr. Katrin Wisniewski, Universität Leipzig, Herder-Institut, Beethovenstr. 15, 04107 Leipzig, E-Mail: katrin.wisniewski@uni-leipzig.de

2 Korrespondenzadresse: Maria Parker, Universität Leipzig, Herder-Institut, Beethovenstr. 15, 04107 Leipzig, E-Mail: maria.parker@uni-leipzig.de

3 Korrespondenzadresse: Prof. Dr. Wolfgang Lenhard, Julius-Maximilians-Universität Würzburg, Lehrstuhl Psychologie IV, Wittelsbacherplatz 1, 97070 Würzburg, E-Mail: wolfgang.lenhard@uni-wuerzburg.de

4 Korrespondenzadresse: Jennifer Seeger, Julius-Maximilians-Universität Würzburg, Lehrstuhl Psychologie IV, Wittelsbacherplatz 1, 97070 Würzburg, jennifer.seeger@uni-wuerzburg.de

Fachrichtung teils deutlich unterscheiden (Heublein et al. 2017). Die Gründe für diese Situation sind bislang kaum erforscht und werden in der Regel in einer Faktorenkombination vermutet, die auf eine lückenhafte sprachliche Vorbereitung, Schwierigkeiten mit dem selbständigen Lernen in einer neuen Studienkultur sowie auf Probleme bei der sozialen Integration zurückgehen könnten (Heublein 2015: 14). Eine Metastudie des Sachverständigenrats deutscher Stiftungen für Integration und Migration (2017: 21-25) spricht von einem "akademischen Kulturschock"; auch eine Bedarfsanalyse zu sprachlichen Herausforderungen in der Studieneingangsphase bestätigt diese Aspekte vorläufig (Bärenfänger, Lange & Möhring 2016). Angesichts der wachsenden Zahlen internationaler Studierender an deutschen Universitäten (Statistisches Bundesamt 2016: 15; DZHW/DAAD 2017: 3) ist diese Situation nicht tragbar und mit hohen psychischen und ökonomischen Kosten verbunden.

Deshalb setzt sich SpraStu zum Ziel, Einflussfaktoren auf Studienverläufe von internationalen Studierenden der Universitäten Leipzig und Würzburg im Längsschnitt detailliert zu untersuchen. Es erhebt ein breites Spektrum sprachlicher Kompetenzen, erstellt ein Lernerkorpus und erfragt auch nicht-sprachliche potenzielle Bedingungsfaktoren für Studienerfolg (Wisniewski 2018). Der Studienerfolg wird mit verschiedenen Indikatoren wie z.B. den pro Semester erworbenen Leistungspunkten empirisch überprüft. Ein Schwerpunkt ist zudem die Untersuchung von Aspekten des selbstregulierten Lernens.

Der Fokus des vorliegenden, methodologischen Beitrags liegt auf der Beschreibung der Erstellung eines ausgewählten Bausteins der letztgenannten Projektsäule, des szenariobasierten Strategiefragebogens ScenEx (*Scenario-Based Executive Metacognition in Studying in a Foreign Language*). ScenEx erhebt deklaratives metakognitives Strategiewissen (Lenhard 2018: 15), d.h. Wissen über die Angemessenheit und Effektivität der Verwendung bestimmter Strategien zur Erreichung von (Lern-)zielen in spezifischen Lernsituationen. Es steht (anders als in anderen Projektinstrumenten) also nicht die Strategieanwendung im Fokus. Zur Erfassung dieses Wissens werden konkrete, für die jeweilige Zielgruppe und Domäne relevante Situationen (Szenarien) und mögliche Strategien geschildert, deren Angemessenheit für die Bewältigung der jeweiligen Situation bewertet werden muss. Wir schließen durch den Vergleich von Bewertungen verschiedener Strategien auf das Wissen. Es handelt sich somit um eine relationale Form von Strategiewissen.

Die Entwicklung eines solchen passgenauen Instruments war zum einen nötig, weil keine zielgruppenadäquaten Instrumente der Fremdsprachenlehr- und -lernforschung vorlagen. Üblicherweise erheben entsprechende Fragebögen die Häufigkeit der Verwendung von Strategien, die explizit dem Sprachenlernen dienen (z.B. "I use rhymes to remember new English words", aus dem Strategy Inventory for Language Learning (SILL), Oxford 1990; vgl. ähnlich Tseng, Dörnyei & Schmitt 2006). Im vorliegenden Fall jedoch steht der Umgang mit so alltäglichen wie herausfordernden

(Bärenfänger et al. 2016) stark sprachgeprägten konkreten Situationen bereits in der Studieneingangsphase im Vordergrund. Studierende benötigen spezifisch sprachbezogene strategische Kompetenzen, um diese zu bewältigen. Das zu entwickelnde Instrument musste damit trotz partieller Übereinstimmungen deutlich spezifischer sein als allgemeinere Strategieinventare.

Nicht nur die mangelnde Passgenauigkeit vorliegender Instrumente, sondern auch ihre nicht zufriedenstellende methodische Qualität legte die Entwicklung eines szenariobasierten Fragebogens wie des ScenEx nahe. Zwar können auch solche Fragebögen viele Dilemmata der Strategieforschung nicht auflösen, aber es gibt zumindest deutliche Hinweise auf ihre Validität. So weisen sie u.a., anders als die verbreiteten allgemeinen Strategiefragebögen (zu Sprachlernstrategien: SILL, Oxford 1990; zu Studierstrategien: *Motivated Strategies for Learning Questionnaire* (MSLQ), Pintrich, Smith, Garcia & McKeachie 1991; VanderStoep & Pintrich 2003; LIST, Schiefele & Wild 1994), robuste Zusammenhänge mit Leistungsindikatoren auf. Szenariobasierte Fragebögen sind in der Sprachlehr- und -lernforschung bislang auch international noch nicht verbreitet, auch wenn im Zuge einer Krise der "klassischen" Lern(er)strategieforschung im Fach (Dörnyei 2005; Tseng et al. 2006; Gao 2007) die Entwicklung von weniger allgemeinen, kontextangepassten Instrumenten als Desiderat formuliert wurde. Bislang wurden unter Rückgriff auf Modelle selbstregulierten Lernens einige wenige aufgaben- bzw. fertigkeitenbezogene, nicht jedoch szenariobasierte Instrumente entwickelt (z.B. Nakatani 2006; Tseng et al. 2006; Rose 2012a, 2012b). Oxford (2017: 317-334) schlägt ein Instrument zur Emotionsregulierung in sprachbasierten Szenarien vor, das allerdings nicht quantitativ auswertbar ist und bislang nicht validiert wurde. Szenariobasierte Fragebögen wurden somit bislang ausschließlich im Bereich der empirischen Bildungswissenschaft, insbesondere der pädagogischen Psychologie, vorgelegt, wo sie ebenfalls innovative Instrumente darstellen (etwa zum Lesestrategiewissen in den Klassen 7-12, Schlagmüller & Schneider 2007; zum Englischen in der frühen Sekundarstufe, Neuenhaus, Artelt & Schneider 2017; oder zu Mathematik in Klasse 5 und 6, Lingel 2016).

Der vorliegende Beitrag geht zunächst auf den theoretisch-methodologischen Hintergrund ein. Vor der Beschreibung der Erstellung von ScenEx wird der Fragebogen kurz in das Gesamtdesign der Projektsäule in SpraStu eingebettet. Angesichts der Neuartigkeit des Erhebungsverfahrens wird dessen Erstellung dann detailliert nachvollzogen. Während der Hauptschwerpunkt dieses Beitrags somit methodologisch-inhaltlicher Art ist, werden doch auch die wichtigsten Ergebnisse der quantitativen Validierungen und der Pilotierung in zusammengefasster Form berichtet. Diese werden zu einem späteren Zeitpunkt und an anderer Stelle ausführlicher referiert werden.

2. Theoretische Einbettung

Die Erforschung von Lern- und Kommunikationsstrategien hat in der Fremdsprachenlehr- und -lernforschung eine lange Tradition (vgl. etwa Rubin 1975; Oxford 1990). Auch in Modellen kommunikativer Sprachkompetenz spielt eine, wenn auch unterdefinierte, "strategische Kompetenz" eine zentrale Rolle (vgl. Wisniewski 2016). Allerdings gelang es hier trotz der regen Forschungstätigkeit nicht, kohärente Modellierungen zu entwickeln. So führten auch terminologische Streitigkeiten bereits auf Ebene des Strategiebegriffs zu einer Krise der Strategieforschung (Dörnyei 2005; Macaro 2006, 2007). Dörnyei (2005: 188) schlug schließlich vor, den Strategiebegriff gänzlich zu meiden und sich auf in der Psychologie verbreitete Modelle der Selbstregulation (SRL) zu beziehen.

Als SRL wird die Anpassung von Gefühlen, Handlungen und Gedanken an die eigenen Lernziele bezeichnet (Zimmerman 2000: 14): "[Self-regulated learners] actively research what they do to learn and how well their goals are achieved by variations in their approaches to learning" (Winne 2010: 472). Grundgedanke von SRL-Modellen ist, dass Lernende einen Ist-Zustand mit einem Soll-Zustand (einem sogenannten Standard) vergleichen und dass dieser Vergleich zu einer Regulation führt (Wirth & Leutner 2008: 104). Es gibt ein äußerst breites Spektrum an SRL-Modellen, auf die hier nicht detailliert eingegangen werden kann (vgl. etwa Pintrich 2000; Zimmerman 2000; Landmann, Perels, Otto & Schmitz 2009; Panadero 2017), die in der Regel jedoch die Hauptkomponenten der Kognition, Motivation und Metakognition (teils auch Emotion) ansetzen.

Lernstrategiewissen ist eine Voraussetzung für SRL und stellt somit nur einen seiner Teilaspekte dar (Neuenhaus et al. 2017). Es betrifft insbesondere den Bereich der Kognition (konzeptionelles und strategisches Wissen über Strategien) und der Metakognition (Steuerung des Lernprozesses bzgl. der Zielsetzung, Planung, Selbstbeobachtung, Reflexion und Anpassung des Lernverhaltens), wird aber auch von der Motivation (d.h. Aktivitäten zur Initiierung und volitionalen Lenkung des Lernens) beeinflusst. Besonders wichtig auch für den hier vorgestellten Fragebogen ist Lernstrategiewissen, d.h. Wissen über Lernstrategien und ihre anwendungsbezogenen Eigenschaften. Lernstrategiewissen kann deklarativ sein (Wissen, dass es Strategien zur Bearbeitung einer Aufgabe gibt), konditional (Bedingungen, unter denen der Einsatz einer Strategie Erfolg verspricht), prozedural (Ausführungswissen) oder relational (Wissen über die Effektivität einer Strategie im Vergleich zu anderen Strategien) (ebd.).

Die Bedeutung einzelner Strategien für Lernerfolg wird somit grundsätzlich relativiert. Dies wurde von der Fremdsprachenlehr- und -lernforschung sehr positiv wahrgenommen; in den letzten Jahren häufen sich hier Studien, die deutliche Bezüge

zu SRL-Modellen herstellen (Gao 2007; Rose 2012a, 2012b; Dörnyei & Ryan 2015; Oxford 2017; Rose, Briggs, Boggs, Sergio & Ivanova-Slavianskaia 2018). Trotz dieser Öffnung des Diskurses auf theoretisch-konzeptueller Ebene liegen allerdings bislang kaum Instrumente vor, die dieser Entwicklung Rechnung tragen (Tseng et al. 2006; Rochecoste, Oliver & Mulligan 2012; Briggs 2015; Teng & Zhang 2016; in Ansätzen auch Gao 2007).

3. Methodische Einbettung

Unabhängig von der jeweils gewählten Methode unterliegt die Messung von Strategien wegen deren schwieriger Beobachtbarkeit grundsätzlich Grenzen (Spörer & Brunstein 2006; Wolters & Won 2017). Es kann es also nicht darum gehen, eine etwaig "beste" Methode herauszustellen. Vielmehr werden zunehmend methodenorientierte Ansätze eingesetzt, die qualitative und quantitative Verfahren kombinieren (Veenman 2005).

Zwei grundsätzliche Unterscheidungen helfen dabei, Methoden zur Strategieerhebung zu kategorisieren. Zum einen lassen sich auf der Ebene des Beobachtungszeitpunkts Offline- und Online-Instrumente differenzieren (Wirth & Leutner 2008: 104). Erstere funktionieren losgelöst von konkreten Lernprozessen und erfassen eher Voraussetzungen für strategisches Handeln (z.B. verschiedene Arten von Wissen über Strategien) als diese selbst. Onlineverfahren hingegen untersuchen Lernprozesse in ihrem Verlauf schätzen ein, ob die Lernenden ihre SRL-Kompetenz beim Lernen tatsächlich einsetzen. Damit zielen sie auf unterschiedliche Komponenten und Phasen der Selbstregulation. Zweitens lassen sich auf Auswertungsebene quantitative und qualitative Vergleichsstandards unterscheiden. Bei einem quantitativen Standard wird in einer Maximalperspektive eine hohe Anzahl/Frequenz von etwas (bspw. die Verwendung einer bestimmten Strategie) als erstrebenswert (Standard) gesetzt. Qualitative Standards hingegen sind an der kontextuellen Angemessenheit, die theorie- und/oder expertenbasiert gesetzt werden muss, orientiert.

Bei der überwiegenden Mehrheit an Erhebungsverfahren zum Strategiewissen handelt es sich um quantitative offline-Fragebögen (z.B. der SILL, Oxford 1990, oder LIST, Schiefele & Wild 1994), die nach der wahrgenommenen Verwendungshäufigkeit von Strategien fragen und dabei eine große Häufigkeit als wünschenswert setzen. Wirth und Leutner (2008: 105) bezeichnen solche Instrumente als mangelhaft valide; sie korrelieren zudem kaum mit Lernergebnissen (Wolters & Won 2017). Methodisch problematisch ist bei dieser Form von Selbstberichten die von den Lernenden verlangte Abstraktions- und Erinnerungsleistung, denn sie müssen über verschiedene

Aufgaben und einen meist nicht spezifizierten Zeitraum verallgemeinern (vgl. Neuenhaus et al. 2017); sie unterliegen sehr stark sozialer Erwünschtheit und produzieren keine Erkenntnisse zum Wissen über die Angemessenheit von Strategien in konkreten Situationen. Gleichzeitig arbeiten zahlreiche Studien ausschließlich mit solchen Verfahren (z.B. Schrader, Helmke, Wagner, Nold & Schröder 2008; Swain, Huang, Barkaoui, Brooks & Lapkin 2009; Wong & Nunan 2011).

Andere Formen von Selbstberichten, die verschiedene Typen metakognitiven Strategiewissens erfassen, sind Lerntagebücher (Perels, Otto & Schmitz 2008), retrospektive Interviews sowie *Stimulated Recalls* (Gass & Mackey 2017). Auch Lautes Denken-Protokolle (Greene, Deekens, Copeland & Yu 2017) werden eingesetzt und versuchen, anders als offline-Verfahren, strategisches Verhalten direkt zu beobachten. Online-Verfahren arbeiten oft computergestützt, z.B. so genannte Mikroanalysen (Cleary & Callan 2017), *learning analytics* (Roll & Winne 2015), *data mining* (Biswas, Baker & Paquette 2017) oder *trace data* (Bernacki 2017).

Die hier fokussierten szenariobasierten Fragebögen kombinieren qualitative und quantitative Vergleichsstandards. Sie sind offline-Instrumente, die Nützlichkeits- bzw. Angemessenheitsurteile über Strategien in bestimmten Situationen (Szenarien) quantifizieren und somit relationales metakognitives Lernstrategiewissen erfassen. Sie beantworten also nicht die Frage nach tatsächlichem Strategieeinsatz in konkreten Aufgaben. Ziel ist eine optimale Auswahl der Antworten, die einem extern gesetzten qualitativen Standard entspricht, der zuvor beispielsweise durch die Befragung von Expertinnen und Experten gewonnen wurde. Voraussetzung hierfür ist der Bezug auf möglichst authentische Szenarien und die inhaltliche Prüfung der Validität der Antwortmöglichkeiten. Solche Instrumente weisen insgesamt eine größere prognostische Validität in Bezug auf Lernergebnisse als Fragebögen mit rein quantitativem Standard auf (Artelt, Demmrich & Baumert 2001; Wirth 2004, 2005). Zum Teil wird diese erhöhte Genauigkeit durch eine Einschränkung des Geltungsbereichs solcher Instrumente möglich. Während solche Verfahren in der empirischen Bildungsforschung zum Einsatz kommen, wird ihr Potenzial in der Fremdsprachenlehr- und -lernforschung trotz des Mangels quantitativer Erhebungsinstrumente (vgl. Cohen 2011) bislang nicht genutzt.

4. Lernstrategiewissen im Projekt SpraStu

Im Folgenden soll der szenariobasierte Fragebogen ScenEx in das Gesamtdesign der "Strategie-Projektsäule" in SpraStu eingebettet werden. Aus Platzgründen können die Verfahren allerdings nur kurz benannt werden.

SpraStu ist längsschnittlich angelegt und begleitet im Rahmen einer Vollerhebung der Bildungsausländer/innen die Teilnehmenden über das gesamte BA-Studium an den Universitäten Leipzig und Würzburg. Die Erhebungen finden einmal pro Jahr je zu Beginn des Wintersemesters statt; es nehmen drei Projektkohorten teil. Die hier berichteten Ergebnisse beziehen sich auf die erste Erhebungswelle zum Beginn des Wintersemesters 2017/2018 (ScenEx-Ergebnisse für N = 130 Bildungsausländer/innen). Ab dem Wintersemester 2018/2019 wird zur Validierung eine muttersprachliche Vergleichsgruppe hinzugezogen, die an ausgewählten Projekterhebungen teilnimmt.

Die Strategieerhebungen führen von einem situationsübergreifenden, allgemeinen Ansatz bis zu einer stark kontextualisierten, prozessorientierten Perspektive. Erstens wird allgemeines studienbezogenes Lernstrategiewissen mit Skalen aus dem quantitativen Offline-Fragebogen LIST (Schiefele & Wild 1994; Wild 2000) erhoben, der den oben genannten Einschränkungen unterliegt, aber weit verbreitet ist und eine Vergleichbarkeit der Ergebnisse mit anderen Studien ermöglicht. Auch Skalen zur Selbsteinschätzung aus dem Gemeinsamen europäischen Referenzrahmen für Sprachen werden verwendet (GeR; Europarat 2018; Skalen "Kompensieren" und "Hinweise erschließen"). Zweitens kommt ScenEx zum Einsatz, der zielgruppenangepasst sprachbezogenes relationales metakognitives Strategiewissen in studententypischen Situationen (Szenarien) erfasst und einen qualitativen Vergleichsstandard setzt. Drittens wird der Strategiegebrauch in zwei konkreten Studiensituationstypen untersucht (klausurbezogene Strategien und Mitschreibstrategien in Vorlesungen). Beide Handlungen hatten Bildungsausländer/innen in einer Bedarfsanalyse (Bärenfänger et al. 2016) als herausfordernd beschrieben. In einem mehrmethodischen Ansatz werden hier Interviews, *Stimulated Recalls* sowie linguistische Analysen von Primärdaten (Klausuren, Mitschriften, Vorlesungen) und spezifische Fragebögen zum Einsatz gebracht.

Die Erhebungsinstrumente insbesondere der zweiten und dritten Ebene sind aufeinander bezogen. Resultate der quantitativ auswertbaren Verfahren (Ebene 1 und 2) können u.a. mit Ergebnissen von Sprachkompetenztests in Verbindung gebracht werden.

5. Entwicklung des Instruments

Die Erstellung von ScenEx im Projekt SpraStu verlief in mehreren Schritten, die im Wesentlichen hier nachgezeichnet werden sollen und die sich an das Vorgehen bei der Entwicklung vergleichbarer Instrumente aus der Psychologie anlehnen (Schlagmüller & Schneider 2007; Lingel 2016; Neuenhaus et al. 2017).

5.1 Schritt 1: Recherche kontextangemessener Anforderungsszenarien

Zunächst wurden Anforderungsszenarien ausgewählt, die typische studienrelevante sprachliche Herausforderungen beinhalten sollten. Hierzu wurde auf die bereits zitierte Bedarfsanalyse von Bärenfänger et al. (2016) rekurriert. Die Herausforderungen sollten fächerübergreifend und explizit sprachlicher (und nicht etwa sozialer) Art sein. Zudem sollte ein gewisses sprachliches Handlungs- und Fertigkeitsspektrum abgedeckt und weitere projektrelevante Themen (v.a. Mitschreiben und Klausuren schreiben) berücksichtigt werden. Ferner sollten empirische Befunde zum jeweiligen Szenario vorliegen.

Schließlich wurden die folgenden sechs Szenarien abgeleitet:

1. Sie sitzen in einer Vorlesung. Der Dozent spricht sehr schnell und kompliziert. Sie haben Schwierigkeiten, ihn zu verstehen. Was tun Sie?
2. In einer Vorlesung erklärt der Dozent etwas. Gleichzeitig versuchen Sie, den Text auf den PowerPoint-Folien zu lesen. Das ist zu viel auf einmal, und Sie merken, dass Sie nicht mehr folgen können. Was tun Sie?
3. Sie sollen einen Text zur Vorbereitung lesen. Der Text enthält sehr viele Informationen. Sie können nicht erkennen, was wichtig ist und was nicht. Was tun Sie?
4. Sie müssen einen Text auf Deutsch für Ihr Studium schreiben, aber Sie haben Schwierigkeiten, Ihre Gedanken auf Deutsch auszudrücken. Was tun Sie?
5. Sie schreiben eine Klausur. Bei einer Multiple-Choice-Frage kommen Sie nicht weiter. Was tun Sie?
6. In einer Klausur kommt eine Frage vor, die Sie beantworten sollen. Die Frage ist aber ziemlich lang und kompliziert formuliert. Sie verstehen nicht genau, was Sie machen sollen. Was tun Sie?

5.2 Schritt 2: Recherche von Antwortoptionen (Strategien)

Nun wurden mögliche passende Antwortoptionen (Strategien) recherchiert, zu denen empirische Studien vorliegen. Für diese Vorauswahl wurden Strategien berücksichtigt, für die eine positive oder negative Verbindung mit (allgemeinen oder fertigungsbezogenen) Sprachkompetenzen nachgewiesen werden konnten, die besonders häufig durch Studierende genutzt wurden, und/oder die wiederholt Gegenstand empirischer Untersuchungen waren. Da die Probandinnen und Probanden (im folgenden

PuP) ein relationales Strategieurteil fällen sollten, war es wichtig, auch dysfunktionale Antwortoptionen zu verwenden. Für das erste Szenario wurden in dieser Arbeitsphase bspw. $N = 17$ Strategien kompiliert, darunter:

- a. Ich konzentriere mich auf das Wichtigste und achte nicht so sehr auf die Details (Goh 1998; Taguchi 2001; Hsueh-Jui 2008; Rahimirad & Moini 2015) (S1A02).
- b. Ich achte auf wichtige Schlüsselwörter (Goh 1998; Vandergrift 2003; Yeldham 2016) (S1A04).
- c. Während ich zuhöre, übersetze ich im Kopf in meine Muttersprache (Goh 2000; Vandergrift 2003; Hsueh-Jui 2008) (S1A14).

Abschließend wurde anhand einer Liste der häufigsten Wörter des Deutschen (Tschirner & Jones 2006) überprüft, dass die Szenarien und Strategien möglichst nur Wörter enthalten, die zu den 3000 häufigsten im Deutschen gehören. Für das Englische entspricht dies in etwa dem Sprachniveau B1 (Milton 2010), während zum Deutschen bislang keine dementsprechenden Studien vorliegen. Bei der Formulierung wurde außerdem auf eine möglichst wenig komplexe Syntax geachtet.

5.3 Dritter Schritt: Empirische Erstellung eines qualitativen Vergleichsstandards

In der Folge musste eine intersubjektiv gesetzte Norm festgelegt werden, anhand derer die Strategieurteile der PuP erst als mehr oder weniger angemessen eingeschätzt werden können ("qualitativer Vergleichsstandard"). Zudem sollte die Anzahl an Strategien pro Szenario auf ca. 7 reduziert werden (vgl. Neuenhaus et al. 2017). Hierzu wurde eine Expertenbefragung durchgeführt, in deren Rahmen $N = 32$ Fachkolleginnen und -kollegen (aus (Fremd-)Sprachdidaktik, Strategieforschung, pädagogischer Psychologie, Diagnostik u.ä.) ein Urteil über die Angemessenheit der Strategien abgaben. Sie beurteilten auf einer sechsstufigen Likertskala (Schulnoten), wie gut die Strategien in ihren Augen für internationale Studierende geeignet sind, um das jeweilige Szenario zu bewältigen. Bei der Ermittlung des Vergleichsstandards wurde folgendermaßen vorgegangen:

1. Die Expertinnen und Experten (im Folgenden EuE) beurteilten, wie gut eine Strategie geeignet ist, um eine spezifische Situation zu bewältigen. Sie vergaben dabei jeweils getrennt nach Szenarien für jede Strategie eine Bewertung, die von 1 = sehr gut bis 6 = sehr schlecht reichen konnte. Verschiedene Strategien konnten die gleiche Bewertung erhalten.

2. Basierend auf den einzelnen Bewertungen wurde für alle EuE eine Matrix der Einzelvergleiche erstellt. Innerhalb jedes Szenarios waren $n!$ (sprich: n Fakultät) Einzelvergleiche möglich.
3. Nun wurde ermittelt, welche Antwortmuster bei mindestens 95% der EuE gleich ausfielen, also bei welchen Einzelvergleichen eine Strategie als klar über- oder unterlegen eingestuft wurde. Dabei kam es nicht auf die Höhe der abgegebenen Beurteilung an, sondern nur darauf, dass das Antwortmuster konsistent war. Wir löschten Einzelvergleiche, in denen Strategien als gleichwertig beurteilt wurden.
4. Das Antwortmuster der EuE diente in der Folge als Blaupause für die Bewertung der Ergebnisse der PuP, wobei immer dann ein Punkt in einem Einzelvergleich vergeben wurde, wenn das Antwortmuster dem der EuE entsprach. Auch hier kam es nicht auf die absolute Höhe der Einstufung einer Strategie an, sondern nur auf den Vergleich. Jeder dieser Einzelvergleiche bildete eine neue Variable, die schließlich mit den statistischen Methoden aus dem psychometrischen Bereich weiter analysiert wurden (siehe Kapitel 6).

Die Auswertung folgte also einer Logik, in der die Strategien eines Szenarios jeweils paarweise zueinander in Beziehung gesetzt wurden. Eben diese paarweisen Vergleiche wurden später analog bei den PuP durchgeführt. Das Ziel unserer Vorgehensweise war, jene Einzelvergleiche herauszufiltern, bei denen EuE eine Strategie unabhängig von der Höhe der Bewertung als zielführender einschätzen und die EuE sehr eindeutig miteinander übereinstimmten. Für eine hohe Ökonomie wurde zudem darauf geachtet, dass nach der Auswahl Strategien übrigblieben, die mehrere Vergleiche ermöglichten. Zudem legten wir großen Wert auf die inhaltliche Abdeckung der Szenarien durch relevante Strategien.

Der untere Schwellenwert für die Aufnahme eines Paarvergleichs wurde mit einer Übereinstimmung von 95% bewusst sehr hoch angesetzt, um möglichst eindeutige Bewertungen zu haben. Der Wert kann im Vergleich zu ähnlichen Instrumenten als ausgezeichnet angesehen werden (Lingel 2016: 80%; Neuenhaus et al. 2017: 70%). Die Bewertung der Einzelvergleiche und der Übereinstimmung der EuE wurden mit einem eigens hierfür entwickelten Computerprogramm durchgeführt, das abhängig vom gewählten Schwellenwert aus den initial möglichen Einzelvergleichen die Vergleiche mit der höchsten Übereinstimmung auswählte. Innerhalb dieser Antwortoptionen wurde zur Erstellung der Pilotfassung außerdem auf eine sinnvolle inhaltliche Abdeckung und ein möglichst breites Bewertungsspektrum geachtet (vgl. exemplarisch Tab. 1). In die Pilotfassung gingen für die sechs Szenarien insgesamt 44 Strategien und 52 Paarvergleiche ein.

6. ScenEx: Pilotierung und Ermittlung der Testgütekriterien

In einem vierten Schritt schloss sich die Pilotierung des Instruments mit $N = 130$ Bildungsausländerinnen und Bildungsausländern an, der ersten SpraStu-Projektkohorte. Ziel der Pilotierung war, Strategien zu identifizieren, die sich bei Bildungsausländerinnen und Bildungsausländern als problematisch erweisen könnten und die aus Sicht psychometrischer Kriterien ungünstige Eigenschaften aufwiesen. Auch vorläufige Informationen über die Testgüte der Einzelvergleiche, der Szenarien und des gesamten Instruments wurden ermittelt.

Die Studierenden beantworteten die Frage nach der Angemessenheit der Strategien analog zu den EuE auf einer sechsstufigen Likertskala. Das anhand der Expertenratings gewonnene Bewertungsschema, also das Muster der Einzelvergleiche, wurde verwendet, um die Einschätzungen der PuP zu bewerten. Jeder Einzelvergleich bildete dabei eine eigene Variable für die folgende Itemselektion. Bei Übereinstimmung des Vergleichs der PuP mit der Experteneinschätzung erhielten diese für diesen Einzelvergleich einen Rohpunkt. Ansonsten wurde der Vergleich mit 0 bewertet. Für den Fall, dass Bewertungen fehlten, wurden alle betreffenden Einzelvergleiche bei dieser Person für die Analyse ausgeschlossen, da kein Vergleich durchgeführt werden konnte. Auch hier ging es also wieder nicht um die absolute Höhe der Einstufung einer Strategie, sondern darum, ob der Vergleich der Angemessenheit der Strategie mit dem Expertenurteil übereinstimmte (relationales metakognitives Strategiewissen). Auf diese Weise entstand ein Bewertungsmaßstab, der abbildet, wie stark eine Person in der Einstufung der situationsspezifischen Nützlichkeit bestimmter Strategien dem Expertenurteil entspricht, unabhängig davon, ob diese Person diese Strategien selbst anwendet oder nicht und ebenfalls unabhängig davon, wie hoch die einzelne Bewertung ausfällt. Zudem ist durch die dichotome Kodierung der Antworten (0 = entspricht nicht dem Expertenurteil versus 1 = entspricht dem Expertenurteil) Intervallskalierung gegeben, sodass die Ergebnisse der Einzelvergleiche aufsummiert werden können und weiterführende statistische Analysen möglich werden.

Die Ergebnisse der Pilotierungsfassung des Instruments variierten zwischen 20 und 52 Rohpunkten. Fünf Teilnehmer/innen erreichten die maximal möglichen 52 Punkte, sodass von einem kleinen Deckeneffekt auszugehen ist. Am unteren Ende der Skala traten dagegen keine Ergebnisse auf, sodass das Testverfahren in diesem Bereich sehr gut differenziert. Die PuP erreichten im Mittel $m = 42$ Punkte, bei einer Standardabweichung von $sd = 6.9$. Die Verteilung ist also leicht schief und der Gipfel zum positiven Bereich hin verschoben. Der Abstand des Skalenmittels vom maximal erreichbaren Wert beträgt 1.45 Standardabweichungen, sodass auch im oberen Bereich eine hinreichend gute Differenzierung gegeben ist.

Die psychometrischen Eigenschaften des Testverfahrens wurden in einem nächsten Schritt auf Ebene der Einzelvergleiche innerhalb jedes Szenarios und für den gesamten Test mittels Verfahren der Item-Response-Theorie (IRT) überprüft und hinsichtlich der Homogenität bewertet. Die Homogenität ist eine Maßzahl, die maximal den Wert 1 erreichen kann und die angibt, wie stark die einzelnen Aufgaben eines Verfahrens (in unserem Fall die Einzelvergleiche) miteinander zusammenhängen und ob sie somit die gleiche Fähigkeit abbilden. Das ist eine notwendige Voraussetzung dafür, die Rohwerte einer Person aufsummieren und als Indikator für deren Fähigkeit bzw. deren Wissen verwenden zu können. Die Trennschärfe gibt an, wie hoch eine Aufgabe mit der Gesamtskala korreliert. Ist diese Trennschärfe sehr gering oder sogar negativ zur jeweiligen Skala, dann misst diese Aufgabe entweder ein anderes Konstrukt, oder die Aufgabe wurde nicht verstanden, war missverständlich formuliert oder es traten sonstige Probleme auf. Solche Aufgaben sollten aus einem Testverfahren entfernt werden. Als Homogenitätsmaß griffen wir auf Cronbachs α und Revelles ω zurück. Die Berechnung der Homogenitätswerte und der deskriptiven Statistiken erfolgte mit Hilfe des Statistikprogramms R mit den Programmbibliotheken psych (Revelle 2018). Für die IRT-Analysen griffen wir auf die Programmbibliothek TAM (Robitzsch, Kiefer & Wu 2018) und für die Überprüfung der Konstruktvalidität mittels konfirmatorischer Faktorenanalysen auf lavaan (Rosseel 2012) zurück.

Insgesamt zeigte sich mit $\omega = .87$ eine hohe Homogenität der Einzelvergleiche des Verfahrens. Auch die IRT-Skalierung mittels eines einparametrischen Rasch-Modells ergab eine gute Reliabilität von 0,834 (*EAP reliability*). Mit Ausnahme von Szenario 1 fand sich in jedem Szenario mindestens ein Einzelvergleich mit einer Trennschärfe kleiner 0.2, in Szenario 5 ("Schreiben einer Multiple-Choice-Klausur") sogar drei niedrige Trennschärfen. Keine der Trennschärfen lag im negativen Bereich. In Szenario 4 und 6 fanden sich Einzelvergleiche mit sehr hohen Trennschärfen, also Vergleiche, die besonders stark zwischen Personen mit hohem und niedrigem metakognitivem Wissen unterscheiden. In der IRT-Analyse konnte nur eine Aufgabe identifiziert werden, deren Passung zum Testverfahren signifikant von den anderen Aufgaben abweicht.

7. Vorläufige Bewertung der Validität des Verfahrens

Die vollständige Validierung des Fragebogens beinhaltet eine ausstehende längsschnittliche Untersuchung im Drei-Kohorten-Design von SpraStu sowie zusätzliche Erprobungen auch bei deutschen Studierenden. Die folgenden Darstellungen stellen deshalb lediglich aus der Pilotierung resultierende vorläufige Einschätzungen der Validität von ScenEx dar. Zur Überprüfung der Validität unseres Strategiefragebogens

gingen wir von den folgenden Hypothesen aus:

1. Die Anforderungen der ScenEx-Szenarien sind stark sprachlich geprägt, so dass das erhobene Strategiewissen in einem niedrigen bis mittleren Bereich im Querschnitt positiv mit Sprachleistungen korrelieren sollte.
2. Unter der Annahme, dass verschiedene Instrumente zur Erfassung strategischen Verhaltens das gleiche Konstrukt valide abbilden, sollten die dazu eingesetzten Verfahren (s.u., LIST, GeR-Skalen und ScenEx) ebenfalls positiv miteinander korrelieren (konvergente Validität). Diese Annahme ist allerdings nur haltbar, wenn jedes Instrument Strategiewissen valide erfasst, was beim LIST so wie auch bei den GER-Skalen diskutiert werden kann (s.o., Kapitel 3; Wisniewski 2014).
3. Die hohe Homogenität der Aufgaben des Fragebogens deutet darauf hin, dass es sich bei dem gemessenen Konstrukt um eine eindimensionale Fähigkeit handelt. Personen haben also ein relativ einheitlich hohes oder niedriges Strategiewissen in verschiedenen Bereichen. Die Szenarien des Fragebogens sollten in der Folge im mittleren bis hohen Bereich miteinander korrelieren (Konstruktvalidität).

Zur Überprüfung von Annahme 1 wurden die ScenEx-Ergebnisse mit Sprachkompetenzerhebungen korreliert, die in SpraStu eingesetzt werden. Wir erhoben hier:

- a. den GeR-bezogenen onSET, einen C-Test des TestDaF-Instituts (<https://www.onset.de/>),
- b. Tests zur Erfassung des produktiven und rezeptiven Wortschatzes (Institut für Testforschung und Testentwicklung e. V. o. J.),
- c. den Reading Proficiency Test (RPT) und den Listening Proficiency Test (ACTFL® 2013, 2014) von ACTFL® (2012),
- d. die Schreibfertigkeit mit Hilfe einer von g.a.s.t zur Verfügung gestellten Test-DaF-Aufgabe.

Ferner verglichen wir die Ergebnisse in den einzelnen Szenarien und dem Gesamttest mit Ergebnissen aus dem LIST (Schiefele & Wild 1994) und einer Selbsteinschätzung auf Basis des GeR (Europarat 2001, 2018; Hypothese 2). Der LIST umfasst Selbstaussagen zur Häufigkeit der Strategieverwendung im Studium. Die beiden GeR-Skalen "Kompensieren" (als "produktive Strategie" geführt, hier Auswahl von 8 Deskriptoren auf B1-C2, Europarat 2018: 78-79) und "Hinweise identifizieren/erschließen" (als "rezeptive Strategie" geführt; hier Auswahl von sechs Deskriptoren auf B1-C1, Europarat 2018: 67) wurden sprachlich leicht vereinfacht zur Selbsteinschätzung vorgelegt und ermöglichen eine Koppelung an das Niveaustufensystem des GeR.

Die Ergebnisse der Korrelationen sind in Tabelle 1 dargestellt. Bei der Überprüfung von Hypothese 1 zeigte sich, dass ausschließlich ScenEx signifikant mit den Sprachleistungen korreliert und auch die Höhe der Korrelation in einem erwartbaren

Bereich liegt. Weder der LIST noch die GeR-Skalen hingen mit den Sprachleistungen zusammen. Zum Teil korrelierten diese sogar negativ mit der Leistung – allerdings auch hier ohne Signifikanz zu erreichen. Da der LIST allgemeines Studienverhalten erfasst, ist die fehlende Signifikanz eher verständlich als bei den GeR-Skalen, die ja auf Fremdsprachenlernende abzielen. Der LIST scheint somit spezifische, sprachliche Herausforderungen der Zielgruppe Bildungsausländer/innen nicht zu differenzieren. Die Strategieskalen des GeR hingegen sind nicht auf den Hochschulkontext bezogen. Ferner sind sie grundsätzlich nur partiell als validiert zu betrachten, denn ihre Skalierung beruht weder auf einer theoretischen Grundlage noch wurden bei der empirischen Validierung Lernendenleistungen oder -urteile mit einbezogen (vgl. Wisniewski 2014, 2016). Die vorliegenden Ergebnisse deuten darauf hin, dass der Zusammenhang der Strategieskalen des GeR mit Sprachkompetenz, dessen empirische Untersuchung noch aussteht, eher indirekter Art sein könnte.

Testet man die Höhe der Korrelationen zwischen LIST, GeR und szenariobasiertem Fragebogen auf der einen und Sprachleistungen auf der anderen Seite, dann hängen letztere signifikant stärker mit dem ScenEx zusammen. Die GeR-Skalen und der List wiesen dagegen keine signifikant positiven oder sogar signifikant negative Korrelationen mit Sprachleistungen auf.

Tabelle 1: Korrelation⁵ von ScenEx mit LIST, GeR-Skalen und den Sprachleistungen im onSET, Tests zum produktiven und rezeptiven Wortschatz, *Listening* und *Reading Proficiency* (ACTFL) und den Ergebnissen der TestDAF Schreibaufgabe

	ScenEx	LIST	GeR (produktiv)	GeR (rezeptiv)
LIST	.030			
GeR (produktiv)	-.004	.084		
GeR (rezeptiv)	.053	.117	.067	
onSET	.257 **	-.093	-.103	.153
Wortschatz (produktiv)	.262 **	-.156	-.030	.114
Wortschatz (rezeptiv)	.235 **	-.117	-.084	.062
<i>Listening Proficiency</i> (ACTFL)	.293 **	-.188*	-.152	.069
<i>Reading Proficiency</i> (ACTFL)	.314 **	-.148	-.077	.074
TestDAF Schreibaufgabe	.228 *	-.026	-.023	.172

5 Die deskriptiv negativen Korrelationen des LIST und der GeR-Skalen (produktiv) sind erwartungswidrig, sollten aber aufgrund der fast durchgehend fehlenden statistischen Signifikanz nicht interpretiert werden.

Hypothese 2 erweist sich als nicht haltbar: Die Korrelationen zwischen den Fragebögen sind durchweg nicht signifikant von 0 verschieden. Sie scheinen somit unterschiedliche Aspekte zu erfassen. Dies spricht jedoch nicht gegen die Gültigkeit von ScenEx. Stattdessen war dieses Instrument unter den verwendeten Verfahren das einzige, für das Validitätshinweise in Bezug auf sprachliche Leistungen vorliegen (vgl. Hypothese 1).

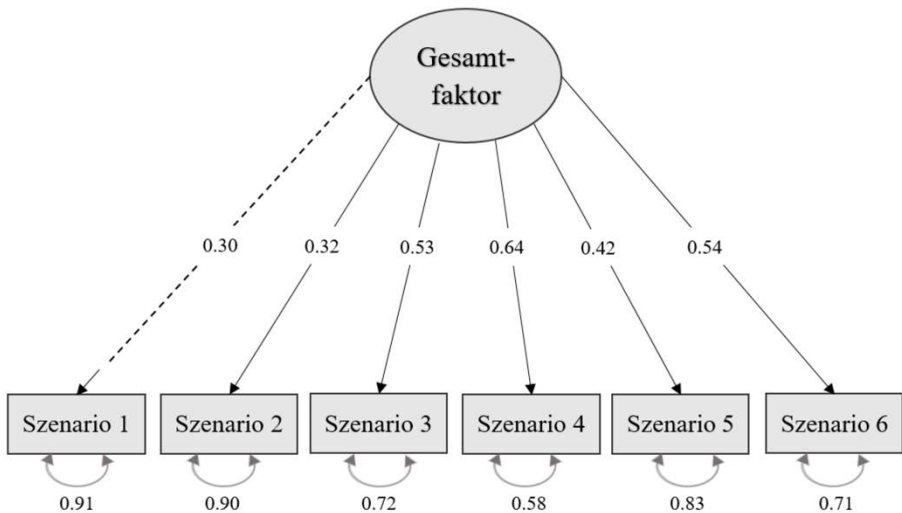


Abbildung 1: Messmodell von ScenEx mit den Ladungen der einzelnen Szenarien auf dem Gesamtfaktor

Die Analyse der internen Struktur des szenariobasierten Fragebogens (Hypothese 3) erfolgte mit Hilfe von Strukturgleichungsmodellen, bei denen basierend auf verschiedenen Beobachtungen auf eine latente Fähigkeit geschlossen wird. Die Höhe der Ladung der einzelnen Beobachtungen auf der latenten Fähigkeit gibt an, wie stark diese die latente Fähigkeit abbilden. In unserem Fall bestanden die einzelnen Beobachtungen im Rohwert, den jede Person in jedem einzelnen Szenario erzielte. Die Ladungen der einzelnen Szenarien auf dem Gesamtwert fallen unterschiedlich hoch aus und reichen von 0.30 (Szenario 1) bis 0.66 (Szenario 4), wobei Szenario 3 bis 6 besonders hoch auf dem Gesamtfaktor laden, die Ladungen von Szenario 1 und 2 dagegen etwas niedriger ausfallen (vgl. Abb. 1; zu den Szenarien vgl. Kapitel 5.1). Die statistischen Gütekriterien für das gesamte Model liegen nicht in einem sehr guten, aber noch in einem akzeptablen Bereich (CFI = .915, RMSEA = .071). Es gibt also Anteile in den Szenarien, die auf einen übergreifenden Faktor zur metakognitiven Lernregulation in unterschiedlichen, für Sprachlernende relevanten Situationen hinweisen. Darüber

hinaus gibt es aber auch Aspekte, die nicht situationsübergreifend betrachtet werden können, sondern spezifische Aufgaben repräsentieren und zwischen verschiedenen Lernenden unterschiedlich gut ausgeprägt sein können.

8. Schluss

ScenEx ist ein neuartiges Untersuchungsinstrument zur Erfassung deklarativen metakognitiven Wissens bei Sprachlernern, das über eine gute Konstruktvalidität und eine sehr gute interne Konsistenz verfügt. Die Besonderheit des Fragebogens liegt einerseits in seiner Konstruktion auf der Basis empirisch als prototypisch aufgezeigter, stark sprachlich geprägter Situationen im Studium von Bildungsausländer/innen und andererseits in der Auswertungslogik, die nicht auf Selbstaussagen zum Strategieeinsatz, sondern auf der Messung deklarativen Wissens basiert. Im Unterschied zu gängigen Verfahren wird nicht eine möglichst häufige Strategieverwendung als erstrebenswert gesetzt. Stattdessen müssen die PuP die Situationsangemessenheit verschiedener Strategien bewerten. In die Auswertung fließt nicht die absolute Höhe in der Einschätzung der Angemessenheit ein, sondern der Vergleich zwischen verschiedenen Herangehensweisen, die anhand eines an EuE gewonnenen Vergleichsmaßstabs bewertet werden. Als Ergebnis erhält man einen Rohwert, der abbildet, wie stark eine Person in ihrem Urteil den Expertinnen und Experten ähnelt.

Konfirmatorische Faktorenanalysen zeigen, dass neben einem starken allgemeinen Faktor zum metakognitiven Wissen vermutlich auch Anteile erfasst werden, die für einzelne Lernsituationen spezifisch sind und nicht innerhalb eines eindimensionalen Modells abgebildet werden können. Die Weiterentwicklung des Instruments wird also einerseits darauf abzielen, die Qualität der Aufgaben weiter zu steigern und zu überprüfen, ob die Bildung von Subskalen innerhalb des Fragebogens sinnvoll ist. Denkbar wäre beispielsweise, zwischen schriftlichen und mündlichen/auditiv rezeptiven Sprachleistungen zu unterscheiden. Zum anderen möchten wir im Projekt SpraStu überprüfen, ob ScenEx auch im Längsschnitt zukünftige Studienleistungen und den Zuwachs an sprachlichen Fähigkeiten vorhersagen kann. Zusammenhänge von ScenEx mit Indikatoren zum Studienerfolg (zur Operationalisierung vgl. Wisniewski 2018) werden im Projektverlauf von SpraStu erhoben und liegen noch nicht vor. Eine erste explorative, stark ausschnitthafte Analyse von Leistungspunkten (für $N = 62$ Leipziger Studierende im Wintersemester 2017/18) liefert jedoch Hinweise darauf, dass ScenEx auch prädiktiven Wert für die Vorhersage von Studienleistungen hat (signifikante Korrelation mit den im folgenden Studienjahr erworbenen Leistungspunkten von $r = .259^*$, $p = .042$; dagegen fehlende Signifikanz bei LIST: $r = .053$, $p = .684$). Da auch Studierende mit deutscher Muttersprache durch die bildungs-

sprachlichen Anforderungen des Studiums häufig herausgefordert werden, planen wir eine Validierung und Adaption des Verfahrens für die Stichprobe monolingual deutscher Studierender.

Die dieser Studie zugrundeliegenden Ergebnisse von 130 Bildungsausländer/innen deuten darauf hin, dass die allgemeinen, bislang verfügbaren Untersuchungsinstrumente LIST und GeR sowohl im (querschnittlichen) Validitätsprüfungen problematische Validitätskennwerte zeigten. In Ermangelung geeigneterer Instrumente setzten wir diese als externe Kriterien ein. Erwartungswidrig korrelierten sie überwiegend nicht signifikant mit Sprachleistungen, und auch untereinander gab es keine signifikante Korrelation der Fragebögen. Dies war bei ScenEx nicht der Fall. Wir halten deshalb zumindest für die Gruppe der Bildungsausländer die Entwicklung fokussierter Untersuchungsinstrumente für unerlässlich.

Limitierungen

Zwar stellt ScenEx ein innovatives und vielversprechendes Instrument dar, das sprachbezogenes metakognitives Strategiewissen in exemplarischen Studiensituationen erfasst, jedoch muss einschränkend verdeutlicht werden, dass dieser und ähnliche Fragebögen die eingangs dargestellten theoretischen und terminologischen Schwierigkeiten der Strategieforschung nicht vollständig zu lösen vermögen. Messmethodisch liegt wie auch bei den anderen Fragebögen ein *Common Method Bias*, also eine Konfundierung der Drittvariable Sprachbeherrschung mit der Beantwortung des Fragebogens, vor. Darüber hinaus wird nicht der tatsächliche Strategieeinsatz in authentischen Situationen untersucht, sondern das Wissen über die Effektivität des Einsatzes. Vorteile gegenüber reinen Selbstbewertungsbögen hat ScenEx jedoch in der Hinsicht, dass hier tatsächliches Wissen geprüft und sozial erwünschtes Antworten vermieden wird. Um der Gesamtheit selbstregulativer Verhaltensweisen Rechnung zu tragen, sind jedoch in jedem Fall *Mixed-Methods-Designs* vonnöten, die auch anderen, etwa motivationalen, Aspekten Rechnung tragen. Und schließlich handelt es sich bei den vorliegenden Ergebnissen um sehr ermutigende erste Schritte, die jedoch durch weitere Validierungen abgesichert werden müssen. Die hier untersuchte Stichprobe erlaubt noch keine Generalisierungen; sie ist sehr heterogen zusammengesetzt. So haben die Probandinnen und Probanden sehr unterschiedliche Ausgangsbedingungen, z.B. (Sprach-)Bildungsbiographien oder Studierenerfahrung.

Fazit

Wir sind überzeugt, mit der vorliegenden Untersuchung eine wichtige und notwendige Weiterentwicklung der Untersuchungsmethodik in diesem Bereich vorzulegen, die die vorhandenen Desiderata in der Verfügbarkeit von Messinstrumenten verkleinert. ScenEx könnte dabei helfen, über Defizite in Sprachkompetenzen hinausgehende Problematiken in der Lernregulation von Studierenden mit nichtdeutscher Muttersprache zu identifizieren, um diesen eine passgenauere Unterstützung im Studium zu ermöglichen.

Eingang des revidierten Manuskripts 24.01.2019

Literaturverzeichnis

- ACTFL® (2012): ACTFL Proficiency Guidelines [Online: https://www.actfl.org/sites/default/files/pdfs/public/ACTFLProficiencyGuidelines2012_FINAL.pdf, 03.09.2018].
- ACTFL® (2013): ACTFL Reading Proficiency Test (RPT): Familiarization Manual & ACTFL Proficiency Guidelines 2012 - Reading [Online: https://www.languagetesting.com/pub/media/wysiwyg/ACTFL_FamManual_Reading_2015.pdf, 04.09.2018].
- ACTFL® (2014): ACTFL Listening Proficiency Test (LPT): Familiarization Manual and ACTFL Proficiency Guidelines 2012 - Listening [Online: https://www.languagetesting.com/pub/media/wysiwyg/ACTFL_FamManual_Listening_2015.pdf, 04.09.2018].
- Artelt, Cordula; Demmrich, Anke & Baumert, Jürgen (2001): Selbstreguliertes Lernen. In: Baumert, Jürgen; Klieme, Eckhard; Neubrand, Michael; Prenzel, Manfred; Schiefele, Ulrich; Schneider, Wolfgang; Stanat, Petra; Tillmann, Klaus-Jürgen & Weiß, Manfred (Hrsg.): *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Wiesbaden: VS Verlag für Sozialwissenschaften, 271-298.
- Bärenfänger, Olaf; Lange, Daisy & Möhring, Jupp (2016): *Sprache und Bildungserfolg: Sprachliche Anforderungen in der Studieneingangsphase*. Leipzig: Institut für Testforschung und Testentwicklung.
- Bernacki, Matthew L. (2017): Examining the Cyclical, Loosely Sequenced, and Contingent Features of Self-Regulated Learning: Trace Data and Their Analysis. In: Zimmerman, Barry J. & Schunk, Dale H. (Hrsg.), 370-387.
- Biswas, Gautam; Baker, Ryan S. & Paquette, Luc (2017): Data Mining Methods for Assessing Self-Regulated Learning. In: Zimmerman, Barry J. & Schunk, Dale H. (Hrsg.), 388-404.
- Briggs, Jessica G. (2015): A context-specific research tool to probe the out-of-class vocabulary-related strategies of study-abroad learners. *International Journal of Applied Linguistics* 25: 3, 291-314.
- Cleary, Timothy J. & Callan, Gregory L. (2017): Assessing Self-Regulated Learning Using Micro-analytic Methods. In: Zimmerman, Barry J. & Schunk, Dale H. (Hrsg.), 338-351.
- Cohen, Andrew D. (2011): L2 learner strategies. In: Hinkel, Eli (Hrsg.): *Handbook of research in second language teaching and learning*. London: Routledge, 681-698.
- Dörnyei, Zoltán (2005): *The psychology of the language learner: Individual differences in second language acquisition*/Zoltán Dörnyei. Mahwah, N.J., London: Lawrence Erlbaum Associates.
- Dörnyei, Zoltán & Ryan, Stephen (2015): *The psychology of the language learner revisited*. London: Routledge.

- DZHW (= Deutsches Zentrum für Hochschul- und Wissenschaftsforschung)/DAAD (= Deutscher Akademischer Austauschdienst) (2017): *Wissenschaft weltoffen kompakt 2017. Daten und Fakten zur Internationalität von Studium und Forschung in Deutschland*. [Online: <http://www.wissenschaftweltoffen.de/publikation>, 27.08.2018].
- Europarat (2001): *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren, beurteilen; [Niveau A1, A2, B1, B2, C1, C2]*. Berlin: Langenscheidt.
- Europarat (2018): *Common European Framework of Reference for Languages: Learning, Teaching, Assessment: Companion Volume with New Descriptors* [Online: <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>, 30.01.2019].
- Gao, Xuesong (2007): Has Language Learning Strategy Research Come to an End?: A response to Tseng et al. (2006). *Applied Linguistics* 28: 4, 615-620.
- Gass, Susan M. & Mackey, Alison (2017): *Stimulated recall methodology in applied linguistics and L2 research*. New York, NY: Routledge.
- Goh, Christine C.M. (1998): How ESL learners with different listening abilities use comprehension strategies and tactics. *Language Teaching Research* 2: 2, 124-147.
- Goh, Christine C.M. (2000): A cognitive perspective on language learners' listening comprehension problems. *System* 28: 1, 55-75.
- Greene, Jeffrey A.; Deekens, Victor M.; Copeland, Dana Z. & Yu, Seung (2017): Capturing and Modeling Self-Regulated Learning Using Think-Aloud Protocols. In: Zimmerman, Barry J. & Schunk, Dale H. (Hrsg.), 323-337.
- Heublein, Ulrich (2015): Von den Schwierigkeiten des Ankommens. Überlegungen zur Studiensituation ausländischer Studierender an den deutschen Hochschulen. *Die Neue Hochschule* 2015: 1, 14-17.
- Heublein, Ulrich; Ebert, Julia; Hutzsch, Christopher; Isleib, Sören; König, Richard; Richter, Johanna & Woisch, Andreas (2017): *Zwischen Studiererwartungen und Studienwirklichkeit: Ursachen des Studienabbruchs, beruflicher Verbleib der Studienabbrecherinnen und Studienabbrecher und Entwicklung der Studienabbruchquote an deutschen Hochschulen*. Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (= DZHW).
- Hsueh-Jui, Liu (2008): A Study of the Interrelationship between Listening Strategy Use, Listening Proficiency Levels, and Learning Style. *ARECLS* 5, 84-104.
- Institut für Testforschung und Testentwicklung e. V. (o. J.): *Wortschatztest Deutsch produktiv* [Online: <http://www.itl-leipzig.de/static/startseite.html>, 30.01.2019].
- Institut für Testforschung und Testentwicklung e. V. (o. J.): *Wortschatztest Deutsch rezeptiv* [Online: <http://www.itl-leipzig.de/static/startseite.html>, 30.01.2019].
- Landmann, Meike; Perels, Franziska; Otto, Barbara & Schmitz, Bernhard (2009): Selbstregulation. In: Wild, Elke & Möller, Jens (Hrsg.): *Pädagogische Psychologie*. Heidelberg: Springer, 49-70.
- Lenhard, Wolfgang (2018): *Leseverständnis und Lesekompetenz: Grundlagen - Diagnostik - Förderung* (2. Aufl.). Stuttgart: Kohlhammer.
- Lingel, Klaus (2016): *Metakognitives Wissen Mathematik: Entwicklung und Zusammenhang mit der Mathematikleistung in der Sekundarstufe I*. Würzburg: Würzburg University Press.
- Macaro, Ernesto (2006): Strategies for Language Learning and for Language Use: Revising the Theoretical Framework. *The Modern Language Journal* 90: 3, 320-337.
- Macaro, Ernesto (2007): Language learner strategies: Adhering to a theoretical framework. *Language Learning Journal* 35: 2, 239-243.
- Milton, James (2010): The development of vocabulary breadth across the CEFR levels. In: Bartning, Inge; Martin, Maisa & Vedder, Ineke (Hrsg.): *Communicative proficiency and linguistic development: intersections between SLA and language testing research*. Eurosla, 211-232.
- Nakatani, Yasuo (2006): Developing an Oral Communication Strategy Inventory. *The Modern Language Journal* 90: 2, 151-168.

- Neuenhaus, Nora; Artelt, Cordula & Schneider, Wolfgang (2017): Lernstrategiewissen im Bereich Englisch. *Diagnostica* 63: 2, 135-147.
- Oxford, Rebecca L. (1990): *Language learning strategies: What every teacher should know*. Boston, Mass.: Heinle & Heinle.
- Oxford, Rebecca L. (2017): *Teaching and researching language learning strategies: Self-regulation in context*. New York/London: Routledge Taylor & Francis Group.
- Panadero, Ernesto (2017): A Review of Self-regulated Learning: Six Models and Four Directions for Research. *Frontiers in psychology* 8, 422.
- Perels, Franziska; Otto, Barbara & Schmitz, Bernhard (2008): Spezielle Auswertungsmethoden der Pädagogischen Psychologie. In: Schneider, Wolfgang; Hasselhorn, Marcus & Bengel, Jürgen (Hrsg.): *Handbuch der pädagogischen Psychologie*. Göttingen: Hogrefe, 712-720.
- Pintrich, Paul R. (2000): The role of goal orientation in self-regulated learning. In: Boekaerts, Monique; Pintrich, Paul R. & Zeidner, Moshe (Hrsg.): *Handbook of self-regulation*. San Diego, California, London: Academic Press, 451-502.
- Pintrich, Paul R.; Smith, David A.F.; Garcia, Teresa & McKeachie, Wilbert J. (1991): *Motivated Strategies for Learning Questionnaire (MSLQ) Manual*. Ann Arbor, Michigan: The University of Michigan.
- Rahimirad, Maryam & Moini, Mohammad R. (2015): The Challenges of Listening to Academic Lectures for EAP Learners and the Impact of Metacognition on Academic Lecture Listening Comprehension. *SAGE Open* 5: 2 [DOI 10.1177/2158244015590609].
- Revelle, William (2018): *psych: Procedures for Psychological, Psychometric, and Personality Research* (R package version 1.8.4).
- Robitzsch, Alexander; Kiefer, Thomas & Wu, Margaret (2018): *TAM: Test analysis modules* (R package version 2.12-18).
- Rochecouste, Judith; Oliver, Rhonda & Mulligan, Denise (2012): English language growth after university entry. *International Journal of Educational Research* 53, 1-8.
- Roll, Ido & Winne, Philip H. (2015): Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics* 2: 1, 7-12.
- Rose, Heath (2012a): Language Learning Strategy Research: Where Do We Go From Here? *Studies in Self-Access Learning Journal* 3: 2, 137-148.
- Rose, Heath (2012b): Reconceptualizing Strategic Learning in the Face of Self-Regulation: Throwing Language Learning Strategies out with the Bathwater. *Applied Linguistics* 33: 1, 92-98.
- Rose, Heath; Briggs, Jessica G.; Boggs, Jill A.; Sergio, Lia & Ivanova-Slavianskaia, Natalia (2018): A systematic review of language learner strategy research in the face of self-regulation. *System* 72, 151-163.
- Rosseel, Yves (2012): An R Package for Structural Equation Modeling. *Journal of Statistical Software* 48, 1-36.
- Rubin, Joan (1975): What the "Good Language Learner" Can Teach Us. *TESOL Quarterly* 9: 1, 41.
- Sachverständigenrat deutscher Stiftungen für Integration und Migration (2017): *Allein durch den Hochschuldschungel: Hürden zum Studienerfolg für internationale Studierende und Studierende mit Migrationshintergrund* [Online: <https://www.svr-migration.de/publikationen/hochschuldschungel/>, 27.08.2018].
- Schiefele, Ulrich & Wild, Klaus-Peter (1994): Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens. *Zeitschrift für Differenzielle und Diagnostische Psychologie* 15: 4, 185-200.
- Schlagmüller, Matthias & Schneider, Wolfgang (2007): *Würzburger Lesestrategie-Wissenstest für die Klassen 7–12: Ein Verfahren zur Erfassung metakognitiver Kompetenzen bei der Verarbeitung von Texten*. Göttingen: Hogrefe.

- Schrader, Friedrich-Wilhelm; Helmke, Andreas; Wagner, Wolfgang; Nold, Günter & Schröder, Konrad (2008): Lernstrategien im Fach Englisch. In: DESI-Konsortium (Hrsg.): *Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie*. Weinheim/Basel: Beltz, 270-282.
- Spörer, Nadine & Brunstein, Joachim C. (2006): Erfassung selbstregulierten Lernens mit Selbstberichtsverfahren: Ein Überblick zum Stand der Forschung. *Zeitschrift für Pädagogische Psychologie* 20: 3, 147-160.
- Statistisches Bundesamt (2016): *Bildung und Kultur: Studierende an Hochschulen. Wintersemester 2015/2016* [Online: https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/StudierendeHochschulenEndg2110410167004.pdf?__blob=publicationFile, 27.08.2018].
- Swain, Merrill; Huang, Li-Shih; Barkaoui, Khaled; Brooks, Lindsay & Lapkin, Sharon (2009): *The Speaking Section of the TOEFL IBT™ (SSTIBT): Test-takers' Reported Strategic Behaviors* [Online: <https://files.eric.ed.gov/fulltext/ED507806.pdf>, 27.08.2018].
- Taguchi, Naoko (2001): L2 Learners' Strategic Mental Processes during a Listening Test. *JALT Journal* 23: 2, 6-31.
- Teng, Lin S. & Zhang, Lawrence J. (2016): A Questionnaire-Based Validation of Multidimensional Models of Self-Regulated Learning Strategies. *The Modern Language Journal* 100: 3, 674-701.
- Tschirner, Erwin & Jones, Randall L. (2006): *The Herder-BYU electronic corpus of contemporary German*. Leipzig: Herder-Institut.
- Tseng, Wen-Ta; Dörnyei, Zoltán & Schmitt, Norbert (2006): A New Approach to Assessing Strategic Learning: The Case of Self-Regulation in Vocabulary Acquisition. *Applied Linguistics* 27: 1, 78-102.
- Vandergrift, Larry (2003): Orchestrating Strategy Use: Toward a Model of the Skilled Second Language Listener. *Language Learning* 53: 3, 463-496.
- VanderStoep, Scott W. & Pintrich, Paul R. (2003): *Learning to learn: The skill and will of college success*. Upper Saddle River, NJ: Prentice Hall.
- Veenman, Marcel V.J. (2005): The assessment of metacognitive skills: What can be learned from multi-method designs. In: Artelt, Cordula & Moschner, Barbara (Hrsg.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis*. Münster/New York/München/Berlin: Waxmann, 77-99.
- Wild, Klaus-Peter (2000): *Lernstrategien im Studium: Strukturen und Bedingungen*. Münster: Waxmann.
- Winne, Philip H. (2010): Bootstrapping learner's self-regulated learning. *Psychological Test and Assessment Modeling* 52: 4, 472-490.
- Wirth, Joachim (2004): *Selbstregulation von Lernprozessen*. Münster: Waxmann.
- Wirth, Joachim (2005): Selbstreguliertes Lernen in komplexen und dynamischen Situationen: Die Nutzung von Handlungsdaten zur Erfassung verschiedener Aspekte der Lernprozessregulation. In: Artelt, Cordula & Moschner, Barbara (Hrsg.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis*. Münster/New York/München/Berlin: Waxmann, 101-127.
- Wirth, Joachim & Leutner, Detlev (2008): Self-Regulated Learning as a Competence: Implications of theoretical models for assessment methods. *Zeitschrift für Psychologie/Journal of Psychology* 216: 2, 102-110.
- Wisniewski, Katrin (2014): *Die Validität der Skalen des Gemeinsamen europäischen Referenzrahmens für Sprachen: Eine empirische Untersuchung der Flüssigkeits- und Wortschatzskalen des GeRS am Beispiel des Italienischen und des Deutschen*. Frankfurt u. a.: Peter Lang.
- Wisniewski, Katrin (2016): Strategische Kompetenz in der Fremdsprache: Ein Forschungsüberblick. In: Brunetti, Simona; Klingebeil-Schieke, Josephine; Pedron, Chiara M.; Piotrowski, Marie-Christin; Ruggieri, Antonella & Schreiber, Rebecca (Hrsg.): *Versprachlichung von Welt - Il mondo in parole: Festschrift zum 60. Geburtstag von Maria Lieber = Il mondo in parole*. Tübingen: Stauffenburg Verlag, 637-658.

- Wisniewski, Katrin (2018): Studienerfolg und sprachliche Kompetenzen bei Bildungsausländer/-innen: Eine Längsschnittstudie an den Universitäten Leipzig und Würzburg. *InfoDaF* 45: 4, 573-597.
- Wolters, Christopher A. & Won, Sungjun (2017): Validity and the Use of Self-Report Questionnaires to Assess Self-Regulated Learning. In: Zimmerman, Barry J. & Schunk, Dale H. (Hrsg.), 307-322.
- Wong, Lillian C. & Nunan, David (2011): The learning styles and strategies of effective language learners. *System* 39: 2, 144-163.
- Yeldham, Michael (2016): Examining the usefulness to learners of processes taught in L2 English listening courses. *International Review of Applied Linguistics in Language Teaching* 54: 1, 43-72.
- Zimmerman, Barry J. (2000): Attaining self-regulation: A social cognitive perspective. In: Boekaerts, Monique; Pintrich, Paul R. & Zeidner, Moshe (Hrsg.): *Handbook of self-regulation*. San Diego, California, London: Academic Press, 13-39.
- Zimmerman, Barry J. & Schunk, Dale H. (Hrsg.) (2017): *Handbook of self-regulation of learning and performance*. London: Routledge.