

## Test zur Einschätzung wissenschaftlicher Widersprüche (TEWI): Messung wissenschaftsbezogenen Reflexionsvermögens in der Sekundarstufe II

Jochen Kramer  · Eva Thomm · Kerstin Oschatz-Grant · Wolfgang Wagner · Rainer Bromme

Eingegangen: 12. Juni 2019 / Überarbeitet: 29. Juni 2021 / Angenommen: 13. Juli 2021  
© Der/die Autor(en) 2021

**Zusammenfassung** Wissenschaftspropädeutik gehört zu den drei zentralen Bildungszielen der gymnasialen Oberstufe. Aufgrund eines Mangels an geeigneten Untersuchungsinstrumenten, gibt es jedoch kaum Studien zur Erreichung dieses Bildungsziels oder seines Einflusses auf den Werdegang junger Menschen. Im vorliegenden Beitrag werden die theoretischen Ansätze zur Wissenschaftspropädeutik kurz erläutert, bestehende Strukturmodelle vorgestellt und als zentrale Elemente die fächerübergreifende Denkweise und die kritisch-reflexive Perspektive auf Wissenschaft herausgearbeitet. Wissenschaftsbezogene Reflexionskompetenz im Umgang

---

Dr. Jochen Kramer (✉)  
Türkische Gemeinde in Baden-Württemberg e. V. (tgbw), Reinsburgstraße 82, 70178 Stuttgart, Deutschland  
E-Mail: [jochen.kramer@tgbw.de](mailto:jochen.kramer@tgbw.de)

Dr. Eva Thomm  
Erziehungswissenschaftliche Fakultät, Bildungsforschung und Methodenlehre, Universität Erfurt, Postfach 90 02 21, 99105 Erfurt, Deutschland  
E-Mail: [eva.thomm@uni-erfurt.de](mailto:eva.thomm@uni-erfurt.de)

Dr. Jochen Kramer · Dr. Kerstin Oschatz-Grant  
Eberhard Karls Universität Tübingen, Tübingen, Deutschland

Dr. Kerstin Oschatz-Grant (✉)  
Pasture Farm, Rushbrook Lane, Tanworth-in-Arden, B94 5HW, Vereinigtes Königreich  
E-Mail: [ksc.schatz@googlemail.com](mailto:ksc.schatz@googlemail.com)

Dr. Wolfgang Wagner  
Hector-Institut für Empirische Bildungsforschung, Eberhard Karls Universität Tübingen, Europastraße 6, 72072 Tübingen, Deutschland  
E-Mail: [wolfgang.wagner@uni-tuebingen.de](mailto:wolfgang.wagner@uni-tuebingen.de)

Prof. Dr. Rainer Bromme  
Institut für Psychologie, Westfälische Wilhelms-Universität, Fliegerstr. 21, 48149 Münster, Deutschland  
E-Mail: [bromme@uni-muenster.de](mailto:bromme@uni-muenster.de)

mit Widersprüchen in der Wissenschaft wird als eine Kernkomponente von Wissenschaftspropädeutik skizziert. Sie bildet auch den Ausgangspunkt für die Entwicklung des Tests zur Einschätzung wissenschaftlicher Widersprüche (TEWI) für das nationale Bildungspanel (NEPS). Im empirischen Teil werden Befunde zur Entwicklung und Pilotierung des TEWI sowie eine Studie zur konvergenten Konstruktvalidierung ( $N=394$ ) des TEWI vorgestellt. Die Ergebnisse weisen den TEWI als reliables Instrument mit zufriedenstellenden Skalierungseigenschaften aus.

**Schlüsselwörter** NEPS · Wissenschaftspropädeutik · Wissenschaftskontroversen · Testinstrument

### **Test for the evaluation of scientific contradictions (TEWI): Measurement of science-related reflective ability in secondary school II**

**Abstract** The aim of scientific thinking (*Wissenschaftspropädeutik*) is one of the three main goals of higher secondary school in Germany. Due to the absence of adequate instruments to measure *Wissenschaftspropädeutik* only few studies have attempted to examine whether the central aims of scientific thinking are achieved at the end of upper secondary school or what importance this might bear in relation to further personal and professional development. This article briefly introduces the theoretical concepts and existing models regarding *Wissenschaftspropädeutik*, extracts the interdisciplinary perspective and illuminate the critical and reflective stance towards science as central elements of the concept. *Scientific reflection competence* in regard to conflicting science claims is defined here as a core component of *Wissenschaftspropädeutik*. It is used as the starting point for the development of the *Test for Evaluating Conflicting Science Claims (TEWI)* for the National Education Panel (NEPS). In its empirical part, the article presents results regarding the development and piloting of the TEWI as well as a study inspecting the construct validity ( $N=394$ ). Results indicate that the TEWI is a reliable test with satisfactory scaling characteristics.

**Keywords** NEPS · Scientific thinking · Conflicting science claims · Questionnaire

Neben der vertieften Allgemeinbildung und der Studierfähigkeit bildet die Wissenschaftspropädeutik das zentrale Bildungsziel der gymnasialen Oberstufe (KMK 1995). Wissenschaftspropädeutik bezeichnet dabei ein komplexes theoretisches Konstrukt mit bildungstheoretischen Wurzeln, das bereits seit den siebziger Jahren Gegenstand bildungswissenschaftlicher Diskussionen ist. Knapp gefasst umgreift der Begriff die Vorbereitung auf einen differenzierten Umgang mit Wissenschaft für das Leben, Lernen und Arbeiten in einer von Wissenschaft durchwobenen Gesellschaft (Huber 2000). Eine verwandte Zielperspektive wird aktuell im Rahmen der Förderung von Kompetenzen im mathematisch-naturwissenschaftlichen Bereich diskutiert (European Commission 2007; OECD 2011, 2019). Diese Kompetenzen werden unter dem Begriff der *Scientific Literacy* zusammengefasst (z. B., Kolstø 2001). Bürgerinnen und Bürger moderner Gesellschaften werden in ihrem persönlichen Leben sowie

in ihrer gesellschaftlichen Verantwortung mit Problemen konfrontiert, die durch moderne naturwissenschaftliche Erkenntnisse und Technologien beantwortet, teilweise aber auch hervorgerufen werden. Das prominenteste Beispiel dafür ist der menschengemachte Klimawandel. Ein adäquates Verständnis und ein angemessener Umgang mit wissenschaftlichen Erkenntnissen, die sich auf solche Probleme beziehen, sind als Qualifikationen für die gesellschaftliche Teilhabe von Bürgerinnen und Bürgern unerlässlich (Kienhues et al. 2018; Tabak 2018).

Obwohl das Konstrukt der Wissenschaftspropädeutik zu den „amtlichen“ Bildungszielen im Hinblick auf die Hochschulreife zählt (KMK 1995), liegen bislang kaum Studien zur Erreichung dieses Bildungsziels und seiner Bedeutung für Studienerfolg vor. Eine Ursache hierfür liegt in einem Mangel an geeigneten Untersuchungsinstrumenten für Wissenschaftspropädeutik (Dettmers et al. 2010). Der vorliegende Artikel gibt einen knappen Überblick über die Ziele von Wissenschaftspropädeutik. Anschließend wird die Entwicklung und Validierung des *Tests zur Einschätzung wissenschaftlicher Widersprüche* (TEWI) für das Nationale Bildungspanel (NEPS; Blossfeld und von Maurice 2011) vorgestellt. Zuletzt wird ein Resümee zur Operationalisierung von Wissenschaftspropädeutik mit dem TEWI in NEPS gezogen und die Bedeutung der Befunde für die weitere Forschung diskutiert.

## 1 Wissenschaftspropädeutik – Zwischen Bildungstheorie und Zeitgeist

Die Aufnahme der Wissenschaftspropädeutik in die KMK Bildungsziele ist im Zusammenhang mit der Diskussion um die Ausrichtung der gymnasialen Oberstufe zu sehen, die mit ihren bildungsgeschichtlichen Wurzeln auf Konzepte der neuhumanistischen Bildungstheorie zurückgeht. Dieser Auffassung folgend besteht das Ziel formaler Bildung darin, auf Wissenschaft vorzubereiten (Hahn 2009). Mit dem Konzept der „Propädeutik“ werden dabei die Differenzen zu der Vermittlung spezialisierten wissenschaftlichen Wissens in dreierlei Hinsicht akzentuiert, nämlich erstens als *fächerübergreifende* Denkweise, zweitens als Betonung einer *kritisch-reflexiven* Perspektive, und drittens als eine zur Wissenschaft *hinführende* und damit in der Tiefe begrenzte Vermittlung. Es geht insbesondere um die Reflexion wissenschaftlicher Erkenntnisse vor dem Hintergrund ihres Entstehungszusammenhanges (Benner 2002; Hahn 2013). In der gymnasialen Oberstufe sollten neben wissenschaftlichen Regeln und Arbeitsweisen ebenso Verfahren, Kontexte, Bedingungen und Folgen von Wissenschaft reflektiert werden (Hahn 2009; Huber 2001, 2005). Die Zielperspektive ist Wissenschaft als soziale Konstruktion von Wirklichkeit begreifbar zu machen, die im sozialen Gefüge entsteht, und deren Aussagekraft an festgelegte Gütekriterien gebunden ist, die auch in den im System agierenden Personen begründet liegen (Oreskes 2019). Bereits die Konzepte zu Wissenschaftspropädeutik der 60er-Jahre (Blankertz 1966; von Hentig 1966) verorteten eine solch kritisch-reflexive Perspektive im fächerübergreifenden Unterricht (Hahn 2013). Nur im Vergleich unterschiedlicher Fachperspektiven würden die Grenzen und Potenziale einzelner Disziplinen sichtbar. Wissenschaftspropädeutik ist in diesem Sinne als ein übergreifendes, der Fachperspektive vorgeordnetes Konzept zu verstehen.

## 1.1 Ziele und Struktur von Wissenschaftspropädeutik

Wissenschaftspropädeutik ist ein unscharfes und normativ aufgeladenes Konstrukt, das in Abhängigkeit von epochalen Auffassungen von Wissenschaft und Bildung unterschiedliche Auslegungen erfahren hat. Früh thematisiert Griese (1983) die „Randlosigkeit“ des Begriffes Wissenschaftspropädeutik, dessen Definition als Untersuchungsgegenstand deshalb eine Herausforderung darstellt (Dettmers et al. 2010). Huber (1997) schlägt eine dreiteilige Gliederung von Wissenschaftspropädeutik vor, die auf der Konzeptualisierung der KMK in den 70er-Jahren basiert. Der wissenschaftspropädeutische Unterricht führe

zur Kenntnis wesentlicher Strukturen und Methoden von Wissenschaften sowie zum Verständnis ihrer komplexen Denkformen,  
zum Erkennen von Grenzen wissenschaftlicher Aussagen und zur Einsicht in Zusammenhang und Zusammenwirken von Wissenschaften,  
zum Verstehen wissenschaftstheoretischer und philosophischer Fragestellungen,  
zur Fähigkeit, theoretische Erkenntnisse sprachlich zu verdeutlichen und anzuwenden. (KMK 1977/1978, S. 561)

Huber argumentiert, dass „es bei Wissenschaftspropädeutik um mehrere Ebenen geht: um das Lernen und Einüben *in* Wissenschaft (Grundbegriffe, -methoden), *an* Wissenschaft (eine Haltung des Immer-weiter-fragens und Gründegebens) und *über* Wissenschaft (kritische Reflexion in größeren Zusammenhängen)“ (Huber 1997, S. 348; Hervorhebungen im Original). Diese Ebenen umfassen nach Huber (2009) Ziele bzgl. der kognitiven Dimension (Kenntnisse, Einsichten), aber auch bzgl. der Ausbildung günstiger motivationaler Orientierung gegenüber Wissenschaft (siehe auch Müsche 2009). Obwohl also das Konzept der Wissenschaftspropädeutik die Befähigung zu *kritischen* Urteilen gegenüber wissenschaftlichen Geltungsbehauptungen umfasst, gehört dazu auch eine positive Haltung zu ‚Wissenschaft‘. Während die erste Ebene das Erlernen wissenschaftlicher Grundbegriffe und Methoden umfasst, ist die zweite Ebene als die Ausbildung einer Geisteshaltung gedacht, die Geltungsbehauptungen als begründungsbedürftig und als – im Grundsatz – immer auch kritisierbar versteht. Auf der dritten Ebene lokalisiert Huber schließlich die Reflexion wissenschaftlicher Erkenntnisse im Hinblick auf ihre Entstehungsbedingungen, Grenzen und Konsequenzen.

Müsche (2009) überträgt die von Huber (1997) spezifizierten Ebenen in ein normatives Strukturmodell wissenschaftspropädeutischer Kompetenz. Die Autorin konzentriert sich auf eine Dimensionierung wissenschaftspropädeutischer Kompetenz, ohne Abstufungen zu modellieren, und unter Ausschluss von motivationalen und affektiven Aspekten. Müsche (2009) geht von den Dimensionen (1) *metawissenschaftliches Wissen*, (2) *Methodenbewusstsein* und (3) *metawissenschaftliche Reflexion* aus. Metawissenschaftliches Wissen umfasst Kenntnisse relevanter Grundbegriffe (z. B. Theorie), wissenschaftlicher Prinzipien (z. B. regelgeleitetes Vorgehen) und zentraler methodischer Begriffe (z. B. Experiment). Im Kontrast zur ersten stehen bei der zweiten Dimension insbesondere prozedurale Aspekte der wissenschaftlichen Arbeitsweise im Fokus. Methodenbewusstsein konzeptualisiert Müsche (2009) als

ein Verständnis der einzelnen Schritte des Forschungsprozesses (z. B. Identifikation einer Problemstellung, Hypothesen formulieren, Untersuchungsweisen auswählen), welches es erlaubt, diese nachzuvollziehen und ggf. anwenden zu können. Diese Konzeptualisierung schließt Einstellungen und Grundhaltungen explizit aus und grenzt sich darin von Hubers zweiter Ebene ab. Die Dimension *metawissenschaftliche Reflexion* ist schließlich eng an die dritte Ebene Hubers angelehnt und umfasst Fähigkeiten, wissenschaftliche Erkenntnisweisen, Aussagen und Befunde zu kontextualisieren und im größeren Zusammenhang zu betrachten; wissenschaftliche Erkenntnismöglichkeiten, Bedingungen und Folgen zu reflektieren; Erkenntniswege und Forschungsmethoden zu beurteilen und dabei beispielsweise methodologisch zu argumentieren; Gültigkeit, Erklärungskraft, Geltungsbereich und Grenzen wissenschaftlicher Aussagen auszuloten und abzustufen; inkonsistente oder widersprüchliche Theorieentwürfe und Befunde gegenüberzustellen und ggf. zu integrieren. Nach Müsche (2009) bezieht diese Dimension Aspekte aus Wissenschaftstheorie, -geschichte und -kritik, Wissenschaftssoziologie und -ethik sowie methodologische Perspektiven ein. Wie bereits dargestellt, lässt sich eine derartige Perspektive nur durch Überwindung der Fachperspektive erreichen (Hahn 2013).

## 1.2 Geteiltes Ziel einer wissenschaftlichen Grundbildung: Reflexion über Wissenschaft als (kontroverser) Diskurs

Wissenschaftspropädeutik soll nicht nur auf Wissenschaft, sondern auch auf ein Leben in einer wissenschaftsbasierten Gesellschaft vorbereiten (Huber 2005). Diese Perspektive berücksichtigt, dass Personen in den meisten Wissensdomänen ihr Leben lang Laiinnen und Laien bleiben werden, und dennoch fähig sein müssen, informierte und begründete Urteile fällen zu können (Bromme et al. 2015; Feinstein 2011; Hahn 2013). Diese Idee von kompetenten Laiinnen und Laien betont weniger die Bedeutung der unmittelbaren Aneignung umfassenden (Vor-)Wissens als vielmehr die Ausbildung von Urteilsfähigkeit und Orientierungswissen darüber, wann man sich als Laiin oder Laie an Expertinnen und Experten richten muss, wie man zuständige Expertise identifizieren kann und welche Aussagen man von Expertinnen und Experten erwarten kann (Bromme 2020; Hahn 2013; Heymann et al. 1990).

Diese zeitgemäße Auslegung wissenschaftspropädeutischer Bildung betont die Fähigkeit, sich im System Wissenschaft orientieren zu können, und bildet eine Schnittstelle zwischen dem älteren, deutschen bildungstheoretisch begründeten Konstrukt der Wissenschaftspropädeutik und dem jüngeren Verständnis von *Scientific Literacy*, einer (natur-)wissenschaftlichen Grundbildung (OECD 2007, 2017, 2019). Insbesondere jüngere Konzeptionen von *Scientific Literacy* betonen Kompetenzen, die zum Verständnis und zur Beurteilung von Diskursen über wissenschaftliche Ergebnisse, auch unter Berücksichtigung wissenschaftlicher Methodik und Praxis, befähigen sollen (Höttecke und Allchin 2020; Kienhues et al. 2018; OECD 2017; Sharon und Baram-Tsabari 2020) – also eine Verstärkung jener reflexiven Aspekte, die im Schwerpunkt des Begriffs der Wissenschaftspropädeutik liegen.

Ansätze verwandter Konstrukte aus der Diskussion um *Scientific Literacy* überlappen mit Teilbereichen von Wissenschaftspropädeutik oder konkretisieren diese: Ansätze zur *Nature of Science* und *Scientific Inquiry* befassen sich mit dem Verständ-

nis prozeduraler Aspekte wissenschaftlicher Arbeitsweisen, ihrer Bedingungen und Grenzen (Lederman 1992, 2007); Ansätze zu *Scientific Reasoning* konzentrieren sich auf die Voraussetzungen und Förderung argumentativer Fähigkeiten (Fischer et al. 2014; Kuhn 1991); Ansätze zum *Epistemic Thinking* betrachten das individuelle Verständnis der Natur, Konstruktion und Begründung wissenschaftlichen Wissens und dessen Rolle in der Auseinandersetzung mit Wissenschaft (Greene et al. 2016). Diese Konstrukte entstammen unterschiedlichen Forschungstraditionen, lassen jedoch einen analogen Wandel erkennen: Sie betonen zunehmend die Förderung reflexiver Fähigkeiten von Schülerinnen und Schülern, um sie insbesondere auf die Auseinandersetzung mit dem wissenschaftlichen Diskurs vorzubereiten (Barzilai und Chinn 2020; Höttecke und Allchin 2020; Sharon und Baram-Tsabari 2020).

Die Relevanz von wissenschaftsbezogener Reflexionsfähigkeiten wird insbesondere dann deutlich, wenn Laiinnen und Laien sich mit *widersprüchlichen wissenschaftlichen Aussagen* auseinandersetzen müssen (Allchin 2011; Bromme et al. 2015; Kuhn 1991). Widersprüche sind zentraler Bestandteil des wissenschaftlichen Diskurses. Wissenschaftliche Erkenntnisgewinnung vollzieht sich als kontinuierlicher Diskurs über wissenschaftliche Geltungsbehauptungen. Neue Erkenntnisse treten im Prozess der wissenschaftlichen Wissensproduktion als Kritik und Alternative zu dem bisherigen Erkenntnisstand in Erscheinung. Der wissenschaftliche Diskurs reguliert diese Konflikte, (z. B. durch Peer Reviews, Gutachten oder Replikationen). Die Zugänglichkeit von Informationen durch das Internet erhöht jedoch auch die Verfügbarkeit von und Konfrontation mit wissenschaftlichen Widersprüchen in der Öffentlichkeit (Stadler und Bromme 2014). Wer im Internet zu aktuellen, wissenschaftsbasierten Fragen Informationen sucht, findet oft heterogene Aussagen. Diese Heterogenität spiegelt häufig den wissenschaftsinternen Diskurs wider, allerdings addieren sich ebenso Aussagen, die nicht wissenschaftlich begründet sind (z. B. pseudowissenschaftliche, verschwörungstheoretische Behauptungen). Die aktuelle Corona-Pandemie veranschaulicht diese Problemlage in all ihren Facetten und in bisher nie dagewesener Weise. Während Wissenschaftlerinnen und Wissenschaftler im Umgang mit Kontroversen geübt sind, stellen sie Laiinnen und Laien vor eine enorme Herausforderung. Die Entscheidung darüber, was sie als am besten begründet betrachten sollten, ist schwierig und erfordert Beurteilungen hinsichtlich der subjektiven Plausibilität der Aussagen sowie der Glaubwürdigkeit der jeweiligen Quellen (Barzilai et al. 2020; Bromme et al. 2015, 2018). Für den Umgang mit Widersprüchen in der Wissenschaft sind daher Fähigkeiten zur Reflexion sowie ein Verständnis der Rationalität wissenschaftlicher Denkweisen und Wissen um die soziale Praxis der Wissenschaften unerlässlich.

### 1.3 Ansätze zur Erfassung unterschiedlicher Facetten von Wissenschaftspropädeutik

Aufgrund seiner Komplexität und seines fächerübergreifenden Charakters entzieht sich das Konstrukt der Wissenschaftspropädeutik einer einfachen Übersetzung in Testinstrumente (Dettmers et al. 2010). Einige Studien haben Subfacetten wissenschaftspropädeutischen Arbeitens in der Oberstufe zu erfassen versucht und dafür Instrumente entwickelt. Sie konzentrieren sich dabei auf Aspekte der Vermittlung

und des Erlernens wissenschaftlicher Arbeitsweisen, die der ersten Ebene der *wissenschaftlichen Grundbegriffe und Methoden* in Hubers (1997) Modell zugerechnet werden können. Basierend auf den Daten aus der *Transformation of the Secondary School System and Academic Careers (TOSCA)* Studie untersuchten Trautwein und Lüdtke (2004, 2007) die wissenschaftspropädeutische Funktion der Oberstufe, indem sie erfassten, welche wissenschaftsnahen Tätigkeiten durchgeführt werden. Es wurde nach der Häufigkeit gefragt, mit der Informationen ausgewählt und integriert, Computer genutzt, Referate angefertigt, Experimente geplant, Diskussionen protokolliert und Literatur beschafft wurde. Zudem erfragten sie die subjektiv erlebte Qualität der Vorbereitung auf ein Studium durch die Oberstufe. Im Evaluationszusammenhang der Schweizer Maturitätsreform von 1995 (EVAMAR I/II) standen Fragen zur Effektivität des Oberstufenunterrichtes im Hinblick auf zwei Kompetenzfelder im Fokus (Eberle et al. 2008; Ramseier et al. 2004): a) die Fähigkeit zum selbstständigen Lernen (Computer- und Mediennutzung, Kritikfähigkeit bei der Anwendung neuer Technologien, Wissen zu Hilfsmitteln) und b) die Fähigkeit zu wissenschaftlichem Arbeiten (wissenschaftlich argumentieren und sich ausdrücken sowie allgemeine Aussagen auf Einzelfälle übertragen können). Diese Kompetenzen wurden von Schülerinnen und Schülern sowie Lehrkräften eingeschätzt.

Bisher existiert unseres Wissens nach keine Studie, die dezidiert die dritte Ebene von Wissenschaftspropädeutik nach Huber (1997) untersucht, und Lernen an und über Wissenschaft adressiert. Jedoch kann auf Instrumente zurückgegriffen werden, die der Erfassung verwandter Konstrukte – nämlich des wissenschaftlichen und epistemischen Denkens – dienen. Beispielsweise untersuchen Zeuch und Souvignier (2015) anhand von Fallvignetten, inwieweit Lehramtsstudierende fähig sind, die Güte von Evidenz beurteilen und evidenzbasierte Schlussfolgerungen ziehen zu können. Das Instrument erlaubt es, basale Fähigkeiten im Umgang mit Wissenschaft zu erfassen, und spiegelt Facetten der zweiten Ebene nach Huber (1997), Lernen und Üben mit Wissenschaft, wider.

Facetten metawissenschaftlicher Reflexion spiegeln sich in Instrumenten wider, die insbesondere die Auseinandersetzung mit Wissenschaftskontroversen erfordern. Basierend auf der Präsentation einer Kontroverse von Expertinnen und Experten lassen Barzilai und Weinstock (2015) Probanden eine Reihe von Items beurteilen, die verschiedene epistemische Dimensionen über die Natur, Konstruktion und Rechtfertigung wissenschaftlichen Wissens (z. B. Sicherheit von Wissen) erfassen und die typische absolutistische, multiplizistische und evaluativistische Sichtweisen widerspiegeln. Thomm et al. (2015) entwickelten einen Fragebogen zur Erfassung subjektiver Erklärungen für Wissenschaftskontroversen (*Explaining Conflicting Scientific Claims Questionnaire*, ECSC). Ebenfalls ausgehend von einer Wissenschaftskontroverse erfasst das Instrument vier verschiedene Ursachenerklärungen (Unterschiede im Forschungsprozess, Komplexität des Themas, Verfolgen von Interessen, Kompetenzunterschiede der Wissenschaftlerinnen und Wissenschaftler), die Laiinnen und Laien zur subjektiven Erklärung und somit der Auflösung des Widerspruchs heranziehen können. Die Beantwortung beider Instrumente erfordert es, dass Probandinnen und Probanden über die Charakteristiken wissenschaftlichen Wissens sowie über Wissenschaft als soziale Praxis reflektieren – wesentliche Merkmale der dritten Ebene nach Huber (1997).

Die vorgestellten Instrumente nähern sich exemplarisch der Messung relevanter Facetten von Wissenschaftspropädeutik an. Allerdings ist ihre Anwendung auf eine bestimmte Domäne oder gar ein spezifisches Thema begrenzt, sodass sie nur bedingt Aussagen über übergreifende Kompetenzen erlauben. Zudem sind sie teils nicht als Test konzipiert. Dennoch können sie inhaltlich sowie methodisch wichtige Impulse für die Testentwicklung geben.

## 2 Entwicklung des Tests zur Einschätzung wissenschaftlicher Widersprüche (TEWI)

Angesichts der Vielfalt und Komplexität des Bildungsziels Wissenschaftspropädeutik, erscheint es uns als unmöglich das Gesamtkonstrukt (im Sinne von Huber 1997) in ein messbares Kompetenzmodell zu überführen. Theoretisch angemessen und praktisch möglich ist es jedoch, die Kompetenzen im Umgang mit Wissenschaft für einzelne typische Problemlagen und Herausforderungen, vor denen Individuen ihrem Alltag stehen, zu spezifizieren und zu messen. Das von uns gewählte Beispiel für eine solche Problemlage ist die *Auseinandersetzung mit widersprüchlichen wissenschaftlichen Aussagen*. Wie aus existierenden Ansätzen abgeleitet (siehe 1.3) stellen wissenschaftliche Widersprüche eine Problemstellung dar, anhand derer man die Ausprägung individueller wissenschafts-propädeutischer Kompetenzen erfassen kann. Sie erfordern den reflexiven Umgang mit Wissenschaft, den wir als Kern des Begriffs der Wissenschaftspropädeutik sehen und als Basis für die Entwicklung eines *Tests zur Einschätzung wissenschaftlicher Widersprüche* (TEWI) für das Nationale Bildungspanel nutzten. Wir bezeichnen die erfasste Kompetenz als wissenschaftsbezogene Reflexionskompetenz.

Der TEWI wurde für den Einsatz im Nationalen Bildungspanel (NEPS) von der NEPS-Arbeitsgruppe *Etappe 5 Gymnasiale Oberstufe* an der Universität Tübingen in Kooperation mit der Universität Münster und der Universität Erfurt mit dem Ziel entwickelt, die wissenschaftspropädeutische Kompetenz bei Abiturientinnen und Abiturienten zu erfassen. Gemäß den Vorgaben des NEPS wurde der TEWI als Papier-und-Bleistifttest konzipiert, der in 30 min als Gruppentestung in Schulklassen der Jahrgangsstufe 12 administriert werden kann.

Für das Instrument wurden Vignetten mit wissenschaftlichen Kontroversen zu zeitgenössischen Problemstellungen als Ausgangspunkt gewählt (Oschatz et al. 2018). Die thematische Breite der Vignetten wurde begrenzt, indem nur Problemstellungen gewählt wurden, die einen – im weitesten Sinne – naturwissenschaftlichen Bezug haben. Das Aufgabenformat wurde in Anlehnung an den ECSC von Thomm et al. (2015) entwickelt. Da das Instrument insbesondere auf die Erfassung *subjektiver* Erklärungen abzielt, erlaubt es jedoch nicht, eine normative Aussage über die Angemessenheit dieser Erklärung und somit über das Ausmaß des Verständnisses der Kontroverse im Sinne eines Testscores zu treffen.



## 2.1 Aufgabenformat

Das gewählte Aufgabenformat besteht aus Stimulustexten, in denen jeweils eine wissenschaftliche Kontroverse geschildert wird. Passend zur Kontroverse werden fünf bis sieben Aussagen zu zentralen Aspekten präsentiert, die in einem forced-choice Format als zutreffend oder unzutreffend zu bewerten sind (vgl. Abb. 1). Die Stimulustexte bestehen aus etwa 300–400 Wörtern. In den Texten wird zunächst der Themenbereich vorgestellt, danach wird der zwischen zwei Wissenschaftlerinnen oder Wissenschaftlern strittige Punkt benannt. Daran anknüpfend werden die Positionen der beiden Forschenden dargestellt. Die Positionsdarstellungen enthalten mehr Informationen als zur Lösung der Aufgaben erforderlich sind. Die im Anschluss an die Kontroverse präsentierten Aussagen benennen mögliche Ursachen für die Kontroverse.

Die TEWI-Aufgaben sind so gewählt, dass sie unterschiedliche naturwissenschaftliche Fachdomänen i. w. S. (Chemie, Medizin, Biologie und Sportwissenschaft) und verschiedene Perspektiven wissenschaftsbezogener Reflexion (methodische, die Datenerhebung, -auswertung, -interpretation betreffende, gesellschaftliche, historische und ethische) abdecken (Thomm et al. 2017). Variiert wurden auch Angaben zu den Wissenschaftlerinnen und Wissenschaftlern (z. B. Status und Berufserfahrung). Allen Aufgaben gemeinsam ist, dass sie das Gegenüberstellen und Integrieren inkonsistenter wissenschaftlicher Befunde oder widersprüchlicher Theorien erfordern, naturwissenschaftliche Inhalte thematisieren, aktuelle und alltagsbezogene Themen zum Gegenstand haben, und multikausale Sachverhalte thematisieren. Damit die Aufgaben auch von kompetenten Laiinnen und Laien gelöst werden können, erfordern sie es *nicht*, komplexe wissenschaftliche Studien zu planen oder Entscheidungen zu treffen, sondern wurden darauf beschränkt, wissenschaftliche Theorien und Erkenntnisse kritisch zu reflektieren. Die Widersprüche, die Gegenstand der TEWI-Aufgaben sind, treten zwischen Wissenschaftlerinnen und Wissenschaftlern, *nicht* bei der Vermittlung wissenschaftlicher Erkenntnisse durch Dritte (Journalistinnen und Journalisten, Schulbücher etc.) auf. Auf Basis dieser Aufgabendefinition wurden zunächst 10 Stimulustexte (Kontroversen) mit insgesamt 94 Aussagen (Items) entwickelt und in kognitiven Vortests<sup>1</sup> überprüft.

## 2.2 Pilotierung

Zur Pilotierung wurden Schülerinnen und Schüler der Jahrgangsstufe 11 in fünf allgemeinbildenden Gymnasien im Großraum Stuttgart eingeladen sowie – um eine größere Stichprobe zu erreichen – Studierende, die am Beginn ihres Studiums (überwiegend im 2. Semester) standen. Die Studierenden wurden im Anschluss an Vorlesungen befragt. Die Teilnahme war freiwillig, wurde mit 10€ vergütet und

---

<sup>1</sup> Kognitive Vortests: Zwei Testversionen mit je fünf Stimulustexten und zugehörigen Items,  $N=4$  Abiturientinnen und Abiturienten, Testdurchführung unter der Bedingung lauten Denkens mit Audioaufnahme und Nachbefragung zur Verständlichkeit der Testinstruktion und weiteren Eindrücken bei der Testbearbeitung.

## KONTROVERSE 1: PESTIZIDE

In der Landwirtschaft werden weltweit Pestizide zur Schädlingsbekämpfung verwendet. In einer Fernsehsendung diskutieren Experten darüber, in welchem Umfang Pestizide weiter genutzt werden sollen.

**Biologe A spricht sich gegen den Einsatz von Pestiziden aus, während Biologe B den Pestizideinsatz befürwortet.**

*Biologe A* ist seit 30 Jahren Professor für Biologie in Berlin und arbeitet in der EU-Behörde für Lebensmittelsicherheit. Er warnt vor dem Einsatz der Pestizide. Besonders gefährlich für Verbraucher seien die langsamen und zeitversetzten Wirkungen, die durch Pestizidrückstände in Nahrungsmitteln verursacht würden. Man vermutet, dass sie die Krebsentstehung begünstigen, das Erbgut verändern, das Immunsystem beeinträchtigen und das Parkinson-Risiko steigern können. Es gebe zwar zahlreiche Studien hierzu aber einen erheblichen Mangel an Langzeituntersuchungen und kaum Studien zur Wechselwirkung mehrerer Wirkstoffe. Aus diesem Grund würden derzeit im Rahmen von EU-Programmen solche Forschungsprojekte finanziell stark gefördert. Solange diese Daten nicht vorliegen und die genannten Befürchtungen nicht ausgeräumt seien, so *Biologe A*, seien die Risiken des Pestizideinsatzes nicht absehbar und es sollte auf den Einsatz von Pestiziden weitgehend verzichtet werden.

*Biologe B* ist seit 10 Jahren wissenschaftlicher Mitarbeiter an dem biologischen Institut an einer Universität. Er findet die Forderung von *Biologe A* nicht gerechtfertigt, schließlich würden Pestizide in Deutschland bereits seit mehr als 50 Jahren erfolgreich zur Schädlingsbekämpfung eingesetzt. Er sieht dagegen Gefahren in der Absetzung der Pestizide: Dadurch könnten sich bisher kontrollierte Schädlinge ungehindert vermehren und neue Probleme verursachen. Als Beispiel führt er die starke Vermehrung der Rötelmaus an und den damit zusammenhängenden Anstieg von Hanta-Virus-Infektionen in Süddeutschland. Die gefährliche Viruskrankung wird von den Mäusen auf Menschen übertragen und kann in Einzelfällen zum Tode führen. Eine Studie von einem seiner Doktoranden habe gezeigt, dass die starke Vermehrung der Mäuse auf dem rückläufigen Einsatz von Pestiziden gegen Nagetiere beruhe. Leider würde dieser Forschungszeitpunkt zu wenig gefördert.

## KONTROVERSE 1: PESTIZIDE

Die Wissenschaftler vertreten widersprüchliche Positionen. Um zu beurteilen, wie gut die Positionen jeweils begründet sind, ist es wichtig, die Positionen genau zu analysieren. Der Aufgabentyp A enthält mehrere Aussagen zu zentralen Aspekten der Kontroverse, die Sie überprüfen sollen. Die Aussagen können mögliche Ursachen für die Kontroverse, die verwendeten Methoden, Erklärungen und Interpretationen der Wissenschaftler betreffen.

Im Folgenden finden Sie eine Reihe von Aussagen zu zentralen Aspekten der Kontroverse *Pestizide*. Bitte beurteilen Sie, ob die getroffenen Aussagen über die *Kontroverse Pestizide* zutreffend oder nicht.

Machen Sie dazu *pro Zeile ein Kreuz*.

	zutreffend	nicht zutreffend
a) Ein Aspekt der Kontroverse ist, dass Wissenschaftler A die Probleme ignoriert, die durch das Absetzen von Pestiziden entstehen.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
b) Der Perspektive von Biologe B sollte mehr Gewicht zugesprochen werden, da er Daten aus einer Studie zur Unterstützung seiner Perspektive anführt, während Biologe A nur die fehlende Datenbasis betont.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
c) Die Kontroverse wäre nicht entstanden, wenn Biologe A die neue Studie zur Vermehrung der Rötelmaus berücksichtigt hätte, die Biologe B anführt.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
d) Die Kontroverse wäre mit Sicherheit auch entstanden, wenn Wissenschaftler B den Problemen des Einsatzes von Pestiziden größere Bedeutung beimessen würde.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
e) Ein Grund der Kontroverse liegt darin, dass Biologe A sich in seinen Aussagen auf ganz Europa bezieht, während Biologe B sich auf den Pestizideinsatz in Deutschland bezieht.	<input type="checkbox"/>	<input checked="" type="checkbox"/>
f) Es ist unwahrscheinlich, dass Biologe A seine Meinung revidieren würde, wenn er die neue Studie zur Vermehrung der Rötelmaus lesen würde.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
g) Für die Kontroverse ist es relevant, dass die beiden Biologen aus unterschiedlichen Perspektiven auf den Pestizideinsatz schauen.	<input checked="" type="checkbox"/>	<input type="checkbox"/>
h) Um die Perspektive von Biologe B nachhaltig zu stützen, ist es nicht ausreichend, eine einzige Studie zu nennen.	<input checked="" type="checkbox"/>	<input type="checkbox"/>

**Abb. 1** TEWI Aufgabentypformat (vgl. Oschatz et al. 2018). Im Unterschied zu den Kontroversen im TEWI sind in dieser Beispielaufgabe acht Aussagen einzuschätzen (im TEWI pro Kontroverse fünf bis sieben Aussagen). Außerdem sind die richtigen Lösungen hier markiert

dauerte maximal 90 min. Insgesamt nahmen  $N=107$  Personen (87 Schülerinnen und Schüler sowie 20 Studierende; 60% weiblich) teil.

Die Pilotierungsdaten wurden zur Itemauswahl verwendet. Als Kriterien zur Itemauswahl wurden psychometrische und inhaltliche genutzt. Die psychometrischen Kriterien, an denen wir uns orientierten, waren relative Lösungshäufigkeiten ( $>0,25$ ), Itemtrennschärfen (positiv) und Faktorladungen (konfirmatorische Faktorenanalysen für kategoriale Indikatoren mit MLR-Schätzer, Ladungen  $>0,30$ ; berechnet mit *Mplus*; Muthén & Muthén 1998–2012). Auf dieser Basis wurden 23 Items ausgewählt (vgl. Online Supplement). Die geschätzte Reliabilität des eindimensionalen Testscores (WLE person separation reliability = 0,71, berechnet mit ConQuest; Adams et al. 2012) und seine Varianz ( $SD^2=0,83$ ) sind zufriedenstellend.

Inhaltlich wurde die gewählte Itemstichprobe von zwei im Themenbereich Wissenschaftspropädeutik promovierenden Expertinnen, die nicht am Entwicklungsprozess beteiligt waren, dahingehend überprüft, ob sie die unterschiedlichen Fachdomänen und Perspektiven wissenschaftsbezogener Reflexion beinhalten. Dies war gewährleistet. Die Items waren jedoch überwiegend positiv gepolt. Da die Pilotierungsstichprobe relativ klein war, haben wir uns entschieden, neun weitere überwiegend negativ gepolte Items mit vorzulegen, von denen zunächst offengelassen wurde, ob sie zur Berechnung des Testscores mit herangezogen werden oder nicht. Insgesamt wurde eine Auswahl von fünf Kontroversen mit je fünf bis sieben (insgesamt 32) Items zusammengestellt.<sup>2</sup>

### 3 Validierungsstudie

Im Mittelpunkt der Validierungsstudie stand die Frage, in welcher Beziehung der TEWI zu verwandten Konstrukten steht, die in einem engeren Sinne der Wissenschaftspropädeutik nahestehen. Im Herbst 2013 wurde deshalb an der Universität Tübingen eine Studie zur konvergenten Konstruktvalidierung des TEWI durchgeführt.

#### 3.1 Validierungsinstrumente, Fragestellungen und Hypothesen

Folgende Konstrukte und Kriterien wurden zur Prüfung der Validität des TEWI eingesetzt:

**Akademische Leistungskriterien** Als externe Validierungskriterien wurden die Noten der Hochschulzugangsberechtigung (HZB; von 1 = *sehr gut* bis 4 = *ausreichend*) und die selbsteingeschätzte Studienleistung im Vergleich zu Kommilitonen und Kommilitoninnen erfragt. Zum Leistungsvergleich wurde eine Skala aus der TOSCA-Studie (Köller et al. 2004) verwendet, die sieben Items umfasst (z. B. „Ich gehöre zu den besten Studenten meines Semesters.“; vierstufige Likertskala von 1 = *trifft überhaupt nicht zu* bis 4 = *trifft völlig zu*).

<sup>2</sup> Nähere Informationen zum Vorgehen bei der Pilotierung können bei den Autorinnen und Autoren angefordert werden.

*Hypothesen zur akademischen Leistung:* Wir gehen davon aus, dass wissenschaftsbezogenes Reflexionsvermögen für akademische Leistungen erforderlich ist und erwarten dementsprechend eine negative Korrelation zwischen TEWI-Werten und HZB-Noten sowie eine positive Korrelation zwischen TEWI-Werten und der selbsteingeschätzten Studienleistung.

**Need for Cognition** In Zweiprozessmodellen der Informationsverarbeitung werden idealtypisch zwei Modi der Verarbeitung unterschieden: ein systematischer Modus zur tiefen Informationsverarbeitung und ein heuristischer Modus zur oberflächlicheren Informationsverarbeitung. In der Validierungsstudie haben wir die Skala *Need for Cognition* des Rational-Experiential Inventory (REI; Epstein et al. 1996; in der Übersetzung von Keller et al. 2000) eingesetzt, mit der die Neigung zum systematischen Modus erfasst werden kann. Wir gehen davon aus, dass a) die Neigung zur systematischen Informationsverarbeitung mit höherer Motivation zu wissenschaftsbezogener Reflexion einhergeht, die eine tiefe Informationsverarbeitung erfordert, b) Personen mit dieser Neigung bereits mehr Gelegenheiten zur wissenschaftsbezogenen Reflexion nutzten und sie c) deshalb ein höheres wissenschaftsbezogenes Reflexionsvermögen aufweisen. Die Skala *Need for Cognition* besteht aus 14 Aussagen (z. B. „Abstrakt zu denken reizt mich nicht“, rekodiert), deren Zutreffen auf einer siebenstufigen Skala einzuschätzen ist (1 = *vollkommen unzutreffend* bis 7 = *vollkommen zutreffend*).

*Hypothese zu Need for Cognition:* Wir erwarten positive Korrelationen zwischen TEWI- und Werten in Need for Cognition.

### **Domänenübergreifendes Entwicklungsniveau epistemischer Überzeugungen**

Wissenschaftsbezogenes Reflexionsvermögen erfordert die Aktivierung von Vorstellungen über Wissen und Wissenserwerb. Diese Vorstellungen werden als epistemologische Überzeugungen bezeichnet (Hofer und Pintrich 1997). Zur Erfassung des Entwicklungsniveaus epistemologischer Überzeugungen wird der Fragebogen FREE (Krettenauer 2005) eingesetzt, bei dem die Befragten zu kontroversen Behauptungen Stellung beziehen sollten, die sieben natur- sowie sieben sozial- und geisteswissenschaftliche Themen betreffen. Zu jedem Thema wurden drei Behauptungen präsentiert, die jeweils die Sichtweise eines Entwicklungsniveaus (absolutistisch, relativistisch und post-relativistisch) zum Ausdruck bringen. Aus einer absolutistischen Perspektive ist Wissen entweder richtig oder falsch. In relativistischer Perspektive wird die Möglichkeit unterschiedlicher Standpunkte anerkannt. Erst auf der postrelativistischen Ebene erkennen Individuen, dass unterschiedlichen Standpunkte aufgrund ihrer Begründung mehr oder weniger Berechtigung zugesprochen werden kann. Die Zustimmung zu den Behauptungen wurde mit einer 6-stufigen Skala (1 = *lehne vollständig ab* bis 6 = *stimme vollständig zu*) abgefragt. Die Zustimmungswerte pro Entwicklungsniveau wurden gemittelt. Zusätzlich wurde eine Präferenz für die post-relativistische Sichtweise berechnet (D-Index, theoretischer Range:  $-1$  bis  $+1$ ,  $D < 0$  bedeutet Präferenz für absolutistische/relativistische Sichtweise,  $D > 0$  Präferenz für postrelativistische Sichtweise; Krettenauer 2005). Wir gehen davon aus, dass Personen mit hoher wissenschaftsbezogener Reflexionskompetenz eine Präferenz für postrelativistische Sichtweisen zeigen.

*Hypothesen zum Entwicklungsniveau epistemischer Überzeugungen:* Wir erwarten positive Korrelationen zwischen dem TEWI-Wert und a) der Skala zur Zustimmung zur postrelativistischen Sichtweise sowie b) dem D-Index. Außerdem erwarten wir negative Korrelationen zwischen TEWI-Wert und den Skalen zu Zustimmungen zu absolutistischen und relativistischen Sichtweisen.

**Umgang mit historischen Quellen und Darstellungen** Ein Ziel der Validierungsstudie war es zu prüfen, inwieweit die Leistungen des naturwissenschaftlich ausgerichteten TEWI mit wissenschaftspropädeutischer Kompetenz in einem anderen Fach korrespondiert. Dabei fiel unsere Wahl aus zwei Gründen auf das Fach Geschichte: Zum einen beschäftigen sich Schülerinnen und Schüler bereits während der Schulzeit mit Geschichte und das Fach Geschichte lebt von Kontroversen zwischen Wissenschaftlerinnen und Wissenschaftlern, die auch Laiinnen und Laien in Medien mitverfolgen können. Zum anderen liegt ein Instrument vor, mit dem zwei Teilkompetenzen wissenschaftsbezogenen Reflexionsvermögens in der Domäne Geschichtswissenschaft gemessen werden können. Es wurde von Bertram et al. (2016) auf Grundlage des geschichtswissenschaftlichen Kompetenzmodells von Körber et al. (2007) entwickelt und umfasst folgende zwei historischen Methodenkompetenzen: a) *Re-Konstruktion* (ReKo) erfordert es, die in historischen Quellen enthaltenen Daten und Fakten in Zusammenhang zu bringen und so eine eigene historische Darstellung zu konstruieren. b) *De-Konstruktion* (DeKo) erfordert es, fertige Darstellungen in ihrer Tiefenstruktur zu analysieren, beispielsweise im Hinblick auf die vermittelten Daten und Fakten, die deutende Geschichtserzählung und die transportierte Botschaft. Zur Erfassung der ReKo werden 22 Aussagen und zur Erfassung der DeKo sechs Aussagen vorgelegt. Als Itembeispiel sei eines aus der DeKo-Skala genannt: „Historiker müssen berücksichtigen, dass Quellen durchaus lückenhaft sein können.“ Das Zutreffen der Aussagen soll mit vier Antwortoptionen eingeschätzt werden: 1 = *bin mir ganz sicher, dass es nicht stimmt*, 2 = *bin mir halbwegs sicher, dass es nicht stimmt*, 3 = *bin mir halbwegs sicher, dass es stimmt*, 4 = *bin mir ganz sicher, dass es stimmt*. Als Treffer werden nur diejenigen gewertet, die die richtige Entscheidung getroffen haben und sich dabei *ganz sicher* waren, falsche und *halbwegs sichere* Entscheidungen werden als nicht gelöst gewertet.

*Hypothese zum Umgang mit historischen Quellen und Darstellungen:* Wir erwarten positive Korrelationen zwischen TEWI-Werten und den Kompetenzen zur Re- und De-Konstruktion historischer Quellen und Darstellungen.

### 3.2 Skalierung des TEWI

Für die Berechnung der TEWI-Werte wurden Itemschwierigkeiten aus einer Skalierung zu Grunde gelegt, die für NEPS durchgeführt wurde (Fischer et al. 2020). Diese Skalierung basiert auf einer Stichprobe von 3897 Abiturientinnen und Abiturienten des Schuljahres 2012/2013, die deutschlandweit erhoben wurde (NEPS, Startkohor-

te 4, Klasse 12, <https://doi.org/10.5157/NEPS:SC4:9.1.1<sup>3</sup>>). Die Skalierung wurde mit dem R-Paket TAM (R Core Team 2019; Robitzsch et al. 2019) durchgeführt. Es wurde ein eindimensionales Raschmodell (Rasch 1960) gewählt. Als Personenparameterschätzer wurden weighted maximum likelihood (WML) Schätzer (Warm 1989) extrahiert. Die Reliabilität des TEWI kann auf dieser Datenbasis als befriedigend angesehen werden (EAP: 0,68, WLE: 0,63).

### 3.3 Stichprobe und Studiendesign

Um sparsam mit Schulzeit umzugehen und auf Grund des leichteren Zugangs wurden zur Validierung Studierende eingeladen, keine Schülerinnen und Schüler. Wir gehen davon aus, dass Studierende im Vergleich zu Schülerinnen und Schülern der gymnasialen Oberstufe über durchschnittlich mehr wissenschaftspropädeutische Kompetenz verfügen, dass aber die Zusammenhänge des TEWI mit anderen Konstrukten, die Wissenschaftspropädeutik nahestehen, davon weitgehend unbeeinflusst sein sollten.

Die Studierenden besuchten zwei Vorlesungen zum Thema psychologische Diagnostik an der Universität Tübingen. Die Vorlesungen richteten sich überwiegend an Bachelor-Studierende in Erziehungswissenschaft (Haupt- und Nebenfach) und Lehramtsstudierende. 394 Studierende beteiligten sich (Vorlesung 1:  $n=320$ ; Vorlesung 2:  $n=74$ ). 73,4% weiblich, 14,6% im ersten Studienjahr, 64,5% im zweiten, 12,5% im dritten und 8,4% im 4. oder höheren Studienjahr. 24,4% gaben an, eines der Fächer zu studieren, die in der Validierungsstudie thematisiert werden (6,3% Biologie, 3,3% Chemie, 0,0% Medizin, 4,1% eine Sportwissenschaft, 13,2% eine Geschichtswissenschaft). Da für die Datenerhebung insgesamt 60 min und davon 20 min für die TEWI-Testung zur Verfügung standen, wurden pro Person nur drei Kontroversen vorgelegt (Multi-Matrix-Design): Die zwei Kontroversen, deren Aufgaben in der Pilotierung am höchsten auf dem TEWI-Score luden, wurden von allen bearbeitet und zusätzlich eine der drei weiteren Kontroversen<sup>4</sup>.

### 3.4 Ergebnisse

Die Korrelationen zwischen TEWI-Score und den Vergleichskonstrukten entsprechen unseren Erwartungen, weichen jedoch auch teils von diesen ab (Tab. 1, Spalte  $r$ ). Wie erwartet stimmen Personen absolutistischen Aussagen umso weniger zu, je

<sup>3</sup> Die Daten des NEPS wurden von 2008 bis 2013 als Teil des Rahmenprogramms zur Förderung der empirischen Bildungsforschung erhoben, welches vom Bundesministerium für Bildung und Forschung (BMBF) finanziert wurde. Seit 2014 wird NEPS vom Leibniz-Institut für Bildungsverläufe e. V. (LifBi) an der Otto-Friedrich-Universität Bamberg in Kooperation mit einem deutschlandweiten Netzwerk weitergeführt.

<sup>4</sup> In beiden Vorlesungen wurden nicht alle Vergleichskonstrukte vorgelegt, sondern folgende in den genannten Reihenfolgen: Vorlesung 1 ECSC-M, FREE, Need for Cognition, Studienleistung, HZB; Vorlesung 2: Connotative Aspects of Epistemological Beliefs (CAEB, Stahl und Bromme 2007), ECSC-M, historischer Kompetenztest, Need for Cognition, Studienleistung, HZB. Die Instrumente CAEB und ECSC-M arbeiten auch kontroversenbasiert und dienen als Ideengeber für die Itementwicklung. Es wurden keine Hypothesen zum Zusammenhang mit dem TEWI formuliert und die Ergebnisse aus Platzgründen nicht berichtet. Sie können bei den Autorinnen und Autoren erfragt werden.

**Tab. 1** Ergebnisse der Validierungsstudie

Konstrukt	<i>n</i>	<i>M</i>	<i>SD</i>	$\alpha^a$	Zusammenhang mit TEWI	
					$r^b$	$\rho^c$
(1) FREE absolutistisch	298	3,07	0,65	0,75	-0,14*	-0,16
(2) FREE relativistisch	298	3,39	0,66	0,74	-0,07	-0,08
(3) FREE post-relativistisch	298	4,31	0,54	0,74	0,18**	0,21
(4) FREE D-Index	298	1,24	0,71	0,67	0,14*	0,17
(5) Geschichte De-Konstruktion	74	0,57	0,28	0,64	0,26*	0,33
(6) Geschichte Re-Konstruktion	74	0,55	0,19	0,80	0,26*	0,29
(7) Need for cognition	228	5,11	0,84	0,84	0,24**	0,26
(8) Studienleistung	271	2,51	0,52	0,84	0,19**	0,21
(9) Hochschulzugangsberechtigung (HZB)	377	2,19	0,56	-	-0,22**	-0,22

\* $p < 0,05$ , \*\* $p < 0,01$  (einseitig)

<sup>a</sup>Reliabilität = Cronbachs  $\alpha$

<sup>b</sup> $r$  = manifeste Korrelation der Konstrukte mit dem TEWI-Score

<sup>c</sup> $\rho$  = operationale Validität (manifeste Korrelation korrigiert um Unreliabilität in den Vergleichskonstrukten)

höher ihr TEWI-Wert ist. Zugleich geht eine höhere Ausprägung der Reflexionskompetenz mit einer stärkeren Zustimmung zu post-relativistischen Aussagen einher. Im Kontrast zu unseren Annahmen zeigt sich jedoch kein Zusammenhang des TEWI-Testscores mit der Zustimmung zu relativistischen Aussagen. Werden alle drei Ausgabendimensionen zu einer Präferenz für post-relativistische Aussagen verrechnet (D-Index), bestätigt sich der erwartete Zusammenhang: Hohe TEWI-Werte gehen mit hohen D-Indizes einher. Höheres Verständnis der Kontroversen geht demnach mit einer fortgeschrittenen epistemologischen Perspektive einher.

Die Hypothesen zum Umgang mit historischen Quellen und Darstellungen haben sich bewährt: je höher die TEWI-Werte einer Person, desto höher ihre De- und Rekonstruktionskompetenzen. Auch Need for Cognition hängt wie erwartet positiv mit TEWI-Werten zusammen. Die Hypothesen zur kriterialen Validität bewährten sich ebenfalls: je besser die HZB und die Einschätzung der eigenen Studienleistung im Vergleich zu anderen, desto höher die TEWI-Werte.

Die bislang betrachteten manifesten Korrelationen des TEWI sind nicht nur vom Zusammenhang der latenten Konstrukte, sondern auch von ihren Reliabilitäten beeinflusst. Um einzuschätzen, wie stark die TEWI-Werte mit den einzelnen latenten Konstrukten im Vergleich zueinander korrelieren, wurden die unterschiedlichen Unreliabilitäten der Korrelate des TEWI deshalb korrigiert (Tab. 1,  $\rho$ -Werte; vgl. Hunter und Schmidt 2004). Die höchsten statistisch bedeutsamen Zusammenhänge des TEWI finden sich mit historischen Kompetenzen (De-Konstruktion  $\rho = 0,33$ , Re-Konstruktion  $\rho = 0,29$ ), gefolgt von Need for Cognition ( $\rho = 0,26$ ), der HZB ( $\rho = -0,22$ ) und den selbst eingeschätzten Studienleistungen im Vergleich zu anderen ( $\rho = 0,21$ ). Danach folgen die Skalen zum Entwicklungsniveau epistemischer Überzeugungen, gemessen mit dem FREE ( $0,16 \leq |\rho| \leq 0,21$ ).

Um die Höhe der Korrelationen des TEWI mit den Vergleichskonstrukten einzuschätzen, nutzten wir die Interkorrelationen der Vergleichskonstrukte (Tab. 2). Der

**Tab. 2** Korrelationsmatrix der Vergleichskonstrukte

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) FREE absolutistisch	–	298	298	298	0	0	156	201	288
(2) FREE relativistisch	0,17**	–	298	298	0	0	156	201	288
(3) FREE post-rel	–0,07	0,25**	–	298	0	0	156	201	288
(4) FREE D-Index	–0,29**	–0,34**	0,39**	–	0	0	156	201	305
(5) Geschichte De-Kon	a	a	a	a	–	73	72	70	70
(6) Geschichte Re-Kon	a	a	a	a	0,33**	–	72	70	70
(7) Need for cognition	–0,22**	–0,13*	0,07	0,12	0,12	0,26*	–	224	221
(8) Studienleistung	–0,12**	–0,13*	0,04	0,05	0,34**	0,13	0,34**	–	263
(9) HZB	0,24**	0,17**	–0,14**	–0,23**	0,01	–0,27*	–0,25**	–0,32**	–

Untere Dreiecksmatrix =  $r$ , obere Dreiecksmatrix =  $n$

\* $p < 0,05$ , \*\* $p < 0,01$  (einseitig)

<sup>a</sup>FREE und historischer Kompetenztest wurden getrennten Stichproben vorgelegt, deshalb liegen keine Korrelationen zwischen ihnen vor

TEWI korreliert im selben Range mit den Skalen der drei Vergleichsinstrumente (FREE, historischer Kompetenztests und Need for Cognition) wie die Vergleichskonstrukte untereinander (Range:  $r = |0,07|$  bis  $r = |0,26|$ ). Mit dem Kriterium Studienleistung korreliert der TEWI höher ( $r = 0,19$ ) als die Skalen des FREE ( $r_{\max} = -0,13^*$ ) und die Skala Re-Konstruktion des historischen Kompetenztests ( $r = 0,13$ ). De-Konstruktion und Need for Cognition korrelieren deutlich höher mit Studienleistung (jeweils  $r = 0,34$ ) als der TEWI. Der Zusammenhang des TEWI mit der HZB beträgt  $r = -0,22$ . Diese wird von der FREE-Skala absolutistisch ( $r = 0,24$ ), dem FREE D-Index ( $r = -0,23$ ), Need for Cognition ( $r = -0,25$ ) und der Re-Konstruktionsskala des historischen Kompetenztests ( $r = -0,27$ ) übertroffen. Während der TEWI mit Studienleistung und mit der HZB ungefähr in derselben Höhe korreliert, unterscheiden sich die Korrelationen mit beiden Leistungskriterien bei den Vergleichskonstrukten deutlicher.

## 4 Diskussion

Übereinstimmend betonen Konzeptionen zur Wissenschaftspropädeutik – sowie zu anverwandten Ansätzen – die Bedeutung von Kenntnissen und Kompetenzen von Laiinnen und Laien, die ihnen die Bewertung von und die Auseinandersetzung mit Wissenschaft als Diskurs ermöglichen. Unser Ziel war es ein Instrument zu entwickeln, das hierfür erforderliche Fähigkeiten erfasst. Anknüpfend an existierende theoretische sowie empirische Vorarbeiten konstruierten wir zu diesem Zweck den TEWI. Die Einschätzung von wissenschaftlichen Kontroversen wurde als Gelegen-



heitsstruktur gewählt, da diese Reflexionsprozesse über Theorie, Methoden sowie soziale Praktiken wissenschaftlichen Arbeitens erfordert. Im Folgenden werden zunächst die Ergebnisse der Validierungsstudie zusammengefasst. Anschließend diskutieren wir ihre Bedeutung für das Nationale Bildungspanel und den Forschungsbereich der Wissenschaftspropädeutik.

#### 4.1 Zusammenfassung der Ergebnisse

Die Befunde der vorliegenden Studie zeigen, dass der für NEPS entwickelte Kompetenztest über zufriedenstellende Reliabilität und Skalierungseigenschaften verfügt und ein eindimensionaler Kompetenzscore gebildet werden kann. Zur Ermittlung der Konstruktvalidität wurde der TEWI-Score zu verwandten Konstrukten aus dem Gegenstandsbereich Wissenschaftspropädeutik und zu akademischen Leistungskriterien in Beziehung gesetzt. Die Ergebnisse dieser Analysen sind vielversprechend:

Es zeigten sich die erwarteten Zusammenhänge mit dem domänenübergreifenden Entwicklungsniveau epistemischer Überzeugungen. So wurde beispielsweise gezeigt, dass hohe TEWI-Werte mit einer Präferenz für post-relativistische Überzeugungen einhergehen. Dieser Befund spricht dafür, dass die Befragten bei der Bearbeitung des TEWI detaillierte und komplexe Sachinformationen berücksichtigen können und sich nicht maßgeblich von vorgefassten Überzeugungen leiten lassen. Sie sind fähig, Merkmale von wissenschaftlichen Aussagen vor dem Hintergrund ihrer Annahmen über die Natur und Genese wissenschaftlichen Wissens abzuwägen und zu gewichten – eine Facette, die die Beantwortung sowohl des TEWI als auch des FREE (wenn auch in unterschiedlicher Weise) erfordern. Die positiven Korrelationen mit einem Kompetenztest zum Umgang mit historischen Quellen und Darstellungen deuten darauf hin, dass wissenschaftsbezogenes Reflexionsvermögen eine domänenübergreifende Kompetenz darstellt. Der erfolgreichen Beantwortung beider Instrumente ist gemein, dass man gegebene Informationen zu Aussagen systematisch analysiert, aufeinander bezieht und hinsichtlich ihrer Passung abwägt, um schließlich zu angemessenen Schlussfolgerungen gelangen zu können. Wie erwartet schnitten Befragte mit höheren Werten in Need for Cognition im TEWI besser ab. Personen, die ein stärkeres Bedürfnis haben, sich gedanklich mit komplexeren Inhalten und Problem zu befassen, waren erfolgreicher in der Bewältigung der vorgegebenen Kontroversen. Neben Konstrukten aus dem Gegenstandsbereich Wissenschaftspropädeutik wurden akademische Leistungskriterien betrachtet: Hier zeigte sich, dass Personen über eine umso bessere Abiturgesamtnote verfügen und ihre Studienleistung als umso besser einschätzten, je besser sie im TEWI abschnitten.

Absolut betrachtet fallen die gefundenen Zusammenhänge klein bis moderat aus. Relativ zu den Vergleichskriterien entsprechen bzw. übertreffen sie die Höhe der Korrelationen mit den domänenübergreifend epistemischen Überzeugungen. Sie erreichen aber nicht das Niveau der Korrelationen, die historischer Kompetenztest und Need for Cognition mit den akademischen Leistungskriterien aufweisen. Insbesondere die Höhe der Zusammenhänge mag auf den ersten Blick ernüchtern, lässt sich jedoch gut mit der Unschärfe des Konstrukts Wissenschaftspropädeutik begründen. Aufgrund der Vielschichtigkeit des Konstrukts war es notwendig, unser Instrument auf die Erfassung einer zentralen Facette zu fokussieren. Diese haben wir in Be-

ziehung zu relevanten, anverwandten Konstrukten gesetzt. Auch wenn diese mit wichtigen anderen Facetten von Wissenschaftspropädeutik korrespondieren, stellen sie dennoch verschiedene Konstrukte dar, die sich von dem Gegenstandsbereich des TEWI abheben. Vor diesem Hintergrund sind sehr enge Zusammenhänge nicht zu erwarten.

## 4.2 Limitationen der Validierungsstudie

Die Überprüfung des TEWI ging mit der besonderen Herausforderung einher, zur Validierung Konstrukte auszuwählen, die anverwandt und/oder relevant für die gemessene Kompetenz sind, und die es zugleich ermöglichen, Aussagen über allgemeine und domänenübergreifende Fähigkeiten zu treffen. Die gewählten Konstrukte stellen allerdings nur eine erste, jedoch nicht erschöpfende Auswahl an Vergleichsmöglichkeiten dar. In weiteren Untersuchungen wäre es wünschenswert, die Zusammenhänge mit weiteren Konstrukten zu überprüfen. Die Ergebnisse erster Anwendungsstudien tragen hierzu bereits bei. So zeigen aktuelle Studien von Lang und Kollegen (Lang et al. 2021, 2020), dass der TEWI-Wert nicht mit individueller Lesefähigkeit korreliert und zur Lösung des TEWI zusätzlich andere Fähigkeiten erforderlich sind.

Eine weitere Einschränkung ist, dass die Validierung nicht an einer Stichprobe von Schülerinnen und Schülern in der gymnasialen Oberstufe durchgeführt wurde. Die stattdessen rekrutierte Studierendenstichprobe dürfte im Hinblick auf wissenschaftspropädeutische Kompetenzen eine positive Vorauswahl darstellen: zum einen, da Schülerinnen und Schüler mit geringer wissenschaftspropädeutischer Kompetenz vermutlich seltener ein Studium beginnen als solche mit hoher wissenschaftspropädeutischer Kompetenz. Zum anderen ist davon auszugehen, dass Studierende über ein ausgeprägteres wissenschaftsbezogenes Selbstkonzept verfügen und in ihrer Studienzeit weitere wissenschaftspropädeutische Kompetenzen erworben haben.

Zu diesen Annahmen passt, dass der TEWI für die Validierungsstichprobe mit Studierenden vergleichsweise leicht ausfiel. Es ist deshalb wichtig, in weiteren Studien zu prüfen, wie ihn Schülerinnen und Schüler bewältigen. Das ist im Rahmen der im NEPS erhobenen Daten bereits möglich. Der NEPS-Datensatz erlaubt darüber hinaus, Zusammenhänge mit Antezedenzen und die prädiktive Validität des TEWI zu prüfen. Darauf gehen wir im Folgenden näher ein.

## 4.3 Relevanz für NEPS

Der TEWI wurde für den Einsatz in Klassenstufe 12 der NEPS Startkohorten 3 und 4 entwickelt. Die erste Befragung beider Startkohorten fand 2010 statt (Startkohorte 3 in Klassenstufe 5, Startkohorte 4 in Klassenstufe 9), anschließend wurden die Teilnehmerinnen und Teilnehmer bis 2020, auch nach Verlassen des allgemeinbildenden Schulsystems, wiederholt erneut befragt. Der TEWI wurde der Startkohorte 4 im Schuljahr 2013 in der siebten Erhebungswelle vorgelegt ( $N_{\text{Kohorte4}} = 3893$ ) und der Startkohorte 3 im Schuljahr 2018 in der neunten Erhebungswelle 9 ( $N_{\text{Kohorte3}} = 1775$ ). Der Einsatz des TEWI in diesen großangelegten Längsschnittstudien ermöglicht es zu überprüfen, von welchen Antezedenzen das Abschneiden im TEWI abhängt,

und er wird es ermöglichen zu untersuchen, welche prädiktive Validität der TEWI im Hinblick auf spätere Bildungsrenditen besitzt. Als Antezedenzen können beispielsweise Merkmale des familiären Hintergrundes (wie Sozialkapital, Kulturkapital, Einkommen), Persönlichkeitseigenschaften, Selbstkonzepte, Interessen, motivationale Variablen und akademische Leistungskriterien betrachtet werden. Als Bildungsrenditen werden subjektive Einschätzungen verschiedener Erfolgsmaße zu Ausbildung, Studium und Beruf zur Verfügung stehen (siehe [www.neps-data.de](http://www.neps-data.de)). Die vorliegende Validierungsstudie ergänzt den NEPS-Datensatz um weitere Wissenschaftspropädeutik nahestehende Konstrukte und Daten zur Lösung des TEWI bei einer Studierendenstichprobe.

#### 4.4 Relevanz für das Forschungsfeld der Wissenschaftspropädeutik

Die in der Forschung diskutierten Schwierigkeiten in der Entwicklung geeigneter Testinstrumente für Wissenschaftspropädeutik spiegeln die Komplexität des Konstruktes wider (Dettmers et al. 2010; Griese 1983). Mit dem vorliegenden Test TEWI wurde ein neuer Versuch unternommen, Aspekte des Bildungsziels der Wissenschaftspropädeutik messbar zu machen. Die Ergebnisse der Validierungsstudie stimmen positiv, dass der TEWI ein Instrument mit zufriedenstellender Reliabilität und zufriedenstellenden Skalierungseigenschaften darstellt und es ermöglicht, wissenschaftsbezogene Reflexionskompetenz als Subfacette von Wissenschaftspropädeutik zu erfassen<sup>5</sup>. Der TEWI schafft damit neue Möglichkeiten für die Forschung zum Bildungsziel der Wissenschaftspropädeutik, wie sie beispielhaft für NEPS ausgeführt wurden. Doch auch über NEPS hinaus, kann die Untersuchung wissenschaftsbezogener Reflexionsfähigkeit zur weiteren Forschung in dem Feld beitragen.

Wie in der theoretischen Einleitung dargestellt, enthält das Bildungsziel der Wissenschaftspropädeutik einen gewichtigen bildungstheoretischen Kern: Die Fähigkeit zur verantwortungsvollen Auseinandersetzung mit wissenschaftlichen Erkenntnissen bezeichnet eine grundlegende Kompetenz heranwachsender Bürgerinnen und Bürger einer demokratischen Gesellschaft. Die Operationalisierung wissenschaftlicher Reflexionskompetenz als Verständnis von Wissenschaftskontroversen wurde systematisch aus den Forschungsdiskursen zu Wissenschaftspropädeutik und Scientific Literacy abgeleitet und umgesetzt – und könnte dabei nicht aktueller sein: In Zeiten der Covid19-Pandemie sind wir als Laiinnen und Laien alltäglich mit neuen, inkonsistenten und auch widersprüchlichen Aussagen von Expertinnen und Experten konfrontiert, und müssen abwägen, welche Informationen wir nutzen und wessen Empfehlungen wir folgen. In drastischer Weise verdeutlicht die Pandemie die Vorläufigkeit, Unsicherheit und Widersprüchlichkeit wissenschaftlichen Wissens. Sie

<sup>5</sup> Auch wenn der TEWI spezifisch für das NEPS entwickelt wurde, ist es möglich, ihn außerhalb des NEPS einzusetzen. Um eine hohe Testsicherheit zu gewährleisten, ist eine Überlassung des TEWI nur auf Antrag beim Leibniz-Institut für Bildungsverläufe e. V. (LifBi) möglich, unter Einhaltung bestimmter Bedingungen, insbesondere der folgenden: a) ein Vergleich der Daten, die im eigenen Vorhaben mit dem TEWI erhoben werden, mit den Daten der NEPS-Studie sind zentral für das eigene Forschungsvorhaben, b) es wird der komplette TEWI eingesetzt und c) die erforderlichen Sicherheitsbedingungen („Testsicherheit“) werden eingehalten.

zeigt auf, dass Wissenschaft ein soziales Unternehmen darstellt und im Diskurs entsteht. Ein Verständnis dessen ist unabdingbar, um als Laiin oder Laie Aussagen sowie Informationsquellen hinsichtlich ihrer Verlässlichkeit und Vertrauenswürdigkeit beurteilen zu können (Chinn et al. 2016; Thomm und Bromme 2016).

Erkennen Laiinnen und Laien die Diskursivität von Wissenschaft, können sie dieses Verständnis nutzen, um sich Widersprüche zu erklären. Angemessenes Wissen um die Ursachen solcher Kontroversen ermöglicht es ihnen (wieder) Sinn und Kohärenz zu erzeugen. Begreifen sie diese Gründe, kann es ihnen helfen, wissenschaftliche Aussagen auch in ihrer Vorläufigkeit und Unsicherheit zu bewerten und zu akzeptieren; mangelt es Personen jedoch an einem grundlegenden Verständnis dessen, fehlen ihnen wichtige Ressourcen, Widersprüche überhaupt bewältigen zu können (Allchin 2011; Barzilai et al. 2020; Thomm et al. 2015). Ein angemessenes Verständnis kann zudem von hoher Bedeutung sein, um aus der heutigen Informationsflut tatsächlich wissenschaftliche (wenn auch kontroverse) Aussagen identifizieren zu können. Folglich schließt sich unmittelbar die übergreifende Forschungsfrage an, in welcher Weise wissenschaftsbezogene Reflexionsfähigkeiten die Akzeptanz und Rezeption von wissenschaftlichen Erkenntnissen beeinflussen können.

**Zusatzmaterial online** Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1007/s11618-021-01057-w>) enthalten.

**Danksagung** Das in diesem Beitrag vorgestellte Instrument wurde im Rahmen des Nationalen Bildungspanels (NEPS) entwickelt, das vom Bundesministerium für Bildung und Forschung (BMBF) gefördert wird. Wir danken dem BMBF für seine Unterstützung.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

**Interessenkonflikt** J. Kramer, E. Thomm, K. Oschatz, W. Wagner und R. Bromme geben an, dass kein Interessenkonflikt besteht.

## Literatur

- Adams, R. J., Wu, M. L., & Wilson, M. R. (2012). *ACER ConQuest 3.0*. Melbourne: ACER. computer program
- Allchin, D. (2011). Evaluating knowledge of the nature of (whole) science. *Science Education*, 95(3), 518–542. <https://doi.org/10.1002/sce.20432>.

- Barzilai, S., & Chinn, C. A. (2020). A review of educational responses to the “post-truth” condition: Four lenses on “post-truth” problems. *Educational Psychologist*, 55(3), 107–119. <https://doi.org/10.1080/00461520.2020.1786388>.
- Barzilai, S., & Weinstock, M. (2015). Measuring epistemic thinking within and across topics: A scenario-based approach. *Contemporary Educational Psychology*, 42, 141–158. <https://doi.org/10.1016/j.cedpsych.2015.06.006>.
- Barzilai, S., Thomm, E., & Shlomi-Elouo, T. (2020). Dealing with disagreement: The roles of topic familiarity and disagreement explanation in reasoning about conflicting expert claims and sources. *Learning and Instruction*. <https://doi.org/10.1016/j.learninstruc.2020.101367>.
- Benner, D. (2002). Die Struktur der Allgemeinbildung im Kerncurriculum moderner Bildungssysteme. Ein Vorschlag zur bildungstheoretischen Rahmung von PISA. *Zeitschrift für Pädagogik*, 48, 68–90. <https://doi.org/10.25656/01:3821>.
- Bertram, C., Wagner, W., & Trautwein, U. (2016). Zeitzeugenbefragungen im Geschichtsunterricht: Entwicklung eines Kurzinstruments für die Wirksamkeitsmessung. In T. Arand & M. Seidenfuß (Hrsg.), *Nachwuchstagung der Konferenz für Geschichtsdidaktik 2012* (S. 191–208).
- Blankertz, H. (1966). Bildungstheorie und Ökonomie. In K.-H. Rebel (Hrsg.), *Pädagogische Provokationen. I. Texte zur Schulreform*. Weinheim: Beltz.
- Blossfeld, H.-P., & von Maurice, J. (2011). Education as a lifelong process. *Zeitschrift für Erziehungswissenschaft*, 14(2), 19–34. <https://doi.org/10.1007/s11618-011-0179-2>.
- Bromme, R. (2020). Informiertes Vertrauen: Eine psychologische Perspektive auf Vertrauen in Wissenschaft. In M. Jungert, A. Frewer & E. Mayr (Hrsg.), *Wissenschaftsreflexion. Interdisziplinäre Perspektiven zwischen Philosophie und Praxis* (S. 105–134). Paderborn: Mentis. [https://doi.org/10.30965/9783957437372\\_006](https://doi.org/10.30965/9783957437372_006).
- Bromme, R., Stadler, M., & Scharrer, L. (2018). The provenance of certainty: Multiple source use and the public engagement with science. In J. L. G. Braasch, I. Bråten & M. T. McCrudden (Hrsg.), *Handbook of multiple source use* (S. 269–284). New York: Routledge.
- Bromme, R., Thomm, E., & Wolf, V. (2015). From understanding to deference: Laypersons’ and medical students’ views on conflicts within medicine. *International Journal of Science Education, Part B: Communication and Public Engagement*, 5(1), 68–91. <https://doi.org/10.1080/21548455.2013.849017>.
- Chinn, C. A., Rinehart, R. W., & Buckland, L. A. (2016). Epistemic cognition and evaluating information: Applying the AIR model of epistemic cognition. In D. Rapp & J. Braasch (Hrsg.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (S. 425–453). Cambridge: MIT Press.
- Dettmers, S., Trautwein, U., Neumann, M., & Lütke, O. (2010). Aspekte von Wissenschaftspropädeutik. In U. Trautwein, M. Neumann, G. Nagy, O. Lütke & K. Maaz (Hrsg.), *Schulleistungen von Abiturienten. Die neu geordnete Oberstufe auf dem Prüfstand* (S. 243–269). Wiesbaden: VS. [https://doi.org/10.1007/978-3-531-92037-5\\_9](https://doi.org/10.1007/978-3-531-92037-5_9).
- Eberle, F., Gehrer, K., Jaggi, B., Kottonau, J., Oepke, M., & Pflüger, M. (2008). *Evaluation der Maturitätsreform 1995 (EVAMAR): Schlussbericht zur Phase II*. Zürich: Universität Zürich.
- Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology*, 71(2), 390–405.
- European Commission (2007). *Science education now: A renewed pedagogy for the future of Europe* (Community Research Expert Group, Bd. 22845). Luxembourg: EUR-OP.
- Feinstein, N. (2011). Salvaging science literacy. *Science education*, 95(1), 168–185. <https://doi.org/10.1002/sc.20414>.
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Neuhaus, B., Dorber, L., Pankofer, S., Fischer, M., Strijbos, J.-W., Heene, M., & Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, 2, 28–45. <https://doi.org/10.14786/flr.v2i3.96>.
- Fischer, L., Rieger, S., Hübner, N., Oschatz, K., Kramer, J., & Wagner, W. (2020). *NEPS technical report for scientific thinking: Scaling results for the Starting Cohorts 3 (wave 9) and 4 (wave 7) in 12th grade* (NEPS survey paper, Bd. 66). Bamberg: Leibniz Institute for Educational Trajectories (LifBi). National Educational Panel Study. [https://www.neps-data.de/Portals/0/Survey%20Papers/SP\\_LXVI.pdf](https://www.neps-data.de/Portals/0/Survey%20Papers/SP_LXVI.pdf). Zugegriffen: 29. Nov. 2021.
- Greene, J. A., Sandoval, W. A., & Bråten, I. (2016). *Handbook of epistemic cognition* (1. Aufl.). New York: Routledge.

- Griese, W. (1983). *Wissenschaftspropädeutik in der gymnasialen Oberstufe* (Dissertation). Oldenburg: Universität Oldenburg.
- Hahn, S. (2009). Wissenschaftspropädeutik in der Sekundarstufe II. Bildungsgeschichtlicher Rückblick und aktuelle Entwicklungen. *TriOS – Forum für schulnahe Forschung, Schulentwicklung und Evaluation*, 4(2), 5–38.
- Hahn, S. (2013). Wissenschaftspropädeutik in der gymnasialen Oberstufe. In D. Bosse, F. Eberle & B. Schneider-Taylor (Hrsg.), *Standardisierung in der gymnasialen Oberstufe* (S. 161–174). Wiesbaden: Springer VS. [https://doi.org/10.1007/978-3-658-00658-7\\_10](https://doi.org/10.1007/978-3-658-00658-7_10).
- von Hentig, H. (1966). Gedanken zur Neugestaltung der Oberstufe. In H. von Hentig (Hrsg.), *Analysen und Modelle zur neuen Schulreform* (Neue Sammlung, Bd. 3, Sonderheft, S. 31–58). Göttingen: Vandenhoeck & Ruprecht.
- Heymann, H. W., van Lück, W., Meyer, M., Schulze, T., & Tenorth, H.-E. (1990). Allgemeinbildung als Aufgabe der öffentlichen Schule. Bilanz einer Diskussion. In H. W. Heymann & W. van Lück (Hrsg.), *Allgemeinbildung und öffentliche Schule: Klärungsversuche* (S. 9–20). Bielefeld: Institut für Didaktik der Mathematik.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88–140. <https://doi.org/10.3102/003465430670010884>.
- Höttecke, D., & Allchin, D. (2020). Re-conceptualizing nature of science education in the age of social media. *Science Education*, 104(4), 641–666. [https://doi.org/10.1007/978-3-030-47260-3\\_2](https://doi.org/10.1007/978-3-030-47260-3_2).
- Huber, L. (1997). Fähigkeit zum Studieren. Bildung durch Wissenschaft. Zum Problem der Passung zwischen Gymnasialer Oberstufe und Hochschule. In E. Liebau (Hrsg.), *Grundlagentexte Pädagogik. Das Gymnasium: Alltag, Reform, Geschichte, Theorie* (S. 333–351). Weinheim: Juventa.
- Huber, L. (2000). Wissenschaftspropädeutik, allgemeine Studierfähigkeit und ihre unterrichtliche Umsetzung in Grundkursen. In Hessisches Landesinstitut für Pädagogik (Hrsg.), *Bildung braucht guten Grund: Beiträge zur Reform der Grundkurse* (S. 17–46). Wiesbaden: HeLP.
- Huber, L. (2001). Stichwort: Fachliches Lernen. Das Fachprinzip in der Kritik. *Zeitschrift für Erziehungswissenschaft*, 4(3), 307–331. <https://doi.org/10.1007/s11618-001-0040-0>.
- Huber, L. (2005). Warum fächerübergreifender Unterricht in der Sekundarstufe II? In L. Huber & K.-J. Tillmann (Hrsg.), *Versuchsschulen und das Regelschulsystem – Bielefelder Erfahrungen* (S. 109–124). Bielefeld: Laborschule Bielefeld.
- Huber, L. (2009). Wissenschaftspropädeutik ist mehr! *TriOS – Forum für schulnahe Forschung, Schulentwicklung und Evaluation*, 4(2), 39–60.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2. Aufl.). Beverly Hills: SAGE.
- Keller, J., Bohner, G., & Erb, H.-P. (2000). Intuitive und heuristische Urteilsbildung – verschiedene Prozesse? Präsentation einer deutschen Fassung des ‚Rational-Experiential Inventory‘ sowie neuer Selbstberichtskaalen zur Heuristiknutzung. *Zeitschrift für Sozialpsychologie*, 31(2), 87–101. <https://doi.org/10.1024//0044-3514.31.2.87>.
- Kienhues, D., Thomm, E., & Bromme, R. (2018). Specificity reloaded: How multiple layers of specificity influence reasoning in science argument evaluation. In F. Fischer, C. A. Chinn, K. Engelmann & J. Osborne (Hrsg.), *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge* (S. 251–270). London: Taylor & Francis.
- KMK (1977). Empfehlungen zur Arbeit in der gymnasialen Oberstufe. *Bildung und Erziehung*, 31, 561–574.
- KMK (1995). *Weiterentwicklung der Prinzipien der gymnasialen Oberstufe und des Abiturs. Abschlussbericht der von der Kultusministerkonferenz eingesetzten Expertenkommission*. Bonn: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.
- Köller, O., Watermann, R., Trautwein, U., & Lüdtke, O. (Hrsg.). (2004). *Wege zur Hochschulreife in Baden-Württemberg*. Opladen: Leske + Budrich.
- Kolstø, S. D. (2001). Scientific literacy for citizenship: Tools for dealing with the science dimension of controversial socioscientific issues. *Science education*, 85, 291–310. <https://doi.org/10.1002/sce.1011>.
- Körper, A., Schreiber, W., & Schöner, A. (Hrsg.). (2007). *Kompetenzen historischen Denkens. Ein Strukturmodell als Beitrag zur Kompetenzorientierung in der Geschichtsdidaktik*. Neuried: ars una.
- Krettenauer, T. (2005). Die Erfassung des Entwicklungsniveaus epistemologischer Überzeugungen und das Problem der Übertragbarkeit von Interviewverfahren in standardisierte Fragebogenmethoden. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37, 69–79. <https://doi.org/10.1026/0049-8637.37.2.69>.
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.

- Lang, F., Kammerer, Y., Oschatz, K., Stürmer, K., & Gerjets, P. (2020). The role of beliefs regarding the uncertainty of knowledge and mental effort as indicated by pupil dilation in evaluating scientific controversies. *International Journal of Science Education*, 42(3), 350–371. <https://doi.org/10.1080/09500693.2019.1710875>.
- Lang, F., Kammerer, Y., Stürmer, K., & Gerjets, P. (2021). Investigating professed and enacted epistemic beliefs about the uncertainty of scientific knowledge when students evaluate scientific controversies. *European Journal of Psychology of Education*, 36(1), 125–146. <https://doi.org/10.1007/s10212-020-00471-8>.
- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, 29, 331–359. <https://doi.org/10.1002/tea.3660290404>.
- Lederman, N. G. (2007). Nature of science: Past, present and future. In S. Abell & N. G. Lederman (Hrsg.), *Handbook of research on science education* (S. 831–879). Mahwah: Lawrence Erlbaum. <https://doi.org/10.4324/9780203824696>.
- Müsche, H. (2009). Wissenschaftspropädeutik aus psychologischer Perspektive – Zur Dimensionierung und Konkretisierung eines bildungstheoretischen Konzeptes. *TriOS*, 4(2), 61–109.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus – Statistical analysis with latent variables. Version 7.0*. Los Angeles: Muthén & Muthén. (1998–2012)
- OECD (2007). *PISA 2006 Science competencies for tomorrow's world* (Bd. 1). Paris: OECD Publishing. <https://doi.org/10.1787/9789264040014-en>.
- OECD (2011). *Education at a glance, 2011: OECD indicators*. Paris: OECD. <https://doi.org/10.1787/eag-2011-en>.
- OECD (2017). *PISA 2015 assessment and analytical framework: Science, reading, mathematics, financial literacy and collaborative problem solving*. Paris: OECD Publishing. <https://doi.org/10.1787/9789264281820-en>.
- OECD (2019). *PISA 2018 assessment and analytical framework*. Paris: OECD Publishing. <https://doi.org/10.1787/b25efab8-en>.
- Oreskes, N. (2019). *Why trust science?* Princeton: Princeton University Press. <https://doi.org/10.1515/9780691189932>.
- Oschatz, K., Kramer, J., & Wagner, W. (2018). *The assessment of Scientific Thinking as metascientific reflection*. Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study. [https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/Kompetenzen/NEPS\\_com\\_st\\_2018\\_fin.pdf](https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/Kompetenzen/NEPS_com_st_2018_fin.pdf). Zugegriffen: 24. Sep. 2019.
- R Core Team (2019). *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing. <http://www.R-project.org/>. Zugegriffen: 31. Dez. 2019.
- Ramseier, E., Allraum, J., Stalder, U., Grin, F., Alliata, R., Müller, S., & Stocker, E. (2004). *Evaluation der Maturitätsreform 1995 (EVAMAR)*. Bern: Eidgenössisches Departement des Innern.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Robitzsch, A., Kiefer, T., & Wu, M. (2019). TAM: Test Analysis Modules. R package version 1.7-56. <http://CRAN.R-project.org/package=TAM>. Zugegriffen: 31. Dez. 2019.
- Sharon, A. J., & Baram-Tsabari, A. (2020). Can science literacy help individuals identify misinformation in everyday life? *Science Education*, 104(5), 873–894. <https://doi.org/10.1002/sce.21581>.
- Stadler, M., & Bromme, R. (2014). The content–source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In D. N. Rapp & J. Braasch (Hrsg.), *Processing inaccurate information: theoretical and applied perspectives from cognitive science and the educational sciences* (S. 379–402). Cambridge: MIT Press.
- Stahl, E., & Bromme, R. (2007). The CAEB: An instrument for measuring connotative aspects of epistemological beliefs. *Learning and Instruction*, 17, 773–785. <https://doi.org/10.1016/j.learninstruc.2007.09.016>.
- Tabak, I. (2018). Functional scientific literacy: disciplinary literacy meets multiple source use. In J. L. G. Braasch, I. Braten & M. T. McCrudden (Hrsg.), *Handbook of multiple source use* (S. 221–237). New York: Routledge.
- Thomm, E., & Bromme, R. (2016). How source information shapes lay interpretations of science conflicts: Interplay between sourcing, conflict explanation, source evaluation, and claim evaluation. *Reading and Writing*, 29(8), 1629–1652. <https://doi.org/10.1007/s11145-016-9638-8>.
- Thomm, E., Barzilai, S., & Bromme, R. (2017). Why do experts disagree? The role of conflict topics and epistemic perspectives in conflict explanations. *Learning and Instruction*, 52, 15–26. <https://doi.org/10.1016/j.learninstruc.2017.03.008>.

- Thomm, E., Hentschke, J., & Bromme, R. (2015). The Explaining Conflicting Scientific Claims (ECSC) questionnaire: Measuring laypersons' explanations for conflicts in science. *Learning and Individual Differences*, 37, 139–152. <https://doi.org/10.1016/j.lindif.2014.12.001>.
- Trautwein, U., & Lüdtke, O. (2004). Aspekte von Wissenschaftspropädeutik und Studierfähigkeit. In O. Köller, R. Watermann, U. Trautwein & O. Lüdtke (Hrsg.), *Wege zur Hochschulreife in Baden-Württemberg: TOSCA – Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien* (S. 327–366). Opladen: Leske + Budrich.
- Trautwein, U., & Lüdtke, O. (2007). Epistemological beliefs, school achievement, and college major: A large-scale longitudinal study on the impact of certainty beliefs. *Contemporary Educational Psychology*, 32, 348–366. <https://doi.org/10.1016/j.cedpsych.2005.11.003>.
- Warm, T. A. (1989). Weighted likelihood estimation of ability in the item response theory. *Psychometrika*, 54, 427–450.
- Zeuch, N., & Souvignier, E. (2015). Zentrale Facetten wissenschaftlichen Denkens bei Lehramtsstudierenden – Entwicklung eines neuen Instruments und Identifikation von latenten Profilen. *Unterrichtswissenschaft*, 43(3), 245–262.