

Matthias von Davier, Eugenio Gonzales, Irwin Kirsch, Kentaro Yamamoto, Editors. 2013. The Role of International Large-Scale Assessments: Perspectives from Technology, Economy, and Educational Research. Dordrecht-Heidelberg-London-New York: Springer.

Erweiterter Review Essay für ‚Magazin erwachsenenbildung.at‘ Nr.23

von Lorenz Lassnigg (lassnigg@ihs.ac.at), Juli 2014

Im Internet unter www.equi.at/dateien/rez-meb23.pdf

Das Buch ist Produkt einer Konferenz im März 2011, die von Educational Testing Service (<https://www.ets.org/>) veranstaltet wurde. ETS ist eine US-basierten international tätige non-profit Organisation, die mit weltweit 2.500 MitarbeiterInnen in 9 Standorten ca. 30 Tests (darunter z.B. den bekannten TOEFL- Test of English as a Foreign Language TOEFL) entwickelt und durchführt, aber auch zur Forschung beiträgt. Die Herausgeber sind in ETS Princeton tätig. Trotz der seit der Konferenz bereits vergangenen Zeitspanne, handelt es sich um ein vielseitiges und sehr informatives Buch, das nicht nur die Vorteile, sondern auch einige Probleme mit den internationalen Large-Scale Assessments (LSAs) sehr qualifiziert diskutiert. Die Probleme haben ja mittlerweile, ausgehend von kritischen US-AkademikerInnen, bereits zu einer politischen Gegenbewegung geführt, die v.a. den politischen Ansatz, aber auch bestimmte pädagogische Folgescheinungen von PISA fundamental kritisieren und Gegenvorschläge machen.¹

Die stark männlich dominierten Beiträge zum Buch (Herausgeber 100%, Co-AutorInnen 80% männlich) gehen grundsätzlich von der positiven Seite aus, und beleuchten sowohl Fragen der Bedeutung und der weiteren Entwicklung dieser Test-Praktiken, als auch politische, ökonomische und pädagogische bzw. erziehungspolitische Themen. Die generelle Botschaft des Buches kann dahingehend zusammengefasst werden, dass die LSAs erstens eine Entwicklung von wachsender Bedeutung nicht nur für die Politik, sondern auch für die Forschung sind, die zweitens bereits wichtige Fortschritte gebracht hat, aber in der Politik inhärent verkürzt wahrgenommen und verwendet wird, und dass drittens für einen ernsthaften Wissenszuwachs in der Forschung die Daten und Informationen nicht ‚hermetisch‘ verwendet

¹ Vgl. <http://www.theguardian.com/education/2014/may/06/oeecd-pisa-tests-damaging-education-academics>; unter den UnterzeichnerInnen finden sich Stephen Ball, Henry Giroux und Diane Ravitch; vgl. die Antworten auf die Kritik: <http://www.theguardian.com/education/2014/may/08/pisa-programme-short-term-fixes>; sowie <http://www.tes.co.uk/article.aspx?storycode=6345258>

werden dürfen, sondern wesentlicher Ergänzungen bedürfen. Die vertretenen Positionen gehen teilweise akzentuiert in unterschiedliche Richtungen, und kritisieren sich manchmal direkt gegenseitig.

Geschichte der LSAs seit den 1960ern und technologische Entwicklungen

Im *einleitenden Kapitel* wird die Geschichte der LSAs kurz von ETS-MitarbeiterInnen resümiert (die Herausgeber mit Marylou Lennon). Die Geschichte geht einerseits auf US Entwicklungen bis zur Eight-Year-Study in den 1930ern zurück, deren Leiter Ralph Tyler (dem die Prägung des Begriffes der Evaluation zugeschrieben wird) in den 1960ern mit der Entwicklung des National Assessment for Educational Progress (NAEP) betraut wurde, das bis heute mit den entsprechenden Weiterentwicklungen eine wichtige Daten- und Informationsbasis in den USA darstellt (<http://nces.ed.gov/nationsreportcard/>; <http://www.nationsreportcard.gov/>). In diesem Rahmen wurde in den 1980ern von ETS erstmals die Item Response Theory (ITR) angewendet. Im Vergleich zur sich ebenfalls parallel entwickelnden psychologischen Testforschung, die die Individuen im Fokus hat, wird bei den LSAs betont, dass hier die Kompetenzen von Gruppen im Vordergrund stehen.

Andererseits wurde bereits Ende der 1950er im Rahmen der UNESCO von den VorläuferInnen der International Association for the Evaluation of Educational Achievement (IEA; http://www.iea.nl/about_us.html)² mit der Durchführung von internationalen Assessments begonnen (erste 12-Länder-Studie Mathematik, Leseverständnis, Geographie, Science und nicht-verbale Fähigkeiten 1959-62; es folgte die ‚First International Mathematics Study 1963–1967 und die First International Science Study 1968–1972 (http://www.iea.nl/brief_history.html)).³ Ab den 1970ern wurde begonnen periodische

² Die IEA wurde ursprünglich eher aus Forschungsinteressen heraus gegründet, heute sind die Mitglieder meistens nationale oder regionale Erziehungsministerien, oder nationale oder regionale Buffer-Organisationen im Erziehungsbereich (z.B. Schweden: Skolverket), oder manchmal auch Forschungsinstitutionen (Finnland: Finnish Institute for Educational Research; Deutschland: German Institute for International Educational Research DIPF).

³ Die mehr als zwanzig teilnehmenden Länder/Regionen der ersten drei Studien waren in etwas wechselnder Zusammensetzung: Australia, Chile, Belgium, Belgium (Flemish), Belgium (French), **England, Finland, France, Germany (FRG)**, Hungary, India, Iran, Italy, Israel, Japan, Netherlands, New Zealand, Poland, **Scotland, Sweden**, Switzerland, Thailand, **United States**, Yugoslavia (fett an allen drei Studien teilgenommen, in der dritten Studie erfolgte bereits eine deutliche Erweiterung der Teilnahme).

Wiederholungsstudien zur Erfassung von Veränderungen zu entwickeln, die ab 1980 zunehmend durchgeführt wurden. Wesentlich auf der Basis der Entwicklungen in NAEP wurde dann ab 1995 mit TIMMS (Trends in Mathematics and Science Studies) in vier-Jahres-Zyklen begonnen, 2001 folgte PIRLS (Progress in Reading Literacy Study). Die OECD als politische Regierungsinstitution ist erst 2000 mit PISA in den Bereich der Jugendlichen operativ eingestiegen, und konnte auf den bisherigen Erfahrungen aufbauen. In den 1990ern wurde unter Beteiligung der OECD mit der Erfassung der Kompetenzen von Erwachsenen mit der International Adult Literacy Study (IALS; 20 Länder in drei Wellen 1994, 1996, 1998; <https://www.ets.org/Media/Research/pdf/RR-01-25-Kirsch.pdf>) begonnen, die von ETS gemeinsam mit Statistics Canada durchgeführt wurde,⁴ gefolgt 2003-08 von Adult Literacy and Life Skills Survey (ALL; 11 Länder) und nun wieder auf größerer Basis von PIAAC.

Neben der Verbreiterung der gemessenen Kompetenzen und der Ausweitung der erfassten Hintergrundvariablen wird v.a. in der neu entwickelten computer-basierten Testung eine wesentliche Veränderung gesehen, die durch zusätzlichen Erfassungsmöglichkeiten (Beispiele, Testverlauf) auch eine vertiefte Kompetenzmessung erlaubt. Dies wird in einem eigenen Kapitel von *Michal Beller* aus Israel vertiefend behandelt. Dieser gibt einen kompakten Überblick über die Entwicklung des Einsatzes von computer-based-Testing (CBT) in den LSAs, wobei auch hier NAEP eine Vorreiterrolle spielte. Es werden aber auch weitergehende Visionen der Verbindung von formativen und summativen Beurteilungen in „full blown digital and online systems“ diskutiert (Beller, 41; als Pilot-Beispiel wird das US-Forschungsprojekt über ‚Cognitively Based Assessments of, for, and as Learning - CBAL‘ angesprochen; <http://www.ets.org/research/topics/cbal/initiative>). Die ‚mega technology companies‘ Cisco, Intel und Microsoft kooperieren auf dieser Linie mit der Universität Melbourne in einem Projekt zur Weiterentwicklung von Beurteilung und Kompetenzen für das 21. Jahrhundert (<http://atc21s.org/>).

Vgl. auch Bobby Naemi, Eugene Gonzalez, Jonas Bertling, Anthony Betancourt, Jeremy Burrus, Patrick C. Kyllonen, Jennifer Minsky, Petra Lietz, Eckhard Klieme, Svenja Vieluf, Jihyun Lee, and Richard D. Roberts (2013) Large-Scale Group Score Assessments: Past, Present, and Future. In: The Oxford Handbook of Child Psychological Assessment, edited by Donald H. Saklofske, Cecil R. Reynolds, and Vicki Schwean. Oxford: OUP, 129-149, DOI: 10.1093/oxfordhb/9780199796304.013.0006

⁴ Vgl. OECD & Statistics Canada (1995) Literacy, Economy and Society: Results of the First International Adult Literacy Survey. Paris.

Politische Nutzung: LSAs als ‚change agent‘?

Das einleitende Kapitel präsentiert auch ein einfaches Modell für ‚evidence-based policy information‘ (Kirsch et al, 6-7), das einem Entwicklungs-Zirkel von ‚(1) policy questions, (2) assessment frameworks and instrumental design, (3) methodological advances und (4) enhanced analysis and interpretation of data‘ (ebd., fig 1.1.) folgt. Das politische Kapitel über LSAs als ‚Change Agents‘ wurde von *Jo Ritzen* verfasst,⁵ der ein international tätiger und bekannter Ökonom war, bevor er in den Niederlanden der längst dienende Minister für Bildung und Wissenschaft (1989-98) wurde und in mehreren Bereichen wichtige Reformen vorantrieb; anschließend war er in der Weltbank tätig und führte dann 2003-2011 im unter seiner Regierung reformierten Governance System die Universität Maastricht als Präsident zu einer hochrangigen Institution. Der Autor greift ausdrücklich auf seine politischen Erfahrungen zurück, auch als Mitglied eines Dreier-Panels bei der OECD zur Analyse der politischen Effekte von PISA 2000-06, aber der Beitrag ist eher enttäuschend. Im ersten Teil wird auf die niederländischen Erfahrungen zurückgegriffen und auf dem Hintergrund einer Reformunwilligkeitsannahme die Bedeutung der Transparenz der Leistungen auf der Ebene der Institutionen betont (Veröffentlichung der Ergebnisse), die über zwei Mechanismen wirken soll: erstens über die Bildungswahlen (Zauberwort ‚Choice‘), und zweitens über den Ehrgeiz der Institutionen gute Werte zu erzielen. Im zweiten Teil werden die widersprüchlichen und tw. inkonklusiven Ergebnisse der OECD PISA-Evaluation⁶ herausgearbeitet, z.B. dass die AkteurInnen v.a. der unteren Politik- und Verwaltungsebenen sich viel mehr Verantwortlichkeiten zuschreiben, als sie dann nach Selbstauskunft tatsächlich annehmen. Auch hier besteht die Hauptbotschaft darin, dass die Ergebnisse der LSAs nur wirksam werden können, wenn sie die praktische Ebene der AkteurInnen auf der Mikroebene tatsächlich erreichen,⁷ und es wird eine wichtige Rolle der Medien bei der Verbreitung der Information betont.

⁵ Dieses Kapitel ist auch als Sample Pages herunterzuladen

http://www.springer.com/cda/content/document/cda_downloaddocument/9789400746282-c2.pdf?SGWID=0-0-45-1339813-p174324927

⁶ OECD (2008) External evaluation of policy impact of PISA. Paris.

⁷ In diesem Zusammenhang verweist er auf einen sehr interessanten Artikel über die Nutzung von Daten von Kevin Lang, der in seinen Schlussfolgerungen feststellt: “[...] I do believe that we can use tests as a trigger for further investigation. Teachers and schools with both low average test scores and low estimated value-added are likely to be weak and, given limited resources, this group is the obvious place to target. Since tests are only very imperfect measures of the desired outcomes of education, we should avoid any automatic consequences for teachers or schools based solely on measures derived from tests. However, in the case of teachers, we know that

Wirtschaftliche Aspekte der kognitiven Kompetenzen

Die Zusammenhänge zwischen Kompetenzen und Wirtschaftswachstum aus ökonomischer Sicht werden von zwei der aktuell renommiertesten Bildungsökonominnen behandelt, die seit einigen Jahren wichtige gemeinsame Publikationen veröffentlicht haben. *Eric Hanushek* hat seit Jahrzehnten versucht, die Fragen der Effektivität und Effizienz der Erziehungs/Bildungsorganisation und -politik aus ökonomischer Sicht zu analysieren, *Ludger Woessmann* hat in den letzten eineinhalb Jahrzehnten die deutsche Bildungsökonomie aus einem (fast) Niemandsland zu einem führenden Mitspieler in der internationalen akademischen Forschung gemacht. Sie betrachten die Kompetenzerhebungen als wesentliche Bereicherung des Verständnisses von ‚Humankapital‘, das ursprünglich nur durch institutionelle Faktoren operationalisiert wurde (v.a. Zahl der absolvierten Schuljahre), und sie fokussieren v.a. auf die Möglichkeiten, die die vergleichenden Daten für die ökonomische Analyse bieten. Diese werden nicht als Ersatz für nationale Studien gesehen, sondern als Komplement. Als zu lösende Probleme werden hervorgehoben:

- Messprobleme (in manchen Ländern ergeben sich sehr geringe Kompetenzlevels, denen zufolge fraglich ist, was in den Schulen gelernt wird)
- Die Identifikation von kausalen Effekten ist in diesen Studien nicht möglich, dafür wären longitudinale (Panel)-Studien erforderlich, die es bisher maximal auf nationaler Ebene gibt
- Der Zusammenhang zwischen Testergebnissen und Einkommen müsste besser verstanden werden.

Nach einem kompakten Überblick über die ‚Explosion‘ der Erhebungen seit den 1960ern, die erst spät für ökonomische Analysen genutzt wurde, behandeln sie die zwei großen Themen der Bildungsökonomie, die ‚Produktionsfunktion‘, d.h. (1) die Frage, wie Inputfaktoren in Ergebnisse transformiert werden, und (2) den internationalen Vergleich der Ergebnisse und deren Verwertung auf unterschiedlichen Aggregationsebenen (‚Lohnfunktion‘,

in most cases, qualitative investigation by principals will confirm the statistical analysis. I expect that the same will prove true if qualified reviewers inspect schools.

The challenge for the next stage of education reform is therefore not to find better statistical methods for identifying low-performers but to provide administrators with the necessary resources to understand the statistical information and the ability and incentives to act on both statistical and qualitative information.“ (Lang 2010, 180-181)

Lang, Kevin (2010) Measurement Matters: Perspectives on Education Policy from an Economist and School Board Member. *Journal of Economic Perspectives* 24(3 Summer), 167-182.

Wirtschaftswachstum). Die Nutzung der Erhebungen über die deskriptiven Vergleiche hinaus, bei der die Bildungsökonomie eine Vorreiter-Rolle gespielt hat, sehen die Autoren immer noch als Entwicklungsaufgabe. Für die beiden Bereiche wird ein Überblick über die vorhandenen Studien gegeben.

- Im ersten Bereich wird v.a. die Bedeutung der Institutionen betont und das Credo der ökonomischen – manche sagen auch der neoliberalen – Politikvorschläge prägnant zusammengefasst: „The clearest and most unique evidence provided by this international work ist that the overall set of educational institutions has a significant impact on student achievement. [...] countries with test-based accountability systems, with more school choice, and with more local decision making or more local autonomy tend to do better.”(Hanushek & Woessmann, 56). Lokale Autonomie macht nur Sinn, wenn sie mit funktionierender Messung von Ergebnissen verbunden ist. Einschränkend wird betont, dass in diesen Vergleichen unbeobachtete kulturelle oder institutionelle Faktoren die Ergebnisse beeinflussen können.

- Im zweiten Bereich, bei der Frage, warum unterschiedliche Volkswirtschaften unterschiedlich rasch wachsen, ist die internationale Variation der entscheidende Punkt, wobei verschiedene Volkswirtschaften zusätzliche Kompetenzen unterschiedlich hoch zu bewerten scheinen. Hier kommen die Kompetenzstudien für Erwachsene (also PIAAC und die Vorläufer IALS und ALL) ins Spiel, die Informationen über altersspezifische Einkommen enthalten. Die beiden Autoren präsentieren anschaulich ihr bekanntes ‚stilisiertes Fakt‘, dass die Einflüsse der Inputs (in Schuljahren) auf das Wachstum verschwinden, wenn die gemessenen Kompetenzen als zusätzlicher Faktor eingesetzt werden. Aufbauend auf diesen Schätzungen projizieren sie in die Zukunft, dass eine Steigerung der Kompetenzen (z.B. von USA auf Finnland) das Wirtschaftswachstum bedeutend erhöhen könnte, wenn die Effekte stabil bleiben. Henry Levin nimmt in seinem Beitrag zum Band (s.u.) mit Bezug auf die Ergebnisse der Forschungsgruppe rund um James Heckman direkt auf diese Ergebnisse Bezug. [...] it appears that that the dramatic and highly publicized extrapolations by Hanushek and Woessmann (2008) of contributions of economic growth of international achievement results among countries overstate the impact of test on economic output, possibly by a large magnitude. Unfortunately, the promise of massive gains in economic output of even modest gains in test scores have been disseminated widely an taken seriously; even though those administering policy are not aware or knowledgeable about the degree to which upward bias is present in the reported results and their policy extrapolations.” (Levin, 82)

Ergänzung durch nicht-kognitive Kompetenzen und Überschätzung der Wirkungen kognitiver Kompetenzen

Eine teilweise direkte Antwort auf die ökonomischen Analysen zu den kognitiven Kompetenzen – und bis zu einem gewissen Grad ein Gegenprogramm dazu – gibt einer der bekanntesten Vertreter der linksliberal polit-ökonomisch orientierten US Bildungsforschung, *Henry Levin*, der 1968-99 in Stanford über Ökonomie und Erziehungs/Bildungsfinanzierung forschte und anschließend an der Columbia ein Zentrum für das Studium der Privatisierung leitet. Das wesentliche Argument dieses Beitrags lautet, dass die kognitiven Kompetenzen einen wichtigen Aspekt darstellen, diese jedoch in ihrer alleinigen Bedeutung wesentlich überschätzt werden, und daher die Erziehungs/Bildungspolitik und -praxis letztlich fehlleiten. „[...] microeconomic studies show that such tests explain only a relatively small portion of the variance of the variance in earnings and supervisory ratings an a minor portion of the of the statistical relation between schooling attainments and economic outcomes. [...] Cognitive achievement is important and should continue to be assessed. But it is a highly incomplete category for measuring student and adult success. This chapter sounds an appeal to consider the potential importance of noncognitive skills and dimensions of human behavior as they comprise important adult competences and the role of schools in developing them.”(Levin, 68) Ein wesentliches Argument zu den empirischen Schätzungen besteht darin, dass die Modelle durch die Nichterfassung der nichtkognitiven Faktoren den ökonomischen Effekt der kognitiven Faktoren überschätzen, weil letztere sich implizit auswirken (‘upward bias‘).⁸ Es wird vorgeschlagen, die Tests durch nichtkognitive Kompetenzen zu ergänzen.

Die Argumentation geht in mehreren Schritten vor. Erstens werden eindruckliche Forschungsergebnisse aus den USA und auch anderen Ländern präsentiert, die die ökonomische Bedeutung der kognitiven Testwerte wesentlich einschränken: Testwerte erklären nur einen kleinen Teil der Einkommen, über die Zeit ist keine steigende Bedeutung der kognitiven Kompetenzen nachzuweisen, und kognitive Testinstrumente erklären nur einen kleinen Teil der Arbeitsleistungen. Zweitens werden Ergebnisse aus verschiedenen

⁸ Aus den präsentierten Ergebnissen zu nichtkognitiven Kompetenzen werden im Einzelnen v.a. drei Gründe für die Überschätzung herausgearbeitet: Erstens besteht ein Korrelation zwischen den kognitiven und den nichtkognitiven Faktoren, so dass die Wirkungen der nicht einbezogenen letzteren in den Effekten der ersteren mit ‚gemessen‘ werden; zweitens sind die beiden Dimensionen nicht nur korreliert, sondern die nichtkognitiven Faktoren tragen zu den kognitiven bei (aber nicht umgekehrt), was die Überschätzung verstärkt; drittens besteht die gleiche Verzerrung nach oben, die Hanushek selbst bei den Produktionsfunktionen nachgewiesen hat ebenfalls bei den Wachstumsmodellen. (Levin, 82)

Studien zur Bedeutung von nichtkognitiven Kompetenzen einerseits für die Schulleistungen und andererseits aus der Sicht der ArbeitgeberInnen aufgrund von Literatur zusammengefasst. Drittens wird aufgrund dazu vorliegender Literatur ein Überblick über bestehende Ansätze der Messung von nichtkognitiven Kompetenzen gegeben.⁹

Der Autor sieht die Verengung auf die kognitiven Kompetenzen, und die aus seiner Sicht massiven ökonometrischen Überschätzungen ihrer wirtschaftlichen Effekte v.a. in ihren politischen und praktischen Auswirkungen als zerstörend an. „Far from being harmless, the obsessive focus on test scores and the omission of the noncognitive impact of schools can provide far reaching damage. [...] pressures are place on the schools through accountability sanctions to raise test scores [...] Schools are pressed to use their time and resources to to improve scores [...] at the expense of other activities and subjects including noncognitive goals. [...] The ‘evidence based’ arguments have led to a singular focus on a cognitive achievement gap in the No child Left Behind legislation, leading schools to narrow their curriculum and focus on test preparation as a major instructional strategy.”(Levin, 82-83)

„Civic Education and Engagement“

Der Beitrag von *Judith Torney-Purta* und *Jo-Ann Amadeo*, die federführend an diesen Studien beteiligt waren, gibt einen ausgezeichneten Überblick über die bisherigen Erhebungen zu zivilgesellschaftlichen Kompetenzen, die seit den 1960ern durchgeführt wurden (größere internationale Studien wurden 1971, 1999-CIVED und 2009-ICCS durchgeführt).¹⁰ Diese Studien haben sowohl kognitive (Wissen) als auch nichtkognitive (Einstellungen) Aspekte erfasst, und der Beitrag gibt einen Überblick über die Vielfältigkeit und Multidimensionalität der Ergebnisse dieser Erhebungen.

In der Zusammenfassung wird einerseits auf die allgemeine Bedeutung der Civic Education sowohl für das Alltagsleben, als auch für die Arbeitswelt und die Demokratie hingewiesen, andererseits wird aber auch herausgearbeitet, dass diese Erhebungen bzw. Aspekte besser in

⁹ Vgl. z.B. das Symposium zu ‘The Noncognitive Determinants of Labor Market and Behavioral Outcomes’, *Journal of Human Resources*, Fall 2008, 43(4), <http://jhr.uwpress.org/content/43/4/729.full.pdf+html>; sowie IZA Workshop on ‘Cognitive and Non-Cognitive Skills’: http://www.iza.org/conference_files/CoNoCoSk2011; es wird zusammenfassend auf die ‘big five’ verwiesen: openness, conscientiousness, extraversion, agreeableness, neuroticism.

¹⁰ CIVED: Civic Education Study; ICCS: International Civics and Citizenship Study

die LSAs integriert werden sollten, auch der Zyklus sollte verkürzt werden. Es wird vorgeschlagen, eine bessere Koordination dieser Studien zu versuchen, und die Verbreitung der Ergebnisse zu verbessern sowie die Auseinandersetzung damit zu vertiefen. „The reports issued by the IEA are inherently complex and are not always easy for those outside the projects to interpret.” (Torney-Purta & Amadeo, 110)¹¹

Assessments und Schulentwicklung

Einen sehr wichtigen Aspekt behandelt der Beitrag von *Eckhard Klieme* mit der Frage nach den möglichen Beiträgen und vor allem auch nach den Grenzen der LSAs für die Schulentwicklung und Effektivität. Es wird hier klar aufgezeigt, dass die aggregierten vergleichenden Analysen zwar sehr wichtig sind, um das Wissen über die Bildungsprozesse und ihre Ergebnisse zu vermehren, dass diese aber nicht ausreichen, um davon direkt praktisch-politische Schlussfolgerungen abzuleiten. Die vergleichenden LSAs werden für die Exploration von Fragen und die Hypothesengenerierung wie auch für die Analyse institutioneller Faktoren als wichtig angesehen. Vor allem drei Arten von Beiträgen werden aus diesen Analysen gesehen: (1) Ableitung von Hypothesen aus den korrelativen und explorativen Ergebnissen, die dann getestet werden müssen; (2) Hypothesen aus der Effektivitätsforschung können in den LSAs getestet werden; (3) die Gültigkeit von Ergebnissen aus der Effektivitätsforschung kann über Länder und Kulturen getestet werden. Für die Hypothesentestung werden Beispiele aus der TIMMS Videostudie und als Erweiterung das Pythagoras-Projekt genannt.¹² Kritiken, die auch sonst schon zum Mainstream gehören, werden hier konstruktiv ergänzt, und es werden Beispiele dargestellt, wie die Erweiterung vor sich gehen kann (DESI und das PISA Schulpanel in Deutschland).¹³

Das zentrale Argument, warum aus den Querschnittsdaten der LSAs mit individuellen SchülerInnentestungen keine direkten Schlüsse für die Schulentwicklung gezogen werden können, besteht darin, dass die Lernprozesse und die Lernorganisation kumulativ

¹¹ Die OECD hat in dieser Richtung bereits Überlegungen laufen:

<http://www.oecd.org/pisa/pisaproducts/Longer-term-strategy-of-PISA.pdf>

¹² Vgl. https://www.mpib-berlin.mpg.de/volltexte/institut/dok/full/Klieme/TIMSS_homepage/html/math11.htm; <http://www.timssvideo.com/>; Pythagoras: http://www.fachportal-paedagogik.de/forschungsdaten_bildung/studie.php?studien_id=1

¹³ DESI Deutsch-Englisch-Schülerleistungen-International <http://www.dipf.de/de/forschung/projekte/deutsch-englisch-schuelerleistungen-international>

funktionieren, und daher der jeweils vorherige Leistungsstand einen wesentlichen Faktor der pädagogischen und organisatorischen Interventionen ausmacht. Dieser bleibt aber unbeobachtet und kann auch statistisch nicht kontrolliert werden, als Ausweg werden Längsschnitterhebungen gesehen. Für diesen Ansatz werden die beiden Beispiele gegeben.

In diesem komplexen Beitrag wird auch ein guter Überblick über wesentliche Aspekte des Zusammenhangs von Ansätzen und Forschungen zur Schuleffektivität einerseits und den LSAs andererseits gegeben. Dies ist interessant, weil hier gezeigt wird, wie die abstrakten Ergebnisse der LSAs mit verschiedenen bildungswissenschaftlichen Modellvorstellungen¹⁴ in Verbindung gebracht werden können, und welche statistischen Probleme hier auch auftreten können. Diese Diskussion gibt einerseits einen Eindruck der doch differenzierten Dimensionalität bildungswissenschaftlicher Vorstellungen, andererseits indirekt auch einen Eindruck von der relativen Armut der österreichischen Diskussion im politischen Raum. In der Forschungsgruppe um Klieme wird ein dreidimensionales Schema von ‚Basic Dimensions of Instructional Quality‘ vorgeschlagen:

- „(a) clear, well-structured classroom management
- (b) supportive, student-oriented classroom climate, and
- (c) cognitive activation with challenging content“ (Klieme, 122)

Auf wenigen Seiten wird dann ein eindrucksvoller Überblick über Herausforderungen im Bereich der Forschung über Schuleffektivität gegeben, der auch die Gewichte der verschiedenen Ebenen behandelt. Wesentliche Punkte sind, dass die Lehrpersonen und die Prozessebene, sowie v.a. die Lernvoraussetzungen wichtiger sind, als die Rahmenbedingungen, und dass Schulen als Handlungseinheiten genommen werden müssen, die externe (politische) Einflussnahmen filtern. Gleichzeitig ist nicht von einer linearen Entwicklung einer Schule auszugehen, sondern diese enthält auch zufällige und erratische Auf- und Abs. An einem Beispiel wird gezeigt, dass individuelle Leistungs-Veränderungen sich in aggregierten Leistungsveränderungen der Schulen nicht zeigen müssen.

Hinsichtlich der Konzeption der LSAs wird ein Wechselspiel zwischen den bildungswissenschaftlichen Modellen und den Erhebungen, sowie zwischen Politik und Forschung betont. Sowohl die Konstrukte und Operationalisierungen als auch die Variablen

¹⁴ V.a. CIPO: Context-input-process-outcome; sowie EER: Educational Effectiveness Research; OTL: Opportunity to Learn; BDIQ: Basic Dimensions of Instructional Quality

werden forschungsbasiert kreiert. Für die Zukunft wird die Unterstützung der Politik für die bessere Interaktion zwischen den Erhebungen und der Effektivitätsforschung gefordert.

Zukunftsansichten: ‚Caveat emptor is the watchword‘ (Braun, 156)

Im abschließenden Kapitel skizziert *Henry Braun*, Direktor am Center for the Study of Testing, Evaluation and Education Policy¹⁵ auf Basis der Kapitel Überlegungen zu zukünftigen Entwicklungen. Es wird von der Differenz zwischen der hohen Publizität der Rangordnungen und den möglichen, aber viel weniger genutzten vertiefenden Analysen einerseits und einer ‚naiven‘ politischen Aktionstheorie – Transparenz, Konsensus, Umsetzung – andererseits ausgegangen. Die Differenz wird nicht aufgeklärt.

Wichtige Punkte sind die Fragen die aus einer weiteren Ausweitung des TeilnehmerInnenkreises in Nicht-OECD-Länder resultieren, sowie die Kompetenz der Medien und PolitikerInnen im Umgang mit den komplexen Instrumenten.

Abschließende Bemerkungen: Gegenwind berechtigt?

Das Buch gewinnt seine Stärke einerseits davon, dass die AutorInnen selbst höchst qualifiziert mit diesen Daten arbeiten und ihre teilweise fundamentale Kritik nicht nur von außen äußern. Die doch breite Vielfalt an angesprochenen Problemen wird sehr gut und einleuchtend, mit einem konstruktiven Zugang dargelegt. Als interessierteR LeserIn bekommt man einen guten Überblick, und in vielen interessanten und strittigen Aspekten auch ein vertiefendes Bild mit vielen sehr guten weiterführenden Hinweisen.

Aus einer Gesamtsicht können vor allem drei wichtige Botschaften festgehalten werden: Erstens die Spannungen zwischen der ökonomischen Analysen und Bewertung der kognitiven vs. nichtkognitiven Kompetenzen, die in den Beiträgen von *Hanushek & Woessmann* einerseits und *Levin* andererseits zum Ausdruck kommen; zweitens die im Beitrag von *Klieme* herausgearbeiteten Grenzen der LSAs für pädagogische Zwecke, die in

¹⁵ Vgl. <http://www.bc.edu/research/csteep/>

der Politik und teilweise auch in der ökonomischen Forschung nicht ausreichend beachtet werden; drittens die Unterstreichung der wissenschaftlichen und gesellschaftlichen Bedeutung der Civic Education, und ihr tatsächliches Schattendasein im Rahmen der LSAs im Beitrag von *Torney-Purta & Amadeo*.

Es ist kein einführendes Buch, aber diskutiert die Probleme doch für ein breiteres Publikum. Der Preis von 109,95 EUR erhöht leider nicht die Zugänglichkeit, aber die Kapitel sind auch gesondert zu erwerben; bei diesem Preis sollte es auch keine Fehler enthalten, was aber doch an manchen Stellen passiert ist.

Ein Schwachpunkt des Buches besteht darin, dass die politische Seite der LSAs nicht adäquat behandelt wird. Dies hängt vielleicht auch mit den zeitlichen Aspekten der Herausgabe zusammen. Die Konferenz, auf der das Buch beruht hat im März 2011 stattgefunden,¹⁶ es ist zwei Jahre später erschienen. In der Zwischenzeit hat sich die politische Auseinandersetzung v.a. um PISA deutlich verstärkt. Die Entscheidung, wesentliche Elemente von PISA 2015 an ein privates multinationales Unternehmen zu vergeben,¹⁷ spielt dabei eine wichtige Rolle, da viele der KommentatorInnen darin einen Tabubruch sehen.

Teilweise klingen die Unterschiede zwischen IEA und OECD an, wobei IEA forschungsbasiert ist, und sich politischer Stellungnahmen enthält, während die OECD eine dezidiert politische Organisation ist. Einleitend wird im Buch festgestellt, dass die LSAs in ihrer Entwicklung von ursprünglicher Nichtbeachtung (IEA) zu einer Überpolitisierung (OECD) gekommen sind. In den Überlegungen zu den Zukunftsperspektiven wird die politische Bedeutung von PISA betont und aufrechterhalten und eine weitere weltweite Ausweitung angestrebt.

Ein sehr wichtiger Aspekt betrifft die Diskussion um die sogenannten kognitiven und nichtkognitiven Kompetenzen, und die Frage der Folgewirkungen der Betonung der

¹⁶ Siehe die Dokumentation, die die Präsentationen und auch qualifizierte Kommentare zu den Präsentationen, alles tw. auf Video enthält http://www.ets.org/sponsored_events/ilsa_conference/presentations_videos

¹⁷ Vgl. zum Unternehmen <http://uk.pearson.com/home/news/2011/september/pearson-to-develop-frameworks-for-oecd-pisa-student-assessment-for-2015.html>; <http://www.pearson.com/about-us/pearson-at-a-glance.html>; <http://www.pearson.com/about-us/board-of-directors.html>; <http://www.pearson.de/main/main.asp?page=about/ueber>; das Unternehmen unterhält auch eine non-profit Stiftung, die starke Verbindungen zur OECD suggeriert <http://www.pearsonfoundation.org/oecd/about.html>; Seitens der OECD wird festgestellt, dass diese Verbindung über interpretiert wird: „[...] it is stated that Pearson is overseeing the Pisa 2015 assessment, which is not the case. Pearson was one of a number of contractors who have been appointed through a competitive tendering process to develop and implement Pisa 2015. Pearson's contract to develop the assessment framework has been completed and has now come to an end.“ (A.Schleicher, The Guardian 8.5.2014).

kognitiven Kompetenzen für die Bildungspolitik und die Praktiken im Bildungswesen. In diesem Punkt bestehen beträchtliche Kommunalitäten zwischen dem beschriebenen und diskutierten Band und den fundamentalen KritikerInnen. In den Zukunftsperspektiven für PISA wird dieser Aspekt angesprochen, aber nicht sehr deutlich.

Ein neueres Buch, PISA, Power and Policy¹⁸ beschäftigt sich dezidiert mit diesen politischen Fragen. Dieses wird in der nächsten Nummer besprochen – Fortsetzung folgt.

¹⁸ Meyer, Heinz-Dieter & Benavot, Aaron, Hg. (2013) PISA, Power, and Policy: the emergence of global educational governance. Oxford: Symposium Books.