

Im Fokus: Potenziale von Large-Scale Assessments und Längsschnittstudien zur Kompetenzentwicklung über die Lebensspanne – analytische und orientierende Perspektiven

1. Einleitung

In den letzten Jahrzehnten hat in nationalen und internationalen Bildungsstudien die Messung von Kompetenzen einen großen Bedeutungszuwachs erfahren. Die internationale Kompetenzdebatte ist mit der Vorstellung verbunden, dass Bildung als Humankapital eine wichtige Ressource für individuelle und gesellschaftliche Entwicklungen darstellt. Dennoch muss man feststellen, dass bis heute keine einheitlichen Begriffe und Auffassungen über den Begriff „Kompetenz“ existieren und auch nicht darüber, wie Kompetenzen gemessen werden und wie sie sich im Lebensverlauf entwickeln. Es ist daher notwendig, in jedem nationalen und internationalen Forschungsprojekt eine klare Definition des Kompetenzbegriffs und der zu messenden relevanten Kompetenzen zugrunde zu legen, um diese dann beim jeweiligen Untersuchungsdesign zu berücksichtigen – wie das beispielsweise in der PIAAC- und CiLL-Studie geschehen ist.

In diesem Beitrag werden die Traditionsstränge der Kompetenzdebatte kurz skizziert, wobei auf zentrale Diskussionspunkte in der aktuellen Debatte sowie auch auf Kompetenzkonstrukte eingegangen wird. Anschließend wird insbesondere das Erkenntnispotenzial von Large-Scale Assessments im Hinblick auf die Kompetenzmessung verdeutlicht. Im Vordergrund des Interesses steht die Reflexion des aktuellen Forschungsstandes, um differenziert die Potenziale von Large-Scale Assessments zur Kompetenzentwicklung über die Lebensspanne zu gewinnen.

2. Tradition der Kompetenzdebatte

In den Sozial- und Erziehungswissenschaften hat die Kompetenzdebatte eine längere Tradition, die bis in die 1970er Jahre zurückreicht (vgl. Edelmann/Schmidt/Tippelt 2012, S. 175). Von besonderer Relevanz für die Bildungsforschung waren die Arbeiten zur Motivationspsychologie von White (1959), der Kompetenzen als Fähigkeiten definiert, die nicht angeboren sind, sondern von Individuen in Lernprozessen erworben werden. Zentral für die aktuelle Kompetenzdebatte sind auch die Studien zur Intelligenz von McClelland (1973), der Kompetenz und Intelligenz differenziert. Eine große Bedeutung haben ebenso die Analysen von Chomsky (1965), der zwischen Kompe-

tenz und Performanz im linguistischen Bereich unterschied (vgl. Klieme/Hartig 2007). Wichtig im deutschsprachigen Raum war Heinrich Roth (1971), weil er in den 1960er Jahren in seiner „pädagogischen Anthropologie“ als oberstes Ziel der Erziehung die Mündigkeit definiert und diese wiederum als Kompetenz für eine verantwortliche Handlungsfähigkeit sieht (ebd., S. 180). Bereits damals wurden individuelle Fähigkeiten, Dispositionen, Verhalten und Urteile differenziert und es wurde zwischen Selbst-, Sach- und Sozialkompetenz im wissenschaftlichen Diskurs unterschieden. Diese Kompetenzbereiche wurden dann von Fend und Weinert weiterentwickelt (vgl. Klieme 2011). Weinert versteht im Rahmen der internationalen Kompetenzdebatte und in seiner Expertise für die OECD Kompetenzen als

die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, so wie die der damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, um die Problemlösung in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können (Weinert 2001, S. 27).

Kompetenzen entstehen offenbar auf der Basis von kontextbezogenen Lernprozessen und konkreten Erfahrungen in sehr bestimmten fachlichen Domänen, und dies unterscheidet den Begriff der Kompetenzen von dem Begriff der Intelligenz, weil Intelligenz vorwiegend als stabile kognitive Grundfähigkeit zu betrachten ist. Kompetenzen von Individuen zeigen also die situative Bewältigung von Anforderungen in einem jeweils definierten Wissens- und Domänenbereich und dies auf einem bestimmten Niveau.

Kompetenzen sind kontextabhängig und basieren auf kognitiven Lernprozessen, die in der Auseinandersetzung des Individuums mit seiner Umwelt stattfinden. Im Unterschied hierzu sind kognitive Grundfunktionen im geringeren Maße erlernbar (vgl. Weinert 2001). Darüber hinaus können allgemein vermittelte generische Kompetenzen (z.B. analytisches Denken) oder domänenspezifische Kompetenzen in einem jeweiligen Lernbereich ausgebildet und modelliert werden.

Im Bereich der theoriegeleiteten Kompetenzmodellierung sind in den letzten Jahren zahlreiche Fortschritte zu verzeichnen, auf die allerdings in diesem Überblicksartikel nicht näher eingegangen werden kann (vgl. Fleischer u.a. 2013; Blossfeld/Roßbach/Maurice 2011). Für die Erfassung von Kompetenzen werden vor allem eindimensionale, kontinuierliche Leistungsskalen (in Form von Rasch-Modellen) verwendet (vgl. Hartig 2007; Leuders/Sodian 2013; zur geeigneten Skalierung vgl. Brennan 2011).

Die Kompetenzerfassung ist – in einem anderen Feld – auch ein Fokus des sogenannten Deutschen Qualifikationsrahmens, der als nationales Pendant des Europäischen Qualifikationsrahmens (EQR) die Vergleichbarkeit und Anschlussfähigkeit von Qualifikationen beschreibt. Der DQR soll durch einen weitgehenden Konsens aller zentralen bildungspolitischen Akteure das lebenslange Lernen sowie die Beschäftigungsfähigkeit fördern. Gleichzeitig steht der DQR für eine stärkere Kompetenzorientierung des deutschen Bildungssystems, wobei Fachkompetenzen sowie soziale

und personale Kompetenzen differenziert werden. Im DQR wird Methodenkompetenz als Querschnittskompetenz aufgefasst und findet aber in der DQR-Matrix keine explizite Erwähnung. Kompetenzen gelten als übergreifende Kategorien, die nach Niveauintikatoren angeordnet sind (vgl. Rein 2010).

3. Erkenntnispotenzial von internationalen Large-Scale Assessments

Aus der Perspektive der Lebensverlaufsforschung wird man sagen müssen, dass in Large-Scale-Studien primär das Kindes- und Jugendalter erforscht wurde. Große Bedeutung erlangte die Kompetenzmessung in den Large-Scale Assessments, insbesondere in den internationalen Schulleistungsmessungen. Dieser politiknahe Bildungsforschungsbereich wurde durch die IEA (*International Association for the Evaluation of Educational Achievement*) und durch die OECD geprägt. Die sogenannte PISA-Studie, die TIMSS-Studie, die PIRLS-Studie stehen jeweils für Programme internationaler Schulleistungsvergleiche, wobei PISA für die Bildungsberichterstattung der OECD (OECD 2006ff.) besondere Bedeutung hat, weil turnusmäßig Leistungsindikatoren zur Verfügung gestellt werden sollen (vgl. Baumert u.a. 2001). TIMSS (*Third International Mathematics and Science Study*) ist dagegen ein Erhebungsprogramm, mit dem die IEA auf periodischer Grundlage international vergleichbare Leistungsdaten ausschließlich im mathematisch-naturwissenschaftlichen Vergleich erhoben hat (vgl. Baumert u.a. 1997). PIRLS (*Progress in International Reading and Literacy study*) und dann auch IGLU, die internationale Grund- und Leseuntersuchung, haben das Leseverständnis im Grundschulalter thematisiert und analysiert (vgl. Bos u.a. 2003). Large-Scale Assessments basieren immer auf repräsentativen Stichproben und ausgewählten Altersgruppen, um dann komparative Aussagen über das Niveau und die Verteilung von Leistungsergebnissen formulieren zu können. Bei den IEA-Untersuchungen, die im vierjährigen Rhythmus (TIMSS) und seit 2001 im fünfjährigen Rhythmus (PIRLS) erhoben werden, spielt immer auch die Erfassung der Lerngelegenheiten eine wichtige Rolle. Man kann sagen, dass im Wettbewerb die OECD ein eigenes Forschungsprogramm (PISA) entwickelt hat, das im dreijährigen Turnus das Leseverständnis sowie mathematische und naturwissenschaftliche Basiskompetenzen erfasst. In den letzten Jahren wurden auch selbstreguliertes Lernen und allgemeine Problemlösefähigkeiten erhoben und analysiert, wobei bei PISA davon auszugehen ist, dass die Basiskompetenzen als eine Voraussetzung für selbstständiges Weiterlernen und den erfolgreichen Übergang in andere Lernphasen im beruflichen und allgemeinen Bereich aufzufassen sind. Ohne an dieser Stelle auf die Schulleistungsmessungen und deren Ergebnisse genauer einzugehen, ist festzuhalten, dass diese internationalen Vergleiche in den letzten Jahren immer wieder auch kritisch diskutiert wurden (vgl. Baumert 2007). Beispielsweise ist es wichtig, dass Wissensbereiche vergleichbare Lerngelegenheiten voraussetzen, wobei gerade beim Leseverständnis die curriculare Validität der

Tests eine geringere Rolle spielt als bei den mathematisch-naturwissenschaftlichen Bereichen, weil auch außerhalb der Schule breite Lernmöglichkeiten existieren. Setzt die Testphilosophie von TIMSS und PIRLS eine transkulturelle curriculare Validität der Aufgaben voraus, so hat die OECD sehr viel stärker die Validität etwa der PISA-Aufgaben für berufliche, zivilgesellschaftliche und politische Teilhabebereiche in den Blick gerückt. Es lässt sich sagen, dass für die kognitiven Anforderungen der PISA-Aufgaben eine transkulturelle unterrichtliche Validität nur eingeschränkt gilt und man auch wenig über die prognostische Validität der Tests in den internationalen Studien weiß. Zudem basieren Large-Scale Assessments zunächst auf Querschnittsuntersuchungen, so dass über den Verlauf der Kompetenzentwicklung keine sicheren Aussagen abgeleitet werden können. Sicher allerdings kann man sagen, dass in diesen Large-Scale Assessments eine Äquivalenz der Testaufgaben notwendig ist, um international fair messen zu können. Dies bedeutet,

dass Personen mit gleicher Fähigkeit Aufgaben länderübergreifend jeweils mit vergleichbarer Lösungswahrscheinlichkeit bearbeiten. Dies sicherzustellen ist eine große Herausforderung, die immer nur annäherungsweise bewältigt werden kann (Baumert 2007, S. 361).

In aktuellen Large-Scale Assessments werden Umweltmerkmale keineswegs einseitig als Umweltdeterminismus interpretiert, sondern es kommt darauf an, welchen aktiven Einfluss Heranwachsende oder jüngere wie auch ältere Erwachsene auf ihre Umwelt nehmen und zwar sowohl in der Auswahl als auch in der Gestaltung ihres jeweiligen Entwicklungskontextes. Auch die Qualität von Beziehungen am Arbeitsplatz oder im Freundeskreis kann Wirkungen auf die Kompetenzentwicklung nehmen. Die Stärken und Potenziale der internationalen Large-Scale Assessments liegen primär auf dem Gebiet der Testkonstruktion, wobei es die auf der Item-Response-Theorie basierende Skalierung ermöglicht, sowohl kriteriumsorientierte als auch an sozialen Normen ausgerichtete Befunde zu interpretieren. Die seit über 20 Jahren existierenden Erfahrungen mit Large-Scale Assessments im Kindes- und Jugendalter haben auch in die PIAAC- und die CiLL-Studie und die dortige Kompetenzmessung von Erwachsenen Eingang gefunden. Während es in PIAAC um alle Kompetenzniveaus geht, zielt die leo. – Level-One Studie (Literalität von Erwachsenen auf den unteren Kompetenzniveaus) primär auf das untere Niveau, den sogenannten Level One und untersucht den Grad der Literalität der deutsch sprechenden Bevölkerung (vgl. Grotlüschen/Riekmann 2012).

4. Kompetenzmessung im Erwachsenenalter: PIAAC und CiLL

Eine der ersten großen Erhebungen auf der internationalen Ebene und somit ein Vorläufer der aktuellen international vergleichenden Studien ist der *International Adult Literacy Survey* (IALS, internationale Untersuchung der Lese- und Schreibfähigkeit

von Erwachsenen), eine Initiative der Organisation for Economic Co-operation and Development (OECD), die von Anfang bis Mitte der 1990er Jahre in mehr als 20 Ländern durchgeführt wurde. Dabei umfasste die Lesefähigkeit insgesamt drei Teildimensionen: „prose literacy“, „document literacy“ sowie „quantitative literacy“ (OECD 2000). Die IALS-Studie betont Zusammenhänge zwischen Lesefähigkeit und bestimmten individuellen und gesellschaftlichen Variablen, wie etwa dem Einkommen, dem Berufsstatus sowie dem Gesundheitsstatus und dem bürgerschaftlichen Engagement.

Das Vorhaben wurde im Rahmen der Studien *Adult Literacy and Life Skills* (ALL-Studie über die Lese- und Schreibfähigkeiten und Lebenskompetenzen Erwachsener) und des Projekts *Programme for the International Assessment of Adult Competencies* (PIAAC, internationale Vergleichsstudie der Kompetenzen Erwachsener) in den letzten Jahren fortgesetzt. Der zentrale Nutzen dieser Studien liegt im Überblick zum Stand der Grundkompetenz des Lesens und Schreibens (literacy) in der Bevölkerung (vgl. Clair 2012; OECD 2005).

Es ist zunächst darauf hinzuweisen, dass sich die Kompetenzbegriffe von PIAAC (Program for the International Assessment of Adult Competencies) und auch von CiLL (*Competencies in Later Life*) von anderen Kompetenzbegriffen in der Erwachsenen- und Weiterbildungsforschung klar unterscheiden. Die Analyse berufsbezogener Kompetenzen wie Sozial-, Fach-, Lern-, Personal- oder Handlungskompetenz, wie man dies in verdienstvollen Studien aus betriebspädagogischen und betrieblichen Weiterbildungskontexten kennt (vgl. Erpenbeck/Rosenstiel 2003), ist nicht Gegenstand der Analyse und auch die Kompetenzkonzeption des Deutschen Qualifikationsrahmens (DQR), die auf Fach- und personale Kompetenzen abhebt, unterscheidet sich vom Kompetenzbegriff und der Kompetenzlogik in der PIAAC- und CiLL-Studie erheblich (vgl. Tippelt 2014).

Unter Kompetenz wird in der PIAAC- und CiLL-Studie die Fähigkeit verstanden, in einer bestimmten Situation Wissen anzuwenden, Werkzeuge zu benutzen, kognitive und praktische Strategien und Routinen einzusetzen, um diese in alltagsnahen Situationen angemessen anzuwenden und die dortigen Probleme zu bewältigen. Gemessen wird bei PIAAC allerdings nicht umfassend der Bereich von Überzeugungen und Werten oder auch zahlreicher weiterer alltagsnaher Kompetenzen (wie Gesundheitskompetenz, zivilgesellschaftliche Kompetenz, interkulturelle Kompetenz), sondern es wird auf einen spezifischen Ausschnitt möglicher Kompetenzen eingegangen, nämlich auf die Lesekompetenz, die alltagsmathematische Kompetenz und die technologiebasierte Problemlösekompetenz. Es lässt sich sagen, dass PIAAC und CiLL den Weiterbildungsdiskurs stärker an die von PISA gemessenen Kompetenzen heranführen, keinesfalls allerdings dadurch den Bildungsdiskurs in der Erwachsenenbildung und in der Erziehungswissenschaft ablösen (vgl. Tippelt 2014; Tippelt/Edelmann 2007).

Für die PIAAC- und die CiLL-Studie sind allerdings operationalisierbare und klar definierte Kompetenzmodelle notwendig, weil man nur so einerseits vergleichend und andererseits altersspezifisch Kompetenzen erheben und interpretieren kann. Kon-

zentriert sich die PIAAC-Studie auf das Kompetenzniveau von erwachsenen 16- bis 65-Jährigen im internationalen Vergleich, so analysiert die CiLL-Studie entsprechende Kompetenzen der 66- bis 80-Jährigen in Deutschland, was durchaus interessante altersbezogene Vergleiche zwischen diesen beiden Studien anregen kann.

Die OECD hat als zentrale Ziele für PIAAC benannt, dass Qualifikationen der erwachsenen Bevölkerung im erwerbsfähigen Alter gemessen werden sollen, um so auf einer empirisch fundierten Grundlage auch politische und soziale Interventionen insbesondere in den Bereichen der Bildungs- und Arbeitsmarktpolitik vorzubereiten. In Deutschland wurde die PIAAC-Studie vom Leibniz-Institut GESIS als Gesamtkoordinator von Anfang 2011 bis 2013, die CiLL-Studie vom Deutschen Institut für Erwachsenenbildung in Kooperation mit der LMU München und der Uni Tübingen von Ende 2011 bis 2014 durchgeführt. Die repräsentative Zufallsstichprobe der 16- bis 65-Jährigen betrug bei PIAAC letztlich 5.465 Probanden, wobei der Ausschöpfungsgrad der Gesamtstichprobe bei 55 Prozent lag, bei CiLL waren dies 1.308 Probanden bei einer Ausschöpfung der Bruttostichprobe von 40 Prozent (vgl. umfangreiche methodische Hinweise in Friebe/Gebrende 2013).

Die gemessenen Kompetenzen wurden wie folgt definiert. *Lesekompetenz* meint die Fähigkeit,

geschriebene Texte zu verstehen, zu bewerten, zu nutzen und sich mit diesen nachhaltig zu beschäftigen, um sich am Leben in der Gesellschaft zu beteiligen, die eigenen Ziele zu erreichen, sein Wissen weiterzuentwickeln und das eigene Potenzial zu entfalten (Rammstedt 2013, S. 33).

Es war also wichtig, den Sinn eines Satzes zu erfassen, längere Textpassagen flüssig zu lesen, um die Inhalte umfassend zu verstehen. Bei *alltagsmathematischer Kompetenz* geht es darum,

sich mathematische Informationen und Ideen zugänglich zu machen, diese anzuwenden, zu interpretieren und zu kommunizieren, um mit mathematischen Anforderungen in unterschiedlichen Alltagssituationen Erwachsener umzugehen (Rammstedt, 2013, S. 47).

Wichtig ist es, die entsprechenden Aufgaben alltagsnah und alltagssprachlich zu formulieren, wobei in den Debatten zur Studie angeregt wurde, bei künftigen Analysen die Tests noch näher am Erlebens- und Erfahrungsraum Älterer auszurichten. Unter *technologiebasiertem Problemlösen* wird letztlich

die Verwendung von digitalen Technologien, Kommunikationswerkzeugen und Netzwerken mit dem Ziel, Informationen zu beschaffen und zu bewerten, mit anderen zu kommunizieren sowie alltagsbezogene Aufgaben zu bewältigen (Rammstedt 2013, S. 61)

verstanden. Beispielsweise geht es darum, virtuelle Formulare zu bearbeiten, E-Mails zu sortieren und zu versenden und auch den Informationsgehalt wie auch die Glaubwürdigkeit von verschiedenen Internetseiten konkret zu beurteilen.

Die Stärke der PIAAC- und CiLL-Studien liegt zweifelsohne in der sozialdifferenzierenden Analyse von Kompetenzen. Ohne an dieser Stelle schon auf die Ergebnisse einzugehen, lässt sich sagen, dass zwischen Männern und Frauen, zwischen Altersgruppen, zwischen Bildungsgruppen, zwischen Gruppen mit sehr unterschiedlichem elterlichen Bildungshintergrund und auch zwischen regionalen Gruppen unterschiedliche Ausprägungen der jeweiligen Kompetenzen diagnostiziert werden konnten.

Wird insbesondere das Alter von Probanden betrachtet, ist es problematisch, die entsprechenden Kompetenzmessergebnisse als Alterseffekte oder als generative Kohorteneffekte zu interpretieren. Mit anderen Worten ist zu klären, ob für den Kompetenzzuwachs oder auch den Kompetenzabfall das biologische Alter verantwortlich ist oder ob die jeweiligen Kohorten sehr unterschiedliche Zugänge zu Bildung und Lerngelegenheiten aufgrund ihrer Generationslagerung (vgl. Mannheim 1928) haben – eine Generationslagerung, die letztlich die Kompetenzergebnisse stark beeinflusst. Diese Fragestellung leitet über zu den Möglichkeiten und Notwendigkeiten längsschnittlicher Untersuchungen.

Außerdem ergibt sich beim technologiebasierten Problemlösen im Vergleich zu anderen Domänen eine etwas geringere Ausschöpfungsquote, was insbesondere für die CiLL-Ältere-Studie zutrifft, bei der diejenigen Personen, die über keine PC-Kenntnisse verfügen oder den Eingangstest nicht bestehen, in dieser Domäne nicht getestet werden können.

5. Longitudinalstudien als erweitertes Erkenntnispotenzial

Für das Erfassen zeitlicher Bezüge und um Aussagen über Entwicklungseinflüsse und Entwicklungsverläufe von Kompetenzausprägungen machen zu können, ist die Wahl des Forschungsdesigns im engeren Sinne von großer Bedeutung. In der sozialwissenschaftlichen Jugend- und Erwachsenenbildungsforschung hat sich der Rückgriff auf Retrospektivdaten als schwierig erwiesen, da die Erinnerungen durch aktuelle Befindlichkeiten überlagert sind und allenfalls Ereignisse, die markant sind (z.B. Arbeitsplatzverlust oder Übergang in das Nacherwerbsalter) zuverlässig im Gedächtnis verbleiben. Erinnerungen an zurückliegende Ereignisse und Lebensbedingungen müssen immer als mentale Repräsentationen gewertet werden, die biografisch Neubewertungen und auch Umdeutungen unterliegen. Besser geeignet, um den sozialen Wandel, das Weiterbildungsverhalten und die Kompetenzentwicklung im Zeitverlauf zu beurteilen, sind replikative Zeitwandelstudien, Längsschnittstudien und schließlich längsschnittliche Kohorten-Sequenzanalysen. Worin liegt nun das jeweils besondere Potenzial dieser verschiedenen Studiendesigns für die Bildungsforschung und insbesondere für die Erwachsenenbildung?

Replikative Zeitwandelstudien sind in der Erwachsenenbildung gut verankert: Solche Studien untersuchen zu verschiedenen Zeitpunkten mit dem gleichen Frageninventar strukturell ähnliche Bevölkerungsgruppen – aber nicht dieselben Personen. Der AES (*Adult Education Survey*) und das zurückliegende Berichtssystem Weiterbildung waren bzw. sind solche Replikationsstudien, mit denen Veränderungen des Weiterbildungsverhaltens und der -einstellungen zeitlich gut untersucht werden können (vgl. BMBF 2013; Gnahs 2010, S. 279ff.). In replikativen Zeitwandelstudien werden zwar Zeitvergleiche durchgeführt, allerdings nicht mit denselben Personen, sondern Replikationsstudien zielen eher auf Generationeneffekte bzw. auf Kohorteneffekte und damit auf den sozialen Wandel ab. Individuelle Veränderungen oder altersbezogene Veränderungen können mit replikativen Zeitwandelstudien nicht aufgedeckt werden. In Replikationsstudien werden unabhängige Stichproben Erwachsener aus zwei oder auch zahlreichen Geburtskohorten miteinander verglichen, wobei die unabhängigen Stichproben in Bezug auf zentrale Merkmale, z.B. das Alter oder die Bildung, sich sehr ähnlich sein müssen, um verglichen werden zu können. Weil in Replikationsstudien auf den sozialen Wandel geschlossen wird, muss immer auch kontrolliert werden, welche besonderen gesellschaftlichen Entwicklungen mögliche Antworttendenzen (z.B. auch Ergebnisse von Kompetenzmessungen) beeinflussen. Es ist bei Replikationsstudien als besonderes Problem hervorzuheben, dass die Messungen, die hier repliziert werden, auch wirklich vergleichbar sein müssen. Wenn zwischen den Erhebungen Kompositionseffekte zugrunde liegen, die Zusammensetzung der Stichprobe also nicht mehr zentralen Kriterien entspricht, so würden Replikationen ohne Ergebnis bleiben, weil dann beispielsweise nicht-vergleichbare Bildungskohorten miteinander in Beziehung gebracht würden. Es gibt tatsächlich im Kontext von Erwachsenenstudien das Problem, dass ältere Jahrgänge mehr Auszubildende und mehr Erwerbstätige, mehr Personen mit weniger Bildungserfahrung beinhalten, als jüngere Altersgruppen, so dass dann bei Replikationsstudien tatsächlich die Schwierigkeit auftritt, dass Kompositionseffekte zu Verzerrungen führen. Es müssen also die Stichproben in jeder Hinsicht methodisch kontrolliert werden, damit gezielte Zeitvergleiche möglich sind, und das Frageninventar darf bei Replikationsstudien nicht verändert oder „modernisiert“ werden.

Worin liegt also das Potenzial von *replikativen* Zeitwandelstudien? Einerseits prägen sehr unterschiedliche historische Ereignisse, soziale Wandlungsprozesse, Veränderungen des Bildungssystems, generell institutionelle Veränderungen die verschiedenen Altersgruppen, andererseits werden diese Faktoren von den Altersgruppen im Erwachsenenalter auch sehr unterschiedlich erlebt. Die hier angesprochenen Prozesse des sozialen Wandels lassen sich in replikativen Zeitwandelstudien analysieren, die zu unterschiedlichen Zeitpunkten einen vergleichbaren Personenkreis (nicht dieselben Personen), etwa einer definierten Altersgruppe, wiederholt vergleichbar befragen oder testen.

Das Potenzial von *längsschnittlichen* Untersuchungen und Längsschnittdaten, die mit wiederholten Erhebungen mit denselben Personen zu unterschiedlichen Zeitpunkten arbeiten, ist anders einzuschätzen, weil hier validere Auskünfte über die Wirkung

von Lernumgebungen und anderen Faktoren auf die personale Entwicklung zu erwarten sind. Solche Längsschnittdaten sind Replikationsdaten oder Querschnittstudien mit einmaliger Datenerhebung bei unterschiedlichen Altersgruppen überlegen, weil einzelne Altersgruppen nicht nur im biologischen Alter differieren, sondern Altersgruppen auch systematisch Unterschiede hinsichtlich ihrer Entwicklungsbedingungen in den verschiedenen biografischen Phasen aufweisen können (vgl. Walper/Tippelt 2010). Längsschnittstudien haben das Potenzial, kausale Hypothesen über die Zusammenhänge von Umweltbedingungen und individuellem Verhalten, Erleben und eben auch der Kompetenzentwicklung zu überprüfen. Dies vor allem dann, wenn in Längsschnittstudien nicht nur individuelle Veränderungen, sondern auch die vorauslaufenden und die nachfolgenden Veränderungen der Umweltbedingungen miterfasst werden, so dass Prozessverläufe genauer beschrieben werden können. Insbesondere Übergangssituationen etwa bei Erwachsenen von Ausbildungseinrichtungen in eine Beschäftigung, Veränderungen von Beschäftigungsformen, Mobilitätsprozessen durch Veränderungen des Wohnortes, aber auch der Übergang in die nachberufliche Lebensphase sind durch solche Längsschnittstudien besser und präziser zu analysieren.

Längsschnittstudien sind aber auch mit einigen typischen Problemen konfrontiert, beispielsweise der bereits erwähnten Konfundierung der Wirkung von Alter und Erhebungszeitpunkt, aber auch den Schwierigkeiten der Stichprobengewinnung und des Dropouts von Untersuchungsteilnehmern über längere Untersuchungszeiträume sowie dem doch erheblich höheren finanziellen Aufwand. Der Dropout, insbesondere bei langfristigen Studien, ist sicher eine besondere Herausforderung in Längsschnittstudien. Auch sind die Intervalle zwischen Messzeitpunkten nicht standardisiert festzulegen, sondern hängen immer von den inhaltlichen Fragestellungen einer Studie ab. Es gibt Veränderungen, die sich relativ rasch und in kurzer Zeit vollziehen und entsprechend schnell abbilden lassen. Es gibt aber auch Veränderungen, die relativ stabile Merkmale umschreiben und die dann nur durch längere Erhebungsintervalle aufgezeigt werden können (z.B. Altersbilder). Vonseiten der befragten Erwachsenen erfordern Längsschnittstudien ein doch sehr hohes Engagement, und neben der motivational bedingten Mortalität im Rahmen von Längsschnittstudien müssen auch dem Schul-, Berufs- und Wohnortwechsel geschuldete Verluste in Längsschnittstudien festgehalten werden. Ausfälle in Längsschnittstudien erschweren immer die Generalisierbarkeit von Ergebnissen, insbesondere dann, wenn diese Ausfälle selektiv bestimmte soziale Gruppen oder bei der Kompetenzmessung auch spezifische Domänen betreffen (z.B. *numeracy* oder Problemlösekompetenzen). Solche selektiven Ausfälle sind besonders deshalb schwierig, weil sie die Stichprobe systematisch verändern und die Repräsentativität in den zu untersuchenden Kohorten leidet. An der Stelle sind Dropout-Analysen sowie Selektivitätsanalysen hilfreich, um genaue Angaben über die Stichprobenentwicklung zu erhalten. Weitere Schwierigkeiten resultieren aus Problemen der Testungseffekte, also dem Einfluss, der direkt von Testungen auch auf die späteren Erhebungen ausgeht. Es kann zu Routinen kommen, es kann zu Übungseffekten

kommen, es kann aber auch zur Ermüdung und zum Motivationsverlust der Befragten kommen, die dann das Antwortverhalten beispielsweise bei der Kompetenzmessung beeinflussen.

Dennoch ist zusammenfassend zu sagen, dass Längsschnittstudien und replikative Zeitwandelstudien Garanten in der Bildungsforschung und in der Erwachsenenbildung dafür sind, über Wandel, Entwicklungen und Veränderungen empirisch belastbare Aussagen machen zu können. Längsschnittstudien gehen meist von klar definierten Stichproben aus und untersuchen diese mit dem gleichen Frageninventar zu verschiedenen Messzeitpunkten. Im Unterschied zu reinen Replikationsstudien erlauben Längsschnittstudien allerdings Aussagen über altersbezogene Veränderungen von Individuen, aber auch von sozialen Gruppen. Auch die Stabilität und die Veränderung von Merkmalen wie Einstellungen, Verhaltensweisen, aber auch von Kompetenzen lassen sich ermitteln, weil man analysieren kann, inwieweit Personen ihre Position innerhalb der Untersuchungsgruppe beibehalten und wie sich mit dem Alter Veränderungen vollziehen. Wenn auch Längsschnittstudien in der Regel durchgeführt werden, um altersbezogene Veränderungen zu verstehen, ist doch hervorzuheben, dass in Längsschnittstudien Alter und Erhebungszeitpunkt konfundieren. Steigt beispielsweise im Verlauf einer solchen Studie die Kompetenz Erwachsener, so mag dies eher aktuellen sozialen Tendenzen geschuldet sein, etwa bestimmten Bildungsmaßnahmen und weniger altersabhängigen Veränderungen. Nur längsschnittliche Kohorten-Sequenz-Analysen, wie sie beispielsweise im NEPS (*National Educational Panel Study*) zugrunde gelegt werden, ermöglichen es, Alter und Erhebungszeitpunkt getrennt zu berücksichtigen, um auf diese Art und Weise Alterseffekte, dann sogenannte Kohorten- oder auch Generationeneffekte und eben auch Zeiteffekte voneinander unterscheiden zu können.

Die komplexeste Form des Untersuchungsdesigns stellen diese längsschnittlichen Kohorten-Sequenz-Modelle zur Analyse von Alterskohorten unter Berücksichtigung von Testzeiteffekten dar. Um die bereits angesprochenen Probleme von Quer- und Längsschnittstudien sowie von replikativen Zeitwandelstudien zu minimieren, wird hier der Versuch unternommen, die verschiedenen Datenerhebungsstrategien zu kombinieren (vgl. Baltes/Reese/Nesselroade 1977, S. 132ff.). Vor allem im entwicklungspsychologischen Terrain – aber für die Erwachsenenbildungsforschung wegweisend – wurde berücksichtigt, dass jedes Verhalten als Ergebnis von Alterseffekten, Kohorten- bzw. Generationeneffekten und Testzeiteffekten interpretiert werden kann (vgl. Schaie 1965). Moderne Kohorten-Sequenz-Analysen erweitern Längsschnittstudien durch den Einbezug mehrerer Kohorten. Genau dies ist die grundlegende Idee des NEPS (National Educational Panel Study), das für den Bildungsbereich und den Bildungsverlauf die validesten Ergebnisse in Bezug auf Alterseffekte, Generationeneffekte und Testzeiteffekte erwarten lässt. Es wird im Folgenden knapp auf das Untersuchungsdesign von NEPS eingegangen.

6. Potenziale für die Erwachsenenbildung in längsschnittlichen Bildungsstudien am Beispiel der Panel-Studie NEPS

Im NEPS (*National Educational Panel Study*) werden Bildungsprozesse und Kompetenzentwicklung von der frühen Kindheit bis ins hohe Erwachsenenalter erhoben und analysiert. Die Daten werden in fünf zentralen Bereichen (sogenannten Säulen) erhoben, die als theoretischer und methodischer Rahmen im Kontext des Lebensverlaufs dienen: (1) Kompetenzentwicklung im Lebenslauf, (2) Bildungsprozesse in lebenslaufspezifischen Lernumwelten, (3) soziale Ungleichheit und Bildungsentscheidungen im Lebenslauf, (4) Bildungsprozesse von Personen mit Migrationshintergrund sowie schließlich (5) Renditen von Bildung. Diese fünf Bereiche gliedern sich je nach Altersgruppe in acht Etappen: (a) Neugeborene und frühkindliche institutionelle Betreuung, (b) Kindergarten und Übergang in die Grundschule, (c) Grundschule und Übergang in die Sekundarstufe I, (d) Wege durch die Sekundarstufe I und Übergänge in die Sekundarstufe II, (e) Wege durch die Sekundarstufe II und Übergänge in den tertiären Bereich oder den Arbeitsmarkt, (f) Übergänge in die berufliche Ausbildung und in den Arbeitsmarkt, (g) Hochschulstudium und Übergang in den Beruf, (h) Bildung im Erwachsenenalter und Lebenslanges Lernen (vgl. Blossfeld u.a. 2011).

Methodisch folgt das Nationale Bildungspanel einem Multi-Kohorten-Sequenz-Design. Es werden mehrere Startstichproben gezogen, die sowohl an die Übergänge im Bildungssystem als auch an die Übergänge zwischen Bildungssystem und Arbeitsmarkt gekoppelt sind. Die repräsentativen Stichproben, die aus Neugeborenen, vierjährigen Kindergartenkindern, Fünftklässlern, Neuntklässlern, Studienanfängern und 23- bis 64-jährigen Erwachsenen bestehen, werden in Studien individuell begleitet, um zu analysieren, welche Kompetenzen sich im Lebenslauf entwickeln, wie Kompetenzen und Entscheidungswege gerade im Hinblick auf die verschiedenen Übergänge im Lebenslauf zusammenhängen und welche familiären Einflüsse bedeutsam sind. Bei der Erhebung wird zwischen formaler und non-formaler Weiterbildung und informellem Lernen differenziert, was auch internationalen Standards entspricht (vgl. Kleinert/Matthes 2010).

In der Erwachsenenkohorte des Nationalen Bildungspanels werden Erwachsene im erwerbsfähigen Alter befragt, um repräsentative Daten über Bildung und Weiterbildung sowie die damit einhergehende Kompetenzentwicklung in diesem Lebensabschnitt zu gewinnen. Hierzu müssen alle Bildungsaktivitäten und zentralen Lernprozesse im Erwachsenenalter möglichst mit den vorhergehenden Entscheidungswegen erfasst werden, die zur Aufnahme der jeweils nächsten Bildungsaktivität geführt haben. Ebenso wird der bisherige Lebensverlauf dokumentiert, da die Entscheidungen untrennbar auf der Basis der jeweiligen Lebensverläufe entstehen (vgl. Bäumeier u.a. 2012). Die Erwachsenen werden zu sprachlichen, mathematischen und naturwissenschaftlichen Kompetenzen und zu informations- und kommunikationsbasierten Problemlösekompetenzen befragt. Darüber hinaus werden Daten zu Metakognition, so-

zialer Kompetenz sowie Merkmalen der Persönlichkeit und der Motivation erhoben. Mit diesem breiten Spektrum an Paneldaten lassen sich u.a. langfristige Entwicklungstendenzen des Lernverhaltens abbilden.

Die Rekrutierung der Studienteilnehmer erfolgte bereits im Jahr 2007 (noch bevor NEPS gegründet wurde), die erste Erwachsenenbefragung fand unter dem Namen „Arbeiten und Lernen im Wandel“ (ALWS, vgl. Matthes u.a. 2012) statt und wurde vom Institut für Arbeitsmarkt- und Berufsforschung (IAB) durchgeführt. Aufbauend auf dieser ersten Befragung wurden die ersten Daten im Rahmen des Nationalen Bildungspanels (NEPS) von November 2009 bis Juni 2010 erhoben. Die Stichprobe bilden Erwachsene im erwerbsfähigen Alter unabhängig von ihrem aktuellen Erwerbsstatus, wobei bereits in der ersten Welle über 11.600 Personen erreicht werden konnten (vgl. Bäumer u.a. 2012). Damit die Befragung der Teilnehmer möglichst kohärent bleibt, erfolgt die Datenerfassung in den jeweiligen Bereichen mittels eines modular gegliederten Lebensverlaufsinterviews. Eine innovative Ergänzung zu den bisherigen Replikationsstudien und den Large-Scale Assessments im Erwachsenenalter besteht in der Erhebung von Merkmalen der Prozessqualität der wahrgenommenen Weiterbildungsangebote, die anhand von drei Dimensionen gemessen wird: (1) Strukturierungsniveau, (2) Ausmaß an Unterstützung bei den Lernarrangements sowie (3) Herausforderung und Aktivierung. Diese Dimensionen wurden im Rahmen der NEPS-Studie erstmalig auf die Weiterbildungsaktivitäten von Erwachsenen transferiert.

Die ersten Ergebnisse der Befragungen der Jahre 2009 und 2010 zeigen, dass mehr als die Hälfte der befragten Erwachsenen Lehrgänge, Kurse oder Weiterbildungen besuchten. Am häufigsten werden Weiterbildungsangebote zwischen dem 35. und 50. Lebensjahr wahrgenommen, da sich viele Erwachsene vor dem 35. Lebensjahr noch in einer Ausbildung befinden und nach dem 50. Lebensjahr Lehrgänge seltener vom Betrieb finanziert werden (vgl. Prusog-Wagner 2013). Im Hinblick auf die Lese- und Rechenkompetenzen zeigt sich ähnlich wie bei den PIAAC-Befunden, dass die jüngeren Studienteilnehmer bessere Ergebnisse beim Lesen und Rechnen erzielen als die Älteren.

Interessant sind jedoch auch die großen Überschneidungen zwischen den beiden Altersgruppen: Den ersten NEPS-Ergebnissen zufolge sind in Bezug auf die Lesekompetenz fast keine Unterschiede zu beobachten. Erst mit 50 bzw. 60 Jahren fallen die Ergebniswerte beim Lesen deutlich zurück, was möglicherweise unter anderem auf die nachlassende Sehkraft zurückzuführen ist (ebd.). Diese Ergebnisse werden allerdings dann durch die CiLL-Daten für das höhere Lebensalter erheblich vertieft. Sicher ist, dass sich das Potenzial der NEPS-Studie und des dort angelegten längsschnittlichen Kohorten-Sequenz-Designs vor allem für das Erwachsenenalter erst in den nächsten Jahren erweisen kann. Insbesondere wird zu prüfen sein, inwieweit die Kompetenzen der verschiedenen Altersgruppen auf biologisches Alter oder auf soziale Kohorten- und Generationeneffekte zurückzuführen sind.

7. Fazit

In Anlehnung an wissenschaftstheoretische Überlegungen ist gerade für diesen Forschungsbereich festzustellen, dass wir empirisch fundiertes Wissen mit großer praktischer Bedeutung haben und dass uns die genannten Studien eine tiefere Einsicht und ein besseres Verständnis des Lernens von Erwachsenen vermitteln können. Allerdings ist es auch so, dass wir „mit jedem Schritt, den wir vorwärts machen, mit jedem Problem, das wir lösen“ (Popper 1969, S. 103), fortwährend neue und ungelöste theoretische und methodische Herausforderungen entdecken.

Literatur

- Baltes, P.B./Reese, H.W./Nesselroade, J.R. (1977): *Life-Span Development Psychology: Introduction to Research Methods*. Belmont
- Bäumer, T./Blossfeld, H.-P./Janik, F./Maurice, J. v. (2012): Lebensverläufe im Längsschnitt. Der Beitrag des Nationalen Bildungspanels für ein Monitoring des Lernens im Erwachsenenalter. In: *DIE Zeitschrift für Erwachsenenbildung*, H. 4, S. 35–39
- Baumert, J. (2007): Internationale Schulleistungsmessung. In: Tenorth, E./Tippelt, R. (Hg.): *Lexikon Pädagogik*. Weinheim/Basel, S. 358–361
- Baumert, J./Artelt, C./Klieme, E./Nuebrand, M./Prenzel, M./Schiefele, U. u.a. (Hg.) (2001): *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen
- Baumert, J./Lehmann, R./Lehrke, M./Schmitz, B./Clausen, M./Hosenfeld, I. u.a. (1997): *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich: Deskriptive Befunde*. Opladen
- Blossfeld, H.-P./Roßbach, H.-G./Maurice, J. v. (Hg.) (2011): *Education as a lifelong process. The German National Educational Panel Study (NEPS)*. In: *Zeitschrift für Erziehungswissenschaft. Sonderheft 14*
- BMBF – Bundesministerium für Bildung und Forschung (Hg.) (2013): *Weiterbildungsverhalten in Deutschland. AES 2012 Trendbericht*. Bielefeld
- Bos, W./Lankes, E.-M./Prenzel, M./Schwippert, K./Walther, G./Valentin, R. (Hg.) (2003): *Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich*. Münster
- Brennan, R.L. (2011): Generalizability theory and classical test theory. In: *Applied Measurement in Education*, H. 1, S. 1–21
- Chomsky, N. (1965): *Aspects of the Theory of Syntax*. Cambridge
- Clair, R.S. (2012): The limits of levels: Understanding the international adult literacy surveys (IALS). In: *International Review of Education*, H. 6, S. 759–776
- Edelmann, D./Schmidt, J./Tippelt, R. (2012): *Einführung in die Bildungsforschung*. Stuttgart
- Erpenbeck, J./Rosenstiel, L. (2003): Einführung. In: Erpenbeck, J./Rosenstiel, L. (Hg.): *Handbuch Kompetenzmessung. Erkennen, verstehen und bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis*. Stuttgart, S. IX–XL
- Fleischer, J./Koeppen, K./Kenk, M./Klieme, E./Leutner, D. (2013): Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms. In: *Zeitschrift für Erziehungswissenschaft*, H. 1, S. 5–22
- Friebe, J./Gebrande, J. (2013): Kompetenzen im höheren Lebensalter – die nationale PIAAC-Erweiterungsstudie „GILL“. In: *Report. Zeitschrift für Weiterbildungsforschung*, H. 3, S. 48–60
- Gnahs, D. (2010): Kompetenzprofile. In: *Deutsches Institut für Erwachsenenbildung (DIE) (Hg.): Trends der Weiterbildung. DIE-Trendanalyse 2010*. Bielefeld, S. 163–169
- Grotlüschen, A./Riekman, W. (Hg.) (2012): *Funktionaler Analphabetismus in Deutschland. Ergebnisse der ersten IEO – Level-One Studie*. Münster

- Hartig, J. (2007): Skalierung und Definition von Kompetenzniveaus. In: Klieme, E./Beck, B. (Hg.): Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International). Weinheim, S. 83–99
- Kleinert, C./Matthes, B. (2010): Forschungsfeld „Lebenslanges Lernen“. Mit neuen Daten Wissenslücken schließen. In: IAB-Forum, H. 1, S. 52–57
- Klieme, E. (2011): Bildung unter undemokratischem Druck? Anmerkungen zur Kritik der PISA-Studie. In: Ludwig, L./Luckas, H./Hamburger, F./Aufenanger, S. (Hg.): Bildung in der Demokratie II: Tendenzen – Diskurse – Praktiken. Opladen, S. 289–302
- Klieme, E./Hartig, J. (2007): Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. Zeitschrift für Erziehungswissenschaft, Sonderheft 8, S. 11–29
- Leuders, T./Sodian, B. (2013): Inwiefern sind Kompetenzmodelle dazu geeignet, kognitive Prozesse von Lernenden zu beschreiben? In: Zeitschrift für Erziehungswissenschaft, H. 1, S. 27–33
- Mannheim, K. (1928): Das Problem der Generationen. Kölner Vierteljahreshefte für Soziologie, H. 2, S. 157–185, 309–333
- Matthes, B./Drasch, K./Erhardt, K./Künster, R./Valentin, M.-A. (2012): Arbeiten und Lernen im Wandel. Teil IV: Editionsbericht. FDZ-Methodenreport 3/2012. Methodische Aspekte zu Arbeitsmarktdaten
- McClelland, D.C. (1973): Testing for Competence rather than for Intelligence. In: American Psychologist, H. 1, S. 1–14
- OECD – Organisation for Economic Co-operation and Development (2000): Literacy in the information Age. Final Report of the international Adult Literacy Survey. Ottawa/Paris
- OECD – Organisation for Economic Co-operation and Development (2005): International Assessment of Adult Skills: Proposed Strategy. COM/DELSA/EDU (2005)4/REV1. 24.10.2005. Paris
- Prusog-Wagner, A. (2013): Neues aus der NEPS-Studie. Ergebnisse für Befragte: URL: www.neps-studie.de/binary_data/363_neues-aus-der-neps-studie-ergebnisse-f-r-befragte-2013.pdf
- Rammstedt, B. (Hg.) (2013): Grundlegende Kompetenzen Erwachsener im internationalen Vergleich. Ergebnisse von PIAAC 2012. Münster
- Rein, V. (2010): Deutscher Qualifikationsrahmen: Transformationskatalysator für Berufs- und Hochschulbildung. In: Swiss Political Science Review, H. 4, S. 821–825
- Roth, H. (1971): Pädagogische Anthropologie. Band II: Entwicklung und Erziehung. Hannover
- Schaie, K.W. (1965): A general model for the study of development problems. In: Psychological Bulletin, H. 2, S. 92–107
- Tippelt, R. (2014): Der pädagogische und bildungspolitische Rahmen der modernen Kompetenzdebatte – Zur Differenzierung von Kompetenzkonzepten. In: BAP (Hg.) (im Ersch.)
- Tippelt, R./Edelmann, D. (2007): Kompetenzentwicklung in der beruflichen Bildung und Weiterbildung. In: Prenzel, M./Gogolin, I./Krüger, H.-H. (Hg.): Kompetenzdiagnostik. Wiesbaden, S. 129–146
- Walper, S./Tippelt, R. (2010): Methoden und Ergebnisse der quantitativ orientierten Kindheits- und Jugendforschung. In: Krüger, H.-H./Grunert, C. (Hg.): Handbuch der Kindheits- und Jugendforschung. Wiesbaden, S. 205–243
- Weinert, F. (2001): Concept of Competence: A conceptual Clarification. In: Rychen, D.S./Hersh Salganik, L. (Hg.): Defining and Selecting Key Competencies. Seattle
- Weinert, F.E. (2001): Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In: Weinert, F.E. (Hg.): Leistungsmessungen in Schulen. Weinheim, S. 17–32
- White, R.W. (1959): Motivation reconsidered: The Concept of Competence. In: Psychological Review, H. 5, S. 297–333