

Bremer, Rainer

Nothing but Evidence - Bildungsforschung aus bildungsfeindlicher Absicht und eine Alternative (Teil II)

Pädagogische Korrespondenz (2011) 44, S. 30-55



Quellenangabe/ Reference:

Bremer, Rainer: Nothing but Evidence - Bildungsforschung aus bildungsfeindlicher Absicht und eine Alternative (Teil II) - In: *Pädagogische Korrespondenz* (2011) 44, S. 30-55 - URN: urn:nbn:de:0111-opus-88309 - DOI: 10.25656/01:8830

<https://nbn-resolving.org/urn:nbn:de:0111-opus-88309>

<https://doi.org/10.25656/01:8830>

in Kooperation mit / in cooperation with:



Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

INSTITUT FÜR PÄDAGOGIK UND GESELLSCHAFT

PÄDAGOGISCHE KORRESPONDENZ

HEFT 44

HERBST 2011

*Zeitschrift für
Kritische Zeitdiagnostik
in Pädagogik und
Gesellschaft*

BUDRICH UNIPRESS OPLADEN & FARMINGTON HILLS, MI

Die Zeitschrift wird herausgegeben vom
Institut für Pädagogik und Gesellschaft e.V. Münster,
im Verlag Budrich UniPress, Leverkusen

Redaktionsadresse ist:

Institut für Pädagogik und Gesellschaft e.V.
Windmühlstraße 5, 60329 Frankfurt am Main, Tel. 069/5973596

Redaktion:

Karl-Heinz Dammer (Heidelberg)
Peter Euler (Darmstadt)
Ilan Gur Ze'ev (Haifa)
Andreas Gruschka (Frankfurt am Main)
Bernd Hackl (Graz)
Sieglinde Jornitz (Frankfurt am Main)
Andrea Liesner (Hamburg)
Andreas Wernet (Hannover)
Antonio Zuin (São Carlos)

Manuskripte werden als word-Dateien an den geschäftsführenden Herausgeber erbeten (a.gruschka@em.uni-frankfurt.de) und durchlaufen ein Begutachtungsverfahren.

Abonnements und Einzelbestellungen:

Institut für Pädagogik und Gesellschaft e.V.
Windmühlstraße 5, 60329 Frankfurt am Main, Tel. 069/5973596
Der Jahresbezugspreis der *Pädagogischen Korrespondenz*
beträgt im Inland für zwei Ausgaben 23,- EURO zzgl. 4,- EURO Versand.
Das Einzelheft kostet im Inland 12,50 EURO zzgl. 2,50 EURO Versand.
Bezugspreise Ausland jeweils zzgl. gewünschtem Versandweg.
Kündigungsfrist: schriftlich, drei Monate zum Jahresende.

Copyright:

© 2011 für alle Beiträge soweit nicht anders vermerkt sowie für
den Titel beim Institut für Pädagogik und Gesellschaft, Münster.
Originalausgabe. Alle Rechte vorbehalten.
ISSN 0933-6389

Buchhandelsvertrieb:

Institut für Pädagogik und Gesellschaft e.V.

Satz & Layout: Susanne Albrecht-Rosenkranz, Leverkusen

Anzeigen und Gesamtherstellung:

Verlag Budrich UniPress Ltd., Stauffenbergstr. 7, D-51379 Leverkusen
ph +49 (0)2171 344694 • fx +49 (0)2171 344693
www.budrich-unipress.de

- 5 **DAS AKTUELLE THEMA**
 Günter Rüdell
 Strukturreform auf Konsenskurs
- 23 **NACHGELESEN**
 Blick aus der Ferne
 Wie die DDR-Pädagogik die BRD-Pädagogik beurteilte
- 30 **AUS WISSENSCHAFT UND FORSCHUNG**
 Rainer Bremer
 Nothing but Evidence – Bildungsforschung aus bildungsfeindlicher
 Absicht und eine Alternative (Teil 2)
- 56 **DOKUMENTATION**
 Ioanna Menhard
 Erfahrungen und Umgang mit dem Bologna-Prozess an einem
 erziehungswissenschaftlichen Institut
- 71 **ERZIEHUNG NEU**
 Christoph Leser
 Das Ende der Erziehung
- 83 **DIDAKTIKUM**
 Jens Rosch
 Eine Fallstudie zum Lernen von Algebra im Schulunterricht
- 104 **AUS DEN MEDIEN**
 Andreas Gruschka
 Die Vermittlung von etwas Unerhörtem und bislang weitgehend
 Ungesehenem
- 111 **IN MEMORIAM**
 Armin Bernhard
 „Standhalten im Dasein“ – In memoriam Hans-Jochen Gamm

Rainer Bremer

Nothing but Evidence – Bildungsforschung aus bildungsfeindlicher Absicht und eine Alternative (Teil II)

Über das, was wahr und was bloße Meinung, nämlich Zufall und Willkür sein soll, entscheidet nicht, wie die Ideologie es will, die Evidenz, sondern die gesellschaftliche Macht, die das als bloße Willkür denunziert, was mit ihrer eigenen Willkür nicht zusammenstimmt. Die Grenze zwischen der gesunden und der pathogenen Meinung wird in praxi von der geltenden Autorität gezogen, nicht von sachlicher Einsicht.
Adorno

I

Wann erscheint der Schein wahrscheinlich?

Wenn von Evidenzen auszugehen ist.¹ Dazu müsste sich Evidenz gegen Nicht-Evidenz durchsetzen. So einfach ist das gar nicht. Die Herstellung von Evidenz benötigt Verfahren, die ihrerseits weder evident noch trivial und vor allem nicht voraussetzungslos sind. Von Evidenz ist vorzüglich dann die Rede, wenn etwas zu beweisen wäre, das aus einem beliebig gespeisten Gestus des Zweifels in Frage gestellt werden könnte. Der Evidenzbeweis benötigt jene Intention, recht zu haben, die denjenigen, der den Beweis führen zu können prätendiert, in die Lage versetzt, diejenigen, die den Zweifel äußern, in der nämlichen Absicht widerlegt, selbst im Recht zu sein. Diese müssen gegen ihren Zweifel etwas einsehen, weil sie es nicht bestreiten können. Was im Streit steht oder stehen könnte, ist auf Unstrittiges zurückzuführen. Wenn es nicht um Rechtsprechung, sondern um Wissenschaft geht, können die

¹ Dieser Beitrag folgt auf einen ersten Teil, der in Heft 42 der Pädagogischen Korrespondenz abgedruckt wurde (S. 34-51). Dessen Schlussabschnitt argumentiert mit der Piagetschen Kategorie der Operationalität und stellt sie in den Zusammenhang mit den Operationsmodi Rezeption und Produktion. Es heißt dort „... [die Rezeption] also eines Auslösers bedarf, um ihr Potential einzusetzen, während Operation ebenfalls ein Potential darstellt, das durch Erfahrung, sprich den Gebrauch operativer Fähigkeiten, wächst. Aber im Gegensatz zur Rezeption, die je nach Intensität des Anlasses bis zu einem zwingenden Grad ausgelöst werden kann, fehlt dem Modus der Operation das Zwangsläufige. Ob jemand sich operationell verhält, ist weitaus mehr von diesem selbst abhängig als die Rezeption vom Rezipienten. Sind hier das Wie und das Was subjektiv gefärbt, bleibt da schon das Ob eine variable Größe. Damit wäre auch das Problem der Empirie solcher Bildungstheorie beschrieben [...] Was an das angenommene Ende, der Ausbildung formal-operationellen Denkens, als adoleszenzreife Intelligenz im operativen Umgang mit abverlangten Einsichten sich anschließen würde, lag für ihn [Piaget] außerhalb der beobachtbaren Entwicklung. Dies wäre das Thema einer empirischen Bildungstheorie. Auf ihrer Grundlage könnte das Kontrastprogramm zur ‚empirischen Bildungsforschung‘ entstehen.“

Dinge noch komplizierter werden. Dann wäre diese Basis des Unstrittigen selbst zu erzeugen.

Das Verlässliche bei Evidenzbeweisen sind die mathematischen Nachweisverfahren. Um sie anzuwenden, braucht es Daten. Und zwar unbezweifelbare. Der Zusammenhang, in den Daten als solche gestellt werden, entsteht aus Methoden. An ihrem Anwendungsfall sind die Daten und deren Erzeugung kritisch. Die Methoden selbst gelten als finalisiert. Es mag neue geben, aber nicht, weil die vorhandenen nicht korrekt wären. Sie erfüllen ihren Zweck vielleicht nicht auf vollkommene Weise. Man könnte sagen, die „Schnittstelle“ zur Wirklichkeit stellt eine „Herausforderung“ dar, für deren Meisterung der Glaube an die Evidenz ebenso maßgeblich wie unnützlich bleibt.

Aus Daten werden Fakten, wenn ein Faktum auf einen einleuchtenden Zusammenhang von Daten zurückzuführen ist. Zusammenhänge, die einleuchten können, sollten berechenbar sein. Stochastische Verfahren behandeln Daten in zufallsverteilten Kombinationen. Erst wenn die Kombinationen in einer Häufigkeit auftreten, die einem zufälligen Zustandekommen eben wegen der nicht mehr zufallskonformen Verteilung entsprechen, gilt der Zufall als ausgeschaltet. Dann sollte man von anderen Gründen ausgehen, die die in Zahlen dargelegte Sachlage aufklären mögen.

Wegen ihrer späteren Verwendung unter stochastischen Prozeduren verlieren die Daten schon vor ihrer Erzeugung jene Unschuld, die der Abbildung empirischer Sachverhalte eigentlich eignen müsste. Der Blick auf die Tatsachen und schon gar ein wissenschaftlich interessierter sollte unbefangener nicht sein können. Sich aus Interesse den Dingen zuzuwenden, setzt jene Erfahrung in Gang, derer der Verstand bedarf, um zur Vernunft zu kommen. Interessierte Zuwendung geht dem Verstehen voraus. Daher sollte die Erfahrung frei sein. Andernfalls entziehen ihre Surrogate dem Verstand jene Tatsachen und Zusammenhänge, aus denen das Komplement zur Erfahrung entsteht, die Erkenntnis.

Die im Namen reiner Evidenz beschworene Dominanz stochastischer Methoden und Prinzipien verdiente bei empirischen Untersuchungen eigentlich Misstrauen. Wer sich vorweg die Skala vorstellt, der sich seine künftigen Daten fügen sollen, wird sich bei der Erzeugung derselben schon zu helfen wissen. Reflektierte Empiriker räumen den virtuell limitierenden Einfluss von randscharf ausgezogenen Dimensionen, vollständigen Matrizen und beliebig substituierbaren Items ein, verweisen aber auf die Alternativlosigkeit solcher Methoden, mittels derer sich schlüssig eine wissenschaftlich gesicherte Aussage von anderen unterscheiden lässt.

Damit greifen sie, bei aller Lauterkeit, doch recht weit in die Daten hinein. Allerdings können sie das nur, weil sie die Reichweite wissenschaftlicher Erkenntnis selbst beschneiden. Ein dazu nicht einmal bekannter Teil der Daten fällt aus der Erhebungsmöglichkeit schlechterdings heraus. Daten, die als methodische Artefakte zunächst erzeugt werden, sind, was die lauterer Empiriker auch noch eingestehen würden, nicht nur Artefakte in Bezug auf ihre Geltung, sondern auch auf ihre Genesis. Es wäre nicht nur einzuräumen, dass

sie erzeugt wurden – und nicht einfach anfielen – und trotzdem als empirisch gegeben gelten, sondern wie sie erzeugt wurden. Die reine Lehre will, dass die Daten nur stochastisch bearbeitbares interpretandum seien. Es soll keines interpretans mehr bedürfen, auch jenes nicht, das ihnen insofern eingeschlossen ist, als sie davon nur ein individueller Ausdruck des Gemessenen sind. Dass ein item, an dem sich Probanden zum Zweck einer vorgeblich nur vergleichenden Schulleistungsuntersuchung bewähren sollen, für die nämlichen Probanden schlicht auch ein Problem ist, an dem sie ihre Fähigkeiten einsetzen müssen, gilt nur als eine Reaktion, die als response vollständig von der item-response-theory erklärt werden kann und auch nur in deren Rahmen zu deuten wäre. Das response eines Probanden ist vorweg kalkuliert und kann sich im multiple choice-Verfahren entweder bestätigen oder wird durch Auswahl einer falschen item-Alternative ausgeschlossen. Die Bedeutung aber, die das item für die Schüler durch die Bearbeitung selbst gewinnt, fällt für die Beurteilung des Testergebnisses im binären Modus von richtig-falsch vollständig aus. Die falschen Antworten sind ohnehin Anathema, aber auch die richtigen – ob einfach oder schwer – zählen nur als das, als was sie hypothetisch konstruiert wurden. Gegen das response als authentisches Resultat einer gedanklichen Bearbeitung durch die Schüler bleibt der Messbefund blind. Wenn aber Authentizität und Evidenz der Probandenleistungen in einer der Bildungsforschung dienenden Untersuchung dermaßen auseinanderfallen, dann scheint die methodologische Konstruktion solcher Evidenz von jener originären Schülerleistung völlig abgewandt, die dem Anspruch nach auch dann alleiniger Messgegenstand bleibt, wenn es im Großen und Ganzen um international verglichene Schulsysteme geht.

II

Ein Fehler muss erscheinen

Das oben umrissene tiefliegende Problem der Substitution von Authentizität durch Evidenz scheint mit den ihrerseits nicht weniger konstruierten Kriterien der Reliabilität und Validität erledigt. Reliabel sind Daten, die nach ihrem provozierten Erscheinen korrekt abgebildet wurden. Eine Frage der Aufnahme, Kodierung, Chiffrierung und maschinellen Auswertung. Auf alten Lochkarten konnte schon mal Fliegendreck ein Ergebnis verfälschen, weil aus dem Fliegenkot auf dem Strichcode eine Antwort wurde, die die Lesemaschine nicht von der der wirklichen Probanden zu unterscheiden in der Lage war. Ernsthafter stellt sich das Problem der Seriosität der „Datenknechte“. Da diese eine zumal schlecht bezahlte Arbeit leisten, könnte ihnen in den Sinn kommen, Formulare in der erwarteten Zahl abzuliefern, ohne die dazu nötige Arbeitszeit aufzuwenden, die mit der sorgfältigen Befragung von richtig ausgewählten Probandengruppen nun einmal anfällt. Das mit Reliabilität verbundene Qualitätsrisiko von Untersuchungen kann nur eine versierte Testleitung eindämmen. Das zählt zur kostenträchtigen Messtechnologie dazu.

Bei der Validität von Messergebnissen geht es hingegen darum, ob auch getestet wurde, was man testen wollte. Erzielen überhaupt die Messkonstrukte Geltung für die Messung jenes Merkmals, für das man die items konstruiert hat? Zunächst sollte man davon ausgehen, dass die Anstrengungen, die es bedeutet, die erwähnten Kriterien einzuhalten, angesichts der großen Summen, die die Auftraggeber dafür springen lassen, wenn nicht mit Leichtigkeit, so doch mit Sicherheit eingehalten wurden. Was da in deutschen Schulen anlässlich ihrer Beteiligung an TIMSS und PISA an Daten anfiel, sollte vom Zweifel an deren Reliabilität und Validität ausgenommen werden können. Auf die Einhaltung solcher Standards verstehen die Chefs und Exekutoren solcher Studien sich gewiss. Nicht ganz ohne Ironie bleibt aber hinzuzufügen, dass diese Kriterien zusammen mit der operativen Schwierigkeit, sie einzuhalten, vielleicht doch mehr mit dem Geschäftsmodell zu tun haben, als die methodologisch normierte Skepsis gegen die beanspruchte Perfektion eingestehen mag. Die Beherrschung der mit dem Reliabilität und Validität benannten Risiken ist am Ende ein weiteres Argument für die nicht nur potentielle, sondern auch faktisch erzielbare Zuverlässigkeit der Methoden empirischer Bildungsforschung sensu TIMSS und PISA.

Das mit Validität angesprochene Problem mag man durch Testreihen beherrschen können, es reicht prinzipiell hinter jeden Evidenzbeweis bis an die Wurzel empirisch geleiteter Erkenntnis zurück. Wer die methodologischen Standards aus Überzeugung, dazu mit viel Geschick und am Ende mit zureichenden Ressourcen einzuhalten vermag, erzeugt Ergebnisse von wissenschaftlicher Geltung. Die Methodologie reicht in alle Details der Untersuchungen hinein, die Schritte vom Design über die item-Konstruktion über die Datenerhebung und -erfassung erfolgen in bedachtsamer Steuerung. Am Ende steht ein entsprechend kontrolliert erschaffenes Ergebnis. Es sollte die Geltungsansprüche einlösen und seinerseits einer empirischen Überprüfung standhalten. Dagegen können sich die Autoren solcher Studien nicht wehren – gerade sie nicht, die jeden Schwur auf Evidenz geleistet haben.

Die Schweizer Bildungsforscher Barbara Stalder, Thomas Meyer und Sandra Hupka-Brunner haben in der Zeitschrift „Die Deutsche Schule“ im Jahr 2008 über eine Studie berichtet, die anlässlich der PISA-Untersuchungen 2000f. in Auftrag gegeben wurde. In der Schweiz hat man PISA nicht nur bestaunt, sondern wollte wissen, auf welche Referenz der Ergebnisse, außer solchen der gleichen Art, mittelfristig man zu deren Einschätzung angewiesen sein könnte.²

2 In einer Fußnote erläutern die Autoren die Zuständigkeiten: „TREE (Transitions from Education to Employment) wird seit 2008 durch den Schweizerischen Nationalfonds (SNF) und die Universität Basel kofinanziert. Von 2000 bis 2007 lag die finanzielle und/oder organisatorische Verantwortung für TREE – neben dem SNF – bei den Kantonen Bern, Genf und Tessin, beim Bundesamt für Berufsbildung und Technologie (BBT) sowie beim Bundesamt für Statistik (BFS). Nähere Angaben zu TREE unter: <http://www.tree-ch.ch>“ (S. 437). So etwas hätte hier auch das BIBB machen können.

Die Möglichkeit, zu der im Folgenden, ausführlich zitierten Quelle vergleichbare Daten zu produzieren, dürfte singular bleiben.³ Das macht sie für die Frage nach einer das Schicksal, die Fähigkeiten und Chancen der sich bildenden Individuen wirklich erfassenden Bildungsforschung noch interessanter als die einzelnen Befunde, die in einem wohlbegründeten Widerspruch zu den auf die Schweizer Ergebnisse sich stützenden Prognosen des PISA-Konsortiums stehen. Ich zitiere ausführlich:

„Die bisher referierten Ergebnisse illustrieren eindrücklich, dass die am Ende der obligatorischen Schulzeit erreichten Lesekompetenzen den Bildungsverlauf auf der Sekundarstufe II nachhaltig beeinflussen. Offen bleibt aber, warum gewisse Jugendliche trotz vergleichsweise schlechter Leistungsvoraussetzungen einen Sek. II-Abschluss erreichen, während dies anderen nicht gelingt. Ebenso kann durch die PISA-Kompetenzen allein nicht erklärt werden, warum gewisse Jugendliche ihre Ausbildung mit einem EFZ abschließen, während andere einen (Berufs-)Maturitätsabschluss erwerben (DDS 2008, S. 443) [...] In diesem Beitrag analysierten wir den Zusammenhang zwischen den PISA-Lesekompetenzen und der Bildungslaufbahn auf der Sekundarstufe II. Die Resultate bestätigen, dass Jugendliche mit mittleren und hohen PISA-Lesekompetenzen häufiger in anspruchsvolle Ausbildungen einsteigen und häufiger eine gymnasiale oder Berufsmaturität erwerben als Jugendliche mit geringeren Kompetenzen. Mehr als die Hälfte der Jugendlichen mit Lesekompetenzen der Stufe ≤ 1 steigen jedoch ohne Verzögerung in eine Berufsausbildung ein, und 60% erreichen einen Abschluss. Die durch PISA formulierte kritische Schwelle, nach der die ‚potenzielle Risikogruppe‘ (Stufe ≤ 1) gegenüber den Jugendlichen auf Stufe 2 oder höher benachteiligt sei, findet in der multivariaten Analyse kaum empirische Unterstützung. Die Chancen, ein Eidgenössisches Fähigkeitszeugnis (EFZ) zu erwerben, sind für Schulabgängerinnen und Schulabgänger mit Stufe 2 kaum höher als für ihre weniger lesekompetenten Kolleginnen und Kollegen, und beide Gruppen bleiben gleichermaßen von anspruchsvolle-

3 Gegenüber der primären PISA-Studie musste die Schweizer TREE-Untersuchung ein ungleich aufwendigeres Verfahren in Gang setzen. Fünfzehnjährige, die noch der Schulpflicht unterliegen, trifft man vormittags in der Regel in der Schule an. Unter Beachtung der Ferienzeiten und Genehmigungsprozeduren dürften sich beim Besuch der Schulen keine Hindernisse einstellen, dort die erwarteten Probanden auch vorzufinden. Bei einer Verbleibsuntersuchung, die sich einen Zeitraum von sieben Jahren vornimmt und damit akzeptieren muss, dass ihre Probanden sich von Fünfzehnjährigen während der Verfolgung und Befragung zu Zweiundzwanzigjährigen weiterentwickeln, ist bereits nach einem Jahr mit dem genau umgekehrten Fall zu rechnen: Die zuvor klassenweise durch PISA erfassten Schüler haben ihre Schulen verlassen, sie sind im besten Fall lokal auf eine folgende gewechselt, die einem der normierten Bildungswege entspricht. Für die Populationsgruppe von maßgeblichem Interesse gilt aber auch dies schon nicht mehr. Abbrecher und solche, die in keine ihrerseits als Bildungsgang geordnete berufliche Laufbahn einmünden, fallen aus dem System heraus. Die Identifizierung ad personam wurde Bedingung der Möglichkeit, eine solche Studie auch bei ungeordnet verlaufenden Bildungswegen durchzuführen. Dazu mussten die Personen nach Namen und Wohnort identifiziert werden, in die Teilnahme an den weiteren Erhebungen einwilligen und trotz einer ungewissen individuellen Entwicklungslage über einen verhältnismäßig langen Zeitraum das Faktum ihrer kontinuierlichen Identifizierbarkeit sicherstellen. Es ist nochmals darin zu erinnern, dass es speziell bei jener durch die Bildungsverwaltung periodisch kontaktierten Gruppe um solche Personen ging, denen die nämliche Verwaltung in ihren Schulen die am wenigsten erfolgreiche Karriere bescherte. Dergleichen würde in Deutschland, wenn nicht schon am Datenschutz, so spätestens an jener bis in die Unterschicht reichenden Paranoia scheitern, die einen Gewinn an Selbstsicherheit bereits dem verspricht, der seine Telefonnummer preiszugeben sich weigert, außer er wird bei Verlosungen danach gefragt.

ren Abschlüssen ausgeschlossen. Der Einfluss des auf Sekundarstufe I besuchten Schultyps sowie des Bildungsstands und des Berufsstatus der Eltern wird vor allem dort wirksam, wo es um die Anschlussfähigkeit des Sek. II-Abschlusses an die Tertiärstufe geht: bei den (Berufs-)Maturitätszeugnissen (S. 445) [...] Die insgesamt doch positive Bilanz zum Schweizer Berufsbildungssystem dürfte auf ein komplexes Zusammenspiel zwischen den strukturellen Gegebenheiten des Arbeits- und Ausbildungsmarktes, dem Konjunkturverlauf, den Kosten-Nutzen-Relationen der Lehrlingsausbildung und den Reformbestrebungen in der Berufsbildung zurückzuführen sein [...] Dennoch stellen sich auch kritische Fragen. Zwar ist das Berufsbildungssystem in der Lage, der Mehrheit der Jugendlichen – auch der leistungsschwachen – den Erwerb eines zertifizierenden Berufsbildungsabschlusses zu ermöglichen. Auch legen unsere Ergebnisse nahe, dass die schulische bzw. soziale Herkunft für den Abschluss einer beruflichen Grundbildung mit EFZ keine statistisch signifikante Rolle spielt, wenn man das Anforderungsniveau außer Acht lässt. Allerdings muss dabei in Rechnung gestellt werden, dass die schulische und leistungsmäßige Situation der Jugendlichen am Ende der obligatorischen Schulzeit ihrerseits bereits ‚herkunftsgeprägt‘ ist – was PISA ja eindrücklich gezeigt hat“ (S. 446).

Die Frage, die sich angesichts der zitierten Schweizer Nacherhebung stellt, gilt weniger der prognostischen Kraft wissenschaftlich gewonnener Erkenntnisse. Da hätten sich die PISA-Leute ja nur die Prognose verkneifen müssen, um fein heraus zu sein. Arbeitsmarkt, Schulsystem und sonst was mögen in der Schweiz eine Absorption von „Risikogruppen“ bewirkt haben, die die PISA-Prognose auszuhebeln vermochte. Dergleichen Faktoren können in Schulleistungsuntersuchungen gar keine Rolle spielen und deshalb auch nicht hinterher einen Einwand gegen deren Vorgehensweise liefern. Der Kern der Prognose hingegen, ein Scheitern der schlecht abscheidenden Probanden wegen ihrer unzureichenden Bildung, messbar am unterdurchschnittlichen Schulerfolg, prognostizieren zu müssen, lässt einen Zweifel an der behaupteten Validität der Messkonstrukte aufkommen, der sich aus einer ganz anderen Quelle speist: Wie konnte eine komplette Kohorte angesichts von PISA-Aufgaben signifikant versagen und doch in einem anderen Milieu des Lehrens und Lernens eine hinreichende Bildung aufbauen, die ihr gestattet, gleichsam unsichtbar unter denen zu verschwinden, die bei der gleichen Untersuchung wesentlich besser abgeschnitten haben? Sie waren offensichtlich – evident halt – bildungsfähig, sie mussten in der PISA-Messung untergehen, sind aber nicht an den Anforderungen gescheitert, denen nicht sich gewachsen zeigen zu können, die PISA-Untersuchung ihnen prognostiziert hat. Vielmehr haben sie diese mit dem gleichen Erfolg wie die besser beurteilte Gruppe gemeistert.

Die Schweizer Ergebnisse treiben ein bislang unbeachtetes Problem hervor: Die psychometrisch bewirkte Bilanzierung von individuellen Bildungsverläufen muss diese gleichsam aufbrechen, indem sie an einem systemisch gesetzten Punkt, dem durchschnittlich-wahrscheinlichen Ende der allgemein verpflichtenden Schule, eingreift, um zu fragen, was bis dato an Entwicklung nach ihrem modellierten Bild, letztlich also psychometrisch nur vermeinter Kompetenzen, herausgekommen ist. Die Konstrukte, mit denen jene erfasst und gemessen werden sollen, sind dem nachgebildet, aus dem Wissen und Können der Schüler durch Unterricht entstanden sein müsste. Die messtech-

nisch organisierte Hypostasierung dieser Kompetenzen verdankt sich einer rückwärts weisenden Orientierung, nicht einer vorausreichenden. Man bricht zwar mittels der Fixierung eines Messzeitpunkts die schulisch initiierten oder begleiteten Entwicklungsverläufe auf und leuchtet gleichsam die Bruchstelle diagnostisch aus, doch dies geschieht anhand der vergangenen, vom je individuellen Bildungsprozess aufgenommenen, eingeschlossenen oder vielleicht auch bloß eingekapselten Inhalte fachlichen Lernens.

Aus der zitierten Schweizer Studie ergibt sich diese rückwärtsgewandte Orientierung als Schwäche. Wer sich berufspädagogisch mit der Entwicklung Jugendlicher zwischen 15 und 20 Jahren beschäftigt, dürfte über die berichteten Resultate nicht staunen. So offenkundig einerseits der bias scheint, mit dem eine blanke Extrapolation des Ergebnisses eines dem Scheitern nahen Verlaufs schulischen Lernens die Prognose des fortgesetzten Scheiterns schlägt, so dringend erscheint es, von diesem bias zu forschungslogischen und darüber hinaus methodologischen Konsequenzen zu kommen. Wenn man in plakativer Formulierung sagen kann, dass nach einem Wechsel von der allgemeinbildenden auf eine berufsbildende Schule in statistisch nennenswertem Umfang – der sich aus einem Abgleich der frühen Abbrecher mit den Erfolgsquoten ablesen lässt – aus schlechten Schülern gute Auszubildende werden, dann muss eine Messung zu jenem Zeitpunkt, an dem systemisch erzwungen und biographisch verbindlich die Entscheidung zwischen fortgesetzter höherer Allgemein- oder beginnender Berufsbildung fällt, von einer paradoxen Konstellation von Faktoren ausgehen, die die individuelle „Reife“ für den anschließenden Bildungsgang ausmachen. Fasst man diese „Reife“ in Kategorien des zurückliegenden Bildungsabschnitts, dann erscheint sie bei „Risikogruppen“ à la PISA unzureichend. Wohl scheint aber etwas vorhanden, das die Individuen in die Lage versetzt, ihre sicher diagnos-

tizierten Defizite nach einem Wechsel in ein anderes Lern- und Anforderungsmilieu auszugleichen, die zuvor vermisste „Reife“ also nachzuholen.

Jene als defizitär klassifizierte „Reife“ dürfte kaum voraussetzungslos zur ausreichenden einfach nur erblühen können. Wir vermuten vielmehr, dass die Faktoren jedes lebendigen Entwicklungsverlaufs, dessen Stand eine Messung sistiert und die dadurch dem Schein nach unterbrochen wird, als ein Ergebnis der Entwicklung nicht zureichend erfassen und darstellen kann. Genau das Element, das die Nachholmöglichkeit der aktuell unzureichend nur ausgedeteten „Reife“ als Potential von Bildung enthält, wird nicht sichtbar, weil es dem Potential der Entwicklung selbst und nicht einem irgend gearteten Ergebnis zuzurechnen wäre. Die Prognose muss daher negativ ausfallen, obwohl das übersehene Element oder der Faktor genau jener ist, der im Laufe einer andauernden Entwicklung jene Prognose im Resultat widerlegt. Dieser Faktor scheint mit den herkömmlichen Mitteln der Schulleistungsuntersuchungen nicht nur nicht erfasst, sondern systematisch verkannt zu werden. Wenn eine Entwicklung als individuelle Bewältigung konkret erfahrener Anforderungen in Gang gehalten wird, dann würde dem ein Untersuchungsansatz strukturell besser gerecht, der nicht auf die Replikation zurückliegender Anforderungen setzt, sondern sich auf voraus liegende richtet.

So groß der Unterschied zwischen einer entwicklungsbilanzierenden und einer entwicklungsprovozierenden Untersuchung ausfallen dürfte, aus Sicht der Bildungstheorie wäre er nur gering. Denn was dann als Messgegenstand zu definieren wäre, ist auch in den traditionell verfahrenen Untersuchungen implizit vorausgesetzt. Auf der Suche nach einem Ansatz, der nicht nur vergleichend die Performanz von Schulabsolventen erhebt, sondern dabei auf ein Vermögen zielt, auf Anforderungen auch unbekannter Art in rationaler oder gebildeter Weise zu reagieren, haben wir uns eine Implikation des konsekutiven Modells von Curricula der Sekundarstufe I zunutze gemacht. Es geht um die jedem fachsystematischen Curriculum zugrunde liegende Idee des Aufbaus. Der Plan zum Aufbau von fachlich systematisierbaren Fähigkeiten muss vor allem eins unterstellen: Eine insofern sinnvolle Ordnung fachlicher Inhalte, als es deren Aneignung durch die Ordnung selbst erleichtert und nicht etwa erschwert wird. Das unterrichtlich bereits bewirkte Wissen und Können versetzt die Schüler in die Lage, das nächste, darauf in systematischer Ordnung folgende ebenso einzusehen oder zu verstehen. Das, was dadurch im Einzelnen bewirkt wird, gilt bildungstheoretisch als genau jenes Moment der Veränderung einer Person, die ihre Fähigkeiten gegenüber einem vorherigen Zustand weiter entwickelt hat. Ob eine solche Entwicklung stattgefunden hat, untersucht die empirische fachdidaktische Forschung. Im Gegensatz zur empirischen Bildungsforschung fühlt sie sich als Fachwissenschaft durchaus zuständig für eine bessere Praxis des Unterrichts. Für sie ist die Idee der individuellen Entwicklung unverzichtbar.

Wenn die Konsekutivität fachsystematischer Curricula auf einer synchronen Struktur des fachinhaltlich geordneten Verlaufplans zum individuellen Wissensaufbau beruht, dann kann man die empirische fachdidaktische

Forschung als eine an der Unterrichtswirklichkeit diachron ansetzende Kontrolle des Lernfortschritts auffassen. Der diachrone Blick auf das faktisch erreichte Verständnis von Schülern legt die für die Funktion der Allgemeinbildung überaus wichtige Implikation in Gestalt jener Operationalität des fachlichen Denkens offen, die ein Curriculum kontrafaktisch unterstellen muss, um seinen systematischen Aufbau zu rechtfertigen – der ja zunächst nur voraussetzen kann, dass Schüler tatsächlich auch in der Logik des curricularen Aufbaus lernen. Insofern unterstellt jedes fachsystematisch aufgebaute Curriculum funktional einen kontinuierlichen individuellen Konzeptaufbau in der Form, dass das mit einem zuvor vermittelten Inhalt bewirkte Verständnis zugleich als jenes Vermögen entsteht, das sich auf den konsekutiv anschließenden nächsten Inhalt bereits erstreckt, ohne ihn damit schon vollständig zu erfassen. Piaget hat das Akkomodation genannt. Reife auf Basis einer fachlich strukturierten Bildung führt zur Vorstellung eines parallel zur inhaltsbezogenen Wissensvermehrung liegenden Konzeptaufbaus. Dies trägt die empirische fachdidaktische Forschung in den Fächern Mathematik, Physik, Chemie und Biologie zur Bildungsforschung bei. Stockt dieser Konzeptaufbau, bleibt es bei Fehlkonzepten, falschen Vorstellungen oder sogar bei bloßer Ignoranz. Dann entfallen die Voraussetzungen, die Piaget in der Unterscheidung zwischen Assimilation und Akkomodation erkannt hat: Wenn sich aus dem vorhandenen Wissen nicht jener Impuls ergibt, eine affine, aber nicht mit den bekannten Antworten identisch zu bedienende Frage zu akzeptieren und zu verstehen, dann „erstreckt“ sich das vorhandene Wissen eben nicht auf das Unbekannte. Der Faden des Verstehens reißt ab – was den Mechanismus von Assimilation und Akkomodation aber unberührt lässt. Schon seinem Begriff nach besteht ein Potential unabhängig von einer aktuell provozierten Entbindung. Es auch aus der Perspektive kommender Anforderungen zu erfassen, führt zu einer großen Schwierigkeit: Wie untersucht man die Bewältigung von Anforderungen, die sich noch gar nicht gestellt haben?

III

Operationalität in Abhängigkeit vom Konzeptaufbau

Es geht um die Fähigkeit eines sich bildenden Subjekts zu lernen. In einer bis heute modernen Formulierung hat die Wissenschaftlergruppe um Herwig Blankertz 1972 den auf Humboldt zurückgehenden Bildungsbegriff reformuliert (vgl. Teil I, PÄK Heft 42). Im Königsberger Schulplan heißt es: „Der Schüler ist reif, wenn er so viel bei andern gelernt hat, daß er nun für sich selbst zu lernen im Stande ist“ (Humboldt 1956, S. 71). Mit „reif“ stellt Humboldt eine Art Epiphänomen der Bildung als Verlauf dar. Höhere und vor allem wissenschaftliche Bildung mag sich in einer beeindruckenden Menge an Wissen niederschlagen, der Mechanismus, nach dem Bildung vorschreitet, ist jedoch keiner der andauernden Akkumulation. Über den „Besitz“ des Wissens hinaus entsteht im Prozess individueller Bildung eine Reife, die zwar nicht mehr oder besseres Wissen unmittelbar selbst bewirkt, aber

dem Lernen eine wachsende, sich anpassende Basis für ein insgesamt steigendes Niveau des Wissens schafft.

Unsere Verwendung der Kategorie Operationalität geht auf das *movens* von Bildung zurück. Unterricht, Belehrung, Instruktion setzen es voraus und sind nicht selbst der Grund der Bildung. In der Wahrheit eines Bildungsprozesses liefert zu lernen das Motiv nicht nur selbst, sondern ist dessen unnaive Voraussetzung. Der Begriff der Operationalität versucht das bildungstheoretisch zu fassen: Eine Person, die darüber verfügt, ist in der Lage – nach Humboldt „reif“ -, die sie interessierenden Dinge mit den vorhandenen Mitteln ihres bis dahin entwickelten Denkens weiter aufzuschließen bzw. das, was durch Erfahrung ihr nicht zugänglich ist, als in welcher Form auch immer bereitgestelltes Wissen sich zu eigen zu machen. Man könnte auch „Bildsamkeit“ sagen, um ein Personenmerkmal dort zu bestimmen, wo es im Bildungssystem auf die „Person“ ankommt. Wir sprechen von Operationalität des Denkens, das im Bildungsprozess selbst eine so bedeutende Rolle spielt, dass man sie unmöglich erst als Ergebnis desselben behandeln kann. Insofern wäre damit zwar ein Merkmal bestimmt, das in einer individuellen Entwicklung geprägt wird. Es ließe sich als Variable aufgreifen, deren Abhängigkeit von diversen Faktoren mit darauf getrimmten Methoden empirisch untersucht werden könnte. Gleichzeitig aber wäre Operationalität auch als das zu behandeln, was bei einer Untersuchung gerade im Sinne der *item-response-theory* anspringen muss, um überhaupt zu einem Ergebnis in Gestalt eines *response* auf ein *item* zu führen.

Wir haben ein Modell entwickelt, das die Kategorie der Operationalität auf einer Funktionsebene verankert. Wir sind der Überzeugung, dass man Operationalität in unserem Sinne abbilden kann, wenn man bei deren Funktion von Konzepten ausgeht, spezifisch von Fachkonzepten z.B. der Physik zur Erklärung von Kraft und Energie im Rahmen der Klassischen Mechanik. Obwohl die empirische fachdidaktische Forschung zu Fehlkonzepten in den Disziplinen Mathematik, Physik, Chemie und Biologie für den Fachunterricht höchst interessante Befunde liefern kann, stehen mit ihr noch keine unproblematischen, geschweige denn fraglosen Daten zur Untersuchung von Operationalität im Sinne einer bildungstheoretisch informierten empirischen Bildungsforschung zur Verfügung.

Wir setzen, anders als die empirische fachdidaktische Forschung, nicht beim Einzelbefund eines entweder „richtigen“ oder typisierbaren „falschen“ Konzepts, sondern bei der Entstehung einer übergreifenden, als Operationalität vorgestellten „Reife“ an. Für die Fächer der Allgemeinbildung gilt, dass die ihrer so denn teilhaftig gewordenen Schüler nach Abschluss der Sekundarstufe I mit ihrem allgemeinen Wissen in Muttersprache, Mathematik und Naturwissenschaft zugleich eine Reife erworben haben, die prinzipiell eine Fortsetzung des Lernens in dann anderen, nicht mehr mit der Allgemeinbildung sich deckenden Zielen ermöglicht. Um zu untersuchen, ob dies sich auch in der Realität der Bewältigung an der ersten Schwelle einlöst, reicht es nicht zu, in Bezug auf das verfügbare Wissen nach richtig oder falsch zu un-

terscheiden. Maßgeblich für eine den Konzeptaufbau erfassende Untersuchung wird eine andere Methode der Unterscheidung sein. Um die binäre Konstruktion zwischen „richtig“ und „falsch“ in Bezug auf Fachkonzepte einerseits und Fehlkonzepte andererseits untersuchungslogisch zu überwinden, haben wir etwas entwickelt, das methodisch auf die bildungstheoretische Differenzierbarkeit zwischen konzepttreu, konzeptgetragen und konzeptfrei setzt.

Abbildung 1: Transformation der nach richtig-falsch binär strukturierten Ebene des Wissens zur Funktionsebene von Operationalität nach den drei bildungstheoretisch möglichen Typen des Konzeptaufbaus.

↓ Ebene des objektiven Wissens ↓		
richtige Lösung ↓	↓ falsche Lösung ↓	
↓ richtig	alltagstheoretisch ↓	↓ abwegig
versuchte Operation		keine Operation
↑ konzepttreu	konzeptgetragen ↑	↑ konzeptfrei
↑ Ebene der Bildungstheorie ↑		

Wir unterstellen in Bezug auf Operationalität drei mögliche Reaktionen, die sich aus den gewählten Lösungen interpretieren lassen: Die Verfügung über das fachlich richtige Konzept bzw. im anderen Fall die Suche nach einer der Problemstellung entsprechenden Antwort bzw. die wie auch immer begründete Weigerung, das Problem anzuerkennen und von sich selbst zu erwarten, dazu einen vernünftigen Zugang zu finden. Wer sich auf die multiple choice-Methode festlegt – z.B. aus untersuchungsökonomischen Gründen –, kann zum Untersuchungsgegenstand Operationalität nicht einfach Fragen formulieren, zu denen dann Antwortmöglichkeiten angeboten werden, die auf mehr oder weniger einfühlsame Weise Probandenreaktionen antizipieren. Nach diesem Schema würde die richtige den falschen Alternativen binär gegenüberstehen. Wonach gefragt wird, könnte im Modus des Wissens richtig entschieden werden, während diejenigen, die diesen Modus nicht beherrschen, durch gedankliche Operationen, also im Modus des Nachdenkens, keinen Zugang zur richtigen Antwort aufdecken könnten.

Das mag bei binärer Konstruktion unproblematisch bleiben und dazu aus den bekannten psychometrischen Gepflogenheiten – Eindimensionalität der Testkonstrukte auch zwingend erscheinen, es genügt aber nicht den Anforderungen einer Untersuchung, die nicht auf die Qualität kognitiver Leistungen zielt, sondern auf die operationelle Qualität von Antworten, die kaum im schulischen Wissensbestand schon vorliegen, sondern die als Lösungen erst zu erzeugen wären. Der Unterschied liegt im Prinzip der Unterstellung eines Wie oder Wieviel und eines Ob. Dass man Intelligenz genau in den Graden

feststellen kann, in denen sie als vermutete Eigenschaft der Untersuchungspopulation verteilt ist, setzt ja nur eine vorhandene Skala voraus, zu der man die Übereinstimmung oder Abweichung der jeweiligen Probandengruppen ermittelt. Wir versuchen hingegen Operationalität als Ausdruck der Bildung bzw. Entwicklung einer Person auf direktem Wege nachzuweisen, indem wir diese Operationalität herausfordern.

Abbildung 2: Paarung der rating-Kategorien mit entwicklungstheoretischen Kategorien.

richtig	alltagstheoretisch	abwegig
generativ adäquat akkomodiert: Fachvorstellung etabliert	generisch falsch assimiliert: Fehlvorstellung persistiert	ignorant agnostisch Vorstellung negiert

Wenn die richtige Lösung höchstwahrscheinlich unbekannt ist, muss sie im Vorgang der item-Bearbeitung entstehen. Das Ergebnis kann richtig, alltagstheoretisch oder abwegig sein.

- Ist es richtig, stand dem Probanden das richtige Fachkonzept zur Verfügung, er konnte es sicher auch in einem Kontext einsetzen, der unter normalen Umständen im Leben von fünfzehn- bis sechzehnjährigen Schulabsolventen unbekannt bleibt. Die richtige Antwort geht auf einen korrekten Strukturierungsvorgang zurück, die Lösung hat vermöge ihrer operativen Qualität generativen Charakter.
- Im Falle einer alltagstheoretischen Antwort gehen wir von einer Anwendung konzeptionell gestützten Wissens aus, das aber falsch assimiliert wurde. Die „falsche Kopie“ eines andernorts richtigen Konzepts bezeichnen wir im Hinblick auf die Strukturierungsleistung als generisch.
- Vom Fall einer im Sinne von EOL (Erhebung operationeller Leitungsfähigkeit) falschen Lösung gehen wir dann aus, wenn die gewählte Antwort von keinem Konzept getragen wird, sondern für Ignoranz oder bestenfalls freie Assoziation spricht, die implizit die Relevanz oder Zulässigkeit der Fragestellung bestreitet. Das bezeichnen wir dann als abwegig.

IV Entwicklung von items für EOL

Gegenüber herkömmlichen psychometrischen Verfahren, die das, was richtig ist und in der Summe für die beste Probandenleistung steht, auf ein Wissen beziehen müssen, dessen Umfang und Qualität auf dem potentiell riesigen Abstand zwischen wissenschaftlich verbürgter Erkenntnis und der schulisch davon reduzierten Kopie beruht, gehen wir in unserer item-Konstruktion von der finiten Rationalität der richtigen Lösung aus. D.h. eine Mathematikaufga-

be ist nicht in das Kontinuum von reduzierbarer oder ebenso beliebig steigerbarer Komplexität mathematischer Fragestellungen eingebettet, sondern ergibt sich zwanglos aus der eigentlichen Frage, deren Lösung z.B. eben Prozentrechnung erfordert. Am anderen Ende liegen dann item-Konstruktionen, die vor dem Hintergrund der jeweiligen Frage- und Problemstellung möglichst keinerlei Sinn stiften und insofern darüber Auskunft geben, ob die Probanden bereit sind, sich auf das Problem überhaupt einzulassen.

Damit sind wir auf ein einheitliches Konstruktionschema festgelegt. Jedes item benötigt einen plot, der eine durchaus umfangreiche Problemexposition verlangen kann. Das Thema des plots kann wissenschaftlich, technisch, ökonomisch oder soziologisch-pädagogisch gewählt werden. Die Kontextualisierung des Problems bis hin zu der den plot resümierenden eindeutigen Frage muss von allem wahrscheinlich Bekannten wegführen. Anschließend bieten wir bis zu sechs, in der Regel vier Antwortalternativen an, die ihrerseits argumentativen Charakter haben oder mindestens ein in der Alltagssprache formuliertes Argument enthalten.

Wir haben 24 Fragen nach dem Muster gestellt, wie es im Weiteren unten an einem Beispiel dargestellt wird. Es lag bei etwa der Hälfte der Fragen ein Akzent auf naturwissenschaftlich-technischen Gegenständen. Dabei wurde in den Fächern Mathematik oder Physik nie das Unterstufenniveau überstritten – es ging um Prozentrechnung oder die basalen Prinzipien der klassischen Mechanik. Die andere Hälfte sprach politische, ökonomische und sozialpädagogische Themen an. In der groben Tendenz folgte die item-Konstruktion in Bezug auf Technik und Naturwissenschaften eher einer schulfachlichen Orientierung, hingegen bei den politisch-moralischen, ökonomischen oder sozialpädagogischen Fragen den Topoi eines vor allem öffentlichen Diskurses über Themen von als durchschnittlich anzunehmender Wichtigkeit. Das heißt, wenn es um Technik oder Naturwissenschaften ging, haben wir auf Fachkonzepte und die ihnen korrelierenden Fehlkonzepte zurückgegriffen – eine Beziehung auf Schul- oder Unterrichtsfachwissen ist kaum zu vermeiden, wenn es um solche plots geht. In den anderen Fällen trat an die Stelle des aus dem Fachunterricht als „richtig“ entscheidbaren plots ein Konflikt zwischen der erfahrungsgemäß politisch wohlgeleiteten, höchstwahrscheinlich sozial akzeptierten Meinung und der konstruierten Perspektive eines dazu im Widerspruch stehenden Sachverhalts.

Das Fehlkonzept, um das herum das folgende, hier beispielhaft illustrierte item konstruiert wurde, darf gängig genannt werden. Der Grund dafür ist zunächst einfach einzusehen: Damit jemand, der auf dieses Konzept baut, damit überhaupt scheitert, müsste er das Gravitationsfeld der Erde verlassen bzw. sich erheblich von ihrer Oberfläche entfernen. Das Gewicht eines Körpers zu seinen unverlierbaren Eigenschaften zu zählen, wird auf der Erde weder durch Anschauung noch Erfahrung widerlegt. Das Fehlkonzept bleibt dadurch – bis auf Belehrungen aus dem Physikunterricht – ungestört.

Wir haben recht plakativ gefragt:

Aufgabe 1: Wie groß ist beim „großen Knall“ der Knall wirklich?

Zu den – in diesem Fall vier – items führt eine künstliche Gesprächssituation, die im Alltag unwahrscheinlich ist. Das setting soll befremden, aber anregen, sich für die als Antworten auf die letztlich komische Fragestellung zu interessieren. Wir haben diesen Reiz so formuliert:

Der Mond hat eine wesentlich geringere Masse als die Erde. Deshalb wiegen die Sachen auf dem Mond weniger als auf der Erde. Jemand beruft sich darauf und behauptet: „Eine Atombombe aus einem Kilogramm Plutonium hat auf dem Mond eine kleinere Sprengkraft als auf der Erde, weil das Plutonium auf dem Mond nicht so viel wiegt wie hier.“

Was würden Sie darauf sagen?

Die folgenden Antwortmöglichkeiten stellen die items dar, von denen wir erwarten, dass die Probanden, angeregt durch die als Frage-Antwort-Situation simulierte Gesprächssituation, sich in die Lage versetzen lassen, nach einer ihnen angemessen scheinenden Antwort zu suchen. Wer noch nie über dergleichen nachgedacht hat, wird angeregt, im Sinne unserer Untersuchungskategorie operationell auf das Problem eingehen. Diese vier Möglichkeiten haben wir angeboten:

Das ist logisch, das stimmt.

Die erste Antwortmöglichkeit bietet eine prompte, in Bezug auf die Leitfrage sozusagen glatte Antwort an, die man, ohne nachzudenken, nehmen könnte. Wir haben sie als abwegig eingestuft. Man müsste das viel spektakulärere Phänomen der Schwerelosigkeit ignorieren, um die Überlegung hinzunehmen, dass ein Gewichtsverlust von einem proportionalen Masseschwund begleitet wird. Dass sich die Körper von schwerelos dahinschwebenden Astronauten in der ISS mangels Masse vollständig auflösen, ist nur schwer anzunehmen. Zumal die ISS oder andere orbitale Flugkörper unter Bedingungen der Schwerelosigkeit fortexistieren, ja sogar heil auf die Erde zurückzukehren vermögen.

Die drei anderen Alternativen hingegen problematisieren genau die im ersten item postulierte Logik.

Das ist falsch, die Masse an Plutonium hat hier mit dem Gewicht nichts zu tun.

Dies ist die richtige Antwort. Sie ist völlig unwissenschaftlich, auch ein wenig ungeschickt formuliert, um zu verhindern, dass gewiefte Schüler mit dem behavioristisch trainierten Feinsinn für terminologische Unterschiede zwischen der Alltags- und Wissenschaftssprache nicht gänzlich ohne nachzudenken auf die richtige Antwort stoßen.

Der Mond hat keine Lufthülle, da ist alles leichter und deshalb ist die Sprengkraft viel höher als auf der Erde.

Im Modell der Leitfrage stellt diese dritte Alternative einen Scheinwiderspruch zur ersten dar. Wer in diesen Zusammenhängen eher ungeübt sich denkend be-

wegt – aber dies immerhin versucht – müsste zwischen dem ersten und dritten item einen aussagelogischen Widerspruch erkennen. Das Angebot besteht darin, sich spätestens an dieser Textstelle in die Ernsthaftigkeit des zunächst ja komisch klingenden Problems verwickeln zu lassen, das die Leitfrage exponiert.

- Die Sprengkraft ist wie auf der Erde, weil die Explosion zwar schwächer ist, aber die Wirkung ist viel größer, weil alles leichter ist.

Dieses item steht für die Kategorie alltagstheoretisch. Das Fehlkonzept liegt in der alltäglich wirksamen Vorstellung einer prästabilierten Balance, eines zur Harmonie führenden Ausgleichs. Das kann sich hier auf eine Scheinlogik stützen, derzufolge es konsequent wäre, nachdem die Aussage des imaginären Fragestellers – fälschlicherweise – akzeptiert wurde, einzuwenden, dass das geringere Gewicht dann für alles gelten müsste, was sich sonst noch auf dem Mond befindet. Nach dieser Scheinlogik erscheint es naheliegend anzunehmen, dass der angeblich gewichtsbedingte Wirkunterschied einer Explosion auf dem Mond zu einer auf der Erde von der damit generell geltenden Bedingung gleichzeitig neutralisiert wird.

V Ergebnisse

Wir haben u.a. im Jahr 2008 an vier Bremer Schulzentren der Sekundarstufe II 24 Fragen mit einem fünfzehnteiligen Bogen eingesetzt. Als Zeitpunkt wurde der Schuljahresbeginn im August plus weitere sechs Wochen gewählt. Die Jahrgangsstufe 11 an beruflichen Schulen nimmt typischerweise alle Absolventen der Sekundarstufe II-Schulen auf, die diese verlassen, die aber noch schulpflichtig sind. Die Ausnahme bilden jene Gymnasienschüler, die in ihrer angestammten Schule verbleiben. Da in den Stadtstaaten an berufsbildenden Schulen auch gymnasiale Oberstufen eingerichtet wurden, entsteht eine einigermaßen erstaunliche Situation, wenn man sich die Bandbreite jener mit Qualifikation oder Kompetenz bezeichneten Fähigkeiten der an die berufsbildende Schule wechselnden Abgangspopulation vergegenwärtigt: Man trifft dort in der Jahrgangsstufe 11 sowohl Abiturienten – die z.B. eine Banklehre beginnen – oder Gymnasiasten – die auf die gymnasiale Oberstufe wechseln – oder abschlusslose Abbrecher von Haupt- oder sogar Realschulen, die in eigens geschaffene Bildungsgänge, in sogenannte Ersatzmaßnahmen einmünden. Die Differenzierung nach Eingangsvoraussetzungen findet zwischen Abiturienten, Realschul- oder Hauptschülern und Schulabbrechern statt. In dieser Spreizung stellt sich die Herkunft der Schüler aus dem dreigliedrigen System dar.

Die berufsbildenden Schulen sind zugleich ein aufnehmendes System. Man kann die Eingangspopulation auch nach den aufnehmenden Bildungsgängen, also nach den Bildungswegen differenzieren, in die die Sekundarstufe I-Absolventen einmünden. Die Typologie dieser Bildungsgänge umfasst gleichfalls vier Varianten: Die gymnasiale Oberstufe, die die leistungsfähigen Realschüler aufnimmt, die Teilzeitberufsschule, die alle Sekundarstufe I-

oder II-Absolventen aufnehmen muss, die eine Lehrstelle gefunden haben, die Schüler von Vollzeitbildungsgängen, die z.B. als Realschüler die Fachoberschule besuchen oder die Erzieher, die eine Vollzeitausbildung absolvieren, und schließlich die Ersatzmaßnahmen wie Berufsvorbereitungsjahr, Berufsgrundbildungsjahr, Werkschule etc.

Wir haben beim Test 2008 an den vier beteiligten beruflichen Schulzentren 2251 aufgenommene Schülerinnen und Schüler erfasst, die zuvor allgemeinbildende Schulen besuchten. Die vier beruflichen Schulzentren unterscheiden sich in ihrer Ausrichtung nach gewerblich-technisch, sozialberuflich und kaufmännisch. Weil die Schulen sich in den Teilzeitberufen ihrer Zuständigkeit nach nochmals unterscheiden, geben nicht nur die Berufsfelder den Ausschlag bei der Zuteilung der Schülerinnen und Schüler, sondern auch noch die Unterscheidung nach industriell und handwerklich.

Um vorweg an eine Eigentümlichkeit zu erinnern, die aufwendigen empirischen Untersuchungen, genauer der Präsentation ihre Ergebnisse eignet: Es kommt meistens heraus, was man schon erwartet hat oder schon wusste. Angesichts des methodischen Aufwands liefert diese Erfahrung ein fast karikaturistisches Bild von evidenzbasierter Forschung. Mit dem Befund etwa, dass am Ende schlechter Unterricht für schlechte Schülerleistungen verantwortlich zu machen ist, dürfte die Immanenz solcher Argumentation kaum zu sprengen geeignet sein.

Uns ist es auf den ersten Blick nicht viel besser ergangen. Wir haben eine Parallelität zu TIMSS und PISA bei den zertifizierten Schulleistungen festgestellt. Die Abiturienten (bei EOL 2008 durchweg Auszubildende in dualen Berufen) haben am besten abgeschnitten, die Hauptschulabbrecher am schlechtesten.

Die Feststellung, die Abiturienten schnitten besser ab als Schüler mit niedrigeren Abschlüssen, basiert auf einer schlichten Auszählungsmethode nach richtigen, alltagstheoretischen und abwegigen Antworten. Die positive Hypothese einer messbaren Qualität dessen, was wir unter dem Personenmerkmal Operationalität verstehen, beruht im Ganzen auf der quantitativen Auswertung des eingesetzten set von itemisierten Problemen und den dazu gebildeten Antwort-ratings. Wir unterstellen auf einigermaßen simple Weise folgendes: Wer unter 24 disparaten Fragen unter Berücksichtigung der dargebotenen Behauptungen und Argumente viele davon richtig beantwortet, hat in einem einfach bestimmbar Umfang auf die Probleme konzepttreu operiert. Für die Antworten von alltagstheoretischer Qualität gilt die Kategorie konzeptgetragen und konzeptfrei für die abwegigen. Dies im Sinne der Hypothese auszuwerten, bedarf keiner sonderlich hohen Datenaggregation. Wir haben stattdessen eindeutige Fallzahlen gebildet, indem die entsprechenden Antworten je Proband ausgezählt und ihm zugeordnet wurden. Der demnach beste Fall war ein Abiturient mit 20 von 24 möglichen richtigen und der schlechteste eine Hauptschülerin im Friseurberuf mit 17 abwegigen Antworten.

Allgemein bedeutsam ist der Sachverhalt ausgeprägter Heterogenität für die eindeutige Korrelation von Schulabschlüssen der Sekundarstufe I und

Testergebnis sowie die schwächere, aber immer noch eindeutige Korrelation zum aufnehmenden Bildungsgang. Dass diese Korrelationen in statistischen Maßstäben nicht auch hochsignifikant sind, hat mit der – konventionell methodologisch betrachteten – Mehrdimensionalität unseres set of items zu tun. Die Implikationen unserer Vorgehensweise lassen sich unter bestimmten Annahmen auf drei Ursachen zurückführen.

(a) Eine Aufgabe, bei der es darauf ankam, aus den absoluten und prozentualen Angaben zu als arbeitslos gemeldeten Personen die Zahl der Beschäftigten zu berechnen, waren nicht einmal die Hälfte der Abiturienten in der Lage zu lösen. In dem Fall bestand die Lösung darin, jene Antwort anzukreuzen, die behauptete, das sei mit einfacher Prozentrechnung zu machen. Als zuständig für die Erklärung dieses eigentlich niederschmetternden Ergebnisses erklärte sich ein Politiklehrer, der darauf hinwies, in seinem Unterricht würden die Schüler erfahren, wie kompliziert die Statistik über Arbeitslosigkeit sei. Daraus ergäben sich statistische Feinheiten, mit denen man die Arbeitslosenzahlen „drücken“ oder, je nach politischem Interesse, auch „erhöhen“ könne. Diese Art von Aufklärung vermochte augenscheinlich das Basiskonzept der Prozentrechnung außer Kraft zu setzen, demzufolge Prozentualwerte ein Ein- oder Mehrfaches des jeweiligen Teilers von Hundert sind. Worauf wir hier gestoßen sind, geht über die Einzelaufgabe hinaus. Während über 80% bei einer anderen Mathematikaufgabe diese Aufgabe lösten, obwohl hier Flächenberechnungen als abschreckendes Beispiel dienten, kamen der Prozentaufgabe gerade Abiturienten nicht annähernd so weit. Der Hinweis, den wir dazu geben können, ist schlicht der, dass im Falle der anderen Mathematikaufgabe kein anderer allgemeinbildender Inhalt interferierte. Es diffundiert bei dieser Aufgabe nichts hinein. Positiv gesprochen: Den Probanden schien nichts den Sinn der Frage zu trüben, hingegen bei jener nach der Zahl der Arbeitsplätze in Form von real Beschäftigten kamen Skrupel hinzu, den Zahlen also solchen zu trauen.

(b) Wir haben danach gefragt, wie ein Auto, das mit ABS ausgerüstet ist, überhaupt zum Stehen gebracht werden kann, wenn das ABS ein Blockieren der Räder verhindert, indem es die Bremsung unterbricht. Die richtige Antwort, dass man bei einer bestimmten Geschwindigkeit die ABS-Steuerung einfach abschaltet, weil man sie unterhalb einer Geschwindigkeit von 6 km/h ohnehin nicht mehr benötigt, ist kaum angewählt worden. ABS wird als nützliche, sicherheitsrelevante Technologie attribuiert. Damit ist sie unproblematisch – und scheint sich auch deshalb von selbst zu erledigen. Das Faszinosum Technik, die List der Vernunft oder einfach nur eine sinnreiche Konstruktion werden als solche gar nicht mehr erkannt. Wer bremst und dessen Auto auch ohne blockierende Bremsen zum Stehen kommt, hat kein Problem mehr. Bei dieser Generation dürfte die Verfügbarkeit von Strom aus der Steckdose schon wegen des Vorhandenseins von Steckdosen garantiert sein.

(c) Wir haben eine Frage zum Widerspruch zwischen den Extrema Vererbungslehre und Behaviorismus gestellt. Die Techniker beantworteten die

Frage genau so wie die Schüler aus sozialpädagogisch-hauswirtschaftlichen Bildungsgängen – in der Regel falsch, weil einseitig. Letzteren war der Sinn ihres Berufs, der ja an die Wirksamkeit des erzieherischen Einflusses doch gebunden bleibt, bei weiten nicht in dem Maße bewusst, wie es ihre Einstellung zum Beruf erwarten ließe. Zum Tragen kamen bei beiden der entweder sozialpädagogischen oder technischen Schulform zugeordneten Gruppen mit fast 50% die als alltagstheoretisch eingestuften Antworten. Nach den in ihrer beruflichen Ausrichtung getrennten Schulen waren keine Unterschiede feststellbar. Nur bei den reinen Erzieherklassen in Entgegensetzung zu allen anderen Probanden beider Schulformen waren es dann 70% gegenüber stabil 50%, die keine richtige, aber immerhin eine alltagstheoretische wählten. Das geschah aber zur Hälfte auf Kosten der richtigen Antworten (minus 10%). Entsprechend fiel das Ergebnis bei den als abwegig eingestuften Antworten ebenfalls besser, um ungefähr 10% niedriger aus (14% vs. 25%). Die höchste Zustimmung erhielt bei den künftigen dezidierten Erziehern die Aussage, nach der im ersten Lebensjahr innerfamiliär jene Entscheidungen fallen, an denen danach durch erzieherischen Einfluss nichts mehr zu ändern sein dürfte.

Fasst man die drei Beobachtungen als Illustration des hier hypostasierten Trends zusammen, dann geht es um einfache Vorurteile, die aber das Misslingen eines Bildungsprozesses indizieren. Wir unterstellen in unseren Aufgaben, dass eine richtige Antwort einen erschließbaren Sinn hat. Das wäre nicht möglich, ohne gleichzeitig die Genese eines Sinns in der Bildung anderer zu unterstellen. Eine Lösung gilt als sinnvoll, die ein als solches anerkanntes Problem beseitigt. Wer kein Problem mehr anerkennt, für den mag alles selbstverständlich erscheinen, aber gewiss noch nicht sinnlos. Damit sprechen wir nicht nur den Sinn an, sondern das Interesse, einen solchen zu erkennen und ihm zu folgen, der der Bildung anderer sich verdankt und diese in der rekonstruierten Lösung gerinnen lässt. Was wir dazu als den das eingangs exponierte Faktum Heterogenität erklärenden Faktor gefunden haben, ist die mehr oder weniger willkürlich aufscheinende Bereitschaft, hinter die Dinge zu steigen. Im ersten Jahr des Besuchs einer berufsbildenden Schule bleibt es Zufall, wenn jemand in einer Technikerklasse sich für etwas interessiert, das im Prinzip jene Technik ausdrückt und Teil der aus ihr zu beziehenden Bildung darstellt, auf die zu erlernen und zu beherrschen er sich beruflich vorbereitet. Bei den Schülern einer sozialpädagogisch-hauswirtschaftlichen Schule war das nicht besser.

Zwar kann man mit Blick auf den voraus liegenden Bildungsverlauf erwarten, dass sich irgendwann, spätestens im Kollegenkreis, ein Speditionskaufmann oder eine Fachkraft für Lagerlogistik doch dafür interessieren wird, mit welcher einfachen Formel man quantitative Verhältnisse prozentual vergleichbar machen kann, dass sich ein Kfz-Mechaniker die Frage stellen wird, ob ein Auto ohne blockierende Räder wirklich zum Stillstand kommen kann und ein Erzieher sich doch noch überlegen dürfte, was er mit seinen an elaborierten Theorien geschulten Interventionen im Verhältnis zum professionellen Anspruch seines Berufs auszurichten vermag.

Aber, da sind die einen, die von Anfang an verstehen wollen, was der weitere Bildungsweg oder schon eine Berufsausbildung ihnen abfordert, und die anderen, die das nicht erfahren haben und dilatorisch behandeln, weil sie solche Erfahrung aufschieben wollen. Was dergleichen für das System der allgemeinen Bildung bedeuten könnte, wäre aus dem Abstand zu ersehen, den der Anspruch auf Bildung durch Unterricht und dessen Ergebnis erzielt. Er ist nicht nur groß, was die Leistungsunterschiede zwischen Abiturienten und Hauptschulabbrechern belegen, er erfüllt sich auch in dem der Selektion dienenden Zweck nicht. Gerade damit wäre eine Heterogenität von Leistungen gar nicht vereinbar.

Die hier vorgeschlagene Deutung ist es hingegen schon, sie erschließt eine Diagnostik, die nicht einfach in der Feststellung schlechter Leistungen und ihrer Unterscheidbarkeit von bessern und besten aufgeht, sondern aus dem Befund gerade der Heterogenität auf die Spur der Ursachen führen will. Desinteresse auf der einen, eine perennierend präoperationale Einstellung zum Weltrest sozusagen, der nach Abzug der eigenen Person zum Beispiel eines Fünfzehnjährigen als Zentrum derselben von ihr übrig bleibt, auf der anderen Seite – beides mündet in die Diagnose einer Erfahrungsunfähigkeit, wie sie prominent Martin Wagenschein als Folge andauernder schlechter Unterweisung durch normalen Schulunterricht demselben gestellt hat.⁴ Unausgesetzte

4 „Die bildenden Begegnungen mit der geistigen Welt sind Anlässe, bei denen man lernt, diese Tugenden selber zu wollen, als Mittel, um dahin zu kommen, wohin die Sache, die uns ergriffen hat und die wir ergreifen, uns zieht und erzieht.“ Wagenschein 1956, S. 61.

Belehrung auch gegen die Fragen, die Schüler überhaupt stellen müssen, damit sie die Antworten verstehen können, machen aus der Welt eine fraglose Sammlung von Fakten, von denen die meisten dazu nicht einmal primär erfahren werden. Dass das in ungeheuerlichem Umfang schiefe geht, weil solches „Schulversagen“ individuell in habitualisierte Bildungsfeindlichkeit mündet, bricht dann irrational auf und durch die Fassade einer garantierten Unterrichtsversorgung der Bevölkerung wieder durch. Die Resultate der aus den Ressourcen dieses Systems finanzierten empirischen Bildungsforschung liefern einen im entscheidenden Befund wortlos bleibenden Abklatsch. Das ruft dann sachlich ebenso zuständige, wie in der Sache naiv argumentierende Opponenten auf den bildungspolitischen Plan.

Ein schon bereits angesprochener Befund hat nach drei Jahren Entwicklung und mehr oder weniger sporadischem Einsatz der originären 24 items uns, als er sich in den ersten Ergebnissen 2008 abzeichnete, zuerst sehr beunruhigt, dann aber zu Hypothesen geführt, die eine Weiterentwicklung von EOL ebenso wünschenswert wie aussichtsreich erscheinen lassen. Der Befund verbirgt sich sozusagen hinter den eindeutigen Ergebnissen wie der Korrelation von Schulabschluss mit dem Testerfolg, der Korrelation von Bildungsgangemündungen zu Abschlussniveaus und auch wie der Zuordnung von Leistungsgruppen zu Berufsfeldern. Auch in kleinen, unbedeutenden Details wie der Anwahl bestimmter Schulen aus dem ihnen vorausseilenden Ruf heraus haben wir in bestätigender Weise herausgefunden, was Eingeweihte wussten und was niemanden an den Schulen überrascht hat – eine erstaunte Reaktion löste höchstens die Erfahrung aus, dass man solche Kleinigkeiten mit einem so grob erscheinenden Instrument in so kurzer Zeit – die klassenweise durchgeführten Tests dauerten höchstens 90 Minuten – herausfinden kann.

Sobald wir Einzelklassen auswerteten, schien, wenn man von einem intuitiven Begriff von Wahrscheinlichkeit ausging, auf den Resultaten ein Makel zu prangen. Die Befunde zu den Einzelklassen zeigten eine unplausible Heterogenität. Ausgehend von der Fiktion, dass sich die Systematik unseres beruflichen Bildungssystems zusammen mit der Reinheit sozialstatistischer Daten wie Alter, Geschlecht und vorherigem Schulbesuch zu einer mindestens latent sichtbaren Homogenität der Leistungen jener in Klassen zusammengefassten Schüler hochschaukelt, erschien uns das, was wir messen, in einer Unordnung ausgeprägt, die sich kaum mit den Untersuchungskategorien vertragen sollte.

Andererseits reagiert der Ansatz von EOL mit einer impliziten Kritik auf die Unterstellung herkömmlicher Schulleistungsuntersuchungen, ihre items taugten in retrospektiver Beziehung auf die schulisch vermittelten Lernerfahrungen ihrer Probanden zu Aussagen über den individuell erreichten Bildungsstand, deren Geltung nicht allein auf der trivialen Gleichsetzung von fachlich-curricular vorgegeben Inhalten und Bildung beruht. Das Kalkül, die schulische Bildung aller sei in den gleichen Inhaltskategorien hypostasierbar, in denen diese als unterrichtliche Anforderung den Schülern gegenübertritt,

erzeugt zwar elegant den Schein nur geringer Heterogenität der nämlichen Leistungen, hängt aber selbst als einer ihrer Messeffekte an solchen Untersuchungen: Nur da, wo tatsächlich eine Entsprechung individueller Leistungen in kohortenhafter Organisation des Unterrichts auf jenes Messmodell stößt, das von einer institutionellen Wirksamkeit inhaltsidentischer und zugleich kohortenhafter Unterweisung ausgeht, könnte sich die Erwartung an Homogenität erfüllen.

Aber genau das steht in Frage. Homogenität ergibt sich als messkonstruktabhängiges Artefakt, Heterogenität besteht hingegen als Ausdruck eines individuellen Bildungsstandes. Gleichwohl bleibt unbefriedigend, dass unsere Untersuchungen jenseits der Hypothesenebene noch keine Erklärung der Heterogenität abwerfen.

Immerhin haben sich Wege eröffnet, sie als Symptom zu steuern. Wir haben zahlreiche Hinweise darauf gefunden, dass sich der sonst Homogenität bewirkende thematische Fokus verschiedener plots nicht nur in Bezug auf die fachlich richtigen items benutzen lässt. In EOL funktioniert das auch mit alltagstheoretischen und sogar abwegigen items. Ein plot exponierte das Problem einer Mutter, die ihr einjähriges Kind in einem Entwicklungsland über eine Strecke von 400km ohne Sicherheitseinrichtung transportieren muss. Diese Frage hat regelmäßig die höchste Bearbeitungshäufigkeit provoziert (96%), aber zugleich die wenigstens richtigen Antworten (10%). Die Konstruktion zielte auf eine physikalisch überlegte Lösung. Insofern hätten wir erwartet, dass all die, die andere technisch-physikalische plots im Sinne eines fachlichen Konzepts lösen, wenigstens das Problem adäquat verstehen, auch wenn sie es mit dem falschen Konzept beantworten. Genau unter dieser Gruppe trug die Frage nach dem sicheren Baby-Transport zur Heterogenität bei. Als wesentlich stärker erwies sich die ethnozentrisch motivierte Weigerung, das Problem, das sich außerhalb Europas überall alltäglich und massenhaft stellt, überhaupt anzuerkennen. Obwohl auch in Deutschland erst seit 1976 eine Anschnallpflicht für Fahrer und Beifahrer besteht, kann sich eine um 1992 geborene Generation schon überhaupt nicht mehr vorstellen, dass es noch nicht auf der ganzen Welt verboten ist, Kinder ohne Sicherungseinrichtung auf dem Rücksitz zu transportieren. Die eigentliche Schwierigkeit bei dieser Frage besteht darin zu erkennen, dass es sich hierbei um ein Problem handelt, das nur mit physikalischem Sachverstand zu entschärfen ist, weil es keine technisch bewerkstelligte Transportsicherheit gibt. Die Nichtverfügbarkeit letzterer als konstitutives Element des plots blockierte die Beschäftigung mit dem Problem.

Durch die mittlerweile große Menge an Testergebnissen kann deren Reproduzierbarkeit unter kontrollierten Bedingungen als gesichert gelten. Daher wären wir in der Lage, jenes Problem der Heterogenität zu umgehen.

Das nur aus dem Grunde einer Absenkung des Grades an Heterogenität der statistischen Homogenität wegen zu tun, halten wir hingegen nicht für sinnvoll. Im Gegenteil, die geschilderten Zusammenhänge, die die Heterogenität erklären können, wären im Sinne von EOL eher so aufzubrechen, dass

wir plots entwickeln, die gerade aus der Konfrontation zwischen obskuren, als Allerweltsweisheiten getarnten Überzeugungen und intellektuell angemessenem Verständnis von bedenkenswerten Problemen dem Messgegenstand Operationalität gerecht werden.

Zum wiederholten Einsatz unseres sets von 24 Fragen gab es im Interesse der explorativen item-Entwicklung keine bessere Alternative. Weil wir keine Skala einsetzen, anhand derer sich die Qualität einzelner Probandengruppen bis hinab zu einzelnen Probanden statistisch feststellen lässt, brauchen wir große Zahlen, um Abweichungen von erwarteten Reaktionen zu identifizieren, zu typisieren und zu interpretieren. Dadurch haben wir die Reproduzierbarkeit unserer Untersuchungsergebnisse nachgewiesen. Ein Defizit ist jedoch geblieben. Infolge mangelnder Mittelausstattung ist uns bisher die wirklich harte Probe auf die leitenden Hypothesen verwehrt. Wir haben letztlich immer im gleichen Milieu getestet. Bremen und Berlin rangieren bei TIMSS und PISA auf den letzten Rängen. Innerhalb dieser Milieus haben wir repräsentative Ergebnisse in Bremen, für die OBF auch in Berlin, weil wir dort eine Totalerhebung durchgeführt haben. Die geschilderten feinen Differenzierungen, die sich dabei noch herausgestellt haben, benötigen eine weitere Belastung, zu der es bislang noch nicht gekommen ist. Unseren bisherigen Ergebnissen wären solche etwa aus Bayern, Baden-Württemberg und besonders der Schweiz gegenüberzustellen. Wenn die für Bremen etwa festgestellte Leistungsdifferenzierung zwischen den Abiturienten in technischen Berufen und Teilnehmern an Ersatzmaßnahmen ohne Hauptschulabschluss stabil etwas aussagt, dürfte ein Vergleich dieser Statusgruppen mit einer Münchner Berufsschule für technische und einer weiteren für hauswirtschaftliche Berufe sowie die gleiche Untersuchung in einem ländlichen bayerischen und baden-württembergischen Raum die notwendige Klarheit schaffen. Abgesehen von den Kosten einer solchen Untersuchung aus der Distanz von Nord- zu Süddeutschland ist es – aus sicherlich verständlichen Gründen – schwer geworden, empirische Untersuchungen in Schulen durchzuführen, für die erst einmal nur das Interesse des nachfragenden Wissenschaftlers noch dazu von einer fremden Universität spricht. Die andere Seite sorgt schon für genug Aufregung.

VI Ausblick

Zwei Dinge bestärken uns in der Fortsetzung des EOL-Ansatzes. Das eine ist eine Folge der parallel betriebenen Weiterentwicklung von bislang freilich noch stärkerem explorativem Charakter. Das andere ist die Entdeckung einer Verwandtschaft der konzeptionellen Basis von EOL mit Bemühungen der australischen Wissenschaftlergruppe um Patrick Griffin, der dem Assessment Research Centre (ARC) der University of Melbourne vorsteht und enge wissenschaftliche Kontakte zu internationalen PISA-Experten und der OECD hat und zu diesen eine damit einhergehende professionelle Nähe hält.

(a) Wir haben die Hypothese von der Verberuflichung des Lernens zur Erklärung seiner Effekte als „nachgeholte Reife“ aufgestellt. Alle statistischen Daten deuten daraufhin, dass die kritische Phase einer Berufsausbildung ihr Beginn ist, während ihr erfolgreicher Abschluss am Ende von drei oder dreieinhalb Jahren eine Regel mit nur geringen Ausnahmen darstellt. Durch eine jahrgangsweise durchgeführte Fortschreibung der Erhebung haben wir als Test der dazu anzupassenden Instrumente ein Modell für EOL II und III entwickelt und in den Schulen eingesetzt, die sich für die Wiederholung von EOL I entschieden haben. Neben neuen, mehr auf die berufliche Fachrichtung bzw. den Schwerpunkt der Schule eingehenden Items haben wir auch jeweils eine Evaluationsaufgabe gestellt. Sie war anhand von Zeichnungen und Photos zu lösen, es gab keine multiple choice-Fragen.

Die Anlage von EOL II und III ist schlicht die, die Verberuflichung des Lernens als einen Konzeptaufbau zu verfolgen, der fachlich entsteht, aber seine curriculare Lenkung und Anreicherung überragt. Uns interessieren dabei vor allem die Persistenz von Fehlkonzepten bzw. die Bedingungen ihrer Auflösung. Das ehrgeizige Ergebnis der Untersuchung sollte die Frage entscheiden, ob eher der betriebliche oder eher der schulische Ausbildungsteil bekannte Fehlkonzepte beseitigt und insgesamt für die überlegene Leistungsfähigkeit des beruflichen Bildungssystems verantwortlich ist.

Unsere Erfahrungen sprechen dafür, dass die bislang eingesetzten Instrumente funktionieren, weil sie belastbare Ergebnisse erbringen. Mit ihnen ist ein erhöhter Aufwand verbunden, der sich dadurch rechtfertigt, dass wir bei den versuchsweise eingesetzten Evaluationsaufgaben deutliche Hinweise auf die Existenz von sogenannten Superkonzepten erhalten konnten, die es gestatten, unterhalb des speziellen Berufs wirksame Konzepte zu identifizieren, die in verschiedenen Domänen aufgebaut, aber nur in einzelnen ausgeprägt werden. Dadurch ergeben sich wahrscheinlich Indikatoren, die berufliche Kompetenzentwicklung in der Weise verstehbar machen, dass hinter dem speziellen Einsatz eines Konzepts dessen domänenübergreifende Bedeutung sichtbar wird. Das hieße z.B., dass man diese Konzepte zwar nur in einer beruflichen Spezialisierung entwickeln kann, diese dann aber eine generellere Bedeutung, eben eine operationelle gewinnen. Für die technischen Berufe gestaltet sich das bislang undramatisch. Man könnte von einem harmonischen Verhältnis der Ergebnisse zu den Hypothesen sprechen.

Das scheint bei den kaufmännischen Berufen ganz anders zu sein. Die Operationalisierung der Frage nach einem Superkonzept haben wir wie folgt vorgenommen: Um nicht aus fachlichen Details oder mathematisch leidlich anspruchsvollen Verfahren jenseits der Prozentrechnung – etwa Zinsrechnung – mühselig eine dann doch generalisierbare Evaluationsaufgabe für das kaufmännisch-verwaltende Berufsfeld zu finden, haben wir ein Grundproblem ökonomischen Handelns mit einer generellen, alltäglichen Erwartung kombiniert: Den gerechten Tausch. Tauschakte sind bei Jugendlichen schon verinnerlicht, gleichzeitig kann man von einer ausgeprägten Sensibilität bei Fragen der Übervorteilung und Täuschung in alltäglichen Geschäften ausge-

hen. Wer sich also nicht übervorteilt sehen will, muss Geschäfte, in die er eintritt, kaufmännisch übersehen, d. h. echte Kosten und den erzielbaren Nutzen erkennen. Wir haben deshalb einen Fall konstruiert, bei dem eine Mischung aus Familie und Wohngemeinschaft mehrere Fahrzeuge hält, die alle für den jeweiligen Zweck wie Fahrt zur Arbeit, zum Studienort, zu Bekannten und Freunden und zum Einkaufen oder zur Haushaltsführung benötigen. Dabei bezeichnen wir die Fähigkeit als Superkonzept, im Äquivalenzprinzip die Gerechtigkeitsnorm eines Tauschakts zu erkennen und es auf Probleme der gerechten Verteilung ökonomisch vernünftig anzuwenden.

Bis auf die ganz wenigen Ausnahmen, die auf einer machbaren, weil aus der Perspektive der Beteiligten konstruierten Regelung beruhte, wurden uns Lösungen präsentiert, die zum überwiegenden Teil insofern auf gar keinem Konzept basierten, als das moralische Problem zwar nicht erkannt, mit der Lösung aber nichtsdestoweniger kräftig moralisiert wurde. Die dritte Gruppe von Lösungen bestand aus grenzwertigen Vorschlägen, bei denen zugunsten eines transparenten Fahrzeuggebrauchs Vorschriften für die Gestaltung von Freizeit, Studium und Beruf gemacht wurden, damit sich die Mobilitätsansprüche aller erfüllen ließen.

Während die in der Wirtschaftspädagogik bekannten Untersuchungen zum moralischen Bewusstsein vom latenten Konflikt zwischen dem professionellen Interesse des Kaufmanns an Profit und den moralischen Grenzen ausgehen, die seiner Verfolgung gesetzt sind, stießen wir bei EOL II für kaufmännisch-verwaltende Berufe mit immerhin 1.150 Probanden auf das augenscheinlich größere und zugleich tiefer liegende Problem, dass die einem gesicherten moralischen Urteil vorausgehende kognitive Leistung der Strukturierung des fraglichen ökonomischen Sachverhalts nach Tauschäquivalenten wesentlich schwieriger zu sein scheint als die Herleitung einer moralisch scheinbar unanstößigen Position. Dass diese dann im Umsetzungsfall zu grellen Ungerechtigkeiten führen würde, hat nennenswerte Teile der Probandengruppen zwischen Hauptschülern, die Verkäufer lernten, und Abiturienten in einer Banklehre gar nicht berührt. Wir sehen darin eine Bestätigung unseres Ansatzes, nicht von Idealen, sondern von den Konzepten zunächst auszugehen, deren Entwicklung im Verlauf eines Bildungsprozesses über die Stadien von Konzeptfreiheit und falsch assimilierten Fehlkonzepten hin zu den tragfähigen Fachkonzepten führt.

Patrick Griffin hat Beobachtungen geschildert, die wohl auch den Psychometrikern zu schaffen machen. Den vergleichsweise hohen Grad an Homogenität ihrer Messbefunde erzielen sie durch Rückgriff auf die Inhalte, an deren Bewährung die Probanden genau zu solchen Gruppen zusammengefasst werden, denen ihr Niveau der Bearbeitung entspricht. Bei einem Projekt, das die amerikanischen Unternehmen Cisco, Microsoft und Intel in Auftrag gegeben haben, sollte er wahrscheinliche, künftig wichtige Anforderung ermitteln und darauf ausgerichtete Testinstrumente für assessments entwickeln. Hier interessiert weniger, wie er zu den Aussagen über künftig bedeutender werdende Anforderungen kam, maßgeblich erscheint das Forschungsergebnis, zu dem seine

Testinstrumente führten. Womit niemand rechnete: dass die Heterogenität der Befunde drastisch anstieg – obwohl seine Testkandidaten dem Reservoir der kommenden Eliten von Cisco, Microsoft und Intel angehören werden und mithin nicht zu den schlechteren Absolventen der einschlägigen amerikanischen Universitäten gezählt haben dürften, fielen die Unterschiede bei den gemessenen Leistungen so hoch aus, dass man angesichts der psychometrisch gewohnten Korridore an Leistungen zu zweifeln begann, ob man das Richtige mit den tauglichen Instrumenten gemessen habe.

Griffins Interpretation dieses Problems konzediert zum einen, dass es sich um den schlichten Effekt handelt, der eintreten muss, wenn man der gebildeten Person selbst näher als deren Bildungsbiographie kommt. Mit anderen Worten, für ihn spricht aus der Heterogenität der Testergebnisse das Produkt einer zwar nicht zufälligen, aber auch nicht genormten oder normierbaren Entwicklung.

Damit eröffnet er zum anderen eine der Untersuchung selbst innewohnende Perspektive, die er den „teachable moment“ nennt. Wer sich der basalen Argumente für Rasch-Skalierung erinnert, dass ein Testergebnis stets zwei Variablen reflektiert, die Leistungsfähigkeit des Probanden und die Schwierigkeit des items, wird vielleicht verblüfft sein zu hören, dass angesichts von test settings, die die item-Schwierigkeit mangels greifbarer Bezüge in durchdeklinierten Fachwissenschaften etc. nicht sicher bestimmen können, man auf Seiten der Psychometriker einzuräumen bereit ist, einen teachable moment zu konstruieren, der besagt, dass sich item-Schwierigkeit und Probandenleistungsfähigkeit für einen Moment so nahe kommen, dass Probanden das Problem einsehen könnten und doch noch der Einsicht so fern sind, dass sie die Lösung des Problems oder die Antwort auf die Frage selbständig nicht finden dürften.

Didaktisch verstanden, ist das der Moment, in dem die Weisheit des Lehrers hilft. Psychometrisch gesprochen, ist das der Moment, in dem die Norm der Raschskala didaktisch sinnvoll zu nutzen wäre: Die beiden Kategorien item-Schwierigkeit und Probandenleistungsfähigkeit liefern in der Vorstellung der Annäherung beider die Legitimation eines belehrenden Eingriffs. Wenn, und nur dann, beides sich so nahe kommt, kann ein Proband lernen.

In EOL bezeichnen wir diesen item-Modus als alltagstheoretisch: Die Imagination des Probanden reicht an das Problem heran, es wird verstanden und dass es verstanden wird, ist nicht trivial, sondern setzt Wissen voraus. Zugleich bleibt dieses Wissen unzureichend. Eine Operation als bloße Assimilation an das Problem führt nicht zur Lösung.

Ein Probandenwissen, das nicht nur das Problem erfasst, sondern zu seiner Lösung führt, überlappt sich nach dem Bild des teachable moments mit der item-Schwierigkeit. Der Proband löst das Problem generativ, er akkomodiert und benötigt über die Konfrontation mit dem Problem hinaus keine Unterstützung. Er kann einer Belehrung entbehren.

Probanden, deren Leistungsfähigkeit sich in Formen mangelnder Operationalität äußert, stehen insofern distanziert zur item-Schwierigkeit, als das Problem schon nicht zum eigenen gemacht werden kann. Außerhalb der Reichweite des eigenen Wissens und eines sich daraus möglicherweise spei-

senden Interesses, bleibt die Konfrontation mit der Frage noch folgenloser als ein Belehrungsversuch fruchtbar sein könnte.

Griffin prognostiziert für die nächsten PISA-Runden eine noch weitergehende Öffnung der Methoden zur Item-Generierung, als sie bislang schon gewagt wurde. Die letzten Modelle, die auf Literacies gingen, haben den Charakter einer bilanzierenden Messung nicht abstreifen können. Erst wenn, so Griffin, die Messung sich auf Anforderungen richtet, die sich durch keine Lernarrangements bislang stellten, werden die PISA-Experten das messen, worauf es eigentlich ankommt. Damit meinte er nicht die Interessen von Cisco, Microsoft und Intel. Er meinte den simplen Anspruch der empirischen Bildungsforschung, in ihren Messungen Bildung überhaupt zu erfassen.

Griffins Rede vom teachable moment trifft sich ein wenig mit dem europäischen Bildungsdenken. Eine psychometrische Weiterentwicklung in seinem Sinne würde jenes Versprechen einlösen, das deutsche Bildungspolitikern heute schon für einlösbar halten, wenn sie nur genug Geld für die empirische Bildungsforschung aufzubringen vermöchten. Es wird aber noch lange dauern, bis die Psychometriker mit Verweis auf ihre Methoden und die damit verbundenen Einsichten in die Bildung ihrer Probanden sagen können: Dem Probanden kann geholfen werden. Wir versuchen das gleiche, von der Seite der Bildung kommend.

Literatur

- Humboldt, Wilhelm von: Der Königsberger Schulplan, in: ders.: Schriften zur Anthropologie und Bildungstheorie, hrsg. von Andreas Flitner, Düsseldorf/München 1956, S. 69-76.
- Gerhard Schaefer (Hg.): Allgemeinbildung durch Naturwissenschaften – Denkschrift der GDNA-Bildungskommission, Kurzfassung, Freising 2010.
- Barbara E. Stalder / Thomas Meyer / Sandra Hupka-Brunner: Leistungsschwach – Bildungsarm? Ergebnisse der TREE-Studie zu den PISA-Kompetenzen als Prädiktoren für Bildungschancen in der Sekundarstufe II, in: Die Deutsche Schule 100. Jg., Heft 4, 2008, S. 436-448.
- Wagenschein, Martin: Wesen und Unwesen der Schule, in: Erziehung wozu?, Stuttgart 1956, S. 49-62.