

Geht's noch? Zur ifo-Studie über die Chancengerechtigkeit in den deutschen Bundesländern

Der zweite PISA-Schock von 2023 hat vielerorts tiefe Verwirrung ausgelöst. Mit einem Male war erkennbar geworden, dass Öffnung und Lockerung doch wohl nicht der richtige Weg sein konnten, denn gerade die Bundesländer in Deutschland mit der striktesten Differenzierung, Sachsen und Bayern, waren von dem allgemeinen Rückgang seit 2012 nicht nur nicht betroffen, sie hatten vielmehr immer weiter zugelegt und waren zum Schluss vor den Krisen 2018 an Schweden und sogar an Finnland davongezogen. Abgefallen waren vor allem die Bundesländer mit dem leichtesten Zugang zu den oberen Bildungswegen, Berlin und Bremen, aber auch Hessen und Nordrhein-Westfalen, auch Baden-Württemberg, das 2011/12 die Verbindlichkeit abgeschafft hatte. In den sozialen Chancen auf gute Leistungen gab es so dabei gut wie keine Unterschiede, ganz im Gegenteil, denn auch die lernschwächeren Kinder waren in stringenten Bundesländern deutlich besser als in den offenen. Das war vorher kaum einmal bemerkt worden. Umso größer war dann das Erstaunen, gerade auch in der Bildungsforschung, als das sichtbar wurde. Nun gibt es aktuell vom ifo-Institut eine Untersuchung („Ungleiche Bildungschancen: Ein Blick in die Bundesländer“, Wößmann et al. 2024), die, wie es aussieht, dieses Bild wieder zurechtrückt. Das Ergebnis lautet, kurzgefasst: In der Chancengleichheit für die Bildung sind Berlin und Brandenburg die Spitzenreiter und ausgerechnet Sachsen und Bayern die letzten. Und schon stehen alle zwischenzeitlichen Bemühungen, das Niveau doch mal etwas anzuheben, wie in Berlin oder Baden-Württemberg und beim Zugang auf das Gymnasium wieder strenger zu werden, wieder in Frage. Denn der Preis dafür kann ja wohl nicht ernsthaft die Wiederkehr eines ständestaatlichen Regimes sein, bei dem nur die besser gestellten nach oben kommen. Wenn es so wäre, gäbe es in der Tat ein erhebliches Problem: Gute Bildung unterläge einem konfliktreichen *Payoff*: Gute Leistungen wären nur mit sozialer Ungleichheit im Zugang zu den oberen Bildungswegen möglich. Und umgekehrt.

Ob das so ist oder ob es auch anders geht, soll hier anhand der ifo-Studie geklärt werden. Zunächst kommt die Frage, ob sich die Befunde dort begründen lassen, dann aber auch darum, was womöglich gefehlt hat oder übersehen wurde, nicht zuletzt über Hinweise auf theoretische Argumente und Untersuchungen, die in dieser Frage nach dem *Payoff* von sozialer Bildungsgleichheit und Leistungsniveau schon weitergekommen sind. Drei Punkte sollen behandelt werden: erstens die zentralen Befunde der Studie, zweitens die Frage an die Angemessenheit der Datengrundlage und der Vorgehensweise bei den Analysen und drittens die Befunde aus einer Untersuchung zu dem Problem, in der alle nötigen und in der in der ifo-Studie fehlenden Informationen berücksichtigt wurden.

1. Anlage der Untersuchung und Befunde

Die Untersuchung bezieht sich auf den Zusammenhang von Bildungsbeteiligung und sozialer Herkunft in den 16 deutschen Bundesländern, auf die soziale „Chancengleichheit“. Die Bildungsbeteiligung ist der Besuch eines Gymnasiums gegenüber allen anderen Zweigen. Die soziale Herkunft wird über den Gymnasialbesuch mindestens eines Elternteils und über das Einkommen gemessen. Es gibt zwei Maße der Chancengleichheit: relativ (Chancenverhältnis oben zu unten), und absolut (Chancendifferenz oben zu unten). Heraus kommt zunächst das bekannte Ergebnis, dass es überall in Deutschland Unterschiede in der Bildungsbeteiligung gibt, aber auch Unterschiede zwischen den Bundesländern. Der hervorstechendste Befund ist,

dass in Berlin und Brandenburg die relative Chancengleichheit in der Bildungsbeteiligung am größten ist und am geringsten in Bayern und Sachsen. Bei der absoluten Chancengleichheit wiederholt sich das Bild nur für Sachsen, das dabei den letzten Platz einnimmt, während Bayern den 8. Platz gleich nach Berlin mit dem 7. und Brandenburg mit dem 6. Platz. Zur Erklärung der Unterschiede werden verschiedene Merkmale der Bundesländer untersucht: durchschnittlicher Gymnasialbesuch, durchschnittlicher sozio-ökonomischer Hintergrund, durchschnittlicher Migrationshintergrund, Bruttoinlandsprodukt, Ausgaben für die Bildung, Stadtstaaten (Berlin, Bremen, Hamburg), Ressourcenausstattung der Schulen, Klassengrößen und, Schüler-Lehrer-Verhältnisse. Keines der Merkmale trägt zur Erklärung der Befunde bei, wonach Berlin und Brandenburg die höchste Chancengleichheit in der Bildungsbeteiligung aufweisen und insoweit das beste Bildungssystem haben, und Bayern und Sachsen das schlechtesten und somit eines der Kernziele jeder Bildungspolitik, die soziale Chancengleichheit, deutlich verfehlen. Daraus wird im weiteren Text, neben einer Reihe von ohne Zweifel wichtigen Verbesserungen und Kompensationen bei den lernschwachen und sozial benachteiligten Kindern, auch die Öffnung und Lockerung des Zugangs zum Gymnasium empfohlen, insbesondere auch die Zusammenlegung der unteren Bildungswege zu „Gemeinschaftsschulen“ bis hin zu Integrierten Gesamtschulen.

2. Die Bewertung der Untersuchung

Der zweite Punkt, die Bewertung der Untersuchung, bezieht sich auf drei spezielle Aspekte: der Mikrozensus als Grundlage einer Untersuchung für die Bestimmung von Effekten der sozialen Herkunft beim Übergang auf das Gymnasium, eine in der Studie zusätzlich herangezogene Untersuchung von Wößmann (2010) zu den Effekten des „längeren gemeinsamen Lernens“ in zwei Bundesländern, Berlin und Brandenburg, und einige Empfehlungen, die es auch schon in einer Übersicht über den Rückgang des Leistungsniveaus seit 2010, gab, dem auch so genannten „traurigen Smiley“.

2.1 Der Mikrozensus als Datengrundlage

Der Mikrozensus bietet sich für Analyse vor allem wegen der Breite der Datengrundlage und der damit hohen Fallzahlen an. Dazu enthält er auch die nötigen Informationen zum sozio-ökonomischen Status der Familien sowie zum Gymnasialbesuch bzw. der Bildungsbeteiligung, auch eine Zuordnung zu den jeweiligen Bundesländern. Weitere Eigenschaften der Bundesländer werden entweder daraus gebildet oder anderen Datenquellen entnommen. Die Stärke des Mikrozensus ist auch die Vollständigkeit, mit der die soziale Ungleichheit abgedeckt wird, nach oben wie nach unten, und es deshalb die Verzerrungen nicht gibt, die ansonsten leicht auftreten. Weitere Informationen gibt es allerdings nicht, etwa alles, was aus den theoretischen Modellen zur Erklärung der Bildungsbeteiligung erforderlich wäre: die zu den primären Effekten der kognitiven Fähigkeiten bzw. der Leistungen vorher, die zu den sekundären Effekten bei den Aspirationen und den Bildungsentscheidungen der Eltern und die zu den tertiären Effekten bei den Bewertungen der Leistungen der Kinder bei der Aufteilung in die Bildungswege durch die Schule, die Noten und insbesondere die Empfehlungen. Ohne deren Berücksichtigung kommt es unvermeidlich zur Überschätzung der Herkunftseffekte und ggf. auch zu deutlichen Verzerrungen (vgl. dazu ausführlich Marks 2014). Dieser „omitted variable bias“ plagt schon die PISA-Studien von Anfang an und ist durch keine noch so trickreiche statistische Simulation oder Kontrolle von „unbeobachteter Heterogenität“ zu beseitigen.

Die in der ifo-Studie berichteten Effekte und Unterschiede bilden also nur den „totalen“ Effekt ab, den man bekommt, wenn sonst nichts einbezogen wird, was die Ungleichheit erklären

könnte, insbesondere die nach der sozialen Herkunft unterschiedliche kognitive Entwicklung in den Familien oder ihre Aspirationen und Einschätzung der Optionen. Es könnte ja auch sein, dass die Gymnasien in Berlin so schlecht sind, dass speziell die ambitionierten Akademiker ihre Kinder davon fernhalten: Umzug ins Umland, selektive Schulwahl dahin, wo es besser zu sein scheint, Privatschule. Auch das würde die statistische „Chancengleichheit“ befördern, womöglich aber auch die Bewertungen der Lehrer nach gewissen Einstellungen. Die Befunde bilden damit allenfalls eine Obergrenze für die Chancenungleichheit. Es ist kein seltener Fehler, der dann gerne hingenommen wird, wenn die Daten fehlen, um es genauer zu untersuchen, und wenn der Befund einfach ins Konzept „passt“.

2.2 Berlin und Brandenburg

In der Studie wird eine dazu analoge Untersuchung von Wößmann (2010) aufgegriffen, nun zu den *Leistungen*. Dabei kommt für Berlin und Brandenburg das gleiche Ergebnis heraus wie in der ifo-Studie für die Bildungsbeteiligung: Die beiden Länder haben den geringsten „sozialen Gradienten“. Diesmal waren es die Daten von PISA 2003. Man hätte dafür auch die Individualdaten aus der PISA-Sondererhebung für Deutschland haben können, aber das damalige PISA-Konsortium hat sie unter Verschluss gehalten. Deshalb war man auf die Analyse der *Aggregatdaten* auf der Ebene der Bundesländer angewiesen, also nur die Mittelwerte. Insofern wäre der Mikrozensus damals schon ein Fortschritt gewesen, weil sich damit die Unterschiede zwischen den Bundesländern bei den individuellen Zusammenhängen bestimmen lassen, auf die es allein ankommt. Auch hier werden weitere Eigenschaften der Bundesländer kontrolliert, die denen aus der ifo-Studie entsprechen. Aber auch nun fehlen wie im Mikrozensus die wichtigsten Bedingungen zur Bestimmung der (System-)Effekte, selbst auf der Aggregatebene: die kognitiven Fähigkeiten und die Leistungen vorher für die so wichtige Kontrolle der Selektivität in die Schulklassen der Sekundarstufe, und nun auch die so wichtige soziale und kognitive Lernumgebung in den Schulklassen. Es sind methodisch die gleichen Probleme wie mit dem Mikrozensus, inhaltlich nur jetzt auf die Besonderheiten des Lernens und der Leistungen in der Sekundarstufe bezogen. Die Folgen sind ganz analog: die Überschätzung der Herkunftseffekte, womöglich auch starke Verzerrungen und nur der „totale“ Effekt als „Obergrenze“. Hinzu kommt in dieser Analyse ein besonders gravierendes Problem noch: Man kann Aggregat-Effekte nicht ohne weiteres als Effekte und Zusammenhänge auf der individuellen Ebene interpretieren. Dazu müssen einige zusätzliche Annahmen gemacht werden, etwa die, dass es (nach dem sog. Kovarianztheorem für Mehrebenen-Analysen wie hier) *keine* Interaktionseffekte zwischen den beiden Ebenen, Individuen und System bzw. Bundesland, geben darf. Aber gerade das soll es ja geben und ggf. gefunden oder widerlegt werden: die *Veränderung* der Effekte der sozialen Herkunft als *Systemeffekt*. Es war das große Manko vieler der frühen Untersuchungen mit den PISA-Daten, etwa schon bei Hanushek und Wößmann (2006), aber auch danach.

Kurz: Auch diese Studie war eigentlich unbrauchbar. Sie widersprach schon damals dem, was man aus den frühen PISA-Sonderauswertungen für die Bundesländer und den IQB-Berichten wusste: Berlin gehörte war Vorletzter bei den Leistungen und letzter in der sozialen Durchlässigkeit bei den Leistungen, und Brandenburg unauffällig in beidem im unteren Mittelfeld. Und Bayern und Sachsen (und zuvor Baden-Württemberg) gerade bei den Kindern im unteren Leistungsbereich deutlich besser als alle anderen, erst recht als in Berlin und Brandenburg.

2.3 Die Empfehlungen

Die ifo-Studie schließt mit einer Reihe von Empfehlungen um das Problem der Chancenungleichheit bei der Bildung zu lösen (S. 8ff.). Es sind zwei Arten: Empfehlungen, die allgemein und unabhängig vom System ohne Zweifel von hohem Nutzen wären und solche die das System betreffen und wo man sich das genauer ansehen muss.

Die *allgemeinen* Empfehlungen beziehen sich auf frühkindliche Bildungsangebote, die besten Lehrkräfte dahin, wo es besonders brennt, Nachhilfeprogramme für benachteiligte Kinder und ein Mentoring für die Navigation durch das Bildungssystem bei den bildungsferneren Schichten. Das würde gewiss manche Schwierigkeiten mindestens mildern helfen.

Auf der *Systemebene* werden die Verschiebung der Aufteilung auf die sechste Klasse (wie in Berlin und Brandenburg) bzw. die Zusammenlegung der unteren Bildungswege zur Zweigliedrigkeit empfohlen. Das wiederum wäre wieder genauer anzusehen (s. unten noch dazu). Denn es gibt auch Hinweise darauf, dass das so eindeutig nicht ist wie es inauguriert wird. Auffällig ist vielmehr etwas anderes: gerade jene institutionellen Regelungen, von denen inzwischen bekannt ist wie wichtig sie für die Erklärung der Unterschiede in Bildungsbeteiligung und Leistungen zwischen den Bundesländern sind, fehlen in den Empfehlungen bzw. bei der Bewertung der Befunde und in den Analysen ohnehin: die Verbindlichkeit Gymnasialempfehlung und die Kontrolle der schulischen Organisation darauf, ob auch wirklich darauf geachtet wird, dass alle richtig und gerecht zugeht.

3. Und so sieht's aus!

Die deskriptiven Befunde der ifo-Studie sind, auch wenn es viele Fragen gibt und es sich nur um totale Effekte handelt, durchaus bedeutsam. Sie verweisen darauf, dass es offenbar Probleme in der sach- und sozialgerechten Passung der Differenzierung geben kann, und allein das veranlasst, das Problem ernst zu nehmen. Denn: Das Konzept des Ability-Tracking *beruht* auf der Annahme, dass es bei der Aufteilung der Bildungswege „wirklich“ und „nur“ nach den „abilities“ geht, wenn es die Effekte geben soll, die davon erwartet werden: weder zu offen, noch zu restriktiv und vor allem nicht sozial selektiv. Es ist das Problem des *Payoff*: Mit jeder Abweichung durch eine Fehlplatzierung nach den kognitiven Fähigkeiten, nach oben wie nach unten, leidet das Potential an Leistungsmöglichkeiten und es steigt die Gefahr von auch leistungsschädlichen sozialen Selektivitäten.

Das konnte in der ifo-Studie nicht untersucht werden, ebenso wenig in früheren Versuchen mit den PISA-Daten, auch lange nicht für die Daten der IQB-Berichte und deren ebenfalls nur deskriptiven Darstellungen, und auch nicht für die diversen Bildungsberichte und Bildungsmonitore seit langer Zeit. Denn in allen diesen Fällen fehlen die wichtigsten Bedingungen und Vorgänge zur Erklärung des Geschehens: die Informationen über die primären, die sekundären und die tertiären Effekte für Bildungsbeteiligung und Leistungen, die man braucht, um den Nachweis zu führen, dass sich die Unterschiede auch als „kausale“ Beziehung begründen lassen und nicht nur eine „Korrelation“ der Beschreibung eines „totalen“ Effekts ohne weitere Kontrollen bleiben. *Jede* praktische Maßnahme hängt daran, auch die, die man für selbstverständlich halten könnte (wie Sprachtests, Mentoring, deren Effekte auch erst einmal zu belegen wären).

Wir gehen deshalb in aller Kürze das durch, was man dazu inzwischen weiß, insbesondere nach Auswertungen mit Daten aus der „National Educational Panel Study“ (NEPS), die einmal

gestartet wurde, um die Unzulänglichkeiten der bis dahin verfügbaren Datengrundlagen, wenigstens für Deutschland und die Bundesländer, zu auszuräumen und zu verlässlicheren Ergebnissen zu kommen als man sie bis dahin erwarten konnte (vgl. die entsprechenden Kapitel und Abschnitte bei Esser 2021, 2023). Zentral ist dabei das Konzept der „Stringenz“: Die Kombination von Verbindlichkeit und Kontrolle, über die die primären Effekte der kognitiven Entwicklung gestärkt, aber die sekundären Effekte der freien Wahl der Eltern und die tertiären Effekte der Fehleinstufungen durch die Schule begrenzt oder ganz aufgehoben werden können. Nach diesem Konzept gehören Bayern, Sachsen und (bis 2012, auf das sich die Analysen beziehen) Baden-Württemberg zu den Ländern mit der höchsten Stringenz (sowohl Verbindlichkeit wie Kontrolle), Bremen, Berlin, Hessen und Nordrhein-Westfalen zu denen mit der stärksten Öffnung und Lockerung (keine Verbindlichkeit, keine Kontrolle), alle anderen dazwischen. Zuerst nun die Befunde zur Bildungsbeteiligung, dann für die Leistungen, und daran anknüpfend dann auch für die Empfehlung zur Umstellung auf die Zweigliedrigkeit und deren Effekten.

3.1 Bildungsbeteiligung

Zur Bildungsbeteiligung in den deutschen Bundesländern hat es seit 2010 etwa eine ganze Reihe spezieller Analysen zu den Effekten der Verbindlichkeit gegeben. Davon gibt es eine, die eine Verringerung der Herkunftseffekte gefunden hat, eine eine Verstärkung, alle anderen dagegen fanden keinen Effekt. Das Problem bei allen diesen Studien war, dass auch sie Lücken bei den zentralen Einflussgrößen auf die Erklärung der Bildungsbeteiligung hatten, seien es die kognitiven Fähigkeiten, die Leistungen vorher oder die Empfehlungen. Einige Untersuchungen benutzten auch den Mikrozensus, dann aber wenigstens mit Blick auf Reformen, Abschaffung oder Einführung der Verbindlichkeit. In keiner der Studien wurden auch die Aspirationen der Familien gemessen, dem wohl wichtigsten Faktor, der den Ehrgeiz der Eltern in den oberen Schichten erklärt.

Nimmt man alles das zusammen, dann zeigt sich ein interessantes Bild: Es gibt nach Kontrolle aller relevanten Bedingungen sonst in der Tat keinen Herkunftseffekt, wohl aber eine stärkere Kopplung des Übergangs auf das Gymnasium an die Leistungen vorher in der Grundschule. Es ist eine deutliche Bestätigung dafür, dass sich stringente Regelungen der Differenzierung, wie in Bayern, Sachsen und (damals) Baden-Württemberg deutlich auf die „Meritokratie“ des Bildungswesens auswirkt und speziell die Leitungsgerechtigkeit fördert, ohne die soziale Chancenungleichheit zu verstärken. Also nachgerade das Gegenteil dessen, was die ifo-Studie ergibt (und verschiedene andere auf der Basis von PISA-Daten mit den gleichen Problemen wie beim Mikrozensus).

3.2 Leistungen

Bei den Leistungen in der Sekundarstufe sieht es ganz ähnlich aus. Hier wurden ebenfalls die für das schulische Lernen relevanten Bedingungen und Vorgänge so gut wie vollständig berücksichtigt, die es weder in den PISA-Daten, noch denen des Mikrozensus gibt: die kognitiven Fähigkeiten der Kinder, ihre Leistungen am Ende der Grundschule und insbesondere auch die Effekte der sozialen und kognitiven Zusammensetzung der Schulklassen. Nach PISA und auch der ifo-Studie hätten die stringenten Regelungen zu keinen Leistungssteigerungen geführt, sondern nur die Herkunftseffekte verstärkt, also auch nun die soziale Chancenungleichheit vergrößert, und die besten Leistungen mit der geringsten sozialen Chancenungleichheit gäbe es in den offenen Systemen bzw. denen mit einem „längeren gemeinsamen Lernen“. Das aber ist gerade nicht der Fall, ganz im Gegenteil: Das Leistungsniveau ist in den stringenten Ländern von Beginn an höher, auch weil sich die Kinder

schon vor dem Übergang besonders anstrengen, und die soziale Ungleichheit wird über den meritokratischen Effekt der Stringenz tendenziell sogar abgeschwächt.

Hinzu kommt ein Effekt, den man kaum hätte erwarten können: Je homogener die kognitive Zusammensetzung der Schulklassen ist, umso besser sind die Leistungen, und das insbesondere bei den Kindern in den unteren Bereichen der kognitiven Fähigkeiten. Das widerspricht wiederum den Befunden der Analyse von Wößmann (2010) für die Leistungen, wonach Berlin und Brandenburg eine höhere soziale Durchlässigkeit für gute Leistungen hätten als alle anderen. Kurz: Mit den nötigen Daten und angemessenen Analysen dreht sich das Bild, das die ifo-Studie, auch insgesamt zu Bildungsbeteiligung und Leistungen, für die soziale Chancenungleichheit zeichnet, geradezu in das Gegenteil.

3.3 Zweigliedrigkeit

Das gilt auch für den Vorschlag in den Empfehlungen am Schluss der ifo-Studie, doch die Zweigliedrigkeit einzuführen, in der sich die Lage verbessere. In der ifo-Studie werden zwei Beiträge benannt: Ammermüller (2013) und Piopiunik (2021). Die Untersuchung von Ammermüller hat die Zweigliedrigkeit nicht zum Gegenstand und bezieht sich auf PISA-Daten, die ohnehin unzureichend wären. Bei Piopiunik sind es Daten der NEPS, aber es gibt allenfalls einen Effekt für das Lesen, für Mathematik und Sachkunde dagegen nicht, und für die Erklärung der Leistungen so wichtigen kognitiven Fähigkeiten und die Leistungen vorher werden nicht berücksichtigt, Schuleffekte auch nicht. Es gibt noch zwei weitere Studien zur Zusammenlegung: (Piopiunik 2014) und (Matthewes 2021), die eine noch mit PISA-Daten und ihren Problemen, die andere mit NEPS-Daten schon und allen Informationen zur Selektivität. Beide finden positive Effekte der Zusammenlegung in den unteren Bildungswegen. Die fehlgeschlagenen Versuche in Berlin und Bremen durch die Zweigliedrigkeit die Probleme dort zu lösen werden nicht erwähnt. Also: „mixed results“.

Die Anzahl der Optionen ist nun nicht alles von Bedeutung bei der Regelung der Differenzierung, auch nicht das Alter bei der ersten Aufteilung. Es muss für *alle* diese Varianten auch *wirklich* ein „Ability“-Tracking sein und eben keines, das alles offen lässt. Keine der Untersuchungen berücksichtigt das, die ifo-Studie schon gar nicht, die nur darauf sieht, wer wo auf das Gymnasium kommt. Tut man das, zeigt sich ein anderes Bild: Die Stringenz als Kombination aus Verbindlichkeit und Kontrolle überlagert alle Variationen in der Anzahl der Optionen von Zwei- und Dreigliedrigkeit, auch die des „längeren gemeinsamen Lernens“ bis zur 6. Klasse deutlich und insbesondere auch das, was in „Integrierten Gesamtschulen“ mit ihren internen Differenzierungen passiert. Den ganz eigenen Effekt der besonderen Anstrengung für gute Leistungen *vor* dem Übergang, über den das Leistungsniveau schon in der *Grundschule* steigt, gibt es bei der Zweigliedrigkeit nicht mehr, ebenso wie bei der freien Wahl von Schulen und Bildungswegen. Die Dreigliedrigkeit fungiert offenbar als weitere Verstärkung der Anreize für besondere Anstrengungen auch schon vorher, das aber auch nur bei Verbindlichkeit und Kontrolle. Die Stringenz überlagert mit ihren „meritokratischen“ Folgen bei Bildungsbeteiligung und Leistungen, wie es aussieht, durchaus mögliche positive Effekte der Zweigliedrigkeit oder des „längeren gemeinsamen Lernens“ deutlich. Die Anzahl der Optionen und das Alter bei der ersten Aufteilung sind, im Vergleich, allenfalls Nebensächlichkeiten. Viel bedeutsamer wäre es, die Differenzierung so zu gestalten, dass die Fähigkeiten und Leistungen (wieder) im Mittelpunkt stehen. In der ifo-Studie ist davon, wie auch sonst oft genug, nichts zu sehen, weder in den Analysen, noch in den Empfehlungen.

4. Fazit

Die ifo-Studie hat viele aufgeschreckt und irritiert, auch Empörung hervorgerufen: Berlin an erster Stelle und Bayern als Letzter im Bildungsranking! Nicht zu glauben! Alles nur Ideologie! hier, Erleichterung und „endlich sagt’s mal einer!“ dort. Und in der Tat: Die Befunde sind, wenigstens in dieser Deutlichkeit, überraschend und provozierend. Und offenbar durchaus auch populistisch beabsichtigt, wenn man in der Fußnote auf der ersten Seite liest, dass BILD das Projekt „finanziell unterstützt hat“, ganz ähnlich wie 2023 schon einmal beim „ifo- Ein Herz für Kinder“-Chancenmonitor“, der auch so als reine bivariate Korrelation aufgebaut war und auch schon kräftig dramatisierend in die Irre geführt hatte. Alles das aber zählt in der Wissenschaft alles nicht, und gerade auch manche Reaktion der Verbände und mancher „PISA-Versteher“, so oder so, zeigen, dass es nicht auf die Passung mit bestimmten Voreinstellungen ankommt, sondern *allein* auf die wissenschaftlich-methodische Haltbarkeit der Studie. Aber da fällt das Fazit nicht schwer: *sub omnibus canonicus!*

Hier noch einmal kurz die Punkte dazu: Der Mikrozensus ist, ebenso wie PISA vorher, ungeeignet zur Analyse der gestellten Fragen, die Befunde geben allenfalls einen „totalen“, rein deskriptiven Zusammenhang wieder, der womöglich völlig anders ausgesehen hätte, wenn man das tut, was man im Methodenstudium der Sozialwissenschaften zuerst lernt: „Kontrolle auf Drittvariablen“. Die Befunde für den Mikrozensus beziehen sich allein auf die Bildungsbeteiligung, also den Zugang zu Zertifikaten. Die aber sind wenig wert, wenn nichts dahintersteht: die Leistungen. Und da sieht es, nach allem, was man etwa schon aus den IQB-Berichten weiß, ganz anders aus: Bayern (und Sachsen) ganz oben, Berlin (und Bremen) ganz unten. Untersuchungen zu den Ländervergleichen bei den Leistungen mit den PISA-Daten, die auch herangezogen werden und Berlin und Brandenburg als die Besten finden, unterliegen den gleichen Fehlern und Verzerrungen wie alle Beiträge mit PISA-Daten vorher: grob irreführend, damals schon. Und der wohl wichtigste Systemaspekt zur Erklärung von Leistungsgerechtigkeit, Effizienz und sozialer Chancengleichheit, die Stringenz in der Organisation der Differenzierung, scheint unbekannt zu sein. Man findet dazu weder etwas in den konzeptionellen Überlegungen, noch in den Analysen, auch nicht in den Empfehlungen am Ende des Beitrags, von denen die allermeisten problemlos auch in das Konzept des stringenten Ability-Tracking passen bzw. die Effizienz der Maßnahmen, wie Mentoring oder Sprachförderung, noch weiter steigern würde.

Vielleicht aber auch noch das: Gerade die *ökonomische* Bildungsforschung hat sich, wengleich mit, wie man heute weiß, unzureichenden Mitteln, schon früh und viel analytischem Sachverstand an die Fragen der Effekte von Bildungssystemen gemacht. Das war ein großes Verdienst. Aber sie hat die Kurve offenbar nicht gekriegt als sich so nach und nach herausstellte, dass es so nicht geht: PISA-Daten ohne die nötigen Konstrukte zur Erklärung von Leistungen und Bildungsentscheidungen, Aggregatdaten-Analysen ohne Berücksichtigung der Fallen von Mehrebenen-Fehlschlüssen, wenn wichtige Kontrollen auf der Mikroebene fehlen, und das unbedachte Festhalten an naiven und theoretisch allenfalls ad-hoc begründeten Vorstellungen der Effekte bestimmter Regelungen der Differenzierung wie das Alter bei der ersten Aufteilung oder die Anzahl der Optionen. Das aber ist besonders irritierend gewesen: Die besondere Stärke der Ökonomie war immer die explizite *theoretische* Modellierung – bevor es daran gehen konnte, die nun erforderlichen Operationalisierungen und empirischen Analysen vorzunehmen. Das gibt es dort so gut wie gar nicht (mehr). Was man sieht sind so gut wie allein von den verfügbaren Daten getriebene Analysen, die sich auf methodische Finessen konzentriert haben, deren Wert unbestritten bleiben kann, dabei aber übersehen haben, dass die wohl entscheidenden Aspekte fehlen, um die Effekte des Ability-Tracking zu bestimmen: Die „Abilities“ und die Schuleffekte aus Unterricht und peer-Interaktionen in Klassen unterschiedlicher Zusammensetzungen nach Niveau und Homogenität darin.

Das *konnte* nicht gut gehen, und die ifo-Studie muss allein deshalb so beunruhigen, weil sie nicht irgendwo wieder in einem Doktorandenprogramm für empirische Bildungsforschung entstanden ist, wo es nicht so auffällt, sondern am ifo-Institut, dem so renommierten, im ifo-Schnelldienst freilich.

Literatur

Ammermüller, A. 2013. Institutional Features of Schooling Systems and Educational Inequality: Cross Country Evidence from PIRLS and PISAe. *German Economic Review*, 14: 190-213.

Esser, H. (2021). „Wie kaum in einem anderen Land?“ Die Differenzierung der Bildungswege und ihre Wirkung auf Bildungserfolg, -ungleichheit und -gerechtigkeit. Band 1: Theoretische Grundlagen. Frankfurt/M. und New York: Campus.

Esser, H. (2023). „Wie kaum in einem anderen Land?“ Die Differenzierung der Bildungswege und ihre Wirkung auf Bildungserfolg, -ungleichheit und -gerechtigkeit. Band 2: Empirische Befunde. Frankfurt/M. und New York: Campus.

Hanushek, E.A., und L. Wößmann. 2006. Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. In: *The Economic Journal*, 116: C63-C76.

Marks, G. N. 2014. Education, social background and cognitive ability. The decline of the social. London: Routledge.

Matthewes, S. H., 2021. Better Together? Heterogeneous Effects of Tracking on Student Achievement. *The Economic Journal*, 131: 1269–1307.

Piopiunik, M. 2014. The effects of early tracking on student performance: Evidence from a school reform in Bavaria. *Economics of Education Review* 42, 12–33.

Piopiunik, M. 2021. How Does Reducing the Intensity of Tracking Affect Student Achievement? Evidence From German State Reforms. CESifo Working Paper 9214.

Wößmann, L. 2010. Institutional Determinants of School Efficiency and Equity: German States as Microcosm for OECD Countries. In: *Jahrbücher für Nationalökonomie und Statistik*, 230: 234–270.

Wößmann, L., F. Schoner und F. Pfaehler. 2024. Ungleiche Bildungschancen: Ein Blick in die Bundesländer. *Ifo Schnelldienst* 5, 2024.