Inhalt

Bernd Groot-Wilken, Kevin Isaac & Jörg-Peter Schräpler

Einleitung: Sozialindices für Schulen – Hintergründe, Methoden und Anwendungen

Ein zentrales Anliegen der Bildungsforschung und -politik ist die Verringerung von Chancenunterschieden im Bildungswesen. Hierbei spielen kollektive Benachteiligungen von bestimmten Bevölkerungsgruppen eine besondere Rolle. Mit der Einführung von Schulleistungsstudien, insbesondere aber nach dem Ergebnis der Schülerinnen und Schüler im deutschen Bildungssystem bei PISA 2000, wurde auch eine schul- und bildungspolitische Diskussion darüber ausgelöst, inwieweit mithilfe von Sozialindices als Steuerungsinstrument sozialen individuellen und kollektiven Benachteiligungen, wie beispielsweise der räumlichen Konzentration sozialer Benachteiligungen in Stadtteilen der Großstädte, adäquat Rechnung getragen werden kann. Ein wichtiges Anwendungsgebiet stellt dabei die Verwendung im Kontext der Vergleichsarbeiten bzw. Lernstandserhebungen dar. In diesem Rahmen wird aus Gründen der Fairness bei Ergebnisrückmeldungen an Einzelschulen mithilfe von Sozialindices angestrebt, den sozialen Hintergrund der Schülerschaft weitgehend zu berücksichtigen. Die Art und Weise, wie dies getan wird, welche Möglichkeiten der Berechnung und Anwendung es derzeit gibt und inwieweit die Indices in diesem Rahmen verwendet werden können, ist bislang noch nicht zusammenfassend dargestellt worden.

Es ist das Anliegen dieses Bandes, einen Überblick der Diskussion über Schulsozialindices im deutschsprachigen Raum zu geben.

Der Beitrag „Sozialindex – Ein Instrument zur Gestaltung fairer Vergleiche" von *Horst Weishaupt* zeichnet diese Diskussion nach, indem er Ansätze für Sozialraumtypologien wie etwa die räumliche Konzentration von sozial benachteiligten Milieus insbesondere in Stadtteilen in Großstädten skizziert, Konzepte für die Berechnung von Sozialindices vorstellt und die Erfahrungen mit sozialindexgesteuerten Ressourcenzuweisungen an Schulen referiert.

Im zweiten Beitrag stellen *Jörg-Peter Schräpler* und *Sebastian Jeworutzki* die Konstruktion des Sozialindex für NRW vor. Hierbei wird u.a. auf ein statistisches Verfahren eingegangen, welches die räumlich-sozialen Unterschiede genau und unabhängig von vorgegebenen kommunalen Abgrenzungen wie Ortsteilen oder Gemeindegrenzen erfasst. Ein wesentlicher Bestandteil des Index ist die räumliche Dichte von SGB-II-Bedarfsgemeinschaften mit Kindern im Schuleinzugsgebiet und der Anteil an Schülerinnen und Schülern mit Migrationshintergrund an den jeweiligen Schulen. In dem

Beitrag wird dieser Index u.a. mit den früheren Standorttypen des Ministeriums für Schule und Weiterbildung (MSW) in NRW evaluiert und Vorschläge zur Weiterentwicklung werden diskutiert.

Ein wesentliches Ziel bei einem fairen Vergleich von Schüler- und Klassenergebnissen bei Vergleichsarbeiten ist es, Lehrkräften eine Ergebnisrückmeldung zu liefern, die zu einem möglichst großen Teil die Ergebnisse ihrer Arbeit widerspiegelt. *Philipp Loesche* und *Ingmar Hosenfeld* entwickeln in ihrem methodischen Beitrag „Beurteilung von fairen Vergleichen anhand eines Rückmeldepassungskoeffizienten" ein Modell der Schülerleistung und den Rückmeldepassungskoeffizienten als Maß für den Einfluss der Lehrkraft auf das Klassenergebnis, welche Aussagen über die Repräsentativität der Klassenleistung für die Arbeit der Lehrkraft im Rahmen eines fairen Vergleichs erlauben. Auf diese Weise kann geprüft werden, inwieweit sich die Repräsentativität der Klassenleistung für die pädagogische Arbeit durch einen fairen gegenüber einem unfairen Vergleich steigern lässt.

In den deutschen Bundesländern werden im Rahmen eines fairen Vergleichs von Vergleichsarbeiten unterschiedliche Adjustierungsverfahren verwendet. *Christiane Fiege* liefert in ihrem Beitrag „Faire Vergleiche bei Vergleichsarbeiten: Möglichkeiten und Grenzen" eine systematische Darstellung aller verwendeten Adjustierungsverfahren. Zudem zeigt sie die Bedeutung des fachspezifischen Vorwissens für die Berechnung fairer Vergleiche anhand einer empirischen Analyse von Daten aus Vergleichsarbeiten auf.

Seit einiger Zeit werden auch in Österreich standardbasierte Schulleistungstests durchgeführt. *Ursula Itzlinger-Bruneforth et al.* berichten in ihrem Beitrag „Fairer Vergleich und Sozialindex in österreichischen Bildungsstandardüberprüfungen" von der Ergebnisdarstellung in Berichten und Rückmeldungen und der Konstruktion des fairen Vergleichs sowie einem davon zu unterscheidenden Index der sozialen Benachteiligung in Österreich.

Dominik Becker et al. berechnen in ihrem Beitrag „Der Sozialindex: Theoretische Fundierung und forschungspraktische Relevanz für die Erfassung der Schülerkomposition von Gymnasien" Sozialindexgruppen auf Grundlage von Befragungsergebnissen aus dem Projekt „Ganz In – Mehr Ganztag mehr Zukunft. Das neue Ganztagsgymnasium NRW" mit der bereits bei KESS angewendeten Methode. Sie kommen zu dem Schluss, dass für die für Gymnasien zu bildenden Indices neue Indikatoren gefunden werden müssen, die dieser selektiven Schulform besser gerecht wird.

Kevin Isaac stellt in seinem Beitrag die Verwendung eines Schulsozialindex im Kontext der Lernstandserhebungen in Nordrhein-Westfalen als Prädiktor zur Vorhersage von Ergebnissen in zentralen Prüfungen dar. Dabei zeigt sich nicht nur die enorme Vorhersagekraft und Relevanz von diagnostischen Tests auf Prüfungen, sondern auch der – unabhängig von den in den Lernstandserhebungen in der Jahrgangsstufe 8 diagnostizierten Kompetenzen bestehende – Einfluss der sozialen Zusammensetzung der Klassen auf die Prüfungsleistungen in Klasse 10.

In Hamburg wurde 1996 erstmalig ein Sozialindex für Schulen eingesetzt. Ziel war es, durch den Einsatz von Sozialindices Schulen in schwierigen Lagen zusätzliche Mittel zur Unterstützung bereitzustellen, um Effekte der Schülerzusammenset-

zung kompensieren und chancenausgleichend wirken zu können. *Klaudia Schulte et al.* berichten in ihrem Beitrag verschiedene Überlegungen zur Weiterentwicklung dieser Indices.

Jan Ammon zeigt in seinem Beitrag die Möglichkeiten auf, die sich mit der Nutzung von Daten aus den Schuleingangsuntersuchungen der Stadt Mülheim ergeben. Die dargestellte Methode ließe sich auf alle nordrhein-westfälischen Kommunen, an denen die Daten verfügbar sind, übertragen.

Die Nutzung von Variablen bei der Bildung von Sozialindices kann gewissen Einschränkungen unterliegen. *Thomas Kemper* beschreibt in seinem Beitrag die Konsequenzen, die sich aus der Änderung des Staatsangehörigkeitsrechts für eine der wichtigsten Ausgangsvariablen, des Migrationshintergrunds der Schülerinnen und Schüler, ergeben. Dies gewinnt auch deshalb an Bedeutung, da diese Variable aus der amtlichen Schulstatistik an vielerlei Stellen zur Berechnung von Sozialindices herangezogen wird.

Mit diesem Band möchten die Herausgeber eine erste Bestandsaufnahme vornehmen. Damit sollen die Diskussion um faire Vergleiche sowohl auf der methodischen als auch administrativen Umsetzungsebene weitergeführt und darüber hinaus Impulse für Weiterentwicklungen gegeben werden.

Jörg-Peter Schräpler & Sebastian Jeworutzki

Der Sozialindex für NRW – Die Bildung von Schulstandortentypen über SGB-II-Dichten und Migrationshintergrund

Zusammenfassung

Im Rahmen der Konstruktion von Sozialindices für faire Leistungsvergleiche zwischen Schulen spielt insbesondere beim Standorttypenkonzept die adäquate Erfassung der räumlich-sozialen Unterschiede eine zentrale Rolle. Diese spiegeln zu einem großen Maße die unterschiedliche soziale Herkunft von Schülerinnen und Schülern und die vorhandenen ökonomischen Ressourcen der Familien wider und können wichtige Indikatoren für unterschiedliche Lernausgangslagen von Schülerinnen und Schülern sein. In NRW wird seit dem Jahr 2009 ein besonderes statistisches Verfahren verwendet, um diese räumlich-sozialen Unterschiede möglichst genau und unabhängig von vorgegebenen kommunalen Abgrenzungen wie Ortsteilen oder Gemeindegrenzen zu erfassen. Mit dem verwendeten KDE-Verfahren werden aus Adressinformationen von SGB-II-Bedarfsgemeinschaften mit Kindern Häufigkeitsdichten und Dichteflächen erzeugt, die zur Beschreibung des näheres Schulumfeldes herangezogen werden können. Die SGB-II-Dichten werden mit amtlichen Schuldaten, wie dem Anteil der Schülerinnen und Schüler mit Migrationshintergrund, zu einem Standort-Index verknüpft. In dem Beitrag wird dieser Index mit einem Referenzindex des Instituts für Schulentwicklungsforschung (IFS) sowie mit den früheren Standorttypen des Ministeriums für Schule und Weiterbildung (MSW) in NRW evaluiert. Abschließend werden Vorschläge zur Verbesserung des derzeitigen Ansatzes diskutiert.[1]
Schlüsselwörter: Sozialindex, Kernel Density Estimation, Schulindex, SGB II Empfänger

1. Einleitung

Zahlreiche Untersuchungen zeigen, dass die Zusammensetzung der Schülerschaft der Schulen in Bezug auf die soziale Herkunft und die zur Verfügung stehenden ökonomischen Ressourcen, selbst innerhalb derselben Schulform, sehr unterschiedlich sein kann (vgl. Baumert et al., 2005; Leist, 2014). Die Zusammensetzung der Schüler-

1 Der Aufsatz basiert auf einem früheren Beitrag von Schräpler (2011).

schaft liegt außerhalb des unmittelbaren Einflusses von Schule und Unterricht, beeinflusst aber als „Hintergrundmerkmal" zusammen mit den kognitiven Grundfähigkeiten der einzelnen Schülerinnen und Schüler die fachspezifischen Schülerkompetenzen erheblich (vgl. Dumont, Neumann, Maaz & Trautwein, 2013; Stanat, 2006). Neben den familiären Merkmalen der Schülerinnen und Schüler können zudem auch die wirtschaftlichen, sozialen und kulturellen Rahmenbedingungen der jeweiligen Schulstandorte einen Einfluss auf die Kompetenzen der getesteten Schülerinnen und Schüler haben. Im Rahmen der PISA-Studie erfolgten die Analysen hierzu auf Ebene der Kreise und kreisfreien Städte (vgl. Baumert, Carstensen, Siegle, 2005). Allerdings sind Städte und Kreise in sich oftmals hochgradig differenziert und es besteht eine kleinräumige soziale Fragmentierung der Wohnbevölkerung (vgl. Jeworutki et al., 2016; ILS NRW, 2003). Mit dieser sozialen Segregation geht euch eine räumliche Trennung zwischen „bildungsnahen" und „bildungsfernen" Bevölkerungsgruppen innerhalb einer Stadt oder einer Gemeinde einher (vgl. Terpoorten, 2005 und 2014). Dies ist besonders bedeutsam, da die Bildungsentscheidungen der Eltern nicht nur von individuellen sozioökonomischen Faktoren wie Status und Beruf, sondern auch von der sozialen und wirtschaftlichen Situation in dem jeweiligen Umfeld beeinflusst werden (z.B. Ernst, 2010; Haunberger, 2007; Terpoorten, 2005 und 2014; Fickermann, 1990).

Vor diesem Hintergrund kann also konstatiert werden, dass unterschiedliche Schulen an unterschiedlichen Standorten auch verschiedene Ausgangssituationen für die Ausbildung der Schülerinnen und Schüler vorfinden. Gleichzeitig wurden im Rahmen der auf die PISA-Studie folgenden Reformen Monitoring-Instrumente wie die Lernstandserhebungen in NRW eingeführt, um es den Schulen durch Leistungsvergleiche zu ermöglichen, Entwicklungspotenziale zu identifizieren. Ein generelles Problem bei solchen Vergleichen ist, dass Unterschiede zwischen den Schulen nicht nur auf die pädagogischen Arbeit von Lehrkräften zurückgeführt werden können, sondern auch durch außerschulische Einflussgrößen des Lernens, die sich dem pädagogischen Einfluss der Lehrkräfte entziehen, beeinflusst werden. Dies können z.B. unterschiedliche individuelle Ausgangsvoraussetzungen oder Kontextbedingungen des Lernens sein (vgl. hierzu ausführlich den Beitrag von Fiege in diesem Band). Um einen fairen Vergleich und realistische Einschätzungen zu ermöglichen, müssen leistungsrelevante Hintergrundmerkmale der Schülerschaft berücksichtigt werden. Fiege unterscheidet im Rahmen einer Auswertung von Vergleichsarbeiten zwischen mehreren Adjustierungsstrategien. Eine Möglichkeit besteht darin, Schulen mit vergleichbaren Rahmenbedingungen etwa in Bezug auf die Zusammensetzung der Schülerschaft und den Schulstandortes zu „Standorttypen" zu gruppieren und Vergleiche innerhalb der jeweiligen Referenzgruppe vorzunehmen. Zur Bildung dieser Standorttypen können aus dem amtlichen Schuldatensatz verschiedene an der Schule erhobene Informationen zur Schülerschaft genutzt werden. Spezifische Merkmale zum Schulstandort oder dem Wohnort der Schülerinnen und Schüler liegen jedoch in der Regel nicht vor. Allerdings können externe Raumdaten mit der Schuladresse verknüpft werden und so zumindest indirekt Auskunft über die Schulumgebung liefern. In diesem Beitrag wird nun gezeigt, dass sich über Häufigkeitsdichten von SGB-II-Empfängern Räume konstruieren lassen, die sich zur Charakterisierung von Schulstandorten nutzen lassen.

Kevin Isaac

Der Sozialindex und die Vorhersagekraft von Lernstandserhebungen in Nordrhein-Westfalen

Analysen zur Relevanz diagnostischer Testverfahren

1. Einführung

In Nordrhein-Westfalen werden mit Lernstandserhebungen seit 2004 standardbasierte Leistungstests durchgeführt. Im Rahmen dieses schulformübergreifenden Diagnoseverfahrens erhalten Schulen für eine aussagekräftige Einordnung ihrer Ergebnisse zusätzlich Vergleichswerte, die auf einem speziell dafür entwickelten Sozialindex basieren. Das sozialräumliche Einzugsgebiet von Schulen hat nicht nur einen gewissen Einfluss auf Schülerleistungen (z.B. Bonsen et al., 2010), sondern dieser wirkt sich gerade in benachteiligten Schulstandorten aufgrund differenzieller Leistungs- und Entwicklungsmilieus stärker aus (Baumert, Stanat & Watermann, 2006; s.a. Itzlinger-Bruneforth et al. in diesem Band). An diesen Schulen scheinen Merkmale wie Unterrichtspraktiken mehr Leistungsvarianz aufzuklären, als in sozial besser gestellten Standorten (vgl. Palardy, 2008).

Bisher existieren wenig empirisch abgesicherte Erkenntnisse über die pädagogische Arbeit von Schulen in solchen herausfordernden Lagen. Daher ist die Untersuchung von Schul- und Unterrichtsentwicklungsmaßnahmen gerade in diesem Kontext ein derzeit wichtiges Forschungsdesiderat (Racherbäumer, Funke, van Ackeren & Clausen, 2013a). Besondere Bedeutung gewinnt diese Frage vor dem Hintergrund der Tatsache, dass bei Lernstandserhebungen weniger die Ergebnisse selbst im Vordergrund stehen, sondern die fachliche Auseinandersetzung mit diesen innerhalb der Schulen. Ergebnisse aus Lernstandserhebungen sollten zunächst nur schulintern interpretiert werden. Die Ursachen von ggf. vorhandenen Abweichungen der schuleigenen Ergebnisse müssen immer im Hinblick auf schulübergreifende Vergleichswerte[1] im Rahmen der kollegialen Unterrichtsentwicklung differenziert aus einer pädagogischen und fachlichen Sichtweise analysiert werden. Unter anderem sind dabei weitere Aspekte zu berücksichtigen, z.B. welche Bereiche bereits Gegenstand des Unterrichts waren oder nach dem schulinternen Lehrplan erst nach der Durchführung der Lernstandserhebungen behandelt werden.

1 Bei den nordrhein-westfälischen Ergebnisrückmeldungen der Lernstandserhebungen erhalten Schulen Kompetenzniveauverteilungen (gestapelte Prozentbalkendiagramme) der Klassen bzw. Lerngruppen im Vergleich untereinander, der Schule, aller Schulen der Schulform und des gleichen Standorttyps (siehe Beispiele unter www.schulentwicklung.nrw.de).

Auch wenn die Ergebnisse aus Lernstandserhebungen somit keine allgemeingültigen, abgesicherten Schlüsse auf die Qualität des Unterrichts oder der schulischen Arbeit insgesamt zulassen, lässt sich zeigen, dass einige Schulen tatsächlich bessere Ergebnisse erreichen, als dies unter den gegebenen Rahmenbedingungen zu erwarten wäre (Burkard, Isaac & Pfeiffer, 2014). Solchen Schulen gelingt möglicherweise eine spezifische Förderung der Schülerinnen und Schüler, in dem u.a. verstärkt der Fokus auf pädagogische Anstrengungen zur Kompensation der Benachteiligung gelegt wird (vgl. Holtappels, 2008; Muijs, Harris, Chapman, Stoll & Russ, 2004).

Aus der Perspektive von Schul- und Unterrichtsentwicklung stellt sich in diesem Kontext die Frage, ob es Schulen unter Berücksichtigung ihrer Standortbedingungen gelingen kann, Schülerinnen und Schüler so zu fördern, dass spätestens bei der Zertifizierung am Ende der Regelschullaufbahn eine zumindest teilweise Entkopplung von der Lernausgangslage erreicht wird. Träfe dies zu, sollten sich keine oder lediglich geringe Zusammenhänge zwischen diagnostischen Tests und Abschlussprüfungen finden lassen. Finden sich entsprechende Zusammenhänge, könnte dies, in Bezug auf die konvergente Validität, auf eine Übereinstimmung der in beiden Tests gemessenen Inhalte hindeuten.

In diesem Beitrag soll die Vorhersagekraft – im Sinne der prognostischen und prädiktiven Validität von Lernstandserhebungen und mithilfe des Zusammenhangs zwischen den Leistungen von Lerngruppen bei Lernstandserhebungen in Klasse 8 und bei Abschlussklausuren der Zentralen Prüfungen in Klasse 10 – über mehrere Jahrgänge hinweg untersucht werden. Dafür sollen Jahrgangskohorten unter Berücksichtigung des Schulstandorts betrachtet werden (Isaac & Groot-Wilken, 2012).

2. Hintergrund und Ausgangslage

Ziel von Lernstandserhebungen

Lernstandserhebungen bzw. VERA[2] sind Diagnoseinstrumente, welche sich dadurch auszeichnen, dass sie Schulen kriteriale Vergleiche durch die Einordnung eigener Ergebnisse in einen durch Kompetenzniveaus definierten kriterialen Bezugsrahmen ermöglichen (Isaac, 2013a; Pant, 2013). In der Sekundarstufe I kommt ihnen die Aufgabe zu, den Erreichungsgrad der in den nationalen Bildungsstandards beschriebenen Kompetenzen, die am Ende der Jahrgangsstufe 10 (Mittlerer Schulabschluss, MSA bzw. Hauptschulabschluss, HSA) in den Fächern Deutsch, Mathematik und der ersten Fremdsprache erwartet werden, bereits zwei Schuljahre zuvor zu überprüfen. Mithilfe eines solchen Diagnosesystems erhalten Fachkonferenzen, Schülerinnen und Schüler sowie Eltern frühzeitig und ohne Zensurendruck Informationen über die in der Klasse vorliegenden Stärken und Schwächen der Schülerinnen und Schüler. Ein verantwortungsvoller Umgang mit den Ergebnissen führt im Idealfall zu unterrichtlichem

2 Der Begriff „Vergleichsarbeiten" (VERA) wird nicht in allen Ländern einheitlich eingesetzt. Die Bezeichnung „Lernstandserhebung" beschreibt den Charakter und die Zielbestimmung des Verfahrens in NRW treffender und wird in diesem Beitrag daher synonym mit „VERA" verwendet.

Handeln, in welchem das Lernen und Lehren gezielt auf in der Klasse vorhandene Kompetenzstände ausgerichtet wird (z.B. Helmke & Hosenfeld, 2003; Isaac, 2010; 2013b; Isaac, Halt, Hosenfeld, Helmke & Groß Ophoff, 2006; Peek, 2001). Die Teilnahme ist in Nordrhein-Westfalen in allen getesteten Inhaltsbereichen verpflichtend.

In einer Frage- und Antwortliste[3] der 67. Amtschefkommission „Qualitätssicherung in Schulen" vom 18.04.2013 wurde die Funktion der Ergebnisrückmeldung bei VERA wie folgt konkretisiert:

- Lieferung differenzierter Informationen über Stärken und Schwächen in der Klasse;
- wichtige Hinweise für die Lehrkraft zur Fokussierung von Inhalten des nachfolgenden Unterrichts;
- Identifikation von Bereichen, in welchen besondere Förderung angebracht ist.

Die Ergebnisrückmeldungen sollen im Sinne eines *Frühwarnsystems* für die Unterrichtsgestaltung vor dem Übergang in die Sek. I bzw. Sek. II genutzt werden.

Sozialindex für Schulen

Die Unterschiede bezüglich der Zusammensetzung der Schülerschaft nach der sozioökonomischen Herkunft und weiteren individuellen Voraussetzungen der Schülerinnen und Schüler führen zu erheblichen Disparitäten nach Standorttypen von Schulen und bestätigen die Existenz einer Bildungssegregation (Terpoorten, 2014). Diese Kontextabhängigkeit von Ergebnissen wirkt sich auch auf die Rezeption der Rückmeldungen aus (Helmke, 2004; Groß Ophoff, 2013). So ist die Intensität der Reflexion von sozialen Vergleichen bei der Ergebnisrückmeldung abhängig von der Fairness und Verständlichkeit der Darstellung der Vergleichswerte (Kühle, 2010). Damit es aufgrund von Ergebnissen aus diagnostischen Tests zu Veränderungen an Schulen kommen kann, sollten den Rezipienten – in erster Line den Fachkonferenzen – die Bedingungen und Einschränkungen bei der Interpretation der Daten klar vermittelt werden.

Einen besonderen Stellenwert gewinnt hierbei die Kontextuierung der ermittelten Schülerleistungen für die Rückmeldung eines sogenannten fair*er*en Vergleichs (Fiege, Reuther & Nachtigall, 2011; Hosenfeld & Isaac, 2006; Isaac & Hosenfeld, 2008). Dem liegt der Anspruch zugrunde, bei vergleichenden Rückmeldungen den Einfluss von Bedingungen mit zu berücksichtigen, die dem direkten pädagogischen Handeln der Lehrkräfte entzogen sind. Damit ist unmittelbar ein Potenzial zur Anregung zentraler Fragen verbunden, z.B. in Bezug auf die Qualität der eigenen pädagogischen Arbeit.

Um an unterschiedlichen Schulstandorten den aufgrund der sozialen Zusammensetzung bedingten Kompetenzunterschieden Rechnung zu tragen und entsprechend aussagekräftige soziale Vergleiche zuzulassen (vgl. Arnold, 2001; Nachtigall, Kröhne, Enders & Steyer, 2008), werden in Nordrhein-Westfalen allen Schulen sog. Standorttypen zugeordnet (Isaac, 2011; Kuthe, Bargel, Nagl & Reinhardt, 1979). Zur Bildung der Standorttypen wurde 2009 ein Sozialindex auf Basis georeferenzierter Sozialraumdaten (vgl. Schräpler & Jeworutzki in diesem Band) entwickelt.

3 http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/2013/2013_04_18-VERA_FragenundAntworten.pdf [Abruf am 15.09.2015]

Zentrale Prüfungen am Ende der Klasse 10

Seit dem Schuljahr 2006/07 werden in Nordrhein-Westfalen der MSA und HSA in einem Abschlussverfahren vergeben, an dem alle Schülerinnen und Schüler teilnehmen. Die Ziele des Prüfungsverfahrens sind größere Transparenz der Anforderungen, bessere Vergleichbarkeit von Leistungen sowie größere Gerechtigkeit bei der Abschlussvergabe.[4] Bisher liegen kaum Erkenntnisse darüber vor, ob sich seit der Einführung Zentraler Prüfungen auch das Leistungsniveau der Schülerinnen und Schüler verbessert hat. Für die Grundkurse in Mathematik gibt es Hinweise auf positive Effekte, die allerdings nicht auf andere Fächer übertragbar sind (Klein, Krüger, Kühn & van Ackeren, 2014).

Den theoretischen Rahmen sowohl für Lernstandserhebungen als auch für die Zentralen Prüfungen bilden die schulformübergreifenden Anforderungen, die in den nationalen Bildungsstandards beschrieben sind. Bei den Aufgaben der Zentralen Prüfungen wird dies durch einen Bezug auf die Kompetenzerwartungen der aktuellen Kernlehrpläne erreicht. Kernlehrpläne greifen ebenfalls die in den nationalen Bildungsstandards enthaltenen schulformübergreifenden Anforderungen auf, setzen Standards für die Ergebnisse von Lernprozessen und berücksichtigen gleichzeitig die Besonderheiten der einzelnen Schulformen und Bildungsgänge. Die Durchführung zentraler Prüfungen sowie diagnostischer, standardorientierter Leistungstests sind eine logische Konsequenz aus einem als *Ergebnis-* oder *Output-Orientierung* bezeichneten Steuerungsparadigma.

Fragestellung

Vor dem Hintergrund der dargestellten Ausgangslage und insbesondere der von der KMK beschriebenen Funktion von VERA (vgl. KMK, 2012) soll der Frage nachgegangen werden, ob Lernstandserhebungen, über ihre diagnostische Bedeutung als Impulsgeber für die Weiterentwicklung des Fachunterrichts hinaus, tatsächlich als Frühwarnsystem dienen können. Untersucht werden der Zusammenhang und die Vorhersagekraft von Lernstandserhebungen auf Prüfungsleistungen unter rechnerischer Konstanthaltung des sozialen Kontexts der Schulen in drei aufeinanderfolgenden Kohorten. Die Berücksichtigung des sozialen Umfelds wird anhand des Sozialindex für Schulen (Schräpler, 2009) vorgenommen.

Die Zusammenhänge zwischen Zensuren und Schulleistungstests sind seit den Arbeiten von Ingenkamp (1989) Gegenstand vielfältiger Untersuchungen gewesen (Baumert, Trautwein & Artelt, 2003; Hochweber & Hochweber, 2010; Klieme, 2003). Die dabei ermittelten niedrigen Korrelationen und teilweise erheblichen Überlappungen von Noten- mit Kompetenzstufen lassen Zweifel an der Objektivität der schulischen Notengebung aufkommen. Als Grundtenor kann konstatiert werden, dass die mittlere Bewertung einer Klasse wenig mit ihrem Leistungsniveau zu tun hat (Thiel & Valtin, 2002). Anders verhält es sich bei der Betrachtung der Validität von Abiturnoten und Studienerfolg. So konnten Blömeke et al. (2014) die prognostische Validität von Abitur- und Examensnoten in Mathematik bei Primarstufenlehrkräften nachweisen.

4 Weitere Informationen online unter www.standardsicherung.schulministerium.nrw.de/zp10 [Abruf am 21.10.2015].

Christiane Fiege

Faire Vergleiche bei Vergleichsarbeiten: Möglichkeiten und Grenzen

Zusammenfassung

Vergleichsarbeiten erheben den Leistungsstand von Schülerinnen und Schülern mittels standardisierter Tests, die einen Vergleich der Schülerleistungen zwischen verschiedenen Klassen ermöglichen. Daraus werden u.a. Aussagen über die Wirksamkeit bzw. Effekte der schulischen Arbeit auf Schülerleistungen – also des Ergebnisses von Lehre und Unterricht – abgeleitet, die Grundlage für Unterrichtsentwicklungsmaßnahmen sein sollen. Ein Problem bei solchen Vergleichen ist, dass Klassenunterschiede nicht nur aufgrund der pädagogischen Arbeit von Lehrkräften zustande kommen können, sondern auch aufgrund außerschulischer Einflussgrößen des Lernens (sog. *Kovariaten*), die sich dem pädagogischen Einfluss der Lehrkräfte entziehen. Solche Kovariaten können einerseits individuelle Ausgangsvoraussetzungen der Schülerinnen und Schüler sein (z.B. ihr Vorwissen), andererseits aber auch Kontextbedingungen des Lernens (z.B. Anteil von Kindern mit Migrationshintergrund im Klassenverband) umfassen. Deshalb können beispielsweise einfache Mittelwertsvergleiche der Testleistungen verschiedener Klassen als unfair angesehen werden. Für faire Vergleiche müssen die außerschulischen Einflussgrößen des Lernens mittels statistischer Adjustierungsverfahren berücksichtigt werden, um diesen Unterschieden Rechnung zu tragen. Hier unterscheiden sich die Bundesländer in ihrem Vorgehen zur Berechnung fairer Vergleiche: Es gibt verschiedene Adjustierungsverfahren, die im Kontext von Vergleichsarbeiten angewendet werden.

In diesem Beitrag werden die Bedeutung und Bedingungen fairer Vergleiche im Kontext der Ergebnisrückmeldung von Testergebnissen aus Vergleichsarbeiten dargestellt. Weiterhin werden die derzeit im Rahmen deutscher Vergleichsarbeiten verwendeten Adjustierungsverfahren, die zur Berechnung potenziell fairer Vergleiche angewendet werden, systematisiert. Schließlich soll die Bedeutung einer zentralen Kovariaten – des fachspezifischen Vorwissens – für die Berechnung fairer Vergleiche anhand einer empirischen Analyse von Daten aus Vergleichsarbeiten aufgezeigt werden.
Schlüsselwörter: Vergleichsarbeiten, faire Vergleiche, außerschulische Einflussgrößen des Lernens (Kovariaten), Adjustierungsverfahren

1. Einführung

Seit den unbefriedigenden Ergebnissen der PISA-Studie von 2000 (Baumert et al., 2001; OECD, 2001), in der Deutschland im Vergleich zu den anderen teilnehmenden Staaten in allen Fächern nur unterdurchschnittliche Leistungen erreichte, zeigte sich ein stark wachsendes öffentliches Interesse an Qualität und Effektivität im Bildungswesen. Vor allem die Leistungen der Schülerinnen und Schüler, d.h. der Output schulischer Arbeit, stehen seither verstärkt im Fokus der Öffentlichkeit. Zur Beurteilung und Sicherung der Qualität schulischer Arbeit beschloss die Kultusministerkonferenz die *Gesamtstrategie zum Bildungsmonitoring* (KMK, 2006), die Maßnahmen zur systematischen und wissenschaftlich fundierten Evaluation von Ergebnissen auf verschiedenen Ebenen des Bildungssystems umfasst. Neben der Teilnahme an internationalen Schulleistungsuntersuchungen, einer gemeinsame Bildungsberichterstattung von Bund und Ländern, der zentralen Überprüfung des Erreichens der Bildungsstandards im Ländervergleich zählt hierzu auch die Durchführung *landesweiter Vergleichsarbeiten* in allen 16 Bundesländern. Die Testergebnisse der Schülerinnen und Schüler in den Vergleichsarbeiten sollen die Lehrperson über den Leistungsstand der Klasse informieren, Stärken und Schwächen in den Leistungen aufzeigen und Grundlage für die Erarbeitung von Maßnahmen zur Unterrichts- und Qualitätsentwicklung sein. Ein gemeinsames, aber wenig explizit formuliertes Ziel landesweiter Vergleichsarbeiten ist die Beurteilung von Unterrichtseffekten auf Ebene einzelner Schulklassen anhand von Output-Daten: Evaluiert wird das Ergebnis von Lehre und Unterricht mittels Schülerleistungen. In Abgrenzung zur Unterrichtsqualitätsforschung geht es im Rahmen von Vergleichsarbeiten nicht um die standardisierte Erhebung und systematische Evaluation von Prozessmerkmalen des Unterrichtsgeschehens (vgl. z.B. Gräsel & Göbel, 2011). Der Begriff Unterrichtseffekt ist im Kontext von Vergleichsarbeiten aus diesem Grund sehr breit (und nicht weiter differenziert) zu verstehen, da keine Erhebung von Unterrichtsprozessmerkmalen stattfindet: Unterricht umfasst in diesem Kontext die Gesamtheit des pädagogischen Handelns bzw. Wirkens in einer Klasse. Um die Effektivität des Unterrichts beurteilen zu können, werden die Testergebnisse einer Klasse mit den Ergebnissen anderer Klassen verglichen. Die Variation der Testergebnisse zwischen den Schülerinnen und Schülern lässt sich jedoch nicht nur auf Effekte des Unterrichts zurückführen, sondern zu einem großen Teil auf weitere Faktoren wie bspw. den sozialen Hintergrund oder das Geschlecht der Schülerinnen und Schüler, die sich dem direkten pädagogischen Einfluss der Lehrkräfte entziehen (vgl. z.B. Isaac & Hosenfeld, 2008; Nachtigall, Kröhne, Enders & Steyer, 2008). Nur wenn die Ergebnisse von Vergleichsarbeiten adäquat analysiert werden, indem die außerschulischen Einflussgrößen des Lernens, d.h. *Kovariaten*, in der Auswertung der Testergebnisse Berücksichtigung finden, können aus den klassenspezifischen Ergebnisrückmeldungen – und zwar im Speziellen aus den sog. *fairen Vergleichen* – belastbare Aussagen über Unterrichtseffekte abgeleitet werden. Um dies zu gewährleisten, müssen jedoch eine Reihe sehr starker Bedingungen bzw. Annahmen erfüllt sein.

Der vorliegende Beitrag gliedert sich in drei Teile, in denen jeweils eine von drei Fragen im Fokus steht. Die Beantwortung dieser drei Fragen soll aus unterschiedli-

chen Perspektiven die Möglichkeiten, aber auch Grenzen fairer Vergleiche im Kontext von Vergleichsarbeiten aufzeigen:

1) *Wie sind faire Vergleiche definiert?* In diesem Abschnitt (Abschnitt 2) werden die Bedeutung sowie die Bedingungen fairer Vergleiche im Kontext von Vergleichsarbeiten dargestellt.
2) *Wie werden derzeit im Kontext von Vergleichsarbeiten (potenziell) faire Vergleiche erstellt?* In Abschnitt 3 geht es um die aktuelle Praxis der Bundesländer bei der Ergebnisauswertung und -rückmeldung im Kontext von Vergleichsarbeiten. Neben der Kategorisierung der derzeit verwendeten Adjustierungsstrategien erfolgt eine Zuordnung der Bundesländer zu diesen Strategien. Zudem werden verschiedene Kriterien diskutiert, hinsichtlich derer sich die verschiedenen Adjustierungsstrategien zur Berechnung potenziell fairer Vergleiche beurteilen lassen.
3) *Welche Bedeutung hat das fachspezifische Vorwissen bei der Berechnung fairer(er) Vergleiche?* Schließlich soll die Bedeutung der in der Schuleffektivitätsforschung wichtigsten Kovariaten – des fachspezifischen Vorwissens – für die Berechnung fairer(er) Vergleiche anhand einer empirischen Analyse aufgezeigt werden. Diese dritte Frage zur Bedeutung des fachspezifischen Vorwissens steht exemplarisch für die umfassendere Frage, welche Kovariaten für faire(re) Vergleiche berücksichtigt werden sollten.

2. Faire Vergleiche – Eine Begriffsbestimmung

2.1 Warum vergleichen wir? – Die Bedeutung von Bezugsnormen

Eine Grundvoraussetzung zur Evaluation von Unterrichtseffekten auf die Leistungen von Schülerinnen und Schülern ist die zuverlässige, valide und objektive Messung von Schülerleistungen sowie von außerschulischer Einflussgrößen des Lernens (vgl. z.B. Hartig, Klieme & Leutner, 2008). Die Erhebung bzw. Messung von Leistungsdaten mittels standardisierter Testverfahren ist jedoch keinesfalls ausreichend, um diese Leistungen beurteilen zu können: Die Lehrkraft einer Klasse kann anhand der erreichten Leistungsstände Fragen wie z.B. „Ist das von meiner Klasse erreichte durchschnittliche Testergebnis von 20 Punkten bei den Vergleichsarbeiten in Mathematik viel oder wenig?“ nicht beantworten. Um eine *Leistungsbeurteilung* zu ermöglichen, bedarf es daher stets des Vergleichs der Testergebnisse mit einem konkreten Kriterium oder einem Standard. Insgesamt lassen sich drei Arten von Vergleichsstandards unterscheiden, die auch als *Bezugsnormen* bezeichnet werden (vgl. Rheinberg, 2001; Watermann, Stanat, Kunter, Klieme & Baumert, 2003; Helmke & Hosenfeld, 2004; Helmke, Hosenfeld & Schrader, 2004): die ipsative, die kriteriale und die soziale Bezugsnorm. Diese stellen jeweils unterschiedliche Informationen des Vergleichs bereit, wobei sich die in diesem Beitrag fokussierten *fairen Vergleiche* insbesondere auf die soziale Bezugsnorm beziehen.

Die ipsative Bezugsnorm. Bei der ipsativen (oder auch individuellen) Bezugsnorm dient die Leistung einer spezifischen Klasse zu einem früheren Zeitpunkt als Referenz. Es werden verlaufsorientierte bzw. entwicklungsbezogene Vergleiche angestellt, d.h. es wird die Leistungsveränderung einer Klasse zwischen mindestens zwei Messzeitpunkten beurteilt. Die zugrundeliegende Frage lautet also: Wie verändert sich die durchschnittliche Testleitung einer Klasse über die Zeit? Wird diese besser, schlechter oder gibt es keine Veränderung?

Die kriteriale Bezugsnorm. Hier dient ein inhaltliches Kriterium der Leistungsbeurteilung. Kriteriale Vergleiche beziehen sich auf die Frage: Wie ist die Leistung einer Klasse im Vergleich zu einem zuvor festgelegten inhaltlichen Kriterium zu beurteilen? So wurden bspw. im Rahmen der PISA-Untersuchung (Baumert et al., 2001) voneinander abgrenzbare Kompetenzstufen inhaltlich definiert. Die Einordnung einer Schülerleistung in eine der Kompetenzstufen beschreibt somit das Kompetenzniveau dieser Schülerin bzw. dieses Schülers anhand charakteristischer Fähigkeiten, die mit dieser Kompetenzstufe assoziiert sind. Diese Vergleichsperspektive hat mit der Einführung der nationalen Bildungsstandards (KMK, 2004) an Bedeutung gewonnen, so dass die klassenbezogenen Rückmeldungen der Ergebnisse aus den Vergleichsarbeiten in vielen Bundesländern auch die Einordnung der Schülerleistungen hinsichtlich der Kompetenzstufen enthalten. Hier werden Mindest- und Regelstandards unterschieden, die jeweils angeben, was eine Schülerin bzw. ein Schüler mindestens bzw. durchschnittlich können soll. Dabei sind die „… von der Kultusministerkonferenz bisher vorgelegten Bildungsstandards […] als Regelstandards definiert" (KMK, 2004, S. 14).

Eine kriteriale Referenz wie die Kompetenzstufenverteilung einer Klasse (d.h. die relativen Häufigkeiten der erreichten Kompetenzstufen aller Schülerinnen und Schüler einer Klasse) ermöglicht jedoch keine Aussagen über Unterrichtseffekte. Um darüber Aussagen treffen zu können, benötigt man entweder Längsschnittinformationen (d.h. den Vergleich mit der Kompetenzstufenverteilung dieser Klasse zu einem früheren Zeitpunkt; ipsative Bezugsnorm), oder man vergleicht die Kompetenzstufenverteilung dieser Klasse mit der einer anderen (vergleichbaren) Klasse. Letzteres ist ein sozialer Vergleich mit kriterialen Normen. Der Nachteil von kriterialen Vergleichen liegt somit darin, dass diese per se keine Informationen über Unterrichtseffekte enthalten. Dies ist nur möglich, wenn man eine weitere Vergleichsdimension berücksichtigt wie bspw. mit der Kombination zwischen kriterialer und sozialer Bezugsnorm.

Die soziale Bezugsnorm. Im Gegensatz zu kriterialen Vergleichen erfolgt die Beurteilung einer Schülerleistung im Rahmen sozialer (oder auch verteilungsorientierter) Vergleiche auf Basis der Leistungsverteilung aller Schülerinnen und Schüler, deren Leistung erhoben wurde. Damit zielen soziale Vergleiche auf die Beantwortung der Frage: Wie ist die Leistung einer Klasse im Vergleich zur Leistung anderer Klassen, die von anderen Lehrkräften unterrichtet wurden, zu beurteilen?

Die soziale Bezugsnorm ist besonders geeignet, um Aussagen über die Unterrichtseffekte in einer Klasse im Vergleich zum Unterricht in anderen Klassen zu ermöglichen, denn ein Effekt ist stets der Unterschied zwischen zwei Bedingungen (hier:

Unterrichtsbedingungen; vgl. Holland, 1986). Mit anderen Worten: Indem eine Lehrkraft die durchschnittliche Leistung ihrer Klasse mit dem Leistungsoutput anderer *vergleichbarer* Klassen, die nicht von ihr unterrichtet wurden, vergleicht, kann sie beurteilen, wie viel besser bzw. schlechter ihre Schülerschaft in Folge des eigenen Unterrichts ist. Die Betonung liegt hier auf *vergleichbaren Klassen*: Die notwendige Voraussetzung für die Interpretation als Unterrichtseffekte sind *faire Vergleiche*, welche systematische und nicht auf den Unterricht attribuierbare Unterschiede zwischen Klassen berücksichtigen. Im folgenden Abschnitt präzisieren wir, was mit *fairen Vergleichen* gemeint ist und wie dieses Ziel der Vergleichbarkeit erreicht werden kann.

2.2 Faire Vergleiche in Vergleichsarbeiten

Bei der Rückmeldung von Ergebnissen aus Vergleichsarbeiten wird auch die soziale Bezugsnorm als Vergleichsstandard zur Beurteilung der Unterrichtseffektivität verwendet. So werden häufig die Testergebnisse einer Klasse mit den Testergebnissen anderer Klassen verglichen. Die Unterschiede zwischen den Ergebnissen beider Vergleichsgruppen lassen sich aber in aller Regel nicht ausschließlich auf die höhere bzw. niedrigere Effektivität des Unterrichts zurückführen, sondern diese sind in hohem Maße von weiteren Einflüssen abhängig (vgl. Isaac & Hosenfeld, 2008; Nachtigall & Kröhne, 2006; Nachtigall et al., 2008). Solche außerschulischen Einflussgrößen des Lernens (Kovariaten) sind bspw. das Vorwissen, der sozioökonomische Status, das Geschlecht, die Herkunftssprache oder die soziale Zusammensetzung der Klasse einer Schülerin bzw. eines Schülers. Kovariaten können demnach sowohl individuelle Schülermerkmale abbilden, als auch Kontextvariablen auf Klassen- oder Schulebene darstellen. Das zentrale Charakteristikum aller Kovariaten ist, dass sie dem pädagogischen Handeln bzw. dem Unterrichtsprozess zeitlich vorgeordnet sind. Somit sind Kovariaten von der Lehrkraft bzw. der Schule nicht beeinflussbar, haben aber ihrerseits einen Effekt auf die Schülerleistung. Es besteht daher mittlerweile ein allgemeiner Konsens, dass für faire Vergleiche statistische Adjustierungsverfahren verwendet werden müssen (z.B. OECD, 2008; Watermann & Stanat, 2004; Wegscheider, 2004). Bei dem Begriff *Fairness* geht es also im Rahmen dieses Beitrags nicht um die Fairness bei der Testung selbst (vgl. Arnold, 1999; Moosbrugger & Kelava, 2007), sondern um die Fairness von Vergleichen der resultierenden Testergebnisse.

Was ist das Ergebnis einer solchen statistischen Adjustierung? Vereinfacht ausgedrückt und bezogen auf die Datenauswertung bei Vergleichsarbeiten zielen statistische Adjustierungsverfahren auf die Beantwortung der folgenden Frage: Welches Testergebnis hätte eine konkrete Klasse unter ansonsten identischen Ausgangsbedingungen erzielt, wenn eine beliebige andere Lehrerin bzw. ein beliebiger anderer Lehrer den Unterricht in dieser Klasse gestaltet hätte (Ceteris-paribus-Klausel; vgl. Mill, 1843). Die Differenz dieses adjustierten Wertes einer Klasse zum beobachteten Testergebnis kann in diesem Fall ursächlich dem Effekt des Unterrichts in dieser Klasse zugeschrieben werden (vgl. Steyer, Partchev, Kröhne, Nagengast & Fiege, in Vorbereitung; Fiege, 2013). Eine solche kausale Attribution ist allerdings nur möglich, wenn die folgenden

zwei (sehr starken) Annahmen erfüllt sind: Es müssen (a) einerseits *sämtliche* Kovariaten, die zusätzlich zum Unterricht das Testergebnis beeinflussen *und* deren Verteilung sich zwischen den Klassen unterscheidet, in die Datenanalyse einbezogen werden (korrekte Kovariatenselektion) und (b) andererseits muss das korrekte statistische Modell zur Analyse der Testergebnisse verwendet werden (korrekte Modellselektion).

Aus den bisherigen Betrachtungen ergeben sich zwei Fragen, die uns in den folgenden zwei Abschnitten beschäftigen werden:

1) Welche statistischen Adjustierungsverfahren werden im Kontext von Vergleichsarbeiten angewendet? Auf welche Weisen versucht man zu fairen Vergleichen zu gelangen?
2) Welche Kovariaten sollten auf welche Weise im Rahmen der Adjustierungsverfahren berücksichtigt werden (Kovariaten- und Modellselektion)?

3. Adjustierungsverfahren bei Vergleichsarbeiten – Wie werden Kontextbedingungen des Lernens beim Vergleich von Testergebnissen berücksichtigt?

Der folgende Abschnitt gibt einen Überblick der derzeit verwendeten Adjustierungsstrategien bei Vergleichsarbeiten sowie von deren Implementierung in den einzelnen Bundesländern Deutschlands. Eine detaillierte Darstellung der im Rahmen der Ergebnisrückmeldung von Testergebnissen aus Vergleichsarbeiten verwendeten Adjustierungsstrategien findet sich bei Fiege, Reuther und Nachtigall (2011). Für eine Einordnung in den internationalen Kontext sei an dieser Stelle auf Fiege (2013) verwiesen. Abschließend werden mögliche Bewertungskriterien dieser Adjustierungsstrategien diskutiert.

3.1 Vier Kategorien von Adjustierungsstrategien bei Vergleichsarbeiten

Bei der Ergebnisauswertung und -rückmeldung von Testergebnissen aus Vergleichsarbeiten werden unterschiedliche statistische Adjustierungsstrategien angewendet, die das Problem der Verzerrung der geschätzten Unterrichtseffekte durch Kovariaten zu berücksichtigen suchen. Die nachfolgend beschriebenen Strategien basieren auf der Kategorisierung nach Fiege et al. (2011)[1]. Die Autoren unterscheiden insgesamt vier Kategorien von Adjustierungsstrategien. Allen Strategien gemeinsam ist, dass sich die Adjustierung jeweils auf die Berechnung eines Referenzwertes bezieht. Unabhängig

1 Die Systematik von Adjustierungsstrategien wurde im Rahmen des Projekts „Faire Vergleiche in der Schulleistungsforschung – Methodologische Grundlagen und Anwendung auf Vergleichsarbeiten" an der Friedrich-Schiller-Universität in Jena erstellt. Dieses Projekt wurde von 2009 bis 2011 vom Bundesministerium für Bildung und Forschung (BMBF) gemäß dem Rahmenprogramm zur Förderung der empirischen Bildungsforschung finanziert. Die Systematik wurde in Anlehnung an Nachtigall et al. (2008) mittels einer Quellenanalyse vorhandener Literatur bzw. von Internetquellen bundeslandspezifischer Vergleichsarbeiten sowie Befragung der Landesinstitute und Ministerien (Stand: Dezember 2009) erstellt.

Klaudia Schulte, Johannes Hartig & Marcus Pietsch

Berechnung und Weiterentwicklung des Sozialindex für Hamburger Schulen[1]

In Hamburg gibt es seit 1996 einen Sozialindex für Grundschulen und weiterführende Schulen mit Sekundarstufe I. Auch einige weitere deutsche Bundesländer wie Bremen, Berlin und Nordrhein-Westfalen nutzen verschieden operationalisierte Indikatoren der sozialen Belastung, um gezielt schulische Ressourcen einzusetzen (Tillmann & Weishaupt, 2015; siehe auch den Beitrag von Weishaupt in diesem Band). Durch den Einsatz von Sozialindices sollen Schulen in schwierigen Lagen mit zusätzlichen Mitteln unterstützt werden, „um Effekte der Schülerzusammensetzung kompensieren und chancenausgleichend wirken zu können: Gleiche Bildungschancen sollen mit ungleichem Mitteleinsatz erreicht werden." (Tillmann und Weishaupt, 2015, S. 7). Auch in Hamburg beschreibt der Sozialindex, seit 2006 (Bos, Pietsch, Gröhlich & Janke, 2006) basierend u.a. auf der Kapitaltheorie von Bourdieu (1982; 1983), die sozialen Rahmenbedingungen der Schulen. Die auf dem Index basierende Zuordnung zu sechs abgestuften Belastungsgruppen hat Auswirkungen auf diversen Ebenen: Auf der einen Seite determiniert der Sozialindex unterschiedliche Ressourcenallokationen (z.B. kleinere Klassenfrequenzen oder höhere Sprachfördermaßnahmen für Schulen mit niedrigeren Indices). Auf der anderen Seite wird der Sozialindex in Hamburg auch in weiteren Zusammenhängen genutzt: bei der Bildung repräsentativer Stichproben im Rahmen von wissenschaftlichen Untersuchungen und Evaluationen (z.B. bei der Auswahl einer repräsentativen Kernstichprobe von Schulen pro Schuljahr für die Schulinspektion), bei der Berechnung und Rückmeldung von Vergleichswerten („fairer Vergleich") für die schulbezogenen Ergebnisrückmeldungen im Rahmen von KERMIT oder bei der Bildung von Vergleichsgruppen im Kontext der Bildungsberichterstattung. Hamburg reagiert damit bildungspolitisch auf den über die Jahre leicht entkoppelten, aber auch aktuell noch beschriebenen Zusammenhang zwischen der sozialen Herkunft und dem Kompetenzerwerb sowie damit verbundenen Bildungschancen in Deutschland, wie auch in den PISA-Ergebnissen gezeigt werden konnte (Klieme et al., 2010).

Im folgenden Artikel werden, nach einer theoretischen Einführung in das zugrunde liegende Konzept der sozialen Belastung, Durchführung und Methode der Berechnung des Sozialindex dargestellt sowie aktuelle Überlegungen zur Weiterentwicklung des Sozialindex aufgeführt, die sich aus den inzwischen langjährigen Erfahrungen mit der Verwendung eines Sozialindex für Ressourcenallokationen ergeben haben.

1 Überarbeiteter und ergänzter Nachdruck von Schulte, Hartig und Pietsch (2014).