

1. Einleitung

Der Kompetenzbegriff ist zentraler Zielbegriff öffentlicher Schulbildung in Deutschland (BMBF, 2007) und die Kompetenzorientierung wird als neues Paradigma für das Lernen in Schule und Arbeitswelt bezeichnet (Max, 1999). Als Teil der Kompetenzorientierung ist ein neuer Modus der output-orientierten, bildungspolitischen Steuerung etabliert worden (Slepcevic-Zach & Tafner, 2012), dessen Grundidee die verbindliche Definition von kompetenzorientiert formulierten Bildungsstandards und deren regelmäßige Überprüfung durch empirische Testverfahren in länderübergreifenden Vergleichsstudien ist (z. B. Stanat et al., 2017). Durch diesen output-orientierten Modus bildungspolitischer Steuerung kommt der Kompetenzerfassung eine besondere bildungspolitische Relevanz zu (Niemann, 2010; Slepcevic-Zach & Tafner, 2012); die Einführung von kompetenzorientiert formulierten Bildungsstandards und deren regelmäßige Überprüfung wird als „zentrales Element“ einer evidenzbasierten Steuerung des Bildungssystems betrachtet (Zuber et al., 2019, XIII) – „the assessment of competencies plays a key role in optimizing educational processes and improving the effectiveness of educational systems“ (Leutner et al., 2017, S. 5). Dementsprechend wurden die Konzeption von Kompetenzmodellen, die Entwicklung von Verfahren zur Kompetenzerfassung sowie die Interpretation und Anwendung von Ergebnissen der Kompetenzerfassung als zentrale Aufgaben der fachdidaktischen und empirischen Bildungsforschung identifiziert (Klieme et al., 2008; Leutner et al., 2017). Die vorliegende Arbeit

betrachtet den zweiten und den dritten genannten Aspekt und dabei insbesondere spezifische Herausforderungen, die sich aus der Komplexität des Kompetenzbegriffs für die Entwicklung von Instrumenten zur Kompetenzerfassung sowie für die valide Interpretation von damit erzielten Ergebnissen ergeben.

Die theoriebasierte Entwicklung von Testaufgaben zur Kompetenzdiagnostik wird von Hartig und Jude (2007) als „deduktive oder rationale Testkonstruktion“ (S. 27) bezeichnet. Hierbei werden auf der Basis eines theoretischen Konstrukts Testaufgaben entwickelt, von Expert*innen hinsichtlich der Passung mit dem theoretischen Konstrukt beurteilt und anschließend empirisch evaluiert (Hartig & Jude, 2007; Jonkisz et al., 2012). In diesem Zusammenhang präsentieren Terzer, Hartig und Upmeyer zu Belzen (2013) die folgenden Schritte für eine systematische Testkonstruktion: (1) die Formulierung der theoretischen Fundierung, (2) die Testkonzeption, (3) die Systematisierung der Aufgabenkonstruktion, (4) die Entwicklung einer Konstruktionsanleitung, (5) die Aufgabenentwicklung, (6) die Aufgabenerprobung und -selektion und (7) die Festlegung des Erhebungsdesigns. Für Schritt 6 schlagen die Autor*innen weitergehend vor, die konstruierten Aufgaben mit Hilfe verschiedener statistischer Maße zu beurteilen (z. B. Schwierigkeit, Trennschärfe, Reliabilität) und auf deren Basis gezielt Aufgaben zu selektieren. Ziel dieses Vorgehens ist es, einen Test zu entwickeln, der eine hohe interne Konsistenz aufweist und gleichzeitig zwischen Personen mit unterschiedlicher Merkmalsausprägung ausreichend gut differenziert (Kelava & Moosbrugger, 2012). Die Forderung einer hohen internen Konsistenz spiegelt die Annahme

eines homogenen zu erfassenden Konstrukts wider (*Unidimensionalität*), weshalb bei mehrdimensionalen Konstrukten die Konsistenzprüfung separat für die einzelnen Dimensionen erfolgt (die dann jeweils als unidimensional betrachtet werden; Blömeke et al., 2015; Hartig & Jude, 2007; Terzer et al., 2013). Goldstein (2004, S. 322) kritisiert dieses Vorgehen der rationalen Testkonstruktion im Zusammenhang mit der PISA-Studie:

One implication of this unidimensionality assumption is that even if a second dimension is present but only expresses itself in terms of a few items, such items are more likely to be regarded as 'dodgy' when only a single dimension is allowed in the analysis.

Im Zusammenhang mit der Erfassung von Kompetenzen, die als komplexe und kontextabhängige Leistungsdispositionen definiert werden (Max, 1999; Rychen & Salganik, 2003; Weinert, 2001a), ist die Berücksichtigung relevanter Dimensionen und Faktoren von zentraler Relevanz. So kann beispielsweise die Selektion von Aufgaben auf der Basis statistischer Maße (und der damit verbundenen Annahme über die Dimensionalität des Konstrukts) zur Vernachlässigung „merkmalsbezogener Teilkompetenzen“ (Prenzel et al., 2002, S. 124) und damit eines relevanten Einflusses unterschiedlicher Kontextmerkmale auf die Aufgabenbearbeitung führen (vgl. Rychen & Salganik, 2003; Shavelson, 2013). Ähnlich diskutiert Taber (2018) das Ziel einer hohen internen Konsistenz von Testaufgaben kritisch, da dies, insbesondere bei der Erfassung

komplexer Konstrukte, auf eine Unterrepräsentation des Konstrukts im Test hindeuten könnte (vgl. AERA et al., 2014; Blömeke et al., 2015; Messick, 1996). Gleichzeitig kann eine hohe interne Konsistenz darauf beruhen, dass „a lot of the variance is due to general respondent-related factors (e.g. intelligence, study diligence, motivation in the subject)“ (Taber, 2018, S. 1292). Dieses Beispiel verdeutlicht, dass im Rahmen der Testentwicklung zur Kompetenzerfassung spezifisch für das zu erfassende Konstrukt geprüft werden muss, inwiefern statistische Maße die anvisierte Interpretation der Testergebnisse unterstützen (vgl. Kane, 2013a; Messick, 1995).

Die Forderung nach der systematischen Erforschung der einer Erfassung zugrundeliegenden Konstrukte (z. B. die Frage nach der Kontextabhängigkeit von Wissenschaftsverständnis; Leach et al., 2000; Urhahne et al., 2011) sowie der Analyse von weiteren kontextuellen Einflussfaktoren auf die Testbearbeitung (z. B. Choi et al., 2014; Levin et al., 1998; Roney & Lehman, 2008) ist nicht erst im Zusammenhang mit Kompetenzerfassung relevant, sondern wird traditionell als Prüfung der Konstruktrepräsentation (*construct representation*) im Zuge der Testevaluation in der pädagogisch-psychologischen Diagnostik diskutiert (vgl. Embretson, 1983; Hartig & Frey, 2012; Mislevy, 2007). Allerdings transportiert der Kompetenzbegriff durch seine Komplexität, unscharfe Extensionalität sowie Kontextabhängigkeit (Franke, 2005; Hartig, 2008; Max, 1999; Weinert, 2001a) besondere Anforderungen an eine valide Erfassung (Klieme et al., 2008; Rychen & Salganik, 2003; Shavelson, 2013).

Vor diesem Hintergrund muss die Erforschung von Grundlagen, Wirkzusammenhängen und möglichen Methoden der Kompetenzerfassung zentraler Teil der naturwissenschaftsdidaktischen Forschung sein: „The measurement of students’ competencies and the analysis of their determinants is one of the most important research questions in a world in which knowledge and competencies are becoming more and more the key resources.“ (Nachtigall et al., 2008, S. 328). Nachtigall et al. (2008) schränken allerdings gleichzeitig ein, dass „an integrated theory of determinants and their causal functioning with respect to students’ competencies is still missing“ (S. 330). In diesem Sinne argumentiert auch Mislevy (2007), dass die Entwicklung von Testinstrumenten aus einer konsequent kognitiven Perspektive den argumentativen Nachweis einer validen Testwertinterpretation stützen kann. Hierbei müssen die Performanz der Person, Eigenschaften des Testinstruments sowie oftmals weitere Faktoren berücksichtigt werden, um ein Testergebnis valide interpretieren zu können. Ähnlich mahnen auch Leuders und Sodian (2013) vor „allzu einfachen Rückschlüssen von Kovariationsstrukturen interindividueller Unterschiede auf die Dynamiken mentaler Prozesse“ (S. 32).

Vor diesem Hintergrund sind die Ziele der vorliegenden Arbeit wie folgt:

- (1) Die theoretische Begründung der für die Kompetenzerfassung besonderen Relevanz einer Auseinandersetzung mit kontextuellen Einflussfaktoren auf die Testbearbeitung (Abschnitte 2, 3).
- (2) Die Ableitung eines Rahmenmodells zu relevanten Einflussfaktoren bei der Kompetenzerfassung in der naturwissenschafts-didaktischen Forschung (Abschnitt 4).
- (3) Die Zusammenfassung der eigenen Arbeiten mit Bezug zum vorgeschlagenen Rahmenmodell (Abschnitt 5).
- (4) Die kritische Diskussion des vorgeschlagenen Rahmenmodells auf der Basis der eigenen Arbeiten (Abschnitt 6).
- (5) Die Ableitung von weiterführenden Fragestellungen für die naturwissenschafts-didaktische Forschung (Abschnitt 7).

2. Kompetenzen

In diesem Abschnitt wird die besondere Relevanz einer Auseinandersetzung mit kontextuellen Einflussfaktoren auf die Testbearbeitung für die Kompetenzerfassung hergeleitet, indem zunächst die Kompetenzorientierung als neues bildungspolitisches Paradigma erläutert, anschließend der Kompetenzbegriff analysiert und darauf aufbauend Konsequenzen für die Kompetenzerfassung abgeleitet werden.¹

2.1 Kompetenzorientierung

Der Kompetenzbegriff ist zentraler Zielbegriff öffentlicher Schulbildung in Deutschland (BMBF, 2007; Lahner, 2011; Veith, 2003; Ziener, 2006). Die Kompetenzorientierung wird als neues Paradigma für das Lernen in Schule und Arbeitswelt bezeichnet (Max, 1999; vgl. Calenda & Tammaro, 2015; Dubs, 2004), das in Deutschland maßgeblich durch die Rezeption der Ergebnisse der ersten Erhebungen des *Programme for International Student Assessment* (PISA; vgl. Klieme et al., 2010) ausgelöst wurde (Lahner, 2011; Niemann, 2010; Veith, 2003). Durchgeführt durch die *Organisation for Economic Cooperation and Development* (OECD) sollen in den PISA-Studien Kompetenzen von Schüler*innen erhoben werden, die „in modernen Gesellschaften für eine befriedigende Lebensführung in persönlicher und wirtschaftlicher Hinsicht sowie für eine aktive Teilnahme am gesellschaftlichen Leben notwendig

¹ Die Analyse des Kompetenzbegriffs ist in Teilen Krell (2013, 2020) entnommen.

sind“ (Baumert et al., 2001, S. 29). Kerngedanke dabei ist, dass in modernen Gesellschaften aufgrund der schnellen technologischen Entwicklung, der zunehmenden Konkurrenz zwischen Staaten und Wirtschaftsräumen sowie der Flexibilisierung des Arbeitsmarktes ein immenser Modernisierungsdruck auf Staaten, Unternehmen und Individuen entsteht, der „über die Ausbildung fachlicher Fertigkeiten und überfachlicher Schlüsselqualifikationen hinaus[gehend], [...] die grundständige Entwicklung und Nutzung individueller Kompetenzen“ notwendig werden lässt (Veith, 2003, S. 17; vgl. Delamare Le Deist & Winterton, 2005; Max, 1999). Die Aufgaben im naturwissenschaftlichen Bereich der PISA-Studien erfassen beispielsweise – orientiert am Konzept der naturwissenschaftlichen Grundbildung (*scientific literacy*; vgl. Gräber et al., 2002; Roberts & Bybee, 2014) – die Fähigkeiten

[...] naturwissenschaftliche Fragestellungen zu erkennen, Phänomene naturwissenschaftlich zu erklären und naturwissenschaftliche Beweise in realen wissenschafts- und technologiebezogenen Situationen zu nutzen, um diese zu interpretieren, zu lösen und entsprechende Entscheidungen zu treffen.

Dieser Ansatz wurde gewählt, um der Art von Kompetenzen Rechnung zu tragen, auf die es in unserer modernen Gesellschaft ankommt, und zwar in vielen Lebensbereichen, vom Erfolg bei der Arbeit bis hin zum aktiven bürgerschaftlichen Engagement. (OECD, 2007, S. 39)

Bildungssysteme, die diese Kompetenzen erfolgreich entwickeln, werden vor dem beschriebenen Hintergrund als bedeutender Faktor für die (zukünftige) wirtschaftliche Konkurrenzfähigkeit von Staaten betrachtet (Hanushek & Woessmann, 2011; OECD, 2010). Die Ergebnisse der PISA-Studien werden daher im Indikatorenprogramm der OECD berücksichtigt (OECD, 2015). Diese volkswirtschaftliche Sichtweise auf Bildung erzeugte, insbesondere aufgrund des relativ schlechten Abschneidens der deutschen Schüler*innen in der ersten PISA-Studie im Jahr 2000 und deren starker medialen Rezeption, einen bildungspolitischen Problem- und Handlungsdruck (Veith, 2003). Es wurde befürchtet, „dass sich negative Bildungsleistungen mittelbar auch auf die wirtschaftliche Performanz Deutschlands im internationalen Wettbewerb niederschlugen“ (Niemann, 2010, S. 69–70). Reaktionen auf diesen „PISA-Schock“ (Knodel et al., 2010) waren unter anderem strukturelle Änderungen im Bildungswesen (z. B. Ausbau der Ganztagschulangebote, Einführung zentraler Abiturprüfungen; Niemann, 2010), die zunehmende Professionalisierung der Lehrkräftebildung (Sälzer & Prenzel, 2017) sowie ein neuer Modus der outputorientierten, bildungspolitischen Steuerung (Slepcevic-Zach & Tafner, 2012; Volante & Fazio, 2017; Zuber et al., 2019). Grundidee hierbei ist die verbindliche Definition von kompetenzorientiert formulierten Bildungsstandards und deren regelmäßige Überprüfung durch empirische Testverfahren in länderübergreifenden Vergleichsstudien: „Die Vergleichsstudien sollten vorrangig auf die Entwicklung grundlegender Kompetenzen ausgerichtet werden,

die die Schülerinnen und Schüler zu einer gleichberechtigten Teilhabe am gesellschaftlichen, wirtschaftlichen, kulturellen und politischen Leben befähigen.“ (KMK, 1997, S. 1).

In diesem Sinne wurden zum Beispiel für die naturwissenschaftlichen Fächer Biologie, Chemie und Physik Bildungsstandards für den Mittleren Schulabschluss verabschiedet (KMK, 2004). Das Institut zur Qualitätsentwicklung im Bildungswesen entwickelt auf der Grundlage der bislang definierten Standards für die Primar- und die Sekundarstufe I Testaufgaben und überprüft in regelmäßigen Abständen, inwieweit die angestrebten Kompetenzen auch tatsächlich von den Schüler*innen erreicht werden (z. B. Stanat et al., 2017).

Das Paradigma der Kompetenzorientierung und die damit in Verbindung stehenden Schulleistungsstudien (z. B. PISA; vgl. Klieme et al., 2010) werden von einigen Autor*innen kritisch diskutiert (z. B. Lahner, 2011; Oser, 2016); zum Beispiel bezüglich des jeweils umgesetzten Testverfahrens (z. B. Goldstein, 2004, 2017; Meyerhöfer, 2006) oder grundsätzlicher als Funktionalisierung von Bildung im Interesse der ökonomischen Wettbewerbsfähigkeit (z. B. Höhne, 2012; Komatsu & Rappleye, 2017; Münch, 2009). Trotz dieser Kritik hat sich das Paradigma der Kompetenzorientierung in Deutschland weitgehend durchgesetzt. So sind wie beschrieben sowohl die Bildungsstandards für den Mittleren Schulabschluss (z. B. für das Fach Biologie: KMK, 2005) als auch die Rahmenlehrpläne der meisten Bundesländer (z. B. in Berlin für das Fach Biologie: SenBJS, 2006a, 2006b) kompetenzorientiert formuliert. Ten-

denziell ersetzt der Kompetenzbegriff etablierte Begriffe wie Wissen, Kenntnisse und Qualifikationen (Max, 1999). Veith (2003) hebt hervor, dass die Attraktivität des Kompetenzbegriffs mit seiner Funktionalität im obigen Sinne zusammenhängt: Aufgrund der schnellen technologischen Entwicklung und der immer komplexer werdenden Arbeitswelt liegt die Bedeutung des Lernens nunmehr insbesondere darin, „die in der Umwelt permanent erzeugte Komplexität prozessual durch Kompetenzentwicklung zu reduzieren“ (S. 321).

2.2 Kompetenzbegriff

Im Folgenden werden unterschiedliche Vorschläge zur Definition des Kompetenzbegriffs skizziert und zusammengeführt. Hierbei werden insbesondere Ansätze vorgestellt, die in der fachdidaktischen Forschung breit rezipiert wurden (Klieme et al., 2008; Weinert, 2001a). Der Kompetenzbegriff war vor seiner Anwendung als Zielkonstrukt für die allgemeine Schulbildung bereits im Kontext der beruflichen Ausbildung etabliert und die zunehmende Popularität des Kompetenzbegriffs in der Erziehungswissenschaft und den Fachdidaktiken kann als eine Art „Ansteckung“ (Max, 1999, S. 114) aus dem Berufsfeld verstanden werden, nach der die Ziele der allgemeinen Schulbildung flexibilisiert und an den Bedürfnissen der Arbeitswelt ausgerichtet werden (Max, 1999). Daher werden ergänzend Definitionen des Kompetenzbegriffs aus diesem Bereich vorgestellt (Franke, 2005; Veith, 2003).