

Thomas JAHNKE, Potsdam

Zur Epistemologie der quantitativen ‚empirischen Bildungsforschung‘

Wie gewinnen die Erkenntnisse der im Titel genannten Forschung ihre Gültigkeit? Sind die benutzten Verfahren dem zu untersuchenden Bereich und Gegenstand angemessen, sind sie angebracht und sinnvoll? Diese Fragen zu beantworten, wäre natürlich eigentlich Aufgabe derer, die sich bestimmter mathematischer und insbesondere statistischer Methoden bedienen, um ihr Vorgehen und die Gültigkeit ihrer Ergebnisse zu rechtfertigen, und nicht derer, die sie befragen oder die Sinnhaftigkeit ihres Einsatzes in Zweifel ziehen.

Um dennoch einer Antwort auf die Spur zu kommen, betrachte ich die beiden Transsubstantiationen, die für die quantitative empirische Bildungsforschung charakteristisch sind: die von der begrifflichen Ebene in die der Daten oder Zahlen, die auch als Operationalisierung bezeichnet wird, und die von den aufbereiteten und mit diversen statistischen Verfahren be- und verarbeiteten Zahlen in die des Sinns, der Interpretation, der Deutung, der Formulierung von „Erkenntnissen“. Ich könnte diese beiden Metamorphosen auch in die gängigen grafischen Darstellungen des sogenannten Modellierungskreislaufs einzeichnen, also den Ort dieser Übergänge markieren, aber ich sehe davon ab, weil ich zum einen das Misstrauen hege, dass es sich bei diesen Darstellungen eher um – möglicherweise etwas aufgeblähte – didaktische Erfindungen handelt als um den Versuch, einen Prozess oder Vorgang erkenntnismäßig sinnvoll zu erfassen und zu strukturieren, zum anderen weil dieser Kreislauf inzwischen sinnwidrig herangezogen wird, um die Bearbeitung einfachster Sachaufgaben ad usum delfini zu beschreiben.

Auch wenn ich periodisch der Kollegenschelte geziehen werde, als sei mein Denken darauf gerichtet, das Lebenswerk dieses oder jenes Kollegen in schlechtes Licht zu rücken, statt es durch meine Kritik zu veredeln, will ich an dieser Stelle doch meiner Verwunderung Ausdruck geben, dass mancher der in der deutschsprachigen Mathematikdidaktik dem Modellieren – in den achtziger Jahren des letzten Jahrhunderts – Glanz und Inhalt verliehen hat, inzwischen in der sogenannten quantitativen empirischen Bildungsforschung tätig, statistische Verfahren nicht als Modellieren kritisch zu durchdenken gewillt ist, so als könnten diese Verfahren – von Kreislauf längst keine Rede mehr – der Realität ihre Erkenntnisse oder gar ihre Wahrheit einfach und faktisch abpressen. Genug der vermeintlichen Publikumsbeschimpfung.

Ich betrachte also einmal die Transsubstantiation, wobei ich die religiöse Konnotation dieses Begriffs nicht ungern in Kauf nehme, der begrifflichen, also sinnhaften Ebene in die der Daten, also die „Operationalisierung“, zum anderen die von den mit diversen statistischen Verfahren bearbeiteten Daten zurück in die Welt und deren angestrebtes Verstehen. Auf dazwischen liegende Vorgehen, das zuweilen unter der recht schlichten Bezeichnung „data processing“ subsummiert wird, werde ich an anderer Stelle ausführlich eingehen. Zu nachvollziehbaren Erkenntnissen kann die sogenannte quantitative empirische Bildungsforschung nur kommen, wenn diese drei Schritte gelingen.

Die Operationalisierung

Hier werden der Übersichtlichkeit und Handhabbarkeit halber zumeist wenige, theoretische Begriffe, prominentes Beispiel sind etwa die fachwissenschaftlichen, die pädagogischen und die fachdidaktischen Kenntnisse von Mathematiklehrkräften, transformiert in Skalenwerte, die sich dichotomisch aus der 0-1-Korrektheit ihrer Beantwortung von sogenannten Items in einem meist nicht allzu langen Fragebogen ergeben. Überprüfbar ist eine solche Operationalisierung nur, wenn die eingesetzten Items veröffentlicht werden; sinnvoll wäre sie nur dann, wenn die Items tatsächlich Auskunft über die betrachteten Kenntnisse geben könnten, diese sich eindimensional quantifizieren ließen und eine gewisse Konstanz aufwiesen oder zumindest näherungsweise als konstant gedacht werden könnten. Mit dieser Konstanz meine ich, dass die Befragten auch unter anderen Bedingungen und Umständen, selbst bei anderen Items und insbesondere auch ohne solche in etwa die gleichen Antwortwerte produzieren würden. Zudem müsste sichergestellt sein, dass die Stichprobe tatsächlich repräsentativ ausgewählt wurde und das Design der Untersuchung insgesamt keinerlei Einflüssen unterliegt, die die Ergebnisse aus welchen Gründen auch immer in eine gewisse Richtung einfärbten.

Der Operationalismus betreibt eine Sinnaustreibung bis zum letzten, dem tautologischen Nullpunkt: die Items müssen lupenrein, durch keinerlei Bias kontaminiert, mit dem übereinstimmen, was sie operationalisieren sollen. Die Validität der Untersuchung muss auf solche klinische Reinheit pochen. Die Kreuze etwa, die verschiedene Probanden setzen, müssen erkenntnisgleich und untereinander identisch sein oder betrachtet werden, sonst ließen sie sich auch nicht verrechnen, ob nun mit simplen oder elaborierten statistischen Verfahren. Ob eine Fünfzehnjährige bei einem PISA-Item liebevoll einen Miniaturglobus mit den Umrissen der Erdteile perspektivisch zeichnet oder ihr Altersgenosse, der an dem Item scheitert, weil er das Wort Hemisphäre nicht kennt, zwischen den verschiedenen Distraktoren anderer

Items beherzt rät, beiden ist in Sachen science literacy in der Summe schließlich die gleiche Kompetenz zu bescheinigen.

Alles ist – oder wird – Zahl

Zahlen sind die universelle Erkenntniswährung der quantitativen empirischen Bildungsforschung. Diese Universalität lässt – und das ist nun wirklich eine assoziative, aber vielleicht doch nicht zufällige Abschweifung – an die Analyse des Geldes im Kapitalismus von Karl Marx denken.

„Mit der ersten Entwicklung der Warenzirkulation selbst entwickelt sich die Notwendigkeit und die Leidenschaft, das Produkt der ersten Metamorphose [Ware in Geld; Th.J.] die verwandelte Gestalt der Ware oder ihre Goldpuppe festzuhalten. Aus bloßer Vermittlung des Stoffwechsels wird dieser Formwechsel zum Selbstzweck. ... Das Geld versteinert damit zum Schatz, und der Warenverkäufer wird zum Schatzbildner.“ (MEW 23/114)

Zurück zu anderen Schatzgräbern! Alles wird in Zahlen transformiert, ob es sich um die erhobenen Daten oder Werte selbst oder deren statistische Bearbeitung und Bewertung handelt, vergleichbar und damit gleich – geschichtslos und aller Umstände entkleidet. Bei PISA finden sich bekanntlich zum Stolz der Konstrukteure so verschiedene Entitäten wie etwa die Itemschwierigkeit und die mathematische Kompetenz der Schüler auf der gleichen Skala wieder. Mehr noch: ein namhafter Psychometriker versicherte jüngst auf einer Tagung, natürlich sei es sinnvoll zu sagen, ein Mensch sei größer, als er klug sei. Man müsse dazu nur untersuchen, wie viele Standardabweichungen seine Körpergröße vom Mittelwert derselben abweiche, und dann die gleiche Untersuchung für seinen Intelligenzquotienten anstellen. Mir scheint, solche Forschung ist nicht mehr weit davon entfernt, auch den Satz „Nachts ist es kälter als draußen“ wissenschaftlich zu belegen.

Sinngebung

Schließlich müssen die statistischen Resultate wieder in Worte umgewandelt werden, die eine Auskunft über den untersuchten Bereich geben. In der Sprache des Modellings geht es nun um die Rückkehr aus der Mathematik in den ‚Rest der Welt‘. Man muss dazu die oben als Sinnaustreibung bezeichnete Operationalisierung wieder rückgängig machen, den numerischen Werten einen Sinn anheften. So heißt es etwa bei PISA, die Kompetenzstufen seien dazu da, die ‚Skalen zum Sprechen zu bringen‘. Schon die Anzahl solcher Stufen (4 bei TIMSS; 5 bei PISA 2000; 6 bei PISA 2009) ist arbiträr, ihre oft hilflose Benennung (z.B. „intermediär“) und Abgrenzung

inhaltlich kaum oder nicht ohne Widersprüche nachzuvollziehen. Entweder es werden leere, nahezu tautologische Aussagen getroffen („Die mathematische Kompetenz von 12% der deutschen Schülerinnen und Schüler ist auf Stufe II.“) oder solche, die sich erkenntnismäßig aus dem Verfahren nicht ergeben. Wenn also schließlich das Hermēneuein ansteht, ohne das auch die empirische Forschung nicht auskommt, da Daten von sich aus nicht reden oder Auskünfte geben, dann ergibt sich entweder Selbstverständliches, was als nun endlich ‚wissenschaftlich‘ erhärtet oder nachgewiesen bezeichnet wird, oder offensichtliche Fragwürdigkeiten, die dann mit einigem Fußnotenaufwand weg zu retuschieren sind, oder blasse Aussagen, die sich kaum unter dem statistisch-technischen Vokabular hervortrauen – in vielen Fällen ohne es überhaupt nicht formulierbar sind. Ich würde das als das Methodendilemma bezeichnen: die Ergebnisse lassen sich von den Methoden nicht trennen, ohne dass der vermeintliche Exaktheit im Nachhinein zur Disposition stellt oder sogar mehr oder minder vorsätzlich negiert. Das scheint mir kaum hintergebar. Endete eine Untersuchung etwa mit dem Resultat, dass die alpha-Kenntnisse von Mathematiklehrpersonen mit ihren beta-Kenntnissen korrespondieren, wäre schon der Gebrauch des Wortes Kenntnis eigentlich übergriffig und unzulässig, aber auch dies zugestanden, könnte man der Sache noch keinen Sinn abgewinnen, so lange man nicht sorgsam erläutert, was unter alpha- und beta-Kenntnissen zu verstehen ist. Zahlen belegen bei solcher Forschung nicht Argumentationen und Worte, sondern Worte oder Begriffe, die nun plötzlich eingebracht werden, bemänteln die Zahlen. Die Globalisierung der Forschung tut ein Übriges, die Bedeutungslosigkeit der Sprache zu befördern. Wenn man z.B. Bildungssystemen, schon der Begriff deutet es an, parametrisiert, sie ihrer nationalen oder lokalen Traditionen aus Gründen der Vergleichbarkeit oder aus anderen ökonomischen oder psychometrischen Motiven entkernt, bleiben Zahlenwerke, die schon ausgeschlossen haben, was sie zu messen versprochen oder vorgaben, und doch Glanz und Härte naturwissenschaftlicher Erkenntnisse präbendieren. Auch statistische Sprechweisen suggerieren zuweilen Erkenntnisse, die sie nicht beinhalten. Wenn man etwa von der ‚aufgeklärten Varianz‘ spricht, dann wird der Statistiker wissen, dass es sich hier um einen wohldefinierten mathematischen Begriff handelt, dessen formale (!) Bedeutung seiner Definition und nicht seiner Bezeichnung zu entnehmen ist, während der Laie wähnt, dass hier eine Varianz ‚aufgeklärt‘ ist oder wurde, was der Experte natürlich gern billigend in Kauf nimmt.