

Nationale und internationale Schulleistungsstudien: was können sie leisten, wo sind ihre Grenzen?

Baumert, Jürgen
Köller, Olaf

Please note:

This paper is a preprint of an article submitted to **Pädagogik 50 (6), 12-18**, therefore there may be minor differences between the two versions.

The copyright of this electronic version remains with the author and the Max Planck Institute for Human Development.

Vorbemerkung

In der Bundesrepublik Deutschland gibt es keine Tradition der dauerhaften Beobachtung von Erträgen institutionalisierter Bildungsprozesse. Dies gilt für den Fachleistungsbereich ebenso wie für die breiten Bereiche der sozial-kognitiven und motivationalen Entwicklung. Ein Hauptinteresse der Erziehungswissenschaft lag in den letzten 10 bis 15 Jahren - mit gutem Grund - auf der Entwicklung und Erprobung von Modellen zur Optimierung der Arbeit in Einzelschulen und dem Entwurf didaktischer Modelle und deren Einführung in die Unterrichtspraxis. Die Selbstvergewisserung über das insbesondere im Unterricht Erreichte und Erreichbare trat demgegenüber in den Hintergrund.

In dieser Situation waren Untersuchungen, die auf einer breiten empirischen Basis die Beschreibung und Analyse der Erträge fachlichen Lernens und schulischer Sozialisationsprozesse in den Mittelpunkt rückten, unzeitgemäße Unternehmungen. Dies gilt sowohl für die 1991/92 begonnene und noch laufende Längsschnittstudie "[Bildungsverläufe und psychosoziale Entwicklung im Jugendalter \(BIJU\)](#)" als auch für die von 1994 bis 1996 durchgeführte "[Dritte Internationale Mathematik- und Naturwissenschaftsstudie](#)" - [Third International Mathematics and Science Study \(TIMSS\)](#). Um so überraschender war die breite Diskussion, die schon die deskriptiven Befunde beider Untersuchungen ausgelöst haben. Allein die Bereitstellung von Basisinformationen über Ertragslagen löste pädagogische Nachdenklichkeit, aber natürlich auch, wie es in einem normativ besetzten Feld nicht anders sein kann, Kontroversen aus. Die Befunde der "Dritten Internationalen Mathematik- und Naturwissenschaftsstudie" zu den Fachleistungen auf der 7. und 8. Jahrgangsstufe sind mittlerweile breit dargestellt (Baumert, Lehmann u.a. 1997; Baumert 1998). Die Darstellung der Ergebnisse für die Sekundarstufe II - TIMSS Untersuchungspopulation 3 - wird in Kürze folgen (Baumert, Bos und Lehmann, in Vorbereitung). Für die stärker analytisch angelegte Längsschnittstudie "Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU)" liegt mittlerweile eine Reihe von Aufsätzen und Monographien vor (s. die Zusammenstellung in MPI 1996; Heyn u.a. 1997; Köller 1998; Schnabel 1998). Der für ein breiteres Publikum gedachte deskriptive Bericht über die schulischen Entwicklungsverläufe wird Ende dieses Jahres erscheinen.

Wir wollen im folgenden weder das bereits Vorgelegte noch einmal ausbreiten noch die ausstehenden Berichte in Einzelheiten vorwegnehmen. Unsere Absicht ist es vielmehr, zu einem besseren Verständnis der bislang veröffentlichten Befunde beizutragen, indem wir die Aussagekraft der vorgelegten empirischen Ergebnisse, aber auch deren Grenzen diskutieren. In einem ersten Schritt wollen wir noch einmal die wichtigsten, teilweise erwartungswidrigen Ergebnisse der beiden Studien zusammenfassen. Im zweiten Schritt werden wir einige Problembereiche der Studien, die auch in der Öffentlichkeit mehr oder weniger sachkundig behandelt worden sind, skizzieren, ohne den Leser mit Fachliteratur oder Statistik zu erdrücken.

Was sind BIJU und TIMSS?

Die Studie "Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU)" ist eine Längsschnittuntersuchung mit zwei unterschiedlichen Altersgruppen, die im Jahre 1991/92 mit der Umstellung der Schulorganisation in den neuen Ländern auf ein gegliedertes Schulsystem begonnen wurde und die Entwicklung von Jugendlichen und jungen Erwachsenen im kognitiven und psychosozialen Bereich in Abhängigkeit von unterschiedlichen familiären und institutionellen Entwicklungsbedingungen untersucht. Die Studie erfaßt Leistungen in mehreren Unterrichtsfächern, einen breiten Kranz von kognitiven, sozial-kognitiven, motivationalen und weiteren Persönlichkeitsmerkmalen, Lern- und Entwicklungsbedingungen institutioneller Art sowie Qualitätsmerkmale von Unterrichtsprozessen.

Die Studie basiert auf für alle Schulformen repräsentativen Stichproben in den Bundesländern Nordrhein-Westfalen, Mecklenburg-Vorpommern, Sachsen-Anhalt und Berlin (Ost und West). Sie wurde in Zusammenarbeit zwischen dem Max-Planck-Institut für Bildungsforschung und dem Institut für die Pädagogik der Naturwissenschaften begonnen. Sie wird zur Zeit am Max-Planck-Institut für Bildungsforschung weitergeführt.

Die "Dritte Internationale Mathematik- und Naturwissenschaftsstudie" (TIMSS) ist eine internationale vergleichende Schulleistungsstudie, die von der International Association for the Evaluation of Educational Achievement (IEA) durchgeführt wurde. Die IEA ist eine internationale Forschungsorganisation, der überwiegend Regierungseinrichtungen der Mitgliedsstaaten angehören. Mit TIMSS wurden die Mathematik- und Naturwissenschaftsleistungen von Schlüsseljahrgängen in der Grundschule und in den Sekundarstufen I und II zum ersten Mal gleichzeitig untersucht.

Anliegen der Untersuchung ist es, Informationen zur Dauerbeobachtung von Schulsystemen zur Verfügung zu stellen. Infolgedessen ist die internationale Studie primär deskriptiv angelegt. In Deutschland wurde die Untersuchung als Längsschnittstudie durchgeführt, die durch gemeinsame Instrumente im Fachleistungsbereich und im nicht primär kognitiven Bereich mit BIJU verzahnt ist. Damit wurden nicht nur die Analysemöglichkeiten verbreitert, sondern es bot sich die Chance der wechselseitigen Bestätigung der Befunde beider Studien.

TIMSS wurde durch Mittel des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie gefördert. Ebenso wurde eine Teilstudie zur politischen Sozialisation im Rahmen von BIJU durch Mittel desselben Zuwendungsgebers finanziert. Weder bei TIMSS noch bei BIJU handelt es sich jedoch um Auftragsforschung. Die Studie wurde allein in Verantwortung der beteiligten Institute und Wissenschaftler durchgeführt. Auch im Rahmen der Genehmigung durch die Kultusministerien, denen die Projektbeschreibungen und alle Untersuchungsinstrumente zur Prüfung vorlagen, gab es keinerlei Auflagen.

Zentrale Befunde aus BIJU und TIMSS

Folgende Ergebnisse aus TIMSS sollen besonders herausgehoben werden:

- Die mathematisch-naturwissenschaftlichen Leistungen von Schülern der 7. und 8. Jahrgangsstufe liegen in der Bundesrepublik unter den durchschnittlichen Leistungen der meisten west-, nord- und osteuropäischen Nachbarstaaten. Die Leistungsunterschiede haben teilweise gravierende Ausmaße.
- Ein erheblicher Prozentsatz der Schüler der untersuchten Altersgruppe erreicht das für einen erfolgreichen Übergang in die berufliche Erstausbildung notwendige Niveau mathematisch-naturwissenschaftlicher Grundbildung nicht.
- Der mathematisch-naturwissenschaftliche Unterricht baut weniger systematisch auf vorher gelernten Inhalten auf. Die Wissenszuwächse sind im internationalen Vergleich gering.
- Es gibt in Deutschland erhebliche regionale Leistungsdifferenzen, die je nach Schulform unterschiedlich ausfallen. In der Regel verbinden sich mit den unterschiedlichen Fachleistungen auch unterschiedliche Leistungsstandards.
- Schulformen unterscheiden sich in ihren Fachleistungen erwartungsgemäß deutlich. Es gibt jedoch erhebliche Überlappungsbereiche der Leistungsverteilungen.

Trotz der deskriptiven Anlage von TIMSS lassen sich einige begründete Vermutungen über Ursache-Wirkungs-Zusammenhänge formulieren:

1. Es gibt keinen einzelnen durchschlagenden Kausalfaktor, der für die teilweise immensen Leistungsunterschiede zwischen den Schülern der Teilnehmerstaaten verantwortlich wäre.
2. Aus den Befunden von TIMSS läßt sich kein strukturelles Argument für oder gegen integrierte Schulformen gewinnen. Dies gilt für beide Argumentationsrichtungen gleichermaßen (z.B. Kraus 1998; Ratzki, in diesem Heft). Strukturdebatten verfehlen die zentrale Problemstellung der Optimierung vorhandener Bildungsgänge.
3. Systematische Erklärungen für Leistungsunterschiede sind wahrscheinlich in folgenden Bereichen zu suchen:
 - in der generellen Wertschätzung von Bildung und schulischem Lernen und der damit verbundenen Bereitschaft, persönliche Ressourcen zu investieren;
 - in der spezifischen Lernkultur eines Schulwesens und in der Bedeutung, die dem kontinuierlichen Wissenserwerb und der damit verbundenen Anstrengung und Ausdauer zugemessen wird;
 - in der gesellschaftlichen und schulischen Wertschätzung bestimmter Fächer
 - und schließlich in der Qualität des Fachunterrichts selbst.

Die Befunde von BIJU und TIMSS stützen sich in vielen Bereichen wechselseitig. Als zentrale Ergebnisse von BIJU seien folgende Punkte genannt:

- Schüler im Einheitsschulsystem der ehemaligen DDR erreichten in allen vergleichbaren Untersuchungsfächern (Mathematik, Naturwissenschaften, Deutsch) tendenziell oder deutlich bessere Leistungsergebnisse als Schüler des gegliederten Schulsystems in den alten Bundesländern. Dies war überwiegend auf Förderungserfolge bei leistungsschwächeren Schülern zurückzuführen.
- Das Gymnasium ist sozial immer noch die homogenste, leistungsmäßig aber mittlerweile die heterogenste Schulform, während etwa die Schülerschaft der Gesamtschulen sozial heterogen, aber leistungsmäßig wider Erwarten relativ homogen ist.
- Unterschiede im Leistungsstand von Schülern lassen sich in einem erheblichen Ausmaß auf institutionelle Merkmale - Schulformzugehörigkeit, Schulzugehörigkeit und in geringerem Maß auch Klassenzugehörigkeit - zurückführen. Diese institutionellen Unterschiede sind überwiegend selektionsbedingt.
- Unterschiede in der Leistungsentwicklung sind ganz überwiegend institutionsbedingt. Im Kernbereich fachlichen Lernens hat die Schule - trotz aller geheimen Miterzieher - praktisch eine Monopolposition.
- Dies gilt nicht für den psychosozialen Bereich, in dem sich unterschiedliche institutionelle Bedingungen nur in sehr begrenztem Maße auswirken. Der Durchgriff auf die Persönlichkeit gelingt offensichtlich in weit geringerem Ausmaß, als dies pädagogischer Optimismus oder pädagogische Befürchtungen unterstellen. Das heißt nicht, daß die Schule für die psychosoziale Entwicklung unwichtig wäre, allerdings sind die Chancen für erfolgreiche pädagogische Interventionen im Bereich fachlicher Förderung größer.

- Im Widerspruch zu weitverbreiteten Annahmen ist das Gymnasium keineswegs eine primär selektionsorientierte Schulform. Vielmehr bietet das Gymnasium sowohl im akademischen als auch im psychosozialen Bereich Schülern unterschiedlicher Eingangsvoraussetzungen die vergleichsweise günstigsten Entwicklungsbedingungen. Die Effekte sind im Leistungsbereich sehr groß, im psychosozialen Bereich klein, aber nicht unbedeutend. Haupt- und Gesamtschule scheinen im Durchschnitt ähnliche Entwicklungsmilieus anzubieten.
- Gute Leistungsentwicklung geht auf der individuellen Ebene tendenziell auch mit wünschenswerter psychosozialer Entwicklung einher. Auf der Schulebene handelt es sich - bei einer Betrachtung über alle Schulformen hinweg - geradezu um Kuppelprodukte: Die akademische und die soziale Kultur einer Schule stützen sich in der Regel wechselseitig, und zwar auch dann, wenn die unterschiedliche soziale und intellektuelle Zusammensetzung der Schülerschaft in Rechnung gestellt wird.

In den folgenden Abschnitten soll der methodische Ansatz, der hinter den BIJU- und TIMSS-Ergebnissen steht, näher beleuchtet werden.

Probleme der Messung von Schulleistungen

Die Vertrauenswürdigkeit von Ergebnissen aus großen nationalen und internationalen Schulleistungstudien steht und fällt mit den gewählten Testaufgaben, ihrer Lehrplan- und Unterrichtsvalidität, der angemessenen Repräsentation der zu erfassenden Fähigkeiten, ihrem Antwortformat und der für die untersuchten Jahrgangsstufen geeigneten Formulierung. Regelmäßige Kritikpunkte an Schulleistungstests beziehen sich auf die curriculare Validität, das Antwortformat und damit oft verbunden die getesteten kognitiven Fähigkeiten und Fertigkeiten (im Überblick s. Baumert, in Druck).

Zum Problem der curricularen Validität

Die curriculare Validität eines Leistungstests kann einmal im Hinblick auf das offizielle Curriculum eines Faches - in Deutschland legt dies der Lehrplan fest - und zum anderen im Hinblick auf tatsächlich im Unterricht behandelten Stoff bestimmt werden. Ein Schüler kann sich nur die Stoffe aneignen, die im Unterricht auch behandelt werden. In Schulleistungstudien tritt fast immer das Dilemma auf, daß keine speziellen Leistungstests für den Unterricht einzelner Klassen konstruiert werden können. Leistungstests werden sich mit ihren Aufgaben daher in der Regel zunächst am Lehrplan orientieren. In TIMSS wurden dementsprechend für die Klassenstufen 7 und 8 etwa 150 Aufgaben in den Mathematiktest und 130 Aufgaben in den Naturwissenschaftstest aufgenommen, mit denen die Stoffgebiete der offiziellen Curricula der beteiligten Staaten möglichst breit abgedeckt wurden. Anschließende Lehrerbefragungen zum Unterricht zeigten, daß die Aufgaben im allgemeinen nicht nur das geplante, sondern auch das unterrichtete Curriculum gut repräsentierten. So ergab sich für die 8. Klasse in Deutschland, daß 89 Prozent der Mathematik-, 83 Prozent der Biologie- und 77 Prozent der Physikaufgaben Stoff thematisierten, der in der Jahrgangsstufe 8 oder vorher in nahezu allen untersuchten Klassen Gegenstand des Unterrichts gewesen war (vgl. Baumert, Lehmann u.a. 1997).

Die gute Korrespondenz von Lehrplan, Unterricht und Testaufgaben hat die Akzeptanz von TIMSS sicherlich gefördert. Außerdem zeigen Faktorenanalysen für die deutsche TIMSS-Stichprobe, daß sich die Mathematikaufgaben aus sechs verschiedenen Sachgebieten (Zahlen und Zahlenverständnis, Messen und Maßeinheiten, Algebra, Geometrie, Proportionalität, Darstellung und Analyse von Daten/Wahrscheinlichkeitsrechnung) im wesentlichen auf eine einzige zugrundeliegende Dimension zurückführen lassen, die man als komplexes mathematisches Fähigkeitssyndrom bezeichnen kann (Köller 1998).

Zum Problem des Antwortformats

Aufgaben in Schulleistungstests sehen in der Regel ein Antwortformat mit Mehrfachwahlen - multiple choice items (MC) - vor, bei dem ein Schüler aus 2 bis 5 Antwortalternativen die richtige Lösung einer Aufgabe herausfinden und ankreuzen muß. Häufig ist kritisiert worden, daß MC-Aufgaben für Schülerinnen und Schüler in deutschen Schulen ungewohnt seien, da Haus- und Klassenarbeitsaufgaben immer ein offenes Format hätten. Aus wissenschaftlicher Sicht ist über lange Zeit auch eingewandt worden, daß Mehrfachwahlaufgaben im wesentlichen nur reproduktive Leistungen, nicht aber produktive Fähigkeiten - schlußfolgerndes Denken, komplexe Operationen oder gar Problemlösen - erfaßten. Vermißt wurde bei MC-Aufgaben auch die Möglichkeit, systematische Fehleranalysen durchzuführen. Als bessere Alternative gelten offene Antwortformate.

Ein Blick in die jüngere methodologische Literatur zeigt allerdings, daß das Antwortformat für globale Leistungsvergleiche, die nicht darauf angelegt sind, Lernprozesse mikroanalytisch zu untersuchen, nachgeordnete Bedeutung hat. Dennoch wurde in TIMSS der Kritik am MC-Format Rechnung getragen, indem zwei Drittel der Bearbeitungszeit für die Lösung von MC-Aufgaben und ein Drittel der Zeit für Fragen mit offenem Format zur Verfügung standen. Die Analysen zeigen, daß das offene Format für die Validität des internationalen Vergleichs keinen Gewinn bedeutet. Offene Aufgaben sind generell nur etwas schwerer, da die Möglichkeit intelligenten Ratens entfällt. Schließlich ist es ein Irrglaube, man könne im MC-Format keine anspruchsvollen Aufgaben entwickeln, mit denen selbständiges Denken, methodische Fähigkeiten - etwa beim naturwissenschaftlichen Experimentieren - oder Problemverständnis erfaßt werden. In TIMSS ist es gelungen, einen großen Teil der mathematischen und naturwissenschaftlichen Aufgaben, die Anwendung des Gelernten in vergleichbaren Kontexten, Transferleistungen, methodisches Know-how oder kreative Problemdefinitionen verlangen, im Mehrfachwahlformat zu gestalten.

Wissenserwerb und das "Andere Lernen": fächerübergreifendes Denken, Anwenden, Forschen, Problemlösen und Selbstorganisation des Lernens

Die beste Voraussetzung für kumulative Lernprozesse und selbständiges erfolgreiches Weiterlernen sind nicht formale Schlüsselqualifikationen, sondern eine solide und gut organisierte Wissensbasis im jeweiligen Schulfach. Damit sind nicht vereinzelte und mechanisch erworbene Kenntnisse gemeint, sondern ein intelligent geordnetes,

in sich vernetztes, in verschiedenen Situationen erprobtes und flexibel anpaßbares Wissen. Dazu gehören Fakten-, Konzept-, Theorie-, Methoden und Prozeßwissen gleichermaßen. Bei steigender Schwierigkeit und Komplexität der kognitiven Anforderungen von Aufgaben- und Problemstellungen nimmt die Bedeutung des spezifischen Vorwissens für deren erfolgreiche Bearbeitung zu. Der Aufbau von intelligentem Wissen ist in der Regel ein langjähriger und übungsintensiver Prozeß, der nicht nur Anstrengung und Ausdauer verlangt, sondern zugleich auch die systematische Schulung von Elementen des "Anderen Lernens" einschließt: also Anwenden, Übertragen, Umstrukturieren und Integrieren. Umgekehrt sind fächerübergreifendes Denken, Anwenden, Forschen, methodisches Vorgehen oder gar die Selbstorganisation des Lernens ohne den Erwerb solider Wissensgrundlagen schwer vorstellbar.

Schulleistungsstudien tun gut daran, diese Zusammenhänge prüfbar zu machen. Im Rahmen von TIMSS läßt sich zeigen, daß in Deutschland bei Aufgaben, die ein selbständiges Anwenden, das Finden kreativer Lösungen, Übertragungsleistungen oder ein prozessuales Verständnis naturwissenschaftlichen Arbeitens verlangen, die Leistungsrückstände gegenüber erfolgreicheren Ländern größer sind, als dies bei Routineaufgaben der Fall ist (in der Schweiz etwa trifft dies, wie Ramseier (1997) zeigen konnte, nicht zu). Wir konnten keine Beispiele - auch nicht in der Gesamtschule - identifizieren, wo Schwächen im Bereich des Basiswissens durch besondere Leistungen im Bereich des Anwendens und Problemlösens kompensiert werden konnten. Dieses Muster wiederholt sich in der BIJU-Studie, wenn man die Verwendung von Lernstrategien - wichtigen Kompetenzen für selbstreguliertes Lernen - analysiert.

Probleme der Erfassung des sozialen Verhaltens und anderer Merkmale

Schule ist ein Unternehmen, das unterschiedliche und teilweise konkurrierende Zielsetzungen gleichzeitig verfolgt. In gelingenden Bildungsprozessen vollzieht sich nach allgemein geteilter Vorstellung die gleichmäßige Entwicklung kognitiver, sozialer und emotionaler Fähigkeiten.

Schulleistungsstudien, die sich auf die Aspekte fachlichen Lernens konzentrieren, setzen sich naheliegenderweise der Kritik aus, ein einseitiges Bild von der Ertragslage der Schule zu zeichnen, da die im engeren Sinne erzieherischen Leistungen der Schule unberücksichtigt blieben. Eingeschlossen ist der Gedanke, daß es dann durchaus Kompensationsmöglichkeiten für geringere Erträge in einem Bereich durch gute Ergebnisse in einem anderen gebe.

Eine zentrale Fragestellung von BIJU ist die Prüfung dieser Annahme. In der Längsschnittstudie wurde deshalb ein breiter Kranz nicht primär kognitiver Variablen erfaßt. Wie wurden diese Merkmale erhoben? Leistungstests liefern Verhaltensstichproben, wenn auch in relativ artifiziellen Situationen. Dennoch kann man zeigen, daß die im Test gemessenen Fachleistungen substantiell mit Außenkriterien korrelieren. Für die Erfassung nicht primär leistungsbezogener Merkmale ist dagegen der Königsweg die Selbstauskunft. Die Befragten werden als Experten für das Nachdenken über die eigene Person ernstgenommen. Solche Selbstauskünfte können trotz aller Antwortverzerrungen relativ zuverlässig mit Fragebögen eingeholt werden. Viele der auf diese Weise erfaßbaren Personenmerkmale sind Ergebnisse des Lernens in sozialen Kontexten, die ihrerseits wiederum verhaltensregulierend wirken. So wissen wir etwa, daß Selbstwirksamkeitsüberzeugungen die Auswahl von Zielen, die Wahl des Anspruchsniveaus, Anstrengung, Ausdauer und das Durchhalten bei Widerständen beeinflussen. Oder: Die Ausbildung prosozialer Motive und die Entwicklung moralischer Urteilskraft sind einerseits vom sozialen Entstehungskontext abhängig und schlagen sich andererseits etwa in der Respektierung von Andersdenkenden oder im Eintreten für Minoritäten nieder. Mit diesen selbstberichteten Personenmerkmalen erfaßt man also sowohl Ergebnisse sozialen Lernens als auch Determinanten sozialen Handelns.

Soziales Verhalten durch direkte Beobachtung zu erfassen, ist im Rahmen von breit angelegten Untersuchungen mit großen Fallzahlen nicht möglich. Aber selbst bei Studien mit kleineren Stichproben oder Einzelfalluntersuchungen ist die Beobachtung nicht immer das zuverlässigste Erhebungsmittel. So wissen wir aus der Unterrichtsforschung, daß Selbstangaben von Schülern über ihre Aufmerksamkeit bessere Prädiktoren für Schulleistung sind als Beobachtungsdaten. Dies gilt generell für seltenes und interpretationsbedürftiges Verhalten. Im Rahmen von BIJU wurde sozial erwünschtes und unerwünschtes Verhalten innerhalb und außerhalb der Schule durch Selbstangaben von Schülern erfaßt.

Schulformen als unterschiedliche Entwicklungsmilieus

Sowohl anhand von TIMSS als auch von BIJU läßt sich zeigen, daß die Leistungsentwicklung an den einzelnen Schulformen unterschiedlich verläuft. Die stärksten Leistungszuwächse zeigen sich auf dem Gymnasium, gefolgt von der Realschule, dann der Gesamtschule und schließlich der Hauptschule. Die unterschiedliche Leistungsentwicklung wird in BIJU besonders deutlich, da es hier möglich ist, einen Entwicklungszeitraum von vier Jahren in den Blick zu nehmen. Die BIJU-Daten zeigen auch, daß sich ähnliche Muster für die sozial-kognitive Entwicklung und für die Veränderung von Einstellungen und Wertorientierungen nachweisen lassen. Im Leistungsbereich sind die Schulformeffekte über die Fächer hinweg stabil und sehr groß; hinsichtlich der sozial-kognitiven Entwicklung sind die Unterschiede systematisch, aber klein. In unserem zweiten Bericht an die an der BIJU-Studie teilnehmenden Schulen haben wir diese Entwicklungsverläufe beschrieben und unter Berücksichtigung der unterschiedlichen Eingangsvoraussetzungen der weiterführenden Schulen als Hinweise auf verschiedene Entwicklungsmilieus interpretiert (Köller 1996). Diese Interpretation hat - jedenfalls für die Gesamtschule - Widerspruch hervorgerufen (Miebach und Virnich 1998; Ratzki, in diesem Heft).

Die Übergangsauslese zu den weiterführenden Schulen am Ende der Grundschulzeit führt zu leistungsmäßig, aber auch sozial sehr unterschiedlich zusammengesetzten Schülerschaften an den Sekundarschulformen. Die Gesamtschule bindet im viergliedrigen Schulsystem - etwa in Berlin oder Nordrhein-Westfalen - eine Schülerschaft, die in ihren sozialstrukturellen und leistungsbezogenen Merkmalen der Schülerschaft der Realschule sehr ähnlich ist. Hinsichtlich der familiären Situation und der Belastung durch Risikofaktoren liegt die Gesamtschülerschaft im Mittel zwischen Haupt- und Realschule.

Worauf nun genau sind die nachweisbaren unterschiedlichen Leistungsentwicklungen, die sich im Gewinnen bis zu mehr als zwei Schuljahren niederschlagen können, und die systematischen, aber kleinen Unterschiede in der

psychosozialen Entwicklung zurückzuführen? Ist die positive Eingangsauslese der Motor für günstige Entwicklung? Spiegeln die Daten nur das Ergebnis unterschiedlicher interner Selektivität der Schulformen wider? Oder lassen sich Entwicklungsmilieus identifizieren, die unabhängig von Übergangsauslese und interner Selektivität Ergebnisse schulischen Handelns darstellen?

BIJU gibt auf diese Fragen klare Antworten. Zunächst ist festzuhalten, daß es keine besonderen Selektionsmuster innerhalb der einzelnen Schulformen gibt - etwa durch die unterschiedliche Handhabung der Klassenwiederholung -, die dazu führen, daß die Leistungsheterogenität der Hauptschul- und Gesamtschulerschülerschaft erhalten bliebe, während die Schülerschaft in Realschule und Gymnasium durch einen positiven Ausleseprozeß im Durchgang durch die Mittelstufe aufgebessert würde. Die Jugendlichen, die über vier Jahre oder länger an dem Längsschnitt teilnehmen, sind eine, wie zu erwarten war, leistungsmäßig leicht positive ausgelesene Subgruppe - und zwar in allen Schulformen in gleicher Weise. Insbesondere kann für die Gesamtschule kein abweichendes Selektionsmuster beobachtet werden.

Wieweit die Unterschiede zwischen den Schulformen institutionelle Fördereffekte abbilden oder nur Folge der unterschiedlichen kognitiven und sozialen Übergangsauslese sind, kann statistisch (durch die multiple Regressionsanalyse) geprüft werden. Wir haben in einem ersten Schritt die Stichproben aus den unterschiedlichen Schulformen hinsichtlich fachlichen Vorwissens in der 7. Jahrgangsstufe, der sprachlichen und sprachfreien Intelligenz, des Sozialstatus der Eltern, der ethnischen Herkunft, der Familienstruktur und der familiären Belastung (Trennung der Eltern, alleinerziehende Elternteile, Berufstätigkeit der Mutter, Arbeitslosigkeit von Vater oder Mutter) angeglichen. Danach haben wir geprüft, ob sich im Längsschnitt zum Ende der 10. Jahrgangsstufe Entwicklungsdifferenzen zwischen Schülern der unterschiedlichen Schulformen nachweisen lassen. Als Bezugspunkt benutzen wir die Gesamtschule, um auf die vorgetragenen Einwände reagieren zu können (die Ergebnistabellen der Regressionsanalysen können von den Autoren bezogen werden).

Für den Vergleich von Haupt- und Gesamtschule ergeben sich nach Kontrolle des Vorwissens sowie der kognitiven und sozialen Variablen keine unterschiedlichen Leistungseffekte zwischen beiden Schulformen: Bei gleichen Eingangsbedingungen wird am Ende der 10. Jahrgangsstufe ein identischer Wissensstand erreicht. Die wichtigsten Einflüsse auf die Leistungsentwicklung üben die kognitiven Variablen Vorwissen und kognitive Grundfähigkeiten aus. Der Einfluß des sozialen Status ist schwach. Ethnische Herkunft und familiäre Situation üben nach Kontrolle der kognitiven Voraussetzungen keinen nachweisbaren Einfluß aus.

Beim Vergleich zwischen Real- und Gesamtschule zeigt sich, daß in der Realschule auch nach Kontrolle kognitiver und sozialer Eingangsvariablen die Leistungsentwicklung günstiger als an der Gesamtschule verläuft. Bei gleichen intellektuellen und sozialen Eingangsbedingungen erreichen Realschüler am Ende der Sekundarstufe I etwa in Mathematik einen Wissensvorsprung von etwa zwei Schuljahren. Noch stärker sind diese Effekte, wenn man Gesamtschule und Gymnasium vergleicht. Bei gleichen intellektuellen und sozialen Bedingungen beträgt der Leistungsvorsprung in Mathematik am Gymnasium mehr als zwei Schuljahre. Es gibt keine Hinweise, daß die ungünstige Leistungsentwicklung durch besondere überfachliche Leistungen kompensiert werden könnte.

In allen Analysen ist der Einfluß der Sozialschicht nach Kontrolle der kognitiven Voraussetzungen relativ gering oder statistisch nicht nachweisbar. Dies weist darauf hin, daß der Einfluß der sozialen Herkunft auf die Leistungsentwicklung innerhalb von Schulformen im Vergleich zu ihrer Bedeutung bei der Übergangsauslese in der Regel überschätzt wird.

In unserem zweiten Bericht an die Schulen haben wir als ein Beispiel für unterschiedliche Entwicklungsverläufe im sozial-kognitiven Bereich die Entwicklung prosozialer Motive beschrieben. Bei dem eingesetzten Erhebungsinstrument handelt es sich um die Kurzfassung eines von Boehnke und Silbereisen im Anschluß an die einschlägige Forschung zum Altruismus entwickelten Fragebogens (Prosocial Motivation Questionnaire (PSMQ)), der sich in deutschen, mittlerweile auch in internationalen Untersuchungen gut bewährt hat (Boehnke 1988). Als Stimuli werden kurze und einfach erzählte Alltagsgeschichten präsentiert, in denen jeweils eine Hilfeleistung vorkommt. Anschließend wird der Befragte gebeten, sich in die jeweiligen Situation zu versetzen und ein für ihn plausibles Helfensmotiv auszuwählen. Der Fragebogen (vgl. Beispiele bei Ratzki in diesem Heft) ist für sozial erwünschte Antworten wenig anfällig und weist eine gute Konstruktvalidität auf, das heißt, die Zusammenhänge mit anderen Merkmalen sozialer Motivation und sozialen Verhaltens sind erwartungsgemäß. Die Kompetenz, die Perspektive eines anderen einzunehmen und die Fähigkeit, mit anderen mitzuempfinden - beides Voraussetzungen für Toleranz, Kooperativität und Hilfsbereitschaft - korrelieren mit dem Altruismusmotiv, das wiederum negativ mit sozial unerwünschten Merkmalen, etwa Ausländerfeindlichkeit zusammenhängt. Ebenso lassen sich substantielle Zusammenhänge zwischen den in dieser Weise erfaßten prosozialen Motiven und Verhaltensmaßen - etwa Devianz und Delinquenz - nachweisen.

Die prosoziale Motivation verändert sich im allgemeinen während des Jugendalters in typischer Weise. Das Egoismusmotiv verliert mit zunehmendem Alter an Bedeutung, während das Altruismusmotiv bedeutsamer wird. Die Ergebnisse der BIJU-Studie belegen nach Schulform differenzierte Entwicklungsverläufe der prosozialen Motivation. Überraschend ist der erwartungswidrige Entwicklungsverlauf an Gesamtschulen, über den wir in unserem 2. Bericht an die Schulen (Köller 1996) berichtet haben. Die Differenzen sind systematisch, allerdings nicht sehr groß. Auch in diesem Fall stellt sich die Frage der Bedeutung der Übergangsauslese. Handelt es sich - wie gelegentlich vermutet wurde (z.B. Ratzki, in diesem Heft) - um Folgewirkungen der mit dem Übergang in die Sekundarschulformen verbundenen sozialen Selektion?

Verfährt man in gleicher Weise wie bei der Analyse der Leistungsentwicklung und gleicht die Stichproben in denselben Merkmalen an, ergibt sich eine klare Befundlage: Die Entwicklung prosozialer Motive ist weitgehend vom außerschulischen Umfeld und den kognitiven Variablen unabhängig. Allein die früheren Ausprägungen der prosozialen Motive weisen substantielle Zusammenhänge mit den Werten am Ende der Sekundarstufe I auf. Für die Schulformen wiederholen sich im wesentlichen - wenn auch abgeschwächt - die Befunde zur Leistungsentwicklung: Es lassen sich keine unterschiedlichen Entwicklungen für Haupt- und Gesamtschüler, wohl aber für Gesamtschüler auf der einen und Realschüler und Gymnasiasten auf der anderen Seite zeigen. Bei Kontrolle der Ausgangsunterschiede fällt das altruistische Motiv der Gesamtschüler in Klasse 10 niedriger aus als

in den beiden anderen Schulformen. Dieses Muster wiederholt sich auch bei Analysen für andere Merkmale sozialer Kognition und sozial relevanten Einstellungen wie negative Einstellung zu Gastarbeitern oder Intoleranz gegenüber Asylbewerbern.

Schlußbemerkung

Nationale und internationale Schulleistungsstudien sind aus den Kinderschuhen herausgewachsen. Large scale assessment ist ein wissenschaftliches Arbeitsgebiet, auf dem es in den letzten 15 Jahren große methodische und inhaltliche Fortschritte gegeben hat. Die Techniken sind elaboriert und ohne Spezialisierung nicht mehr zugänglich. Unter den Sachkundigen besteht weitgehend Einigkeit darüber, daß es mit den entwickelten Instrumenten grundsätzlich möglich ist, Leistungen in schulischen Kernfächern auch im internationalen Vergleich zuverlässig zu erfassen.

In der Regel sind die großen nationalen oder internationalen Schulleistungsstudien, die zur Dauerbeobachtung von Bildungssystemen dienen, deskriptiv angelegt. In dieser Funktion liefern sie nützliches Orientierungswissen und wirken sogar Augen öffnend - jedenfalls wenn man es allzu lange vermieden hat, sich über Ertragslagen Klarheit zu verschaffen. Man muß jedoch auch mit Nachdruck auf die Grenzen dieser Studien hinweisen. Die Studien sind für die Entwicklung didaktischer Maßnahmen oder deren Überprüfung wenig geeignet. Dazu bedarf es differenzierterer und spezifischerer Instrumente. Und schon gar nicht sind sie für die Evaluation oder gar Steuerung einzelner Schulen brauchbar.

Werden große Schulleistungsuntersuchungen als Längsschnittstudien mit wenigstens zwei Messungen zu unterschiedlichen Zeitpunkten angelegt, so steigt ihre Aussagekraft deutlich: Die Bedeutung von Schule für die kognitive und soziale Entwicklung kann jetzt von Effekten der unterschiedlichen Eingangsbedingungen auf Seiten der Schülerinnen und Schüler getrennt werden. Sie erreichen allerdings Grenzen, wenn es um die Mikroanalyse individueller Lern- und Entwicklungsprozesse geht.

Literatur

- Baumert, J.: TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich - Anlage der Studie und ausgewählte Befunde. In: J. List (Hg.), TIMSS: mathematisch-naturwissenschaftliche Kenntnisse deutscher Schüler auf dem Prüfstand (S. 13-65). Köln 1998
- Baumert, J.: Internationale Schulleistungsvergleiche. In: D. H. Rost (Hg.), Handwörterbuch "Pädagogische Psychologie". Weinheim 1998 (in Druck)
- Baumert, J.; Lehmann, R., u. a.: TIMSS - Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Opladen 1997
- Baumert, J.; Bos, W.; Lehmann, R. (Hg.): TIMSS - Mathematisch-naturwissenschaftlicher Unterricht am Ende der Schullaufbahn. Opladen (in Vorbereitung)
- Boehnke, K.: Prosoziale Motivation, Selbstkonzept und politische Orientierung: Entwicklungsbedingungen und Veränderungen im Jugendalter. Frankfurt a.M. 1988
- Heyn, S.; Schnabel, K.; Roeder, P. M.: Von der Options- zur Realitätslogik. Stabilität und Wandel berufsbezogener Wertvorstellungen in der Statuspassage Schule - Beruf. In: A. Meier, U. Rabe-Kleberg, K. Rodax (Hg.), Jahrbuch Bildung und Arbeit 1997 (S. 281-305). Opladen 1997
- Köller, O.: Die Entwicklung der Schulleistungen und psychosozialer Merkmale während der Sekundarstufe. In: Max-Planck-Institut für Bildungsforschung, Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU). 2. Bericht für die Schulen (S. 13-24). Berlin: Max-Planck-Institut für Bildungsforschung 1996
- Köller, O.: Zielorientierungen und schulisches Lernen. Münster/New York 1998
- Kraus, J.: Bildungspolitische Schlußfolgerungen aus der TIMSS-Studie Deutschland - 15 Thesen. In: J. List (Hg.), TIMSS: mathematisch-naturwissenschaftliche Kenntnisse deutscher Schüler auf dem Prüfstand (S. 66-97). Köln 1998
- MPI - Max-Planck-Institut für Bildungsforschung: Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU). 2. Bericht für die Schulen. Berlin 1996
- Miebach, C. H.; Virnich, P.: Ein schlechtes Zeugnis für die Gesamtschule in Nordrhein-Westfalen? Neue deutsche Schule, 50. Jg., Heft 2/1998, S. 9-10
- Ramseier, E.: Naturwissenschaftliche Leistungen in der Schweiz. Amt für Bildungsforschung. Bern 1997
- Schnabel, K.: Prüfungsangst und Lernen. Empirische Analysen zum Einfluß fachspezifischer Leistungsängstlichkeit auf schulischen Lernfortschritt. Münster, New York, München, Berlin 1998