

Detlef H. Rost

Interpretation und Bewertung pädagogischer und psychologischer Studien

4. Auflage

Eine Arbeitsgemeinschaft der Verlage

Brill | Schöningh – Fink · Paderborn

Brill | Vandenhoeck & Ruprecht · Göttingen – Böhlau Verlag · Wien · Köln

Verlag Barbara Budrich · Opladen · Toronto

facultas · Wien

Haupt Verlag · Bern

Verlag Julius Klinkhardt · Bad Heilbrunn

Mohr Siebeck · Tübingen

Narr Francke Attempto Verlag – expert verlag · Tübingen

Psychiatrie Verlag · Köln

Ernst Reinhardt Verlag · München

transcript Verlag · Bielefeld

Verlag Eugen Ulmer · Stuttgart

UVK Verlag · München

Waxmann · Münster · New York

wbv Publikation · Bielefeld

Wochenschau Verlag · Frankfurt am Main



Der Autor

Dr. Detlef H. Rost war von 1981 bis 2010 Universitätsprofessor für Pädagogische Psychologie und Entwicklungspsychologie am Fachbereich Psychologie der Philipps-Universität Marburg und anschließend bis 2020 Gastprofessor.

Von 2013 bis 2020 war er zudem Gastprofessor an der Faculty of Psychology der Southwest University Chongqing. Seit 2020 ist er dort Professor am Center for Mental Health Education.

Detlef H. Rost

Interpretation und Bewertung pädagogischer und psychologischer Studien

Eine Einführung

4., überarbeitete und
erweiterte Auflage

Verlag Julius Klinkhardt
Bad Heilbrunn • 2022

Online-Angebote oder elektronische Ausgaben zu diesem Buch
sind erhältlich unter utb.de und elibrary.utb.de

Die Deutsche Bibliothek – CIP-Einheitsaufnahme

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

2022.Kk. © by Julius Klinkhardt.

Das Werk ist einschließlich aller seiner Teile urheberrechtlich geschützt.

Jede Verwertung außerhalb der engen Grenzen des Urheberrechtsgesetzes ist ohne Zustimmung
des Verlages unzulässig und strafbar. Das gilt insbesondere für Vervielfältigungen, Übersetzungen,
Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen.

Satz: Kay Fretwurst, Spreewald.

Einbandgestaltung: Atelier Reichert, Stuttgart.

Druck und Bindung: Friedrich Pustet, Regensburg.

Printed in Germany 2022

Gedruckt auf chlorfrei gebleichtem alterungsbeständigem Papier.

utb-Band-Nr.: 8518

ISBN 978-3-8385-8789-9 **digital**

ISBN 978-3-8252-8789-4 **print**

Inhalt

1 Die Beurteilung empirischer Studien	9
1.1 Funktionen der Forschung	20
1.2 Forschung und Theorie	25
1.3 Forschung und Praxis	32
1.4 Empirikritik	33
1.5 Verständnisprobleme	34
1.6 Forschungsberichte	37
2 Die Problemdarstellung	43
2.1 Wissenschaftliches Arbeiten	48
2.2 Stand des Wissens	49
2.2.1 Narratives Review	49
2.2.2 Meta-Analyse	52
2.3 Replikationen	61
2.4 Theorien	67
2.5 Problemformulierungen, Operationalisierungen	68
2.5.1 Operationalisierung	69
2.5.2 Nicht-reaktive Messung	73
2.6 Forschungsethik, Forschungskandale	74
3 Die Methodik	91
3.1 Hypothesen, Falsifizierung	91
3.2 Nullhypothesen, Alternativhypothesen	93
3.3 Feldstudie, Laborforschungen, retrospektive Studien	99
3.4 Störfaktoren, Ausreißer	105
3.5 Beobachtungsverfahren, Inhaltsanalysen	112
3.6 Nullhypothesen, α -Fehler, β -Fehler	119
3.7 Stichproben	125
3.7.1 Zufallsstichprobe	127
3.7.2 Nicht-zufällige Stichprobe	130
3.7.3 Weitere Stichprobenprobleme	133
3.7.4 Sonderfall Klumpenstichprobe	136
3.8 Extremgruppen, Regressionseffekte, Veränderungsmessungen	141
3.9 Gültigkeit von Experimenten	150
3.9.1 Internale Validität	150
3.9.2 Externale Validität	155

3.10	Versuchspläne	157
3.10.1	Schwache Designs	159
3.10.1.1	Einzelfallstudie	160
3.10.1.2	Zeitreihen-, „Experiment“	161
3.10.1.3	Intensive Versuchspläne der Verhaltensmodifikation.	162
3.10.1.4	Schwacher Ein-Gruppen-Plan ohne Vortest, mit Behandlung, mit Nachtest	164
3.10.1.5	Schwacher Ein-Gruppen-Plan, mit Vortest, mit Behandlung, mit Nachtest	165
3.10.1.6	Schwacher Zwei-Gruppen-Plan, ohne Vortest, mit Behandlung, mit Nachtest	166
3.10.1.7	Schwacher Zwei-Gruppen-Plan, mit Vortest, mit Behandlung, mit Nachtest	167
3.10.2	Starkes Designs (mit Randomisierung).	169
3.10.2.1	Starker Zwei-Gruppen-Plan mit Vortest, mit Behandlung, mit Nachtest	169
3.10.2.2	Starker Drei-Gruppen-Plan mit Vortest, mit Behandlung, mit Nachtest	171
3.10.2.3	Starker Zweigruppenplan, mit Behandlung, ohne Vortest.	172
3.10.2.4	Starker Drei-Gruppen-Plan, ohne Vortest, mit Behandlung, mit Nachtests	173
3.10.2.5	Starker Solomon-Vier-Gruppen-Plan	174
3.11	Entwicklungsstudien	175
3.11.1	Querschnitt	175
3.11.2	Längsschnitt.	176
3.11.3	Inhaltliche und methodische Probleme	178
3.12	Variablen, Indikatoren	181
3.13	Kausale Beziehungen	184
3.13.1	Korrelation.	185
3.13.2	Multiple Korrelation, Regressionsanalyse und kanonische Korrelationsanalyse	187
3.13.3	Kovariation und Ursache	193
3.14	Interviews, Fragebogen	202
3.15	Tests, Skalen	205
3.15.1	Objektivität	211
3.15.2	Reliabilität	212
3.15.2.1	Testwiederholungsreliabilität, Paralleltestreliabilität	213
3.15.2.2	Testhalbierungsreliabilität, Konsistenzanalyse	214
3.15.3	Validität	218
3.15.3.1	Inhaltsvalidität, Augenscheinvalidität, curriculare Validität.	220
3.15.3.2	Übereinstimmungsvalidität	221
3.15.3.3	Kriteriumsvalidität	221
3.15.3.4	Differentielle Validität	221
3.15.3.5	Konstruktvalidität	221
3.15.4	Normierung.	223
3.15.5	Ökonomie	223
3.16	Hilfsdaten, Vergleichsdaten	224
3.17	Voruntersuchungen	230

4 Die Ergebnisse	231
4.1 Fehlende Daten	233
4.2 Weitere Aspekte	241
4.3 Hilfsmittel Statistik	244
4.4 Skalentypen	245
4.4.1 Nominalskala	246
4.4.2 Ordinalskala	247
4.4.3 Intervallskala	248
4.4.4 Verhältnisskala	249
4.5 Spezifische Verfahren	249
4.5.1 Parametrische vs. nicht-parametrische Verfahren	251
4.5.2 Varianzanalyse	254
4.5.3 Kovarianzanalyse	259
4.5.4 Dimensionsanalyse	263
4.6 Signifikanz, Relevanz, Power, α -Fehler-Kumulierung	276
4.6.1 Signifikanzniveau	277
4.6.2 Effektstärke	281
4.6.3 Power-Analyse	291
4.6.4 α -Fehler-Kumulierung	296
4.7 Tabellen, Abbildungen	304
5 Die Diskussion	309
6 Das Literaturverzeichnis	311
7 Eine erboste Nachbemerkung	317
Literaturverzeichnis	321
Anhänge	365
Anhang 1: Entscheidungsfelder im Forschungsprozess	365
Anhang 2: Kurze Checkliste zur Beurteilung empirischer Untersuchungen	366
Anhang 3: Übersicht über die verschiedenen Skalentypen	369
Anhang 4: Pädagogisch-psychologische Fachzeitschriften	370
Anhang 5: Kontroversen in der Wissenschaft	373
Stichwortregister	377

1 Die Beurteilung empirischer Studien

Kommen wir gleich zur Sache: Empirische (= erfahrungswissenschaftliche) Untersuchungen in der Erziehungswissenschaft, Pädagogischen Psychologie und angewandten Entwicklungspsychologie (aber auch in anderen psychologischen Teildisziplinen wie Wirtschaftspsychologie, Klinische Psychologie, Arbeitspsychologie, Differentielle und Diagnostische Psychologie) sowie in weiteren sozialwissenschaftlichen Fächern (wie Soziologie oder Politikwissenschaft) zu verstehen und kritisch zu bewerten, das ist wahrlich nicht leicht. Empirische Methoden kann man sich nicht mal schnell *en passant* (= im Vorbeigehen) reinziehen. Wie überall, so gilt auch hier der Satz: „Ohne Fleiß kein Preis“. Das ist die weniger angenehme Botschaft. Und nun folgt die gute Botschaft: Empirische Methoden sind nicht so kompliziert und schwierig zu verstehen, wie es viele glauben. *Forschungsmethodik ist kein Herrschaftswissen. Angst und Ehrfurcht vor Gedrucktem, insbesondere vor empirischen Zeitschriftenartikeln, sind unangebracht.*

Viele Methodenbücher strotzen nur so vor Formeln und mathematischen Ableitungen. Nicht wenige Studenten fassen solche Lehrbücher nur ungern an, zumal sie oft furztrocken geschrieben sind. Dieses Buch ist etwas anders. Es bedient sich einer ziemlich lockeren Sprache. Es bringt nur ganz vereinzelt ganz einfache Formeln. Es verzichtet auf mathematische Ableitungen, da es nicht um mathematisch-statistische Beweise, Umformungen und Rechenwege geht. Man kann grundlegende methodische Prinzipien und Konzepte empirischer Forschung auch verstehen, wenn man kein Mathefreak oder Statistikguru ist, wenn diese in einer einfachen Sprache und anhand von Beispielen erläutert werden. Zu vielen Themen finden sich kommentierte einführende und weiterführende Literaturhinweise für diejenigen, die sich etwas mehr mit den jeweils angesprochen Methoden beschäftigen wollen. Also: keine Angst. Ran an den Methodenspeck.



Literaturhinweise – Statistikschonkost für Anfänger:

Ja, es gibt sie tatsächlich, die verständlichen und vergnüglich zu lesenden Einführungen in die (Alltags-)Statistik. Die nachfolgenden Bücher verdienen alle das Prädikat „sehr empfehlenswert“:

- Novizen mit einer angehenden oder bestehenden Phobie (= extreme Angst) für Zahlenkram führen sich zuerst *Achtung: Statistik* (2015) und *Achtung: Mathe und Statistik* (2018) von Christensen und Christensen zu Gemüte. Ohne Formelkram und Fachbegriffe präsentiert jedes Buch 150 kurze alltagsnahe Kolumnen zum Nachdenken und Schmunzeln. Der Lustgewinn ist hoch.
- Das Buch *So lügt man mit Statistik* (Krämer, 2015a) ist von ähnlichem Kaliber und echt gut, ebenfalls für blutige Laien geeignet. Zehn unterhaltsame Seiten täglich vor dem Zubettgehen lesen! In drei Wochen ist man dann klüger und geht statischem Blödsinn, wie er oft in Zeitungen und anderen Massenmedien verzapft wird, nicht mehr auf den Leim. – Wer ergänzend eines (oder besser noch: in alle) der ähnlich gelagerten Bände *Lügen mit Zahlen* (2012) und *Die Zahlentrickser* (2019) von Bosbach und Korff (2011) oder *Statischer Unsinn. Wenn*

Medien an der Prozenzhürde scheitern (Quatember, 2015) liest (oder besser noch: alle), ist mit geringem Aufwand an Gehirnschmalz ein Alltags-Statistikexperte geworden.

- Gesteigert wird der Lerneffekt durch Krämers *Statistik für alle* (2015b). Anschaulich und vergänglich werden 101 wichtige Begriffe sehr erklärt. – *Statistik schön einfach. Eine Einführung in 50 Bildern* (Spiegler, 2020) ist eine prima Ergänzung dazu.
- Die beiden programmierten Texte für Anfänger von Gotkin & Goldstein (1972, *Grundkurs in Statistik*, 1972) und McCollough und van Atta (1970, *Statistik programmiert*) eignen sich zum Selbststudium der deskriptiven (= beschreibenden) Statistik. Sie sind schon lange vergriffen. Mit etwas Glück findet man sie noch in Universitätsbibliotheken. Ausleihen, kopieren, anfangen!
- Wer wissen will, wie man mit Statistik tricksen und täuschen kann, kann sich Huff (2010, englisch) reinziehen. – Nett zu lesen und verständlich ist auch die Neuauflage des Buchs von Salkind (2014, englisch). Es hält, was sein Titel verspricht.

Das vorliegende Buch ist als Essenz aus meinen universitären Lehrveranstaltungen erwachsen. Es soll helfen, empirische Studien besser lesen, besser verstehen und besser bewerten zu können. Ich habe es mit der Zielsetzung geschrieben, dem in empirischer Methodik weniger bewanderten Leser einige Leitlinien zur Lektüre erfahrungswissenschaftlicher Untersuchungen an die Hand zu geben. Diese Leitlinien können darüber hinaus auch helfen, eigene empirische Arbeiten auf Schwachstellen hin abzuklopfen und besser zu planen. Es ist weder schädlich noch schändlich, wenn sich die vielen Sozial-, Familien-, Gesundheits- und Bildungspolitiker, Kultusministerialbeamten, Mitarbeiter von Schulaufsichtsbehörden, Schulleiter, Lehrkräfte, aber auch Studenten etc. mit elementaren Prinzipien empirischer Forschung so weit vertraut machen, dass sie verstehen, was sie lesen, damit sie etwas fundierter über Forschungsergebnisse reflektieren und diskutieren können. Auch dabei könnte die Lektüre dieses Buchs eine gewisse Grundlage bereitstellen. Und es will denjenigen Psychologen, Pädagogen und Sozialwissenschaftlern von Nutzen sein, deren Auseinandersetzung mit Forschungsmethodik (sprich: deren Studium) schon weiter zurückliegt und die das Bedürfnis haben, früher Gelerntes und inzwischen fast Vergessenes wenigstens in exemplarischen Grundzügen aufzufrischen.

Obwohl nicht als primäre Zielgruppe in den Blick genommen, greifen auch Masterstudenten und manche Doktoranden der Psychologie und Erziehungswissenschaft zu diesem Buch, wenn es darum geht, eine empirische Qualifikationsarbeit zu planen, auszuwerten und abzufassen. Das Buch ist voller Beispiele, getreu der Devise *verba docent, exempla trahunt!* (= Wörter belehren, Beispiele wirken!). Ich habe die meisten in pädagogisch-psychologischen Zusammenhängen angesiedelt. Das hier Gesagte ist nahtlos und ohne Mühe auf andere Inhaltsbereiche (Entwicklungspsychologie, Sozialisationsforschung, Klinische Psychologie, Arbeits-, Betriebs- und Organisationspsychologie, Sozialpsychologie, Sonder- und Sozialpädagogik usw.) übertragbar. Die methodischen Probleme und Lösungsmöglichkeiten bei der Evaluation (= empirische Erfolgs- bzw. Misserfolgskontrolle) eines klinisch-psychologischen Therapieprogramms differieren kaum von denjenigen, die sich bei der Effektivitätskontrolle einer pädagogisch-psychologischen Fördermaßnahme stellen.



Kasten 1.1 – „Gendergerechte“ Sprache:

- Ach ja: Die umständliche parallele Nennung von männlichen und weiblichen Personen (Forscher und Forscherinnen, Professorinnen und Professoren, Autoren und Autorinnen, Empirikerinnen und Empiriker, Studenten und Studentinnen, Verfasserinnen und Verfasser und so weiter) findet sich in diesem Buch nicht. Solche Doppelnennungen machen den Text nicht

lesbarer. Wenn da „Student“ oder „Autor“ steht, dann sind Frauen (= weibliche Studenten oder weibliche Autoren) und Männer (= männliche Studenten oder männliche Autoren) sowie Personen, die sich nicht als Frau oder Mann bezeichnen möchten, gleichermaßen gemeint. Sprachwissenschaftlich spricht man in diesem und vergleichbaren Fällen vom generischen Maskulinum. Denn: Das *grammatikalische Geschlecht* ist etwas völlig anderes als das *biologische Geschlecht*. Die Auffassung, bei der Verwendung des generischen Maskulinums würden Mädchen und Frauen sowie Personen, die sich einem anderen als dem männlichen oder weiblichen Geschlecht zugehörig verstehen, nicht mitbedacht, ja systematisch unterdrückt, teile ich nicht. Wenn in den Tagesthemen von einem „Germanistenkongress“ berichtet wird, nimmt niemand an, dort wären nur Männer zugegen, aber keine Frauen, obwohl in Deutschland Frauen knapp mehr als 50 % der Germanistiklehrstühle inne haben. Wenn in der Hessenschau berichtet wird, im Burgwald wäre erstmals ein Wolf gesichtet worden, meint wahrscheinlich kaum jemand, es sei auf jeden Fall ein männlicher Wolf (= Rüde) gewesen, es habe sich aber auf keinen Fall um ein weibliches Tier (= Fähe) gehandelt.

- Welcher Unsinn mittlerweile unter der Flagge der „gendergerechten“ Sprache segelt, wird nicht nur am Online-Duden deutlich, der allen Ernstes *Bösewichtin* aufführt. Die Wirtschaftsverwaltung des Fachbereichs Psychologie der Philipps-Universität Marburg schickt E-Mails mal an *Liebe MitgliederInnen*, mal an *Liebe Mitglieder:innen*. Anne Will verwendete 2018 *Mitglieder und Mitgliederinnen*. Dass „Mitglied“ ein Neutrum ist und damit alle möglichen Geschlechtsidentitäten umfasst, scheint weder der Verwaltung des psychologischen Fachbereichs der Uni Marburg noch Anne Will bekannt zu sein. Deutlicher kann man sprachliche Inkompetenz nicht dokumentieren. Ein renitenter Professor bemerkte dazu in einer universitären Gremiensitzung, er unterstütze sehr das Bemühen um eine „geschlechtsgerechte“ Sprache und regte an, ob man nicht besser „Liebe Mit-und-ohne-Glieder“ verwenden sollte. Annalena Baerbock, Spitzenfrau der Grünen, setzte in der Talkshow von Anne Will noch einen drauf, als sie *Steuerinnenzahler* verwendete. Die Spitze sprachlicher Verblödung leistete sich der öffentlich-rechtliche Fernsehkanal *Phoenix*, der ein YouTube Video zum Triell von Baerbock, Laschet und Scholz mit *Kanzler:inkandidat:innen* untertitelte (31.08.2021).
- Urs Bühler schrieb am 17.2.2021 in der NZZ (*Neue Züricher Zeitung*) in seinem Kommentar *Der Duden heisst die Bösewichtin willkommen*: „Die Sprache ist keine Notdurftanstalt, in der es jede Tür einem Geschlecht zuzuordnen gilt. Sie kann die Verhältnisse dieser Welt weder passgenau abbilden noch verbessern, doch dient sie im Alltag in erster Linie der Kommunikation. Diese soll ohne unnötige Komplikationen und Verunstaltungen stattfinden.“ Dem ist kaum etwas hinzuzufügen.
- Gnadenlosen Verfechtern der „genderisierten“ Sprache empfehle ich, den klugen Essay *Genus ist nicht sexus* des Philosophen Richard Schroeder zu lesen, der am 17. Dezember 2018 in der WELT erschienen ist (https://www.welt.de/print/die_welt/debatte/article185627376/Essay-Genus-ist-nicht-sexus.html). Und den ausgewogenen Artikel *Die Zerstörung des Deutschen* des renommierten Sprachwissenschaftlers Peter Eisenberg in der Welt vom 10. August 2021 (<https://www.welt.de/kultur/plus233017189/Gendern-und-Standardsprache-Die-Zerstoe-rung-des-Deutschen.html>) sowie das Buch des Musikers und Komponist Fabian Payr (2021, *Von Menschen und Mensch*innen*). Payr erläutert darin *21 gute Gründe, mit dem Gendern aufzuhören*. Seine Überlegungen sind differenziert und bedenkenswert.
- Das Gender-Sternchen (z. B. *Forscher*innen*), den Gender-Doppelpunkt (z. B. *Forscher:innen*), das große Binnen-I (z. B. *ForscherInnen*) und den Gender-Unterstrich (z. B. *Forscher_innen*) werde ich nicht benutzen, wenn ich Forscher (unabhängig von ihrem Geschlecht) meine, auch nicht *Forschende*. *Studierende* verwende ich nicht, wenn ich Studenten (unabhängig von ihrem

Geschlecht) meine. Das Partizip Präsens in seiner substantivierten Form ist nur gerechtfertigt, wenn man Personen bezeichnen will, die gerade jetzt ein bestimmtes Verhalten zeigen. Kurt und Laura, die sich im Freibad sonnen, sind in dieser Situation keine Studierenden, auch wenn sie an einer Uni eingeschrieben sind. Sitzen Laura und Kurt dagegen in der Unibibliothek und lesen Lehrbücher, dann sind sie in dieser Situation Studierende. Ein Ertrinkender säuft gerade im Wasser ab – setzt er seinen Fuß auf das Land, ist er kein Ertrinkender mehr.

- Was die „empirische Genderforschung“ betrifft, ist sie nicht selten in sich widersprüchlich. Gender-Studien postulieren Geschlechtsunterschiede als soziales (nicht als biologisches) Phänomen. In vielen Gender-Untersuchungen wird aber nur das biologische Geschlecht erfragt – das soziale oder psychologische Geschlecht wird häufig nicht gemessen, obwohl dafür diverse Fragebogen verfügbar sind.
- Das amtliche Regelwerk zur Standardisierung der deutschen Orthografie stellt eine weithin verpflichtende Richtschnur für Schulen etc. dar. Es kennt kein Binnen-I, keinen Gender-Doppelstrich, keinen Gender-Unterstrich und kein Gender-Sternchen.



Deutschsprachige Bücher zu Geschlechtsunterschieden:

Die Frage, ob sich Jungen von Mädchen und Frauen von Männern in psychologischen und biologischen Variablen unterscheiden, und wenn ja, wie und warum, wird in der Öffentlichkeit seit vielen Jahren mit gesteigertem Interesse verfolgt. Seit Jahren gibt es zu diesem Thema heftige Auseinandersetzungen, zum großen Teil ideologisch-emotional unterbaut. Vergessen wird häufig, dass empirisch belegte Geschlechtsunterschiede nichts darüber aussagen, *wie* diese entstanden sind. Sowohl genetisch-biologische wie auch sozialisationsbedingte Faktoren und vor allem die Interaktion dieser Merkmalsbündel sind zur Erklärung in Betracht zu ziehen. Diesen Faktoren kommen aber variablen-spezifisch unterschiedliche Gewichte zu. Wer sich unvoreingenommen über das Thema informieren möchte, sollte um die vielen einseitigen populärwissenschaftlichen Machwerke (ich erspare mir, an dieser Stelle welche zu nennen) einen ziemlich großen Bogen machen und stattdessen zu soliden Abhandlungen greifen. Davon gibt es genug. Ich führe nachfolgend in chronologischer Reihung exemplarisch einige deutschsprachige Sach- und Fachbücher an:

- Bierhoff-Alfermann, D. (1977). *Psychologie der Geschlechtsunterschiede*. Kiepenheuer & Witsch.
- Merz, F. (1975). *Geschlechtsunterschiede und ihre Entwicklung*. Hogrefe.
- Keller, H. (1979). *Geschlechtsunterschiede. Psychologische und physiologische Grundlagen der Geschlechterdifferenzierung*. Beltz.
- Bierhoff-Alfermann, D. (1989). *Androgynie. Möglichkeiten und Grenzen der Geschlechterrollen*. Westdeutscher Verlag.
- Pool, R. (1995). *Evas Rippe. Das Ende des Mythos vom starken und vom schwachen Geschlecht*. Knaur.
- Lauterbacher, S., Güntürkün, O. & Hausmann, M. (2007). *Gehirn und Geschlecht. Neurowissenschaft des kleinen Unterschieds zwischen Mann und Frau*. Springer.
- Maccoby, E. (1998). *Psychologie der Geschlechter. Sexuelle Identität in den verschiedenen Lebensphasen*. Klett-Cotta.
- Pinker, S. (2008). *Das Geschlechterparadox. Über begabte Mädchen, schwierige Jungs und den wahren Unterschied zwischen Männern und Frauen*. DVA.
- Eliot, (2010). *Wie verschieden sind sie? Die Gehirnentwicklung bei Mädchen und Jungen*. Berlin Verlag.

- Athenstaedt, U. & Alfermann, D. (2011). *Geschlechterrollen und ihre Folgen. Eine sozialpsychologische Betrachtung*. Kohlhammer.
- Bischof-Köhler, D. & Znyi, N. (2022). *Von Natur aus anders. Die Psychologie der Geschlechtsunterschiede* (5. Aufl.). Kohlhammer.
- Meyer, A. (2015). *Adams Apfel und Evas Erbe. Wie die Gene unser Leben bestimmen und warum Frauen anders sind als Männer*. Bertelsmann.

Ganz klar: Einerseits besteht ein großes Bedürfnis zur Erforschung systematischer Tendenzen in den Prozessen der Entwicklung und Entwicklungsoptimierung, Erziehung, Unterrichtung, Prävention und Behandlung, weil unser Wissen über diese Phänomene noch zu gering ist. Andererseits ist aber die komplexe Entwicklungs-, Erziehungs- und Unterrichtswirklichkeit mit ihren unterschiedlichsten Institutionen und mit ihren vielen Faktoren erfahrungswissenschaftlich auch nicht annähernd vollständig zu erfassen. Hinzu kommt, dass die vielen von der Öffentlichkeit und der Bildungsverwaltung im Anschluss an die großen vergleichenden Schulleistungsstudien wie TIMSS (*Third International Mathematics and Science Study*), PISA (*Programme for International Student Assessment*), IGLU (*Internationale Grundschul-Lese-Untersuchung*) oder DESI (*Deutsch-Englisch Schülerleistungen International*) geforderten Reformen im Bildungsbereich in ihren Auswirkungen so objektiv wie möglich kontrolliert und evaluiert werden müssen (Sälzer, 2016; eine komprimierte Übersicht zu den diversen internationalen Schulleistungsvergleichsstudien geben Baumert & Stanat, 2018).

Noch heute leidet die (anwendungsbezogene) sozialwissenschaftlich-empirische Forschung, verstanden als „systematische Deutung sozialer Erscheinungen“ (Atteslander, 1995, S. 13), nicht nur, aber auch in der empirischen Pädagogik und Pädagogischen Psychologie unter einer Vergangenheit, die man durch unzulänglich formulierte Forschungsfragen, die in unzulängliche Versuchspläne gezwängt und unzulänglich ausgewertet wurden, kennzeichnen kann. Gerade in Deutschland hat die empirische Pädagogik immer noch nicht den Stand, den sie verdient, obwohl in den ersten beiden Jahrzehnten des letzten Jahrhunderts die experimentelle Pädagogik (Meumann, 1914) und experimentelle Didaktik (Lay, 1910) das Feld für die damalige Zeit ziemlich gut bestellt hatten. Aber es ist in den letzten 20 Jahren besser geworden.

In der Folge von PISA & Co wurden an Universitäten mehr und mehr Professuren für Erziehungswissenschaft mit quantitativ arbeitenden Forschern, oft aus der (Pädagogischen) Psychologie kommend, besetzt – und das ist gut so. Seit etwa 20 Jahren ist nämlich die Einsicht in die Notwendigkeit empirischer Studien im Bereich von Unterricht und Erziehung enorm gewachsen, sogar bei den Fachdidaktikern, auch wenn deren methodischen Standards hier und da noch verbesserungsbedürftig sind: Im negativen Extremfall erschöpft sich empirische fachdidaktische Forschung in der Sammlung einer Unmenge von Daten ohne präzise Fragestellung (die manchmal im Nachhinein darüber gestülpt wird). Als Auswertung werden die Daten schließlich Zahl für Zahl, Korrelation für Korrelation, verbal abgearbeitet. Spötter haben dafür das Schlagwort „Prozentzahlbelletristik“ geprägt. Die nachfolgende Episode belegt, wie fachdidaktische Forschung nicht laufen sollte.

► **Beispiel 1.1:** Von einem Professor einer hessischen Universität (den Universitätsnamen verschweige ich höflich, ich nenne den Kollegen „Herrn Möttke“) wurde ich vor einigen Jahren angerufen: „Guten Tag, Herr Rost, wir haben uns ja längere Zeit nicht mehr gesehen. Ich hoffe, es geht Ihnen gut?“ – „Ja, ich kann nicht klagen.“ – „Jetzt, wo Ihre Lehrverpflichtung um 50 % reduziert worden ist, haben Sie vermutlich mehr „Luft“ zur Forschung, wenn ich das so sagen darf. Ich möchte Sie gern

zu einer Mitarbeit an einem umfangreichen Forschungsprojekt einladen.“ – „Danke, Herr Möttke. Um was geht es denn?“ – „Äh, ich habe im letzten Jahr zusammen mit meiner Mitarbeiterin religionspädagogische Daten an einer ziemlich großen Stichprobe von Jugendlichen im Großraum Fulda erhoben. Es sind fast 2600 Fälle. Es ist bestimmt ein interessanter Datensatz. Damit kann man doch viel machen. Er muss nur noch ausgewertet werden.“ – „Das ist ja nett, dass Sie an mich denken. Können Sie etwas konkreter werden? Welche Fragestellung verfolgt das Projekt?“ – „Es hat natürlich etwas mit unserem Fach zu tun, versteht sich. Genauer: Es soll fachdidaktisch sein. Präziser wissen wir das selbst noch nicht. Wir haben erst einmal dafür gesorgt, dass wir jede Menge an Daten haben. Wir arbeiten ja selbstverständlich quantitativ. Eine passende Fragestellung fehlt noch, äh, deshalb rufe ich an, das würden wir gerne mit Ihrer Hilfe in Angriff nehmen.“

Können Sie sich die Daten einmal anschauen und eine relevante Fragestellung entwickeln?“

Der Telefonhörer fiel mir beinahe aus der Hand, ich schluckte. Es schoss mir durch den Kopf: Wie absagen, ohne zu beleidigen? Nach einer kurzen Pause dann: „Nun, sie wissen, dass ich Pädagogischer Psychologe und Entwicklungspsychologe bin, aber kein Fachdidaktiker. Von Religionspädagogik habe ich überhaupt keine Ahnung. Aber wichtiger: Obwohl ich jetzt weniger lehre, bin ich voll ausgelastet. Meine parallele Tätigkeit an der Südwest-Universität in Chongqing kostet viel Zeit. Zum anderen sitze ich gerade an der Vorbereitung der Neuauflage eines meiner Bücher. Das hat hohe Priorität. Mir fehlt schlichtweg die Zeit, mich zusätzlich in einem größeren Forschungsvorhaben zu engagieren, leider. Ich bitte um Verständnis.“ Herr Möttke: „Natürlich, selbstverständlich, das kann ich verstehen. Das ist aber schade“. – „Danke, nochmals, für das großzügige Angebot. Alles Gute für Ihr Forschungsvorhaben.“

Ich legte auf. Gott sei Dank. Das war geschafft. Ein Erleichterungsseufzer entrang sich meiner Brust. Vielleicht war er zu laut: Kollege Prof. Pfeffer, der sein Zimmer nebenan hat, stürmte ins Zimmer: „Ist alles OK mit Ihnen? Ich habe ein Stöhnen gehört.“ Ich lachte und klärte ihn dann bei einer Tasse Kaffee auf.

Nicht selten werden unzureichende statistisch-methodische Kenntnisse vornehm hinter der Bezeichnung „*qualitative Forschung*“ versteckt. Die Verwendung des Attributs „qualitativ“ im Zusammenhang mit „Forschung“ suggeriert, es handle sich um eine qualitativ hochwertige Forschung. Zu oft, viel zu oft ist leider das Gegenteil der Fall. Und mit dem Herausstreichen von „qualitativ-empirisch“ wird suggeriert, „quantitative Forschung“ sei so etwas wie banale Fliegenbeinzählerei, klar doch, minderwertiges Herumstochern in oft wenig relevanten Datenhaufen, sei eben Forschung ohne (besondere) Qualität.

Hermann Schöler, emeritierter und für seine kritische Zunge bekannter Psychologieprofessor der Pädagogischen Hochschule Heidelberg, hatte 2008 in einem Vortrag vor Sprachheilpädagogen *Zum Unsinn der Förderdiagnostik* beiläufig auch zur „qualitativen Forschung“ kernig Stellung genommen. Seine Ausführungen sind erfreulich klipp und klar. Sie treffen einen wunden Punkt vieler „qualitativ“ arbeitenden Pädagogen: Qualitative Forschung „wird als [...] Totschlagwort verwendet. [...] Auch hier ist es keine reine Sprachschlamperei, denn ich gehe mal davon aus, dass alle versuchen, nur qualitativ hochwertige Forschung durchzuführen. ‚Qualitative Forschung‘ ist ein Unsinnswort – auch wenn es zwischenzeitlich eine Reihe von Lehrbüchern dieses Titels gibt und sogar eine eigene Zeitschrift mit diesem Titel existiert. In der Forschung wird mit quantitativen oder qualitativen *Methoden* oder – wie ich meine – zumeist beiden gearbeitet. Auch hier wird ein Gegensatz impliziert, nämlich die sogenannte quantitative Forschung, sie wird abqualifiziert, sozusagen in die Ecke zum Fliegenbeinzählen gestellt, sie erzeugt keine Qualitäten. Auf der anderen Seite stehen die qualitativ hochwertigen Resultate der quantitativen Forschung. Solche Begrifflichkeiten werden nicht zufällig oder sprachschlamperisch gewählt, sondern sehr gezielt zur Diskriminierung anderer Positionen.“ Klare Worte, die nicht an Klar-

heit zu wünschen übrig lassen. Wo Hermann Schöler recht hat, hat er recht. Das gefällt manchen „qualitativen Forschern“ vermutlich nicht. Aber damit müssen sie leben. Inzwischen verwildern die Sitten immer mehr. So ziemlich alles, was nicht ausschließlich in den Hirnen von Forschern entstanden ist, segelt unter „empirisch“. Das zeigt sich auch auf den regelmäßigen Tagungen der *Arbeitsgruppe für Empirische Pädagogische Forschung* (AEPF). Gegründet wurde die AEPF Ende der 1960er Jahre, um die *quantitativ-empirische* Forschung im Bereich der Pädagogik zu stärken. Heute segeln dort unter dem Schlagwort „Empirie“ mehr und mehr „qualitative“ Beiträge, die eigentlich nicht auf solch eine Tagung gehören. Ein Beleg? Bitte sehr, gern:

► **Beispiel 1.2:** Unter dem Titel „Milieuspezifische Einflüsse auf die Übergangentscheidung von der Grundschule auf die weiterführende Schule“ (Klinge et al., 2012) wurden auf der 77. Tagung der *Arbeitsgruppe für empirische pädagogische Forschung* (AEPF) an der Universität Bielefeld zwei (von nur elf) „leitfadengestützte narrative Interviews“ mit hohem Aufwand qualitativ-interpretativ ausgebeutet. Mit einem Selbstbewusstsein ohne gleichen wurde aus den Interviews rausgepresst, was pressbar war, wurde längs und quer interpretiert und hoch und runter spekuliert, schön mit PowerPoint Folien aufgehübscht und im Brustton der tiefen Überzeugung präsentiert. Der Lehrer Dieterich Maurer machte nach dem Vortrag in der Kaffeepause seinem Ärger Luft: „Mann, das war vielleicht was! Nervtötend! Geschichtenerzählerei! Ich brauche jetzt dringend ein Butterplätzchen“ – und steckte sich gleich zwei in seinen Mund. „Schade, dass nicht Weihnachten vor der Tür steht. Da gibt es leckere Spekulatius. Würde so gut zum eben Gehörten passen. Spekulatius und Spekulationen – irgendwie treffend, oder?“

Wenn qualitative Forschung sich für etwas eignet, dann zum einen zur Hypothesenbildung (Hypothesen sind Vermutungen über Beziehungen zwischen Variablen bzw. Gesetzesvermutungen; auch Effektvermutungen sind im Grunde Zusammenhangshypothesen). Zum anderen eignet sie sich zur Illustration prototypischer Befunde (= Darstellung besonders typischer Beispiele). Sie ist dagegen ein unzulängliches Werkzeug zur systematisch-empirisch basierten Erkenntnisgewinnung oder gar Erkenntnisüberprüfung. Forschung muss immer sauber und methodenbewusst angelegt und durchgeführt werden. Das gilt für jede Art von Forschung, egal, ob sie sich quantitative oder qualitativ nennt. Im Übrigen hat die These, dass es „kein methodisches Paradigma der qualitativen Forschung gibt, welches ein sinnvoll definierbares und in sich geschlossenes eigenständiges Konzept darstellt“ (J. Rost, 1998, S. 35), viel für sich. Und J. Rost beschließt seine Ausführungen mit der kernigen Feststellung: „Die Intentionen qualitativer Forscher sind sehr gut nachvollziehbar, die methodische Umsetzung ist oft fragwürdig, und die erkenntnistheoretische Position allzu oft unreflektiert“ (S. 41). Das Zitat ist inzwischen mehr als 20 Jahre alt. Aber es trifft auch heute noch ins Schwarze.



Literaturhinweise – qualitative Forschung:

Wenn viele Autoren an einem Werk beteiligt sind, ist es erwartungstreu, dass die Qualitäten der einzelnen Beiträge schwanken – von sehr gut bis gerade noch ausreichend. Das sollte man bei den weiter unten angeführten Sammelbänden beachten.

- An deutschsprachigen Einführungen ermangelt es nicht. Ich beschränke mich auf eine Auswahl: Cropley (2019, *Qualitative Forschungsmethoden. Eine praxisnahe Einführung*), Friebertshäuser und Seichter (2013, *Qualitative Forschungsmethoden in der Erziehungswissenschaft. Eine praxisorientierte Einführung*), Mayring (2016, *Einführung in die qualitative Sozialforschung*), Flick (2016, *Qualitative Sozialforschung. Eine Einführung*) und Kunz et al. (2012, *Qualitativ*

Forschen als Schlüsselqualifikation). – Auch Rosenthals (2015) *Interpretative Sozialforschung* ist zu sog. qualitativen Forschung zu zählen.

- Woods (2006, englisch), Wolcott (2008, englisch), Silverman (2017, englisch) und Yin (2016, englisch) wollen qualitativ arbeitende Forscher anregen, vernünftig zu schreiben und nehmen Forschungsnotizen an die Hand – vom Anfang bis zum Ende einer Arbeit.
- Umfassend informieren das von König und Zedler (1995) edierte zweibändige Werk *Bilanz qualitativer Forschung* und das *Handbuch qualitative Forschungsmethoden in der Erziehungswissenschaft* (Friebertshäuser et al., 2013). – Dazu passt das bewährte Lehrbuch von Lamnek und Krell (2016, *Qualitative Sozialforschung*).
- Umfangreich ist der von Flick et al. (2008) herausgegebene Wälzer *Qualitative Forschung*. Ich rate zu diesem dicken Taschenbuch. Bei seinem günstigen Preis-Leistungs-Verhältnis kann man auch mal einige weniger gelungene Beiträge verschmerzen. – Ein zweibändiges Handbuch haben May und Mruck (2020, *Qualitative Forschung in der Psychologie*) publiziert. – Ein entsprechendes Handbuch gibt es auch für die Sozialarbeit (Bock & Miethe, 2010, *Handbuch Qualitative Methoden in der Sozialen Arbeit*).
- Es lohnt sich immer noch, den zwar schon älteren, aber gut austarierten Artikel von Treumann zu lesen (1985, *Zum Verhältnis qualitativer und quantitativer Forschung. Mit einem methodischen Ausblick auf neuere Jugendstudien*). – Das ergänzt man dann mit dem Buchbeitrag *Qualitative vs. quantitative Forschung* von Wolf (1995). – Wer es etwas ausführlicher wissen möchte, schmökert im dünnen Bändchen *Quantitative und Qualitative Forschung im Vergleich* herum (Wichmann, 2019).
- Geballtes Wissen für Fortgeschrittene bieten zwei dicke Handbücher. Jeder, der qualitativ „qualitativ“ arbeiten will, sollte zumindest eines der beiden zur Hand haben. – Das von Denzin und Lincoln edierte, in der 5. Auflage (2018, englisch) vorliegende Buch zur qualitativen Forschung umfasst in 42 Kapiteln rund 1000 Seiten. – Das ähnlich geartete Handbuch von Leavy (2020, englisch) hat 40 Kapitel mit zusammen 1200 Seiten.
- Mey hat 2005 *Qualitative Entwicklungspsychologie* auf den Markt geworfen. Nicht wenige Beiträge darin sind von bescheidener Qualität. Dann hilft es auch nicht weiter, wenn man im Titel „qualitativ“ unterbringt. Dadurch werden schwache Beiträge auch nicht qualitätsvoller. – Ein guter Rat: Ein vernünftiges umfassendes Lehrbuch wählen, das sich der quantitativen Tradition der Entwicklungspsychologie verpflichtet fühlt. Da weiß man, was man hat. Es gibt viele. Exemplarisch nenne ich Siegler et al. (2021), *Entwicklungspsychologie im Kindes- und Jugendalter* (2021), Berg (2019), *Entwicklungspsychologie* sowie Schneider und Lindenberger (2018), *Entwicklungspsychologie*.
- Der Vollständigkeit halber erwähne ich die *Zeitschrift für Qualitative Forschung*. Ihre Nennung bedeutet nicht, dass ich sie regelmäßig lese oder zu ihrer regelmäßigen Lektüre rate. Reinschnuppern kann man ja mal. Dann aber mit einer besonders kritischen Lesehandlung, versteht sich.

Es gibt Forschungsfragen, zu deren Beantwortung unterschiedliche methodische Ansätze, zu meist qualitative und quantitative, miteinander kombiniert werden. In der einschlägigen Literatur steht für ein solches integratives Vorgehen *multimethodischer Ansatz*, *Methodenkombination*, *Methodentriangulation*, *Methodenmix* und was noch so alles an Begriffen herumgeistert. *Triangulation* (nicht zu verwechseln mit der soliden Messtechnik in der Geodäsie) ist dabei eine aufgeblasene Bezeichnung, hinter der sich nicht mehr verbirgt als die banale Erkenntnis, dass man gut beraten ist, in einem Forschungsprojekt möglichst mehrere Methoden bzw. Ansätze bzw. Vorgehensweisen bzw. Theorien bzw. Personengruppen bzw. Messzeitpunkte usw. miteinander

zu kombinieren, will man besonders aussagekräftige Resultate erhalten. Das ist weder eine neue Erkenntnis noch eine Erfindung der „qualitativen Forschung“.

Triangulation kann vielleicht eine Kombination von Verfahren sein (wie die Untersuchung der Persönlichkeit von jugendlichen Straftätern mit einem Fragebogen, einem projektiven Test und einem Interview). Das kann auch eine Kombination von Datenquellen bedeuten, beispielsweise die getrennte Befragung von Vätern, Müttern und Kindern bei der Analyse von Familien mit hochbegabten Kindern und Jugendlichen, wie es im Marburger Hochbegabtenprojekt realisiert wurde (Rost, 1993a, 2009a). Und das kann eine Kombination von Theorien meinen (wie die Verknüpfung behavioraler (= verhaltensbezogener), kognitiver und tiefenpsychologischer Ansätze bei der Erforschung von Ängstlichkeit). Oder es zielt auf eine Kombination von Disziplinen ab (z. B. gemeinsame soziologische, psychologische und erziehungswissenschaftliche Analysen der Resultate der eben erwähnten Schulleistungsvergleichsstudien TIMSS, PISA, IGLU). Oder ganz generell eine Kombination qualitativer und quantitativer Methoden.

Letztendlich sind das, wie man auf den ersten Blick erkennt, alles nur Selbstverständlichkeiten guter empirischer Forschung. Im eben erwähnten Marburger Hochbegabtenprojekt wurden bei diversen Fragestellungen nicht nur unterschiedliche Datenquellen genutzt (Väter, Mütter, Lehrer, Kinder), sondern auch unterschiedliche Erhebungsverfahren verwendet (Tests, Fragebogen, standardisierte Interviews). Hätten die Marburger Forscher Eindruck schinden wollen, hätten sie das als eine „Super-Triangulation“ (gewissermaßen als „trianguläre Triangulation“) bezeichnen können. Das haben sie – vernünftigerweise – nicht getan. Meinen manche „qualitativen“ Forscher, durch Begriffskauderwelsch, durch schwülstige Wortklingeleyen qualitätsarme Studien aufwerten zu können? Man kann einfache Sachen auch einfach und verständlich ausdrücken.



Literaturhinweise – Triangulation:

- Flick (2011) führt ausführlicher in die *Triangulation* ein. – Krüger und Pfaff (2008) informieren über *Triangulation quantitativer und qualitativer Zugänge in der Schulforschung*.
- Für Pädagogen bietet sich zur Vertiefung der Sammelband von Lüdemann und Otto (2019, *Triangulation und Mixed-Methods. Reflexionen theoretischer und forschungspraktischer Herausforderungen*) an. – Treumann (2005) zeigt in seinem Buchbeitrag *Clusteranalytische Auswahl prototypischer Fälle für qualitative Untersuchungen. Zur triangulativen Verknüpfung quantitativer und qualitativer Methoden* den Nutzen dieses Vorgehens am Beispiel des Medienhandelns Jugendlicher auf. – Ergänzend empfiehlt sich die Lektüre des Buchkapitels von Ganguin et al. (2017, *Methodentriangulation in der medienpädagogischen Forschung – Von agonalen Paradigmen zu einer methodologischen Synergie*).
- Zur Integration quantitativer und qualitativer Ansätze in der Psychologie und in den sozialen Wissenschaften liegen Lehrbücher vor von Todd et al. (2004, englisch) und Hall (2020, englisch). Hier werden nicht nur theoretische und methodische Ansätze diskutiert, es finden sich auch praktische Beispiele aus diversen Anwendungsgebieten. – Einen Kontrastierung dieser zwei Ansätze unternimmt Wichmann (2019) mit ihrem Buch *Quantitative und Qualitative Forschung im Vergleich*.
- Schon früh wurden Stimmen laut, die Schwierigkeiten, Probleme, Fallstricke und Missinterpretationen bei der Triangulation bzw. beim Methodenmix thematisierten (Sim & Sharp, 1998, englisch; Massey, 1999, englisch; Hammersley, 2008, englisch). – In diesem Zusammenhang leist man auch das Kapitel von Settineri (2015, *Forschst Du noch, oder triangulierst Du schon?*).

Die Chance, nach der „Wende“ auf die in den neuen Bundesländern frei werdenden bzw. neu eingerichteten Professuren für Pädagogik vermehrt empirisch arbeitende Forscher zu berufen, wurde damals ziemlich leichtfertig vertan. Erst die oben erwähnten großen internationalen Bildungsstudien konnten in der Öffentlichkeit, bei Politikern und Bildungsadministratoren sowie – man mag es kaum glauben – auch in den universitären Pädagogikinsti- tuten die Einsicht in die Notwendigkeit empirischer pädagogisch-psychologischer Forschung wecken. Es ist heute viel leichter als noch vor 50 Jahren, erfahrungswissenschaftliches Arbeiten dauerhaft in der Erziehungswissenschaft zu verankern. Was im Vergleich zur Psychologie an manchen Unis immer noch nicht Standard ist: eine angemessene, solide Ausbildung von Pädagogen in empirischer Forschungsmethodik, in Statistik und Versuchsplanung. Dies würde Lehrer und Pädagogikdo- zenten befähigen, erfahrungswissenschaftliche Arbeiten zu verstehen, gründlich zu analysieren und kritisch zu bewerten. Es muss sich ja nicht jeder Pädagoge Statistik und Forschungsmetho- dik selbst forschend betreiben.

Die Psychologie hatte es nach dem 2. Weltkrieg ebenfalls schwer, wieder empirisch-experimen- telles Arbeiten als selbstverständliche Forschungsstrategie zu etablieren. Erneut erstarkende geisteswissenschaftlich-verstehende Ansätze wurden aber schon in den 1960er Jahren zugunsten einer erfahrungswissenschaftlichen Ausrichtung überwunden. Dazu haben Peter R. Hofstätter mit seinem 1953 erstmals erschienenen Statistiklehrbuch *Quantitative Methoden der Psychologie* (spätere Auflagen hat er zusammen mit Dirk Wendt verfasst) und Heinrich Düker 1959 mit der Initiierung der *Tagung experimentell arbeitender Psychologen* (TEAP) entscheidende Meilensteine gelegt. Heute dominiert weltweit ein empirisches Verständnis von Psychologie: Das Verhalten (im weitesten Sinne verstanden, also sowohl beobachtbare Verhaltensweisen als auch innere Phä- nomene wie Denken, Fühlen, Erleben) von Menschen und Tieren wird mit den Methoden des Experiments, der Beobachtung, des Tests, des Fragebogens, des (standardisierten) Interviews etc. unter Zuhilfenahme quantitativer Methoden erforscht.

Traditionelle Unzulänglichkeiten setzen sich fort: Politiker stellen Fragen an die Bildungsforschung, auf die nur in Ausnahmefällen solide geantwortet werden kann, *weil pädagogisch-psychologische Projekte im Allgemeinen und größere Bildungsstudien im Besonderen nicht angelegt worden sind, ex post* (= im Nachhinein) *vorgebrachte Fragen adäquat zu beantworten*. Empirische Forschung hat nämlich nichts mit Hellsehen zu tun.

Ungeachtet dessen wurden und werden größere Bildungsstudien von Politikern und Journalisten (und leider auch von einigen Wissenschaftlern) so ungehemmt ausgebeutet, dass ihnen eigentlich die Schamröte ins Gesicht steigen sollte. Eine *ex post* Interpretation jagt die andere. Finnland hatte in der ersten PISA-Studie besonders gut abgeschnitten, Deutschland nicht. So weit, so schlecht. Politiker und Pädagogen reisten nach Finnland, um sich dort „gute“ Schulen und „guten“ Unterricht anzusehen (als gäbe es in Deutschland keine guten Schulen, keinen guten Unterricht).

Die professionellen Pädagogen, die Erziehungswissenschaftler und die Bildungspolitiker sind sich außerordentlich uneins, fragt man sie, was denn eine „gute“ Schule ausmacht. Deshalb suchte man Unterschiede zwischen Deutschland und Finnland, und je mehr man suchte, desto mehr fand man. Jeder machte das für die konstatierten Leistungsunterschiede verantwortlich, was ihm besonders gut in seine pädagogische bzw. politische Weltanschauung passte, was seine (Vor-)Urteile bestätigte. Egal ob schwarz, rot, grün oder gelb gefärbte Bildungspolitiker, sie alle kehrten ziemlich zufrieden vom Finnlandbesuch zurück, jeder fühlte sich in seiner (von den anderen abweichende) Meinung bestätigt, obwohl alle Ähnliches gesehen hatten.

Und es wurden mehr oder weniger abstruse Interpretationen gewagt, spekulative Schlussfolgerungen gezogen: Ändert die Einführung eines Zentralabiturs (von Kultusministerien als Konse-

quenz von PISA im Aufbau) bzw. das Überstülpen allgemeiner Bildungsstandards („am Ende des 10. Schuljahrs soll der Schüler einen Lebenslauf schreiben können“) – früher treffend Lehrpläne genannt – auch nur einen Deut an der offensichtlich gewordenen mangelnden Lesekompetenz bzw. an der Benachteiligung von Schülern unterer sozialer Schichten? Wird sie durch die Verkürzung des Gymnasiums auf acht Jahre besser? Natürlich nicht. Ursprünglich waren G-8-Klassen als Schulversuch für Hochbegabte gedacht, sie bezeichnete man als *D-Zug Klassen*. Dann wurden sie von den Kultusministern plötzlich zur pädagogischen Wunderwaffe gegen vieles, was bildungspolitisch falsch gelaufen war, erklärt. G-8 erwies sich aber als keine Wunderwaffe, sondern als Rohrkrepiere. G-8 war für mindestens 50 % der Schüler eine größere Belastung, viele sinnvolle Freizeitaktivitäten wurden gestoppt, viele Eltern protestierten.

G-8 war wohl auch nicht sonderlich förderlich für die Intelligenz. Der Verdacht erhärtet sich, dass die Verkürzung der Schulzeit das intellektuelle Wachstum unserer Jugendlichen bremst (Bergold et al., 2017, 2018). Jeder zusätzliche Schulmonat scheint, so legt es die Forschung nahe, die Intelligenz um etwa 0.3 bis 0.6 IQ-Punkte zu steigern (Rost, 2013, S. 444–452). Aufgrund des enormen Drucks der Öffentlichkeit sind die Kultusminister vieler Länder zu G-9 zurück gerudert und haben G-8 entweder ersatzlos entsorgt oder lassen Schulen und Eltern die Wahl zwischen G-8 und G-9. Lediglich Rheinland-Pfalz hatte in dieser Frage von Anfang an vernünftig reagiert und G-8 erst gar nicht eingeführt.

Leider beteiligen sich nicht wenige Erziehungswissenschaftler engagiert bis begeistert, gewissermaßen fröhlich pfeifend und singend, an der PISA-Kaffeersatzleserei im Nachhinein. Sogar manche Bildungsforscher (deren Namen will ich ausnahmsweise einmal verschweigen) verbiegen sich wie ein Schilfrohr im Wind und machen sich zum Büttel der Politik, weil sie an staatliche Knete wollen: Diese edlen Leute betonen vor allem, was die jeweiligen Politiker hören wollen. Unangenehme Botschaften werden in kleingedruckten Fußnoten entschärft, manchmal sogar ganz verschwiegen. Diese edlen Leute spielen Schwächen der Versuchsanlage herunter oder übergehen sie. Diese edlen Leute versuchen, Planungsproblem sowie Operationalisierungsschwierigkeiten (= Probleme der adäquaten Umsetzung von Konzepten in Messverfahren) mit Massenerhebungen und hochkomplexen statistischen Methoden, die nur eine Minderheit ihrer Kollegen versteht, zu überspielen. Warum das alles? Klar, diesen edlen Leuten geht es eigentlich nicht um Erkenntnisgewinn, um Wissenschaft. Es geht ihnen eigentlich nur um *money*. Man will den staatlichen Schotter abkassieren, um Kollegen im Wettbewerb um Drittmittelinwerbungen (so bezeichnet man heute diese besondere Art des Imponiergehaves) zu übertrumpfen. Man will dem Universitätspräsidium, welches nicht selten eine Universität mit einem gewinnorientierten Wirtschaftsunternehmen wie eine Bonbonfabrik verwechselt, zeigen, was man doch für ein toller Geldhamster ist. *Pecunia non olet* (= Geld stinkt nicht) oder, wie im Film *Cabaret* gesungen, *Money makes the world go round*.

Dazu passt, dass etliche Voraussetzungen der hochkomplexen Auswertungsmethoden nicht überprüft werden (und häufig auch nicht erfüllt sind), dass die Erfassung der zentral interessierenden Variablen manchmal naiv bis unzureichend ist. (Eine Variable ist ein Merkmal, das mehrere Ausprägungen annehmen kann, wie Schulensuren, Einkommen, Lesefähigkeit, Sozialstatus oder Intelligenz, siehe 2.5 und 3.12). Querschnittliche Daten werden lustvoll längs-schnittlich vergewaltigt, die Interpretation und Diskussion der Befunde sprechen aus politischen Gründen manche nahe liegende Alternativen nicht an. Aber auch bei besserer methodischer Anlage erlauben derartige Studien keinerlei Aussagen über die Ursache der konstatierten Leistungsdifferenzen zwischen den (Bundes-)Ländern. Das wird von fast allen Politikern leider übersehen, und leider vergessen das manchmal auch einige Urheber dieser Studien, wenn sie entsprechende Interviews geben oder Vorträge vor Nicht-Fachleuten halten.

Um es klar auszudrücken: *Aus PISA, IGLU und den anderen querschnittlichen Massenerhebungen lassen sich keine Ursache-Wirkungs-Zusammenhänge extrahieren. Und auch keine belastbaren didaktisch-methodischen Ratschläge für guten Unterricht.* Zur Klärung wären Längsschnittstudien hilfreich (siehe 3.11.2), besser noch gezielte randomisierte Experimente (siehe 3.10.2). Weil starke Designs organisatorisch nur schwer umzusetzen sind, wären unter Umständen auch sehr gut kontrollierte Quasi-Experimente hilfreich. Von „gut kontrolliert“ kann aber bei PISA & Co keine Rede sein. Das kümmert viele Bildungspolitiker wenig. Locker „entdecken“ sie eine Ursache nach der anderen, und manche Pädagogen und Erziehungswissenschaftler reihen sich ebenso locker ein. Schlimmer noch: „Bevor die Ergebnisse von PISA 2000 und Gründe über das ungünstige Abschneiden bei den Lesekompetenzen ... genauer untersucht worden waren, wussten einige Gruppierungen in der Politik und Öffentlichkeit, was zu tun ist“ (Becker & Schubert, 2006, S. 279).

Der Trend geht zur größeren Innovationsbereitschaft beim Formulieren von Forschungsproblemen und bei der Entwicklung adäquater Versuchspläne, und das ist gut so. Ein weniger pedantisches Beharren auf herkömmlichen Ansätzen und eine vermehrte Zuwendung zu solider erfahrungswissenschaftlich-quantitativer Methodik führen zu einem ständig zunehmenden Verständnis von Forschungsprozeduren. Solche Erfahrungen lassen für die Zukunft eine größere Fruchtbarkeit empirischer Forschung für Theorie und Praxis erhoffen, wenn die Möglichkeiten der empirischen Forschung realistisch eingeschätzt und die Grenzen der jeweiligen Versuchspläne und Untersuchungsanlagen (siehe 3.10) beachtet werden.

1.1 Funktionen der Forschung

Tag für Tag stehen wir vor dem Problem, Entscheidungen mit längerfristigen Konsequenzen zu treffen. Dabei dürfen wir die Ergebnisse wissenschaftlicher Untersuchungen, auch wenn sie noch so unvollkommen sind, nicht ignorieren. Gute, das heißt einfallsreiche und solide Forschung, kann uns bei unseren alltäglichen Überlegungen, was geschehen soll, helfen, indem ihre Resultate zum Verständnis der Beziehungen zwischen den komplexen Ereignissen in einer nicht leicht durchschaubaren Welt beitragen.

Wir müssen jedoch die spezifischen Begrenzungen jedes Forschungsprojektes kennen, um seinen Nutzen für Wissenschaft und Praxis abzuschätzen. Viele Erziehungswissenschaftler, Bildungspolitiker, Lehrer, Psychologen und Praktiker verlangen nach direkt umsetzbaren, also konkret verwertbaren Ergebnissen. Dieser Anspruch ist verständlich. Er ist aber nicht immer angemessen, weil theoriebezogene Forschung das Bindeglied von Konzepten zu praktischen Erfahrungen darstellt. Eine gute und empiriegestützte Theorie, also ein sinnvolles Geflecht von Hypothesen und Regeln, hilft, die Vielzahl von Fakten und Befunden zu ordnen, zu komprimieren, zu integrieren, zu erklären, Unwesentliches zu ignorieren und neue Zusammenhangshypothesen zu formulieren, die anschließend überprüft werden können. Ein bekanntes Bonmot drückt dies so aus: „Nichts ist praktischer als eine gute Theorie“ – und man sollte ergänzen: „vor allem dann, wenn sie kontinuierlich auf den Prüfstand der Realität gestellt wird!“

Die Wichtigkeit der empirischen Bewährung von Überlegungen des theoretisierenden Verstandes zeigt sich sehr schön an der bereits erwähnten PISA-Studie, die ich lediglich exemplarisch zitiere – viele andere Fälle könnten aufgeführt werden. Bei PISA nimmt die Erfassung der Lesekompetenz einen prominenten Platz ein. Lesekompetenz wird dort als Leseverständnis konzipiert und wie folgt umschrieben: „geschriebene Texte unterschiedlicher Art in ihren Aussagen, ihren Absichten und ihrer formalen Struktur zu verstehen und in einen größeren sinnstiftenden Zusammenhang

Viele Methodenbücher sind leider staubtrocken geschrieben und zudem noch voll von mathematischen Ableitungen und Formeln. Dieses Studienbuch ist bewusst anders. Statt Kaffeesatzleserei bietet es praxisrelevantes Wissen. Es zeigt außergewöhnlich verständlich, wie man empirische Untersuchungen liest, sachkundig interpretiert und kritisch bewertet. Es bedient sich dabei einer lockeren und alltagsnahen Sprache. Formeln kommen praktisch nicht vor. Grundlegende empirische Methoden und Konzepte kann man nämlich auch verstehen, wenn man kein Mathefreak oder Statistikguru ist.

Die über 100 Beispiele stammen aus dem konkreten Forschungsalltag. Der Blick auf Problemzonen empirischer Forschung wird durch mehr als 160 Fragen geschärft. Diese eignen sich zugleich gut als Leitlinien für eigene empirische Studien. Zu den einzelnen Themen gibt es außerdem kommentierte ein- und weiterführende Literaturhinweise.

Die 4. Auflage ist in allen Teilen umfassend überarbeitet und erweitert worden.

Dies ist ein utb-Band aus dem Verlag Klinkhardt. utb ist eine Kooperation von Verlagen mit einem gemeinsamen Ziel: Lehr- und Lernmedien für das erfolgreiche Studium zu veröffentlichen.

ISBN 978-3-8252-8789-4



9 783825 287894



QR-Code für mehr Infos und Bewertungen zu diesem Titel

utb.de