Trautwein, Ulrich; Bertram, C.; Borries, B. von; Brauch, N.; Hirsch, M.; Klausmeier, K.; Körber, A.; Kühberger, C.; Meyer-Hamme, J.; Merkt, M.; Neureiter, H.; Schwan, S.; Schreiber, W. u.a.: *Kompetenzen historischen Denkens erfassen. Konzeption, Operationalisierung und Befunde des Projekts „Historical Thinking – Competencies in History“ (HiTCH)*. Münster: Waxmann Verlag 2017. ISBN: 978-3-8309-3598-8; 144 S.

Rezensiert von: Manuel Köster, Historisches Institut, Universität zu Köln

Der Diskurs um Kompetenzen historischen Denkens war einer der zentralen Gegenstände geschichtsdidaktischer Auseinandersetzungen der letzten zehn Jahre. Dies gilt insbesondere für den Bereich empirischer Forschung: Während einerseits zahlreiche Arbeiten existieren, die sich auf eines der deutschsprachigen Kompetenzmodelle beziehen, verstehen sich andere Studien als Beschreibung weiterer historischer Kompetenzen oder grenzen sich bewusst vom Kompetenzparadigma ab.¹ Der vorliegende, von insgesamt 18 Autoren aus Geschichtsdidaktik und empirischer Bildungsforschung verantwortete Band markiert dem Urteil der Verfasser zufolge einen „Meilenstein“ (S. 13) kompetenzorientierter geschichtsdidaktischer Empirie. Das Hauptziel des hier dokumentierten Projektes HiTCH (Historical Thinking – Competencies in History) bestand darin, „einen Test zu entwickeln, mit dem sich die Kompetenzen historischen Denkens von Schülerinnen und Schülern in Large-Scale-Assessments erfassen lassen“ (S. 11). Hier wird also die Entwicklung eines Instrumentes zur Kompetenzmessung beschrieben, nicht die Ausprägung historischer Kompetenzen in einer Schüler/innenpopulation.

Die Publikation richtet sich gleichermaßen an Adressaten in Geschichtsdidaktik und empirischer Bildungsforschung. Daraus erwächst die schwierige Aufgabe, die Darstellung so zu gestalten, dass sie für beide Rezipientenkreise anschlussfähig ist. Aus geschichtsdidaktischer Perspektive kann diese Herausforderung als weitgehend gemeistert bezeichnet werden. So wird in einem ersten Kapitel zunächst das Kompetenzmodell der

Gruppe FUER-Geschichtsbewusstsein² erläutert und kontextualisiert, wobei regelmäßig vor allem die Orientierungsfunktion historischen Denkens betont wird. Beim anschließenden, notwendig selektiven Überblick über den Stand empirischer Forschung bleibt allerdings offen, anhand welcher Kriterien die präsentierten Studien ausgesucht wurden.³ Gerade der knappe internationale Vergleich arbeitet die Spezifika des deutschen Kompetenzdiskurses aber sehr transparent heraus.

Die beiden Folgekapitel beschreiben den Entwicklungs- und Evaluationsprozess des HiTCH-Tests. Wie immer bei quantitativer Forschung bestand auch hier die zentrale Herausforderung für die Forscher/innen aus Deutschland, Österreich und der Schweiz darin, empirischen Gütekriterien gerecht zu werden, ohne dabei die Testlogik der Logik des zu erhebenden Konstrukts überzustülpen. Die eigentlichen Aufgaben, die die insgesamt 2.853 Schüler/innen der neunten Klassen aus 52 Schulen aller Regelschulformen in der Haupterhebung bearbeiten mussten, werden hier aus Gründen der Testsi- cherheit nicht offengelegt. Allerdings präsen- tieren die Autoren strukturell vergleichbare Aufgaben (S. 64–75), denen es mehrheitlich gelingt, Teiloperationen historischen Denkens zu isolieren und zu testen. Die Aufgaben zielen dabei jeweils auf einen der vier Kompetenzbereiche des FUER-Modells (Fragekompetenzen, Methodenkompetenzen, Orientierungskompetenzen, Sachkompetenzen). Sie tun dies jedoch in unterschiedlicher Intensität. Die historischen Frage- und Orientierungskompetenzen werden mit deut-

¹Für einen Überblick vgl. Manuel Köster / Holger Thünemann, Geschichtsdidaktische Forschung im Schatten von PISA, in: Zeitschrift für Didaktik der Gesellschaftswissenschaften 6,2 (2015), S. 34–55.

²Vgl. Andreas Körber / Waltraud Schreiber / Alexander Schöner (Hrsg.), Kompetenzen historischen Denkens. Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik, Neuried 2007.

³Hier wäre etwa auch eine Auseinandersetzung mit der von Schönemann, Thünemann und Zülsdorf-Kersting durchgeführten Analyse von Abiturklausuren lesenswert gewesen, die ihren Zugang dezidiert als Alternative zur Kompetenztestung verstehen. Vgl. Bernd Schönemann / Holger Thünemann / Meik Zülsdorf-Kersting, Was können Abiturienten? Zugleich ein Beitrag zur Debatte über Kompetenzen und Standards im Fach Geschichte, Berlin 2010, v.a. S. 34–36, 126f.

lich weniger Aufgaben angesprochen als die Methoden- und Sachkompetenzen⁴, obwohl die Orientierungsfunktion historischen Denkens in der Vorstellung des FUER-Modells besonders hervorgehoben wurde. Dies mag ein Hinweis auf die Tatsache sein, dass diese weniger leicht mittels standardisierter, geschlossener Items zu erheben sind – zumindest solange man, wie es hier völlig zurecht geschieht, die Fähigkeit zur historischen Orientierung testet (hier eher in der analytischen Variante als Identifikation unterschiedlicher Orientierungen in bestehenden Aussagen operationalisiert), nicht das Vorhandensein spezifischer Orientierungen. Bedingt durch die Tatsache, dass der historische Denkprozess hier in Einzeloperationen zergliedert wird und der Test zudem von inhaltlichem Vorwissen unabhängig sein muss, lassen sich einzelne der vorgestellten Aufgaben allerdings wohl auch durch kombinatorische Fähigkeiten oder durch reines Textverstehen⁵ lösen.

Nach zwei Pilotierungsdurchgängen und mehreren flankierenden Validierungsmaßnahmen wurden in der Haupterhebung insgesamt 15 Aufgaben mit 106 Items evaluiert.⁶ Diese Aufgaben adressieren jeweils das mittlere, konventionelle Niveau der Kompetenzausprägung. Zwar bezeichnen Trautwein und Kollegen die Testung eines transkonventionellen Niveaus mittels geschlossener Aufgaben als „nicht ausgeschlossen“ (S. 122), in der vorliegenden Form könne der Test „[q]ualitative Umschläge“ (ebd.) zu elaborierteren Formen historischen Denkens jedoch nicht erfassen. Tatsächlich erscheint zweifelhaft, ob die dafür notwendige Reflexivität und die damit einhergehende Idiosynkrasie historischen Denkens sich anhand eines auf einem Richtigkeitsstandard beruhenden Instrumentes erfassen lassen. Die psychometrische Qualität der Testaufgaben wurde anhand in der Geschichtsdidaktik bislang nicht annähernd erreichter methodischer und testtheoretischer Standards bestimmt. Es ist den Autoren hoch anzurechnen, dass die Darstellung dieser Prüfverfahren auch für ein geschichtsdidaktisches Publikum über weite Strecken nachvollziehbar bleibt und die hier zur Anwendung kommenden, sehr voraussetzungsvollen Konzepte kontextualisiert

und erläutert werden.

Die Autoren kommen zu dem Schluss, „dass das HiTCH-Instrument objektiv, reliabel und hinreichend valide ein Konstrukt erfasst, welches auf Grundlage einer dahinter stehenden Theorie als ‚Kompetenz historischen Denkens‘ bezeichnet werden kann“ (S. 117). Allerdings war es in Pilotierung und Hauptstudie nicht möglich, die Leistungen der Lernenden einzelnen Kompetenzbereichen empirisch zuzuordnen (vgl. u.a. S. 79, 94, 97). Stattdessen wurde die Varianz in der Schüler/innenleistung – neben Aufgabenspezifika, die „einen nicht komplett vernachlässigbaren Teil der Varianz“ (S. 97) erklären – auf einen „Generalfaktor Historisches Denken“ zurückgeführt. Auch wenn „die empirische ‚Bestätigung‘ der unterschiedlichen Kompetenzbereiche“ (S. 120) nicht beabsichtigt war, ist dieses Ergebnis doch bemerkenswert. So werden hier zwar unterschiedliche Erklärungen für diesen Befund angeführt – neben der engen Verflechtung der Kompetenzbereiche (S. 79f.) auch die „leichte Schiefelage“ (S. 199) in der Verteilung der Aufgaben auf die Kompetenzbereiche –, grundsätzliche Einwände gegen FUER und HiTCH werden aber nicht erörtert. Dabei wären auch zwei andere Ursachen denkbar. Zum einen könnte der Umstand, dass hier nur der „Generalfaktor“, nicht aber die Teilkompetenzen empirisch abgebildet werden konnten, darauf hinweisen, dass historisches Denken zu komplex ist, um mittels geschlossener Aufgaben erhoben zu werden, dass es sich also entweder nur analytisch, nicht aber empirisch in Teiloperationen zergliedern lässt, oder dass dafür qualitative Verfahren geeigneter sind. Zum anderen wäre vorstellbar,

⁴ Anhand der Angaben im vorliegenden Band ist nicht zweifelsfrei zu erkennen, ob sich eine oder zwei der 15 Aufgaben auf die Orientierungskompetenzen beziehen (vgl. S. 91, 95). Aufgrund der Tatsache, dass die Tabelle auf S. 91 offenbar fehlerhaft ist (so wird z.B. der Aufgabeninhalt der Aufgabe zur Orientierungskompetenz mit „Anordnen von Textbausteinen zu logischen Aussagen zum Thema Hexenverfolgungen“ beschrieben), ist eher von zwei Aufgaben auszugehen.

⁵ Die mittels HiTCH erhobene historische Kompetenz korreliert denn auch hoch mit Lesekompetenz und verbaler Intelligenz, was aber laut den Autoren ebenso für andere Kompetenztests gilt (vgl. S. 106–110).

⁶ Nach der Evaluation umfasst der Test nun 91 Aufgaben.

dass sich Teilkompetenzen historischen Denkens im Large-Scale-Modus zwar erheben lassen, dass aber das FUER-Modell diese Teilkompetenzen nicht angemessen modelliert.

In einem bilanzierenden Kapitel deuten die Autoren die Grenzen des HiTCH-Tests selbst an, verweisen aber vor allem auf dessen weitere Einsatzmöglichkeiten, etwa als abhängige Variable in Interventionsstudien. Durch eine geplante Weiterentwicklung des Tests sollen unter anderem die Ungleichverteilung der Aufgaben auf die Kompetenzbereiche behoben und neue Altersklassen angesprochen werden. Das ursprüngliche Vorhaben⁷ eines europaweiten Vergleichstests im Sinne eines Geschichts-PISA dagegen wird zwar als „wünschenswert“ (S. 128) bezeichnet, aber aktuell offenbar nicht aktiv verfolgt.

Insgesamt setzt die vorliegende Publikation neue methodische Standards im Bereich der „Messung“ historischen Denkens, sie sensibilisiert aber zugleich für die Grenzen derartiger Forschungszugänge. Der hier eingeschlagene Weg sollte unbedingt weiter verfolgt werden, verspricht er doch verlässliche Einblicke in spezifische Teilbereiche historischen Denkens. Eine Verabsolutierung des in diesem interdisziplinären Kooperationsprojekt vertretenen Konzepts empirischer Forschung, wie sie als Postulat gelegentlich anklingt, erscheint dagegen als unnötige Verengung von Forschungsperspektiven.

HistLit 2018-1-149 / Manuel Köster über Trautwein, Ulrich; Bertram, C.; Borries, B. von; Brauch, N.; Hirsch, M.; Klausmeier, K.; Körber, A.; Kühberger, C.; Meyer-Hamme, J.; Merkt, M.; Neureiter, H.; Schwan, S.; Schreiber, W. u.a.: *Kompetenzen historischen Denkens erfassen. Konzeption, Operationalisierung und Befunde des Projekts „Historical Thinking – Competencies in History“ (HiTCH)*. Münster 2017, in: H-Soz-Kult 13.03.2018.

⁷ Vgl. Waltraud Schreiber, *Geschichte denken, nicht pauken*, in: *Agora 2* (2012), S. 26–27.