



Netzwerk Berliner
Schülerfirmen

Fördern in Kooperation



LABEL - Ergebnisse

LABEL 8 - 10

Werkstattberichte dienen der Profilbildung in der sonderpädagogischen Förderung. Mit ihnen soll im „Netzwerk Berliner Schülerfirmen“ (NBS) die Diskussion zur systematischen Qualitätsentwicklung unterstützt werden. Es handelt sich dabei um Arbeitsfassungen, die unter Berücksichtigung der Hinweise, Anregungen und Diskussionsergebnisse aller am „Netzwerk Berliner Schülerfirmen“ Beteiligten überarbeitet werden und in Buchpublikationen münden.

Senatsverwaltung für Bildung, Jugend und Sport, Referat Sonderpädagogische Förderung (Hg.):

LABEL 8 - 10 - Bericht zur Erhebung der Lernstände
Autoren: Lehmann, Rainer; Nikolova, Rouminiana; Peek, Rainer

Im Internet: <http://www.wir-in-Berlin.de/nbs/label-bericht2004.pdf>

Alle Rechte vorbehalten.
Nachdruck mit Quellenangabe erlaubt,
Zusendung von Belegexemplaren erwünscht.

Koordination der Fachlich-Pädagogischen Leitung:
Dipl. Päd. Rainer Eßer, SenBJS III B 12
eMail: rainer.esser@senbjs.verwalt-berlin.de
Senatsverwaltung für Bildung, Jugend und Sport

Wiss. Begleitung des Netzwerks: Prof. Dr. Gerhard H. Duismann,
Prof. Dr. Klaus Hasemann, Dr. Helmut Meschenmoser

Der vorliegende Werkstattbericht steht in Beziehung zu dem Vorhaben „Erhebung der Lernausgangslagen arbeitsrelevanter Basiskompetenzen im Förderschwerpunkt Lernen (LABEL 8 - 10)“, das gemeinsam mit der Humboldt-Universität 2002/2003 durchgeführt wird.

Evaluationsgruppe der Humboldt-Universität zu Berlin:
Institut für Erziehungswissenschaften, Abteilung Empirische Bildungsforschung
und Methodenlehre Prof. Dr. Dr. Rainer Lehmann,
Roumiana Nikolova, Dr. Rainer Peek

Koordination der Datenerhebung:
Christa Neumann-Kieslich, Gerd Migulla

Träger: Arbeit-Schule-IntegrationsGesellschaft e.V., Berlin
Lindower Str. 18, 13347 Berlin

Das Projekt „Netzwerk Berliner Schülerfirmen“ wird mit Mitteln des Europäischen Sozialfonds (ESF), des Landesarbeitsamts Berlin-Brandenburg sowie des Landes Berlin gefördert.

Dem Projektbeirat gehören u.a. an: Vertreter/innen der Handwerkskammer Berlin, der Industrie- und Handelskammer Berlin, des Landesarbeitsamtes Berlin-Brandenburg, der Stiftung der Deutschen Wirtschaft, der Deutschen Kinder- und Jugendstiftung, der Gewerkschaft Erziehung und Wissenschaft, der Elternschaft, der Schulen im Netzwerk, der Schulträger und der Schulaufsicht.

Inhaltsverzeichnis

Vorwort von Senator Klaus Böger	5
Zur Erhebung der Lernausgangslage im Netzwerk Berliner Schülerfirmen	7

Bericht der Evaluationsgruppe

1	Ziele der Untersuchung	11
2	Anlage und Durchführung der Untersuchung	14
2.1	Erhebungsinstrumente	14
2.1.1	Fachleistungstest	14
2.1.2	Problemlösetest	16
2.1.3	Test zum schlussfolgernden Denken	17
2.1.4	Schülerfragebogen	17
2.1.5	Schülerakte	18
2.2	Datenstruktur	18
3	Befunde zu den Fachleistungen und Einstellungen der Schülerinnen und Schüler	23
3.1	Fachleistungen	23
3.1.1	Leseverständnis	24
3.1.2	Wortkenntnis	28
3.1.3	Informationsentnahme aus Karten, Tabellen und Diagrammen	30
3.1.4	Mathematik	33
3.1.5	Gesamtergebnis in den Fachleistungen	36
3.2	Problemlösen	39
3.3	Einstellungen, Einschätzungen und Werthaltungen	41

4	Differenzierungen in den kognitiven Leistungen	46
4.1	Unterschiede in den Leistungen zwischen Klassenstufen	46
4.2	Unterschiede in den Leistungen zwischen Schulen	47
4.2.1	Unterschiede in den Leistungen zwischen Schulen allgemein	47
4.2.2	Unterschiede in den Leistungen zwischen Schulen nach ihrer Teilnahme am Netzwerk Berliner Schülerfirmen	56
4.3	Unterschiede in den Leistungen zwischen Schulklassen	60
4.3.1	Unterschiede in den Leistungen zwischen Schulklassen allgemein	60
4.3.2	Unterschiede in den Leistungen zwischen Schulklassen nach ihrer Teilnahme am Netzwerk Berliner Schülerfirmen	62
4.4	Leistungsunterschiede zwischen Mädchen und Jungen	66
4.5	Leistungen von Jugendlichen mit Migrationshintergrund	70
4.6	Sozialräumlich bedingte Leistungsunterschiede	74
5	Differenzierungen in den Selbsteinschätzungen der Jugendlichen	77
5.1	Fachleistungsbezogenes Selbstkonzept	77
5.2	External attribuierende Erfolgsdeutung	79
5.3	Eigene Arbeitstugenden	80
5.4	Flexibilität und Belastbarkeit	80
5.5	Technikfreundlichkeit	81
6	Zusammenfassung	83
7	Literaturverzeichnis	87
8	Glossar	88

Vorwort von Klaus Böger

Senator für Bildung, Jugend und Sport

Der Senat von Berlin hat der Förderung von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf stets besondere Bedeutung beigemessen. Sonderpädagogische Förderung verwirklicht für Schülerinnen und Schüler beziehungsweise für Kinder und Jugendliche mit sonderpädagogischem Förderbedarf das Recht auf eine ihrer persönlichen Begabung und ihrem persönlichen Leistungsvermögen entsprechende schulische Bildung und Erziehung.

Sie soll den Betroffenen ein möglichst hohes Maß an schulischer und beruflicher Eingliederung, gesellschaftlicher Teilhabe und selbstständiger Lebensgestaltung ermöglichen. Dazu sind gemeinsame Anstrengungen aller Beteiligten bei der systematischen Entwicklung arbeitsrelevanter Basiskompetenzen mehr denn je erforderlich.

Durch das neue Schulgesetz und die künftigen Rahmenlehrpläne erhalten die Schulen noch mehr als bisher Gestaltungsspielräume, um den aktuellen Bedürfnissen der Zielgruppe entsprechend den Unterricht zu flexibilisieren, die Förderung zu individualisieren und somit Leistung zu fördern. Komplexe Lernarrangements wie Schülerfirmen, erweiterte und zusätzliche Betriebspraktika, themenfeldorientierter Unterricht, jahrgangsübergreifendes handelndes Lernen werden mit großen Hoffnungen zur Verbesserung der Chancen der Heranwachsenden verbunden.

Mit Mitteln des Europäischen Sozialfonds konnten im Rahmen des Projektes „Netzwerk Berliner Schülerfirmen“ mit dem Träger Arbeit-Schule-IntegrationsGesellschaft e.V. wichtige Schritte der Qualitätsentwicklung gemacht werden. Die Erfahrungen aus dem „Netzwerk Berliner Schülerfirmen“ bestätigen die bekannte Bedeutung der systematischen Qualitätssicherung durch

- individuelle Förderpläne,
- aktuelle und zielgruppenspezifische Rahmenlehrpläne,
- Schulprogramme und schulinterne Curricula,
- kontinuierliche interne und externe Evaluation.

Diese Qualitätssicherung erfordert zuverlässige Daten über die Lernausgangslagen der Schülerinnen und Schüler. Im Rahmen des „Netzwerk Berliner Schülerfirmen“ konnte erstmals im deutschsprachigen Raum ein large scale assessment für die Zielgruppe der Jugendlichen mit sonderpädagogischem Förderbedarf Lernen durchgeführt werden. Mit rund 2.500 Teilnehmerinnen und Teilnehmern an 45 Berliner Sonderschulen handelt es sich um die erste Vollerhebung in einem Bundesland.

Hierfür ist vor allem den engagierten Lehrerinnen und Lehrern zu danken, ohne deren tatkräftige Mitwirkung die Erhebung der Daten nicht möglich gewesen wäre.



Mit Unterstützung der Forschungsgruppe der Humboldt-Universität konnte in Kooperation mit der Wissenschaftlichen Begleitung des Netzwerkes ein Instrumentarium zur Erhebung der „Lernausgangslage arbeitsrelevanter Basiskompetenzen im Förderschwerpunkt Lernen in den Klassen 8 – 10 (LABEL 8 – 10)“ zusammengestellt und ausgewertet werden.

Die vorliegenden Ergebnisse liefern empirische Informationen für die weitere Qualitätsentwicklung in der sonderpädagogischen Förderung, die über Berlin hinaus von überregionaler und internationaler Bedeutung sind. Dieser Bericht mit ersten Daten und vorsichtigen Interpretationen bietet differenzierte Hinweise zu den derzeitigen Lernständen in den Leistungsbereichen Leseverständnis, Wortverständnis, Informationsentnahme, Mathematik und Problemlösen sowie Einstellungen der Schülerinnen und Schüler zu Lernen, Arbeit und Technik.

Die empirischen Befunde machen deutlich, dass einige herkömmliche Annahmen über die Zielgruppe der Schülerinnen und Schüler mit dem sonderpädagogischem Förderbedarf zu überdenken sind. Offensichtlich wird, dass auch in den Schulen mit dem Förderschwerpunkt Lernen von einer überraschend großen Heterogenität auszugehen ist. Dies erfordert zielgruppenspezifische Rahmenlehrpläne, die es durch gestufte Kompetenzerwartungen und –zertifizierung vermögen, die kontinuierliche individuelle Förderplanung und deren Evaluation zu unterstützen. Wesentlich dabei ist es, die Jugendlichen zunehmend in die Lernplanung einzubeziehen, um das notwendige selbstständige lebenslange Lernen anzubahnen.

Im Rahmen von LABEL 8 – 10 wurden bereits im Mai 2003 für jede Lerngruppe klassen- und personenbezogene Rückmeldungen mit ausführlichen Hinweisen zur schulinternen Auswertung übergeben. Sie bieten den Lehrerinnen und Lehrern konkrete Unterstützung für die Weiterentwicklung von Unterricht und Förderkonzepten auch über die beteiligten Klassen hinaus.

Im Vergleich zwischen den einzelnen Schulen werden erhebliche Unterschiede in Bezug auf die Lerndispositionen, Leistungen und Einstellungen deutlich. Wenn in einigen Schulen die festgestellten Leistungen deutlich über den Erwartungswerten liegen, spricht dies für die intensiven Bemühungen der dort beteiligten Sonderpädagoginnen und Sonderpädagogen zur Qualitätssicherung. Gleichzeitig wird deutlich, dass es erhebliche Chancen gibt, vorhandene Stärken und Potenziale dieser Schülerschaft noch besser zu nutzen. LABEL bietet darüber hinaus die große Chance der Selbstvergewisserung und stellt für die notwendige schulinterne Evaluation einen wichtigen Referenzrahmen.

Eine erste Befragung der Schulleitungen im Oktober 2003 lässt darauf schließen, dass viele Schulen die Chancen von LABEL bereits umfassend in allen Schulstufen und Lernbereichen für die Qualitätssicherung nutzen. Mit Kreativität und Engagement wurden inzwischen vielfältige Maßnahmen zur Qualitätssicherung umgesetzt und weitere geplant.

Durch weitere Auswertungen sowie die Dokumentation von beispielhafter Schulpraxis, Erarbeitung von Materialien und Fortbildungsmaßnahmen zur Unterstützung bei der schwierigen, aber lohnenden Arbeit wird der Prozess der kontinuierlichen Qualitätsentwicklung auch künftig weiter befördert.

Zur Erhebung der Lernausgangslage im Netzwerk Berliner Schülerfirmen

Intentionen

Die aktuellen Veränderungen der Arbeits- und Lebenswelt, der Wertvorstellungen und Einstellungen, sowie der sich daraus ergebenden Anforderungen an Schulabgänger, machen es erforderlich, das bisherige Handeln in Schulen nicht aufgrund von Mutmaßungen und individuell höchst unterschiedlichen Erfahrungen, sondern auf der Basis gesicherter Erkenntnisse zu reflektieren. Um die notwendigen Schlussfolgerungen aus Maßnahmen ziehen zu können, mit denen versucht wird, dem Wandel gerecht zu werden, müssen die zu diesem Zweck zu erhebenden Daten empirisch abgesichert sein.

Die Suche nach Verfahren, die eine aussagefähige Feststellung der Ausgangslage des Projektes „Netzwerk Berliner Schülerfirmen (NBS)“ erlauben könnten, führte zur Anknüpfung an die Erfahrungen, die Lehmann et al. 1997 mit einer Untersuchung in Hamburger Schulen gemacht hatten. Das dort eingesetzte Untersuchungsinstrumentarium bildete die Grundlage für die Erhebung der Lernausgangslage der am NBS beteiligten Schülerinnen und Schüler. Die einzelnen Schritte zur Entwicklung des neuen, mit LABEL 8-10 bezeichneten Verfahrens (Adaptation, Modifikation und Ergänzung um projektspezifische Aspekte, Anwendung, Auswertung und Interpretation) erfolgten in enger Kooperation zwischen der Abteilung Empirische Bildungsforschung und Methodenlehre des Instituts für Allgemeine Pädagogik der Humboldt Universität zu Berlin (Prof. Dr. Dr. R. Lehmann, R. Nikolova und Dr. R. Peek), der wissenschaftlichen Begleitung und der fachlich-pädagogischen Begleitung des Projektes, der Senatsverwaltung für Bildung sowie der Lehrerinnen und Lehrer der beteiligten Schulen.

Im NBS bildet das Lernarrangement „Schülerfirma“ die inhaltlich-didaktische Grundlage des Unterrichtes in der Oberstufe der Schule für Lernbehinderte, nunmehr der „Schule mit dem Förderschwerpunkt Lernen“. Auf dieser Grundlage werden für die einzelne Schülerin und den einzelnen Schüler Förderpläne erstellt, die sich auf die Vermittlung bzw. Aneignung sogenannter „arbeitsrelevanter Basiskompetenzen“ (vgl. Duismann 2002) zentrieren. Die Erfassung des individuellen Status eben dieser Kompetenzen war wesentli-

ches Ziel der im Herbst 2002 an den Schulen im NBS durchgeführten Untersuchung.

Um die Ergebnisse durch einen tragfähigen Zielgruppenvergleich abzusichern, wurde das Instrumentarium gleichzeitig bei Schülerinnen und Schülern der Klassen 8-10 auch der Berliner Schulen für Lernbehinderte eingesetzt, die zum Untersuchungszeitpunkt nicht in das Projekt NBS einbezogen sind. Auf diese Weise ergab sich erstmals die Möglichkeit, Daten zu gewinnen, die auf mehrfache Weise bei der Gestaltung der Oberstufe der Berliner Schule für Lernbehinderte hilfreich sind und einen wichtigen Beitrag für die weitere Planung der Schulentwicklung leisten können. Die vorliegende Auswertung der Daten von LABEL 8-10 bildet eine wesentliche Grundlage für die weitere Arbeit und liefert Kriterien für die im NBS entwickelten und erprobten innovativen Maßnahmen.

Nutzung der Ergebnisse

Die aus LABEL 8-10 gewonnenen Daten dienen in erster Linie der Rückmeldung. Alle, die in irgendeiner Weise in den Bildungsprozess einbezogen sind, erfahren auf diese Weise etwas über die Qualität ihres Tuns:

- die Schülerinnen und Schüler,
- die Lehrerinnen und Lehrer,
- die Schulen,
- die Schulaufsicht.

Schülerinnen und Schüler

Die Schülerinnen und Schüler wollen wissen, wo sie mit ihren Fähigkeiten und Fertigkeiten in der Schule stehen; sei es im Vergleich mit eigenen früheren Leistungen, sei es durch Vergleich mit anderen Gleichaltrigen, sei es aus der Unterschiedlichkeit ihrer Ergebnisse in einzelnen Bereichen. Die Daten bieten eine zuverlässige Grundlage für die individuelle Förderplanung der Schülerinnen und Schüler sowie deren kompetente Beratung durch ihre Lehrerinnen und Lehrer, durch Praxisbegleiter und Praktikumsbetreuer.

Lehrerinnen und Lehrer

Die Daten erlauben den Lehrerinnen und Lehrern der in die Untersuchung einbezogenen Schülerinnen und Schüler, sich ein fundiertes Urteil über das

eigene unterrichtliche und erzieherische Handeln und einige seiner Effekte zu bilden. Zwar lassen sich mit einem solchen Verfahren keineswegs alle Aspekte des komplexen Lehrer- und des daraus resultierenden Schülerverhaltens erfassen, jedoch vermögen die daraus gewonnenen Daten durchaus konkrete Anhaltspunkte für selbstkritische Reflexion zu liefern.

Zu den Rückmeldungen aus der Untersuchung zählen aber nicht nur Testwerte über Leistungen in den Kulturtechniken, sondern auch ein Bündel von Informationen über die Einstellungen der Schülerinnen und Schüler, im Hinblick auf das Projekt NBS insbesondere zu „arbeitsrelevanten Basiskompetenzen“.

Schließlich können die Daten aus LABEL 8-10 auch eine Basis für das Erstellen und Fortschreiben von Förderplänen bilden und hilfreich bei der Erprobung neuer Formen der Dokumentation von Schülerleistungen sein.

Schulen

Die Schulen brauchen empirisch gesicherte Grundlagen zur Einschätzung ihrer eigenen Entwicklung auf dem Hintergrund des Vergleichs mit anderen Schulen, ohne damit Rangordnungen herzustellen. Auf diese Weise kann – zum Nutzen des eigenen Unterrichts - der Austausch bisher oft isolierter Systeme angeregt werden. Leistungsdaten können Anstöße geben zu fragen: „Wo stehen wir, wo wollen bzw. sollten wir stehen? Was ist bei uns gut, wo müssen wir vielleicht etwas ändern?“

Schulaufsicht

Die Schulaufsicht braucht Rückmeldungen über die Entwicklung des Unterrichts in den Schulen, u.a. über die durchschnittliche Leistungsfähigkeit der Schülerinnen und Schüler, über die von diesen erworbenen Kompetenzen, hier insbesondere über die "arbeitsrelevanten Basiskompetenzen“ über allgemeine Effekte des Unterrichts und wegen des Ansatzpunktes der Untersuchung LABEL 8-10 die Anschlussfähigkeit der Schülerinnen und Schüler in den Oberstufen der Schulen für Lernbehinderte, bzw. der Schule mit dem Förderschwerpunkt Lernen.

Die Schulbehörde braucht für die Schulentwicklung, für die Ausarbeitung von Bildungs- und Rahmenplänen, zur Lösung von Raum- und Ausstattungsproblemen sowie zur Ermittlung des Personal- und Fortbildungsbedarfs empi-

risch gesicherte Erkenntnisse. Nur so kann die Nähe zur Praxis gesichert und steuernd auf sie eingewirkt werden.

Ausblick

Mit der Vorlage des zusammenfassenden Berichts der Erhebung der Lernausgangslage mit dem Instrument LABEL 8 - 10 in den Klassen 8 bis 10 aller Berliner Schulen mit dem Förderschwerpunkt Lernen ist ein erster Schritt in Richtung auf eine Verbesserung der Qualität von Unterricht unternommen worden. Damit die möglichen Effekte eintreten, gilt es, die Ergebnisse auf den unterschiedlichen Ebenen aufzunehmen und zur Grundlage von Diskussionen und Konzeptentwicklung für zu verändernde Praxis zu machen.

¹ Werkstattberichte aus dem „Netzwerk Berliner Schülerfirmen“

- Praktika in Betrieben. Berlin 2002
- Förderplanung arbeitsrelevanter Basiskompetenzen. Berlin 2002
- Zum Umgang mit den LABEL-Rückmeldungen. Berlin 2003
- Strategien der Schulen zum Umgang mit LABEL-Befunden, Teil 1. Berlin 2004

Im Internet zum Herunterladen unter : <http://www.wir-in-Berlin.de/nbs>

Bericht der Evaluationsgruppe

1 Ziele der Untersuchung

Im Kontext der Diskussion um Fragen der Standardsicherung im Bildungswesen und um die Weiterentwicklung von Schule wird in zunehmendem Maße die Notwendigkeit gesehen, lernbiografisch entscheidende Kompetenzen des Bildungssystems in den Blick zu nehmen (vgl. dazu Klieme et al. 2003). Aktuelle Beispiele solcher Evaluationsstudien im allgemeinbildenden Schulbereich sind die Untersuchung *Aspekte der Lernausgangslage und der Lernentwicklung (LAU)*, die die Lernstände und Lernentwicklungen von Hamburger Schülerinnen und Schülern am Ende der Grundschulzeit in Klassenstufe 4, am Ende der Beobachtungsstufe in Klassenstufe 6, und dann Ende Klassenstufe 8 bzw. 10 in den Blick nimmt (vgl. Lehmann et al. 1997, 1999, 2001), die *Internationale Grundschul-Lesestudie (IGLU / PIRLS)*, die sich ebenfalls auf erreichte Lernstände am Ende der Grundschulzeit richtet (vgl. Bos et al. 2003) oder die Berliner Untersuchung *Erhebungen zum Lese- und Mathematikverständnis – Entwicklungen in den Jahrgangsstufen 4 bis 6 (ELEMENT 4-6)*, in der es um das Aufdecken von Besonderheiten in der Lernentwicklung von Schülerinnen und Schülern in der sechsjährigen Grundschule, sowie in den Anfangsjahren der grundständigen Gymnasien geht (vgl. Stryck & Lehmann 2003).

Mit der Untersuchung *Lernausgangslagen arbeitsrelevanter Basiskompetenzen im Förderschwerpunkt Lernen – Klassenstufen 8 bis 10 (LABEL 8-10)* werden in Berlin – bundesweit erstmalig – für die Abschlussjahrgänge der Schule für Lernbehinderte in Form eines *large scale assessment* Fragen zur Situation von Schülerinnen und Schülern vor dem Übergang in berufsqualifizierende Maßnahmen in den Blick genommen.

Die Untersuchung LABEL 8-10, die im Kontext reformorientierter Programme wie der Einrichtung von Schülerfirmen und der Weiterentwicklung dieses Konzeptes steht (zum Netzwerk Berliner Schülerfirmen vgl. ausführlich DUISMANN / ESSER / HASEMANN / HÜBNER / MESCHENMOSER 2001; MESCHENMOSER 2003), zielt im Hinblick auf ausgewählte arbeitsrelevante Basiskompetenzen (vgl. DUISMANN 2002) darauf, in systemischer Perspektive Informationen über die besondere Situation von Schülerinnen und Schülern gegen Ende ihrer Schulzeit an der Schule für Lernbehinderte bereit zu stellen. Die Untersuchung LABEL 8-10 hat fünf zentrale Zielstellungen:

1. die Bilanzierung erreichter Lernstände der Schülerinnen und Schülern in der Oberstufe der Schule für Lernbehinderte in Berlin (Klassenstufen 8, 9, 10) in ausgewählten schulischen Fachgebieten, namentlich sprachlicher Art (Wortkenntnis und Leseverständnis), im Umgang mit diskontinuierlichen Texten (Karten, Tabellen und Diagrammen) und in der Mathematik als Schlüsselkompetenzen zur Teilhabe an gesellschaftlichen Prozessen, und zwar insbesondere am Erwerbsleben,
2. die Beschreibung von Einstellungen derselben Schülerinnen und Schüler hinsichtlich ihrer Zukunftsorientierung in den Bereichen Lernen und Arbeit,
3. die Erfassung fächerübergreifenden Kompetenz Problemlösen, gemessen über Aufgaben mit offener Problemstruktur in einer Teilstichprobe von Schülerinnen und Schülern,
4. eine Differenzierung der bilanzierten Lernstände und Einstellungen nach Bezirken, Schulen, Klassenstufen und Klassen innerhalb Berlins und schließlich
5. das Aufdecken von Zusammenhängen zwischen erreichten Lernständen und entwickelten Einstellungen auf der einen Seite und individuellen Schülermerkmalen (Alter, Geschlecht, Herkunftssprache, kognitive Voraussetzungen, Schullaufbahn Daten) sowie spezifischen schulischen Angebotsstrukturen (Praktika, und insbesondere Schülerfirmen) auf der anderen.

Entsprechend diesen Zielstellungen ist der weitere Bericht aufgebaut: Im Abschnitt 2 wird zunächst die Anlage der wissenschaftlichen Untersuchung skizziert. Dabei werden die eingesetzten Evaluationsinstrumente vorgestellt und ihre Auswahl begründet. Es werden Kennwerte der untersuchten Schülerpopulation mitgeteilt und Informationen zur Organisation der Datenerhebung und zu den eingesetzten Verfahren der Datenaufbereitung und -auswertung gegeben.

Im Abschnitt 3 werden Befunde zu den erreichten Lernständen in den Testbereichen Wortkenntnis, Leseverständnis, Informationsentnahme aus diskontinuierlichen Texten und Mathematik sowie für einen zusammengefassten Gesamttestwert mitgeteilt, wobei sich die Aussagen auf die getestete bzw. befragte Schülergruppe der Acht-, Neunt- und Zehntklässler insgesamt beziehen. Auch für die Aufgaben zum Problemlösen und die Einstellungsmerkmale werden in Abschnitt 3 noch keine Differenzierungen nach Schulen, Klassen,

Klassenstufen, Geschlecht, Migrantenstatus, Schülerfirma oder anderen Kriterien vorgenommen.

Im Abschnitt 4 wird der Frage nachgegangen, inwieweit zwischen bestimmten Schülergruppen systematische Unterschiede in den Lernleistungen bestehen, wobei die soeben genannten Gruppierungsmerkmale den Abschnitt gliedern. Der Abschnitt 5 behandelt analog die Ergebnisse hinsichtlich der schul- und zukunftsrelevanten Einstellungen, Einschätzungen und Werthaltungen der untersuchten Jugendlichen.

In Abschnitt 6 schließlich werden die wichtigsten Befunde der Untersuchung zusammengefasst.

2 Anlage und Durchführung der Untersuchung

In diesem zweiten Berichtsabschnitt werden die Auswahl und der Aufbau der Erhebungsinstrumente erläutert (2.1), und es werden Informationen zur Datenstruktur einschließlich der Teilnahme- und Bearbeitungsquoten mitgeteilt (2.2).

2.1 Erhebungsinstrumente

Im Rahmen der Untersuchung wurden ein standardisierter Schulleistungstest, ein neu entwickeltes Instrument zur Erfassung der Fähigkeit, einfache Probleme zu lösen, und ein Schülerfragebogen eingesetzt. Diese Instrumente wurden ergänzt durch einen weitgehend sprachfreien Test zum schlussfolgernden Denken. Darüber hinaus wurden den amtlichen Schülerakten bestimmte demographische und schullaufbahnbezogene Angaben entnommen. Im Folgenden werden diese Erhebungsinstrumente genauer beschrieben. Für die eingesetzten Tests und für die Einstellungsskalen des Fragebogens werden Kennwerte zur Güte der Messverfahren nachgewiesen.

2.1.1 Fachleistungstest

In der Untersuchung LABEL 8-10 galt es als besondere Herausforderung, ein Instrumentarium zur Bilanzierung von *Fachleistungen* in den Kompetenzbereichen Sprache (Wortkenntnis und Leseverständnis), Umgang mit diskontinuierlichen Aufgaben (Karten, Tabellen, Diagrammen) und Mathematik auszuwählen. Hinsichtlich der Auswahlkriterien galten folgende Vorüberlegungen:

- Es musste auf ein bereits vorliegendes Instrument zurückgegriffen werden. Für die Entwicklung eines spezifischen, auf die Erfordernisse der besonderen Situation von Schülerinnen und Schüler der Schule für Lernbehinderte abgestimmten Fachleistungstests waren weder die zeitlichen noch die finanziellen Ressourcen vorhanden.
- Es sollte ein erprobtes Instrument eingesetzt werden, das in der vergleichend angelegten Untersuchung unter standardisierten Bedingungen als Gruppentest einsetzbar ist.
- Dem Instrument sollte ein Testkonzept zugrunde liegen, das den Standards und Anforderungen moderner Schulleistungsstudien genügt.

Zur Bestimmung des erreichten Leistungsstandes im Leseverständnis, in der Informationsentnahme aus Karten, Tabellen und Diagrammen und in Ma-

thematik wurde nach ermutigenden Ergebnissen aus Vorerprobungen auf Unterskalen des *Hamburger Schulleistungstest für 4. und 5. Klassen – HST 4/5* (MIETZEL & WILLENBERG 2000) zurückgegriffen¹. Dieser Test wurde für die Hamburger Untersuchung *Aspekte der Lernausgangslage in fünften Klassen* (LAU 5; vgl. Lehmann et al. 1997) entwickelt. Nach Experteneinschätzungen sind die in den genannten Unterskalen formulierten Anforderungen so gestellt, dass sie im Bereich der Oberstufe der Schule für Lernbehinderte zur Bilanzierung von grundlegenden fachbezogenen Kompetenzen eingesetzt werden können. Es darf hiernach vermutet werden, dass sie curricular und im Hinblick auf den Übergang in berufsvorbereitende und -qualifizierende Maßnahmen auch prognostisch valide sind.

Der Untertest *Leseverständnis* besteht aus insgesamt 27 Aufgaben, die sich auf zwei Prosa- und zwei Sachtexte beziehen. Die Aufgaben decken ein breites Spektrum von Schwierigkeiten ab: Angefangen mit Aufgaben, die vergleichsweise einfache Fertigkeiten und Fähigkeiten abrufen, werden auf den höheren Anforderungsniveaus komplexere Aufgaben gestellt, die zunehmend eigenständige mentale Operationen erfordern. Insgesamt geht es um

- das Benennen einzelner, wörtlich vorgegebener Textinformationen auf ausdrückliche Frage hin,
- das Benennen von Textinformationen, die unter einschränkenden Bedingungen identifiziert oder erschlossen werden müssen,
- das Erschließen von nicht ausdrücklich genannten, aber implizierten Sachverhalten und Beziehungen,
- das Bewerten und Beurteilen von Sachverhalten und Beziehungen und
- das Erkennen einer angedeuteten Handlungsabsicht, des Hauptthemas eines Textes oder der wesentlichen Aussage in einem Absatz oder Abschnitt².

Bezogen auf sprachliche Inhalte wurden den Schülerinnen und Schülern zusätzlich zum Untertest *Leseverständnis* Aufgaben gestellt, die das Bedeutungswissen von Wörtern messen. Die 14 Aufgaben des Untertests *Wortkenntnis* bestehen aus einem hervorgehobenen Wort, das in einen kurzen sprachlichen Zusammenhang gestellt ist, und vier weiteren Wörtern, von de-

¹ In 10 Berliner Schulen für Lernbehinderte erfolgte im Februar 2002 eine Stichprobe (N=180), die von MIGULLA ausgewertet wurde.

² Zur eingehenden Beschreibung der hier angedeuteten Kompetenzstufen im Leseverständnis vgl. BAUMERT ET AL. 2001; OECD / STATISTICS CANADA 1995; LEHMANN ET AL. 1995.

nen eines synonym zum hervorgehobenen Wort gebraucht werden kann und von den Schülerinnen und Schülern entsprechend zugeordnet werden soll.³

Der Untertest *Informationsentnahme aus Karten, Tabellen und Diagrammen* des *HST 4/5* erfasst Fähigkeiten im Umgang mit diskontinuierlichen Texten. Insgesamt 16 Aufgaben prüfen das Kartenverständnis. Die Schülerinnen und Schüler sind hier aufgefordert,

- sich anhand einer Karte zu orientieren,
- Himmelsrichtungen zu bestimmen,
- bestimmte Orte auf einer Karte zu lokalisieren und
- Wegstrecken nachzuvollziehen.

Fünf zusätzliche Aufgaben erfordern auf der Grundlage von grafischen Darstellungen mit mathematischen Inhalten (Tabellen, Diagrammen)

- eine genaue Wahrnehmung,
- die Kenntnis grafischer Repräsentationen von mathematischen Sachverhalten und
- das Erkennen identischer Beziehungen in unterschiedlich dargebotenen Formen.

Ein vierter Untertest – *Mathematik* – spiegelt ein Spektrum ausgewählter mathematischer Teilgebiete wider, deren Beherrschung in der Oberstufe der Schule für Lernbehinderte erwartet werden kann. Im Bereich *Zahlenverständnis* werden Kenntnisse der Ziffern- und Stellenwerte, Kenntnisse des Zusammenhangs zwischen Zahlwort und Zahldarstellung, Grundkenntnisse der Rechenoperationen und Fähigkeiten im Umgang mit Balkendiagrammen und Zahlenstrahlen gefordert, im Bereich *Größen* werden Kenntnisse der Maßeinheiten überprüft, und der Bereich *Rechnen* umfasst Aufgaben mit überschaubaren Sachverhalten zur Prüfung der Rechenfertigkeit.

Die Aufgaben sind durchgängig im Multiple-Choice-Format mit jeweils vier Distraktoren pro Aufgabe gestaltet; die Testzeit für die vier Untertests betrug 75 Minuten, die auf zwei Vormittage verteilt waren.

2.1.2 Problemlösetest

Die Testkomponente „Probleme lösen“ wurde speziell für die vorliegende Studie entwickelt (Adling/Belusa/Duismann/Meschenmoser/Peek). Dabei

³ In der Untersuchung wurde außerdem ein Untertest *Sprache* eingesetzt, der jedoch wegen der erkennbaren Schwierigkeiten der Schülerinnen und Schüler mit der Aufgabenstellung und der damit einher gehenden unzureichenden Reliabilität der Ergebnisse keine eingehenden Analysen zuließ.

handelt es sich um drei Teile zu den Themen *Fahrradtour*, *Geburtstagsfeier* und *Wandertag*. Inhaltlich ging es beispielsweise beim Wandertag um die Planung und dafür

- ein Ziel zu finden,
- den Tagesablauf zu planen,
- die Fahrtkosten und Fahrzeiten auszurechnen und
- Überlegungen zur Besorgung der Eintrittskarten anzustellen.

Viele der Aufgaben können so interpretiert werden, dass es sich dabei um Anforderungen an die höheren Stufen des Leseverständnisses handelt, also darum, Implikationen zu erschließen. Dies spielt insbesondere dann eine Rolle, wenn es darum geht, die Konsequenzen bestimmter Handlungsoptionen zu beurteilen. Besondere Bedeutung besitzen auch die Fragen nach der Sequenzierung der erforderlichen Arbeitsschritte. Für den Bereich der Allgemeinen Schulen liegen entsprechende Erfahrungen bereits aus der Hamburger Lernausgangslagenuntersuchung (LEHMANN ET AL. 1999) vor.

2.1.3 Test zum schlussfolgernden Denken

Der Fachleistungstest wurde durch den weitgehend sprachfreien *Culture Fair Intelligence Test – CFT 20* in einer Kurzform (WEISS 1998) ergänzt. Dieser Test zielt auf die Messung der sog. flüssigen Grundintelligenz und erfasst Aspekte des schlussfolgernden Denkens, also sehr allgemeine kognitive Lernvoraussetzungen. Die Aufgaben sind in vier Subtests (*Reihen fortsetzen; Klassifikationen; Matrizen; topologische Schlussfolgerungen*) untergliedert. Die einzelnen Subtests bestehen aus zeichnerisch dargestellten und nach Schwierigkeiten geordneten Einzelaufgaben im Multiple-Choice-Format. Seine figurale Gestaltung macht diesen standardisierten Test vom Grad der Beherrschung der deutschen Sprache sowie von fachgebundenem Wissen und Können nahezu unabhängig und gestattet es, *Diskrepanzen* zwischen dem kognitiven Potenzial einer Schülerin bzw. eines Schülers und ihren oder seinen fachgebundenen Fähigkeiten zu erkennen. Die Durchführung des CFT 20 in seiner Kurzform mit insgesamt 46 Aufgaben erfordert ca. 35 Minuten.

2.1.4 Schülerfragebogen

Mit Hilfe eines Schülerfragebogens wurden Einstellungen und Selbsteinschätzungen der Schülerinnen und Schüler gegenüber den Bereichen Lernen und Arbeiten gemessen. Der Fragebogen enthielt insgesamt 110 Aussagen u.a. zu den Themen *Selbstwirksamkeit*, *leistungsbezogenes Selbstkonzept*, *Belastbarkeit*, *eigene Arbeitstugenden* und *Affinität zur Technik*, die jeweils auf einer vierstufigen Skala eingeschätzt werden sollten.

2.1.5 Schülerakte

Als Quelle amtlich verfügbarer Informationen wurde der in den Schulen vorhandene *Schülerbogen* genutzt. Neben Daten zur persönlichen Situation der Schülerinnen und Schüler (Alter, Geschlecht, Nationalität, Herkunftssprache) ging es hier vor allem um Rahmendaten zur Schullaufbahn: Zurückstellung, Einschulungsdatum, Wiederholung einer Klasse, Informationen zur Feststellung sonderpädagogischen Förderbedarfs, Schulwechsel, insbesondere Zeitpunkt des Übergangs an die Schule für Lernbehinderte, Noten des letzten Zeugnisses für die Fächer Deutsch, Mathematik, Technisch-naturwissenschaftlicher Unterricht und Arbeitslehre sowie die Anzahl der im letzten Schuljahr registrierten Fehltage.

Bei den Informationen aus den Schülerbögen wurden insbesondere auch solche Daten aufgenommen, die Aussagen über absolvierte Praktika und Tätigkeiten im Bereich von Schülerfirmen zuließen (Berufsfelder, Zeiträume und Organisationsformen).

2.2 Datenstruktur

Es wurde angestrebt, mit dem Fachleistungstest, dem CFT 20 und dem Schülerfragebogen sowie mit den Angaben aus den Schülerbögen Informationen für sämtliche Schülerinnen und Schüler aus achten, neunten und zehnten Klassen in Berliner Schulen für Lernbehinderte zu erhalten. Die tatsächlich erreichten Schülerinnen und Schüler dieser Zielgruppe umfassten insgesamt 41 Schulen mit 201 Klassen und 2.239 Schülerinnen und Schülern⁴. Die folgende Abbildung 1 zeigt, wie sich die einbezogenen Schulen auf die ehemaligen 23 Stadtbezirke verteilen. Zusätzlich zur Angabe der Anzahl der Schulen pro Bezirk ist in der Legende jeweils die Anzahl der in die Untersuchung einbezogenen Schülerinnen und Schüler eingetragen.

⁴ Über die 41 Schulen für Lernbehinderte hinaus haben sich Lerngruppen mit dem sonderpädagogischem Schwerpunkt Lernen beteiligt, deren Schulen ebenfalls am *Netzwerk Berliner Schülerfirmen* teilhaben. Diese Lerngruppen befinden sich an einer Schule für geistig behinderte Schülerinnen und Schüler, an zwei Schulen für körperbehinderte Schülerinnen und Schüler und einer Berufsschule mit sonderpädagogischer Aufgabe. Die Werte der Lerngruppen dieser vier Schulen bleiben in diesem Bericht aufgrund der gegenüber den Schülerinnen und Schülern der Schulen für Lernbehinderte sehr unterschiedlichen Voraussetzungen unberücksichtigt.

Abbildung 1 Verteilung der in die Untersuchung einbezogenen Schulen ($N = 41$) sowie aller erfasster Schülerinnen und Schüler (Klassenstufen 8, 9, 10: $N = 2.239$; Klassenstufe 11: 62), nach Bezirken



<p>1 Wedding 2 Schulen, 134 Schüler Mitte 1 Schule, 40 Schüler Tiergarten 1 Schule, 24 Schüler</p>	<p>4 Charlottenburg 1 Schule, 42 Schüler Wilmersdorf 1 Schule, 65 Schüler</p>	<p>8 Neukölln 6 Schulen, 285 Schüler</p>
<p>2 Friedrichshain 1 Schule, 47 Schüler Kreuzberg 1 Schule, 56 Schüler</p>	<p>6 Zehlendorf 1 Schule, 43 Schüler Steglitz 1 Schule, 49 Schüler</p>	<p>9 Treptow 1 Schule, 89 Schüler Köpenick 1 Schule, 72 Schüler</p>
<p>3 Pankow 2 Schulen, 98 Schüler Weißensee 1 Schule, 40 Schüler Prenzlauer Berg 2 Schulen, 104 Schüler</p>	<p>7 Schöneberg 1 Schule, 64 Schüler Tempelhof 2 Schulen, 73 Schüler</p>	<p>10 Marzahn 3 Schulen, 192 Schüler Hellersdorf 2 Schulen, 251 Schüler</p>
		<p>11 Lichtenberg 2 Schulen, 104 Schüler Hohenschönhausen 3 Schulen, 223 Schüler</p>
		<p>12 Reinickendorf 3 Schulen, 144 Schüler</p>

Die folgende Tabelle 1 zeigt die Fallzahlen für die berücksichtigten Klassen sowie für die entsprechenden Schülerinnen und Schüler nach Klassenstufen. Gezählt sind hier alle Schülerinnen und Schüler, für die mindestens zu einem der obligatorischen Erhebungsinstrumente (Fachleistungstest, CFT 20, Schülerfragebogen, Schülerakte) Angaben vorliegen.⁵

Tabelle 1 Stichprobenstruktur: Anzahl der berücksichtigten Klassen sowie der Schülerinnen und Schüler nach Klassenstufe

	Klassen	Schülerinnen und Schüler
Klassenstufe 8	95	1.073
Klassenstufe 9	84	933
Klassenstufe 10	22	233
<i>insgesamt</i>	<i>201</i>	<i>2.239</i>

In der folgenden Tabelle 2 sind die Schülerinnen und Schüler differenziert nach den Merkmalen Geschlecht, Herkunftssprache und Schülerfirma für die drei in die Analysen einbezogenen Klassenstufen aufgelistet.

Tabelle 2 Stichprobenstruktur: Anzahl der Schülerinnen und Schüler in den Klassenstufen 8, 9 und 10, nach Geschlecht, Herkunftssprache und Beteiligung am Netzwerk Berliner Schülerfirmen

	Klassenstufe			
	8	9	10	<i>insgesamt</i>
nach Geschlecht⁶				
Mädchen	367	315	80	762
Jungen	559	503	124	1.186
nach Herkunftssprache⁷				
deutsch	772	718	190	1.680
nicht deutsch	197	154	27	378
nach Beteiligung am Netzwerk Berliner Schülerfirmen				
Schülerfirma	73	184	149	406
nicht Schülerfirma	1.000	749	84	1.833

⁵ Zusätzlich haben sich sieben 11. Klassen mit insgesamt 62 Schülerinnen und Schülern an der Untersuchung beteiligt. Aufgrund der niedrigen Quote insgesamt und der besonderen Situation dieser Klassenstufe in der Schule für Lernbehinderte bleiben diese Daten im Weiteren unberücksichtigt.

⁶ Für 291 Schülerinnen und Schüler liegt die Angabe zum Geschlecht nicht vor.

⁷ Für 181 Schülerinnen und Schüler liegt die Angabe zur Herkunftssprache nicht vor.

Für eine Teilgruppe der Schülerinnen und Schüler gilt, dass sie zusätzlich einen fachübergreifenden Test zum Problemlösen bearbeitet haben. Die folgende Tabelle 3 zeigt, wie viele Schülerinnen und Schüler das waren.

Tabelle 3 Datenstruktur der Erhebungen im Test *Problemlösen*: Anzahl Schülerinnen und Schüler in den Klassenstufen 8, 9 und 10, nach Geschlecht, Herkunftssprache und Schülerfirma

	Klassenstufe			<i>insgesamt</i>
	8	9	10	
nach Geschlecht⁸				
Mädchen	76	36	14	126
Jungen	122	52	41	215
nach Herkunftssprache⁹				
deutsch	140	58	56	254
nicht deutsch	46	31	2	79
nach „Schülerfirma“				
Schülerfirma	23	21	40	84
nicht Schülerfirma	187	69	20	276

Die Information der beteiligten Schulen über die Ziele und Verfahren der Untersuchung und die Organisation der Datenerhebung (Versorgung der Schulen mit Test- und Befragungsmaterialien; Beratung der Testleiterinnen und Testleiter) erfolgte durch das Referat Q2 des ehemaligen Landesschulamts der Senatsverwaltung für Bildung, Jugend und Sport des Landes Berlin.

Die Datenerhebung in den Schulen fand zwischen dem 23. September und dem 5. November 2002 statt, wobei dieser Zeitraum zwei Ferienwochen enthielt. Für die Durchführung des Fachleistungstests und des CFT 20 waren *Ambulanzlehrkräfte* verantwortlich, die bezirklich organisiert an Schulen, an denen sie selbst nicht unterrichteten, eingesetzt waren. Es handelt sich dabei um ausgebildete Sonderpädagogen.

Die Tests wurden jeweils in einer Woche an zwei Unterrichtstagen im Klassenverband durchgeführt. Der Einsatz des Schülerfragebogens lag in der Verantwortung der *Lehrkräfte für das Fach Arbeitslehre*, unterstützt durch die *Klassenlehrkräfte*. Die Entnahme von Informationen aus den Schülerbogen erfolgte ebenfalls durch die Klassenlehrkräfte, gegebenenfalls mit der Unterstützung der jeweiligen Schulleitung. Für die Durchführung der beiden letztgenannten Untersuchungsteile war als zeitliche Vorgabe lediglich der gesamte

⁸ Für 19 Schülerinnen und Schüler liegt die Angabe zum Geschlecht nicht vor.

⁹ Für 27 Schülerinnen und Schüler liegt die Angabe zur Herkunftssprache nicht vor.

effektive Erhebungszeitraum (4 Wochen) gesetzt. Die nachfolgende Tabelle 4 zeigt die Teilnahme- bzw. Bearbeitungsquoten für die obligatorischen Bestandteile der Untersuchung. Angesichts dieser hohen Quoten kann insgesamt von einer soliden Datengrundlage gesprochen werden, auf die sich die weiteren Analysen stützen.

Tabelle 4 Bearbeitungsquote der Instrumente (Schülerinnen und Schüler), nach Klassenstufe (in Prozent)

	Schulleistungstest						
	Lesever- ständnis	Wort- kenntnis	Informa- tions- entnahme	Mathe- matik	CFT 20	Frage- bogen	Schüler- akte
Klassenstufe 8	91	91	92	90	89	89	94
Klassenstufe 9	93	93	94	92	92	92	97
Klassenstufe 10	95	95	98	97	97	90	97
<i>insgesamt</i>	92	92	94	91	91	90	96

Die Konstruktion bzw. Zusammenstellung der Erhebungsinstrumente lag in der Verantwortung der wissenschaftlichen Forschungsgruppe an der Humboldt-Universität zu Berlin und der wissenschaftlichen Begleitung des Projekts „Netzwerk Berliner Schülerfirmen“. Die Datenerfassung, -aufbereitung und -auswertung sowie die Erstellung des vorliegenden Endberichts über die Untersuchung waren Aufgabe des Autorenteam¹⁰.

¹⁰ Mit Rückmeldungen von klassenbezogenen Ergebnissen an die Schulen wurde – unter inhaltlicher Verantwortung der Forschungsgruppe an der Humboldt-Universität zu Berlin und der wissenschaftlichen Begleitung des Projekts „Netzwerk Berliner Schülerfirmen“ – ein auf EDV-Auftragsarbeiten spezialisiertes Unternehmen (Firma Daten-Werk, Petershagen und Berlin) beauftragt. Diese Rückmeldungen wurden mit ausführlichen Hinweisen (DUISMANN/MESCHENMOSER 2003) im Mai 2003 an die beteiligten Schulen versandt.

3 Befunde zu den Fachleistungen und Einstellungen der Schülerinnen und Schüler

3.1 Fachleistungen

Der Auswertung der Fachleistungstests liegt ein statistisches Verfahren zugrunde, das im Zusammenhang großflächig angelegter Schulvergleichsuntersuchungen zum Standard geworden ist. Nach Maßgabe der probabilistischen Testtheorie (vgl. ROST 1996) wurden die gestellten Aufgaben der fachbezogenen Untertests unter Nutzung des sog. „inparametrischen Rasch-Modells“ skaliert. Im Unterschied zu klassischen Methoden der Testauswertung wird bei diesem Verfahren nicht vorausgesetzt, dass alle gestellten Aufgaben gleich schwierig sind. Vielmehr wird das Leistungsniveau (die *Fähigkeit*) einer Schülerin oder eines Schülers unter Berücksichtigung der *Schwierigkeit* der gelösten Aufgaben ermittelt. Das Verfahren erlaubt es u.a.,

- die Einschlägigkeit von Testaufgaben für die theoretisch unterstellte, gemeinsame Fähigkeitsdimension empirisch zu prüfen und davon abweichende Items auszuschließen sowie
- den Schwierigkeitsgrad von Testaufgaben und die Fähigkeit von Personen auf demselben Maßstab abzubilden.

Bei dem Skalierungsverfahren wird die Fähigkeit einer Schülerin bzw. eines Schülers über Wahrscheinlichkeiten definiert. Dabei steigt generell die geschätzte Fähigkeit eines Probanden mit der jeweiligen Gesamtzahl der richtigen Antworten. Umgekehrt hängt die Festsetzung der Aufgabenschwierigkeiten von der jeweiligen Anzahl richtiger Lösungen ab: Je weniger Schüler eine Aufgabe gelöst haben, um so größer ist deren in diesem Testmodell geschätzte Schwierigkeit. Wenn sich nun die Antwortmuster der Schülerinnen und Schüler in guter Näherung so darstellen lassen, dass die Aufgaben stets nur bis zu einer für die einzelne Schülerin bzw. für den einzelnen Schüler bestimmbaren Schwierigkeit gelöst werden, können für alle Aufgaben und Fähigkeitsgruppen die Lösungswahrscheinlichkeiten berechnet werden.

Innerhalb des Rasch-Modells lässt sich die Bewältigung von Aufgaben eines bestimmten Schwierigkeitsgrades als diejenige Leistung interpretieren, die von einer Schülerin bzw. einem Schüler des entsprechenden Fähigkeitsniveaus mit einer genau festgelegten Wahrscheinlichkeit erbracht wird. Grundsätzlich kann diese Entsprechung für beliebige Wahrscheinlichkeiten definiert werden. Bei der konventionellen Festlegung auf die Wahrscheinlichkeit von $p = 0,65$ kann man indessen davon ausgehen, dass bei der Projektion von Aufgabenschwierigkeiten und Schülerfähigkeiten auf eine und dieselbe Skala eine

plausible Mindestanforderung formuliert ist, oberhalb derer die Aussage „Der Schüler i ist kompetent, die Aufgabe j und gleich schwierige Aufgaben derselben Domäne zu lösen“ allererst sinnvoll ist.

Übersteigt nun die auf solcher Grundlage geschätzte Fähigkeit einer Schülerin oder eines Schülers die Schwierigkeit einer Aufgabe, so wird sie bzw. er diese mit mindestens $p = 0,65$ und im Übrigen um so eher lösen, je größer die Differenz zwischen Fähigkeits- und Schwierigkeitsparameter ist. Unterschreitet umgekehrt die geschätzte Fähigkeit des Schülers oder der Schülerin die Aufgabenschwierigkeit, so wird er oder sie um so eher scheitern, je größer die Differenz ist.

Die Ergebnisse der nach dem Rasch-Modell ausgewerteten Untertests wurden für die weitere Darstellung linear so transformiert, dass die durchschnittliche Schülerleistung auf 100 und die Standardabweichung auf 20 gesetzt sind¹ (zum Verfahren der Rasch-Skalierung vgl. ausführlicher auch DEUTSCHES PISA-KONSORTIUM 2001, S. 517ff.).

3.1.1 Leseverständnis

Die Eigenschaften der Aufgaben im Leseverständnistest und die bei den Schülerinnen und Schülern gemessenen Kompetenzen im Leseverständnis sind also über die Rasch-Skalierung zugleich miteinander verschränkt und aufeinander projizierbar. Die Gesamtskala für das Leseverständnis umfasst aufgabenseitig unterschiedliche Textformen (Prosa- und Sachtexte) und Schwierigkeitsgrade², schülerseitig unterschiedliche Kompetenzgrade oder Kompetenzstufen. Dabei ist es möglich und sinnvoll, die Verteilung der Aufgabenschwierigkeiten derjenigen des gemessenen Leseverständnisses grafisch gegenüberzustellen.

¹ Die Wahl der Transformationsregel unterliegt im Rahmen der Rasch-Skalierung einzig Gesichtspunkten der Zweckmäßigkeit. Die bei der Rasch-Skalierung zunächst nicht festliegenden Mittelwerte werden üblicherweise auf einen „runden Wert“ gesetzt; als „Maßeinheit“ dient dann meist eine Größe, die von der Standardabweichung als Kennwert für die Streuung abgeleitet ist. Dabei wird angestrebt, dass die Fähigkeitsparameter im Rahmen der gegebenen Messgenauigkeit als ganze positive Zahlen dargestellt werden können.

² Von den insgesamt 27 Aufgaben zum Leseverständnis erwiesen sich zwei Aufgaben auf Grund ihrer zu hohen Schwierigkeit und der zu niedrigen Trennschärfe als ungeeignet. Sie bleiben deshalb bei den weiteren Analysen unberücksichtigt. Die Skala *Leseverständnis* umfasst somit für die weitere Berichtslegung 25 Items.

Abbildung 2 Schülerfähigkeiten im Vergleich mit den Schwierigkeiten der Testaufgaben im LABEL 8-10-Leseverständnistest, insgesamt und differenziert nach Textformen

Verteilung der Schülerleistungen

Testaufgaben

Beispielaufgaben zum Sachtext „Mücken“

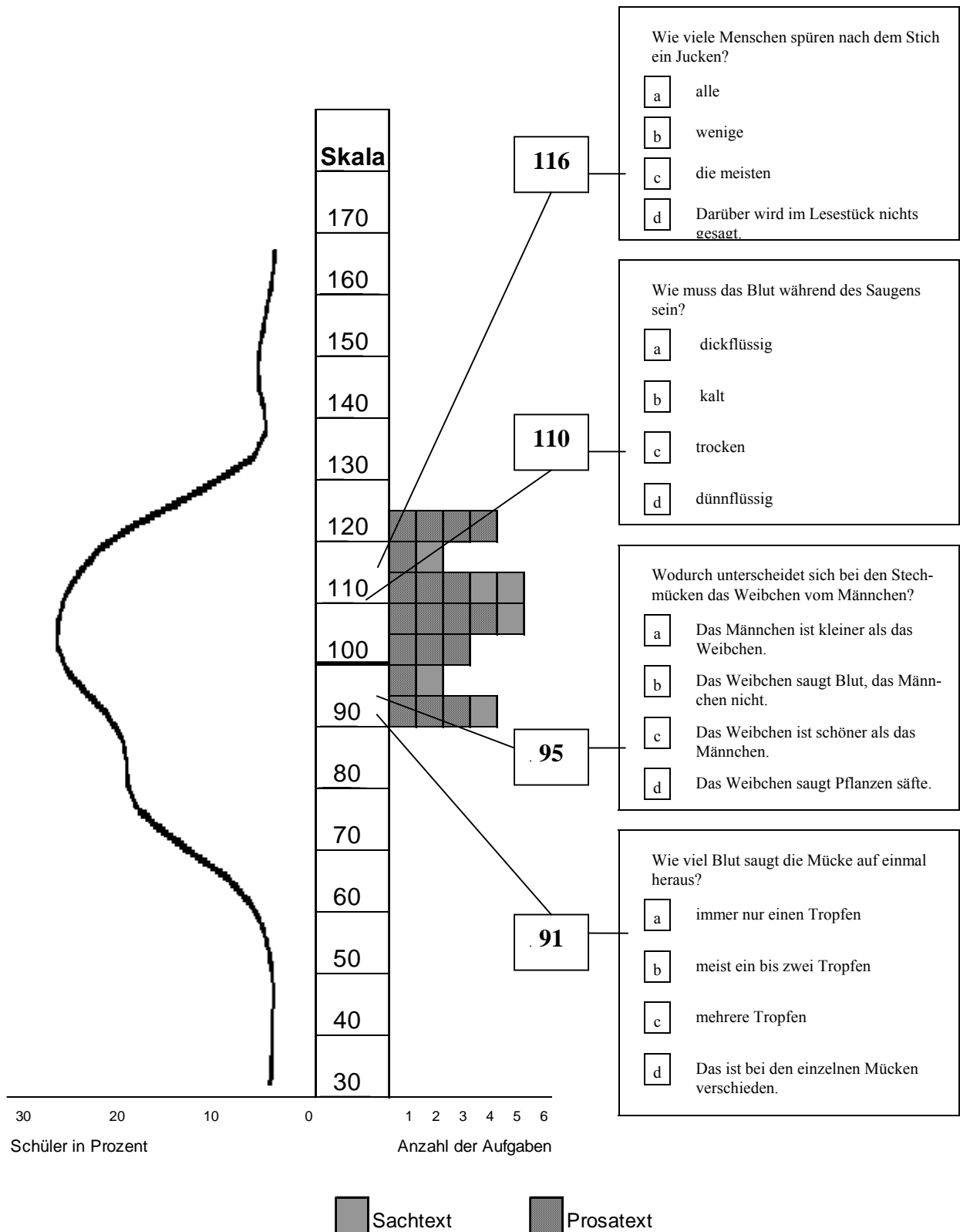


Abbildung 2 zeigt das Ergebnis für die Gesamtgruppe der getesteten Schülerinnen und Schüler, wobei die insgesamt 25 Aufgaben, die sich auf vier Einzeltexte beziehen, mit ihrem Rasch-Skalenwert ausgewiesen sind.

Hinsichtlich der Aufgabenschwierigkeiten hat sich bei einem gesetzten Mittelwert von 100 eine vergleichsweise geringe Reichweite ergeben, die sich von 90 bis 122 Skalenpunkten erstreckt. Die Aufgaben, die sich auf die beiden Sachtexte Kerzen und Mücken beziehen, waren für die untersuchten Schülerinnen und Schüler im Vergleich zu den Prosaaufgaben geringfügig einfacher, aber ähnlich breit über die Skala gestreut.

Das Leseverständnis ist – wie auf der linken Seite der Grafik ersichtlich – keineswegs normalverteilt. Zwar ist die für Normalverteilungen typische Konzentration der Werte im mittleren Bereich zu erkennen, wobei einige Schülerinnen und Schüler extrem schwache Verständnisleistungen (Minimum = 34 Skalenpunkte) gezeigt haben, während andere gute und sehr gute Werte (Maximum = 163) erzielen konnten. Gleichwohl ist diese Tendenz durch eine zusätzliche Häufung (gleichsam durch einen „Nebengipfel“) im unterdurchschnittlichen Leistungsbereich charakterisiert. Diese Erscheinung, die in anderen Leistungsbereichen noch deutlicher aufgetreten ist, bedarf noch der näheren Analyse.

Der Mittelwert für das Leseverständnis liegt mit dem definierten Skalenpunkt 100 (Standardabweichung: 20) etwas niedriger als die mittlere Aufgabenschwierigkeit (110). Der eingesetzte Test war also für die Zielgruppe keineswegs generell zu schwer, sondern durchaus angemessen. In den Extrembereichen, insbesondere im unteren Teil des Leistungsspektrums der Zielgruppe, diskriminiert der Test jedoch nicht sonderlich gut.

Bei der Testentwicklung wurden gleichwohl Schwierigkeitsunterschiede erfasst, die einen beträchtlichen Anforderungsbereich abdecken, der von einfachen Dekodierungsaufgaben bis hin zum schlussfolgernden Umgang mit Texten reicht. Dies bestätigt sich empirisch, wenn man die in den einzelnen Items jeweils geforderten Leistungen näher untersucht. Die Tendenz von einfachen Aufgaben des (Wieder-) Erkennens zu komplexeren mentalen Operationen ist deutlich. Immer erscheinen solche Aufgaben am leichtesten, die mit dem Auffinden eines Wortes, einer kürzeren Passage oder auch eines Symbols lediglich einen direkten Abgleich zwischen Text und Frage verlangen. Die Schwierigkeit steigt in dem Maße, wie komplexere Randbedingungen präsent gehalten werden müssen und wie der Gehalt umfangreicherer Passagen zunächst verdichtet und gegebenenfalls weiterverarbeitet werden muss.

In der Abbildung 2 sind – beispielhaft – aus dem im Folgenden abgedruckten Sachtext „*Mücken*“ vier Aufgaben wiedergegeben, die das Prinzip der Hierarchie zunehmend schwierigerer Anforderungen verdeutlichen. Die Frage: „*Wie viel Blut saugt die Mücke auf einmal heraus?*“ (Beispielaufgabe 1) erfordert auf einer niedrigen Anforderungsstufe (Rasch-Skalenwert 91) die Fähigkeit, eine gegebene Formulierung aufzufinden, die wortwörtlich im Text steht. 67,7 Prozent der getesteten Schülerinnen und Schüler haben diesen Fähigkeitsstufe der einfachen Informationsentnahme übertroffen, 32,3 Prozent lagen dagegen darunter.

Beispieltext „Mücken“

Wenn uns eine Mücke umschwirrt, verjagen wir sie. Das ist eigentlich nicht immer nötig, denn nicht alle Mücken stechen. Nur die Weibchen der Stechmücken saugen Blut, die Männchen dagegen ernähren sich von Pflanzensäften. Trotzdem sehen wir uns vor, denn nur wenigen Menschen macht ein Mückenstich nichts aus. Bei den meisten schwillt die Haut um den Stich herum an. Die Mücke sticht zuerst mit einem scharfen Stchrüssel zu. Mit einer Saugröhre im Stechrüssel saugt sie ein bis zwei Tropfen Blut heraus. Eine Flüssigkeit am Ende der Saugröhre verhindert, dass dieses Blut dick wird oder gar eintrocknet. Diese Flüssigkeit vertragen die meisten Menschen nicht. Eine winzige Spur davon, die unter unserer Haut zurückbleibt, lässt die Stelle schon anschwellen. Fast immer spüren wir gleichzeitig ein leichtes Jucken.

Die etwas anspruchsvollere Beispielaufgabe 2 („*Wodurch unterscheidet sich bei den Stechmücken das Weibchen vom Männchen?*“) mit dem Rasch-Skalenwert 95 verlangt die Fähigkeit, aus dem Text einen Sachverhalt zu rekonstruieren, der über das einschränkende „*nur*“ erschlossen werden muss. Diesen Skalenwert haben 58,6 Prozent der untersuchten Schülerinnen und Schüler erreicht oder übertroffen; dagegen haben 41,4 Prozent den Raschwert unterschritten, der für die Schwierigkeit der Beispielaufgabe steht.

Die dritte Beispielfrage: „*Wie muss das Blut während des Saugens sein?*“ (Skalenwert 110) steht für ein Fähigkeitsniveau, das noch von 32,4 Prozent der in die Studie einbezogenen Schülerinnen und Schülern übertroffen, hingegen von 67,6 Prozent nicht erreicht worden ist. Hier ist verlangt, über sprachlich gegebene Ausschlusskriterien („*... verhindert, dass dieses Blut dick wird oder gar eintrocknet.*“) die richtige Lösung zu erschließen.

Ein Beispiel für eine Aufgabe, die mit dem Rasch-Skalenwert 116 für die Schülerinnen und Schülern besonders anspruchsvoll war, liefert die Frage: „*Wie viele Menschen spüren nach dem Stich ein Jucken?*“ (Beispielaufgabe 4). Die richtige Antwort erfordert – über die vorgelagerten einfacheren Leseverständnisstufen hinaus – die Fähigkeit, aus der sprachlichen Formulierung „*fast immer*“ auf den Sachverhalt „*die meisten*“ zu schließen; noch schwieriger wäre die Schlussfolgerung aus der Umkehrformulierung „*nur wenige Menschen*“. Immerhin wurde der entsprechende Rasch-Skalenwert noch von 22,2 Prozent der getesteten Schülerinnen und Schüler überschritten, wenn auch von 77,8 Prozent nicht erreicht.

Mehr als ein Viertel der Schülerinnen und Schüler bewältigt demnach Aufgaben, in denen lediglich nach einer einfachen, wörtlich im Text enthaltenen Information gefragt ist, nicht mit der notwendigen Sicherheit ($p \geq 0,65$). Andererseits hat fast ein Viertel, nämlich 22,2 Prozent der Schülerinnen und Schüler von Schulen mit dem Förderschwerpunkt Lernen, Testwerte erzielt, nach denen sie durchaus befähigt sind, auch komplexere Textstrukturen inferenziell zu entschlüsseln, etwa durch die Berücksichtigung von Umkehrrelationen.

3.1.2 Wortkenntnis

In der gleichen Form wie die Werte zum Leseverständnis wurden die Schülerangaben zum Untertest *Wortkenntnis* mit insgesamt 12 Aufgaben skaliert. Die Raschskala umfasst auf der Aufgabenseite ein Spektrum, das vom leichtesten Item mit dem Schwierigkeitswert 92 bis zum schwierigsten mit dem Wert 123 reicht³. Die Schülerinnen und Schüler waren hier gefordert, zu dem fett gedruckten Wort aus einer kurzen Aussage unter den darunter angebotenen Ausdrücken denjenigen herauszufinden, der die gleiche Bedeutung hat, also ein Synonym darstellt. In der Abbildung 3 sind wiederum die Verteilungen der gemessenen Schülerfähigkeiten und der Aufgabenschwierigkeiten einander gegenüber gestellt und dazu Beispielaufgaben genannt, die für unterschiedliche Schwierigkeitsniveaus stehen.

³ Angesichts der rein definitiven Festlegung des Mittelwerts auf 100 und der Standardabweichung auf 20 bei allen hier einschlägigen Skalen lässt sich von den skalierten Werten nicht unmittelbar auf die relative Schwierigkeit der Untertests für die untersuchten Probanden schließen.

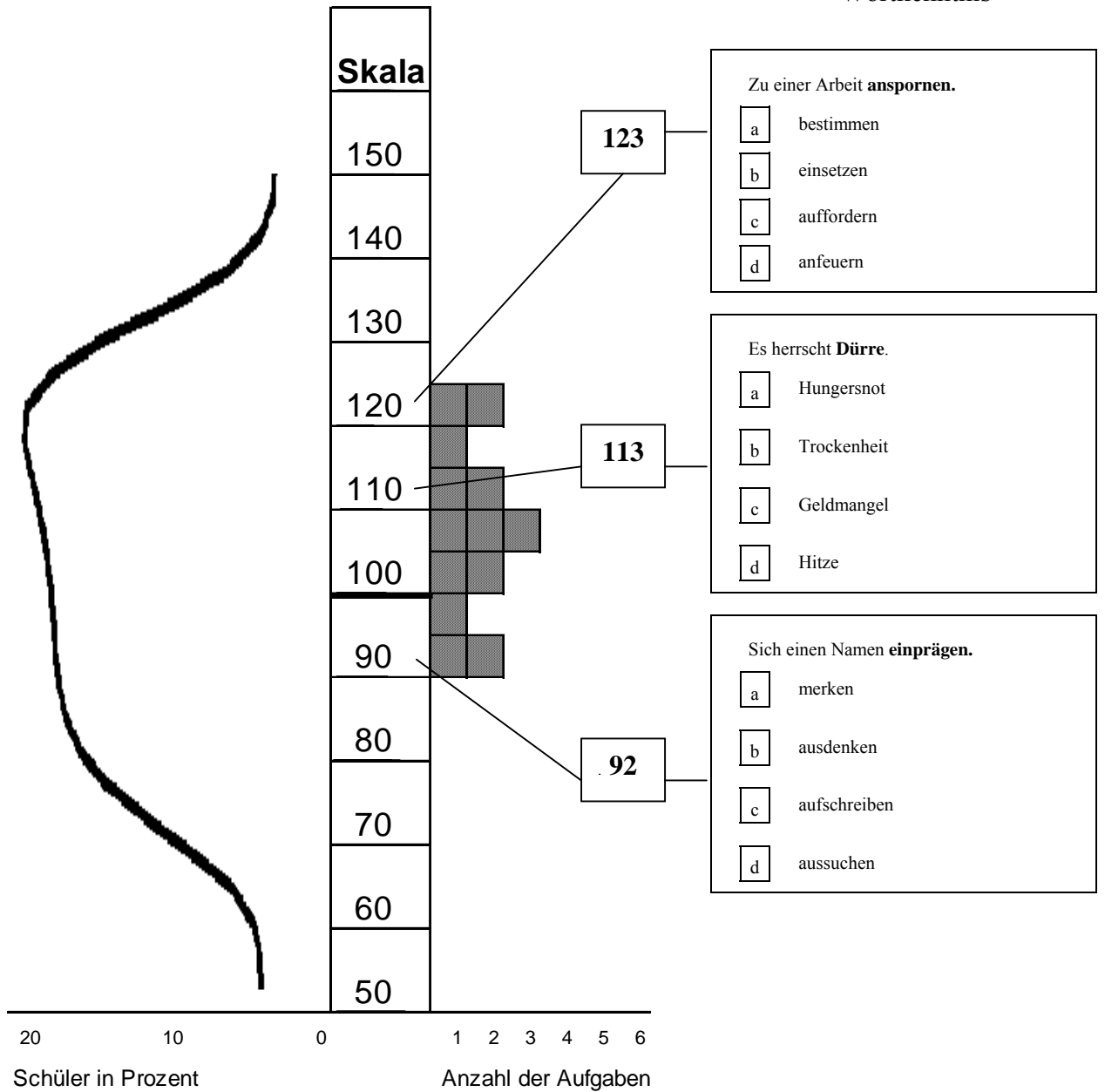
Abbildung 3

Beispielaufgaben des Untertests *Wortkenntnis*

Verteilung der Schülerleistungen

Testaufgaben

Beispielaufgaben zur Wortkenntnis



Gesamtstichprobe (N = 2.071)

Die Schwierigkeit dieser Aufgaben dürfte in erster Linie von der relativen Häufigkeit der Synonyme im Stamm oder in der Lösung der Aufgabe bestimmt sein, ggf. auch von der Suggestivität der in Frage stehenden Metapher (z.B. „einprägen“ statt „merken“). Die Schwierigkeit kann erhöht sein durch häufig auftretende Assoziationen, etwa im Falle von „Dürre“ statt des Synonyms „Trockenheit“ die in Meldungen häufig damit verbundene Konsequenz „Hungersnot“. Angesichts dessen ist es nicht sonderlich plausibel, in diesem Falle von „Kompetenzstufen“ zu sprechen. Gleichwohl bleibt es bemerkenswert, dass bei diesem Leistungsaspekt nicht etwa eine Normalverteilung zu beobachten ist, sondern dass ähnlich wie im Falle des Leseverständnisses Ansätze zu einer zweigipfligen Leistungsverteilung zu erkennen sind.

Der korrelative Zusammenhang zwischen der in diesem Testteil erfassten *Wortkenntnis* und dem *Leseverständnis* der Schülerinnen und Schüler ist mit $r = 0,67$ substantiell und positiv: Je sicherer die Kenntnis von Wortbedeutungen, desto besser ist das Leseverständnis ausgeprägt. Damit wird die konstitutive Bedeutung jener Teilkompetenz für das sinnverstehende Lesen von Texten unterstrichen.

3.1.3 Informationsentnahme aus Karten, Tabellen und Diagrammen

Dem angemessenen Umgang mit den sog. „diskontinuierlichen Texten“, also grafisch und tabellarisch dargestellten Informationen, wird im Kontext der Diskussion um erwünschte arbeitsrelevante Basiskompetenzen große Bedeutung beigemessen. Im Fachleistungstest wurden deshalb den Schülerinnen und Schülern Aufgaben zu Karten, Plänen, Diagrammen und Tabellen gestellt, die auf die Feststellung entsprechender Fähigkeiten angelegt sind.

Von den insgesamt 16 gestellten Aufgaben erwiesen sich unter teststatistischen Gesichtspunkten vier als ungeeignet für die untersuchte Schülergruppe. Sie bleiben aufgrund ihrer zu hohen Schwierigkeit und der zu niedrigen Trennschärfe bei den weiteren Analysen unberücksichtigt. Die Skala *Informationsentnahme* umfasst somit für die weitere Berichtslegung 12 Items. Die Teststruktur auf der einen Seite und die Fähigkeitsverteilung der Schülerinnen und Schüler auf der anderen lassen sich wiederum anschaulich darstellen, wenn man beide über die definitorisch auf $p = 0,65$ festgelegte Entsprechung von Aufgabenschwierigkeit und Personenfähigkeit einander gegenüber stellt.

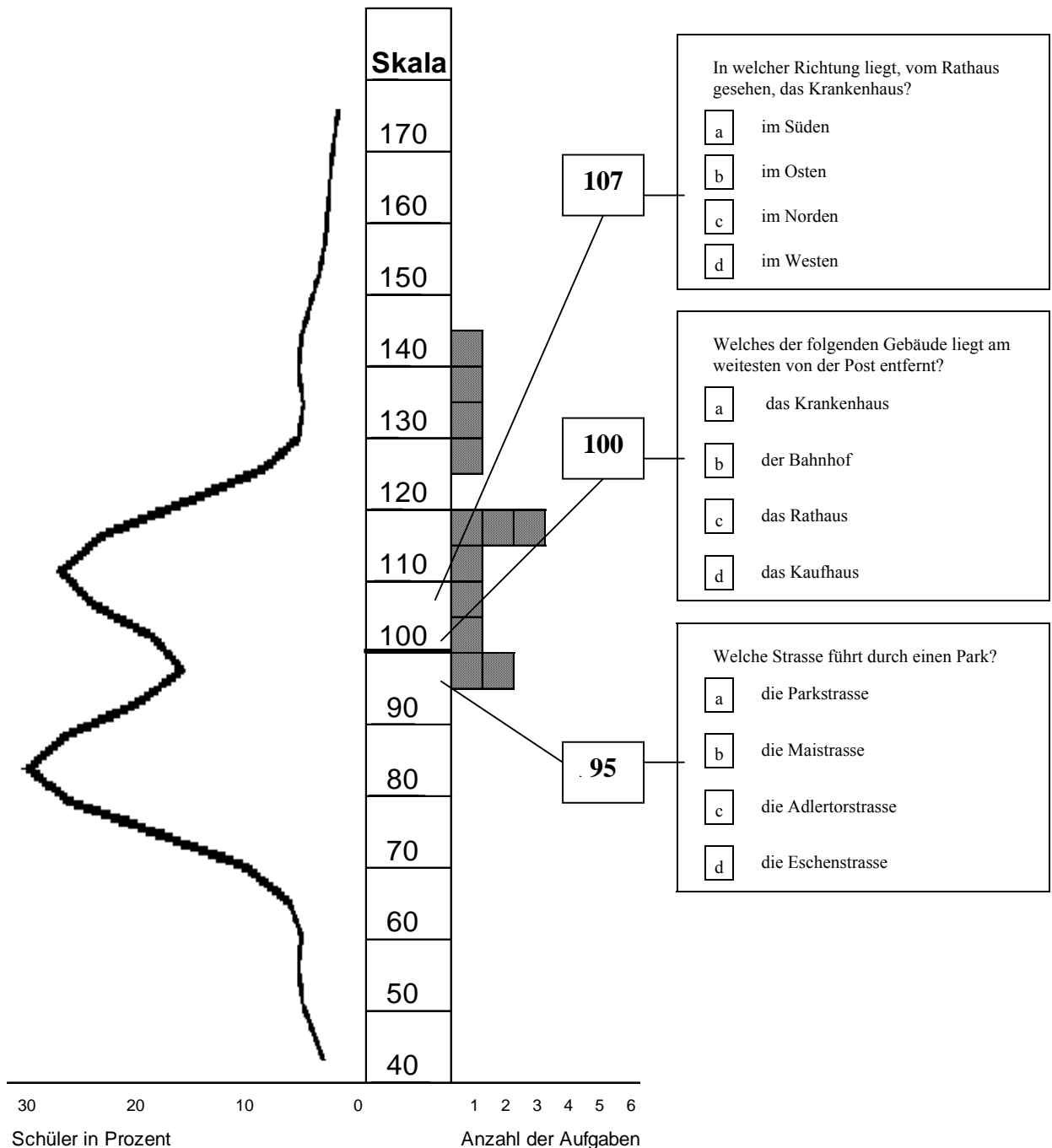
Die gegenüberliegenden Verteilungen in der Abbildung 4 zeigen, dass die Aufgaben für die Schülerinnen und Schüler insgesamt eher schwierig waren.

Abbildung 4 Verteilung der Schülerleistungen im Untertest *Informationsentnahme* im Vergleich mit den Schwierigkeiten der Testaufgaben

Verteilung der Schülerleistungen

Testaufgaben

Beispielaufgaben zur Informationsentnahme aus einem Stadtplan¹

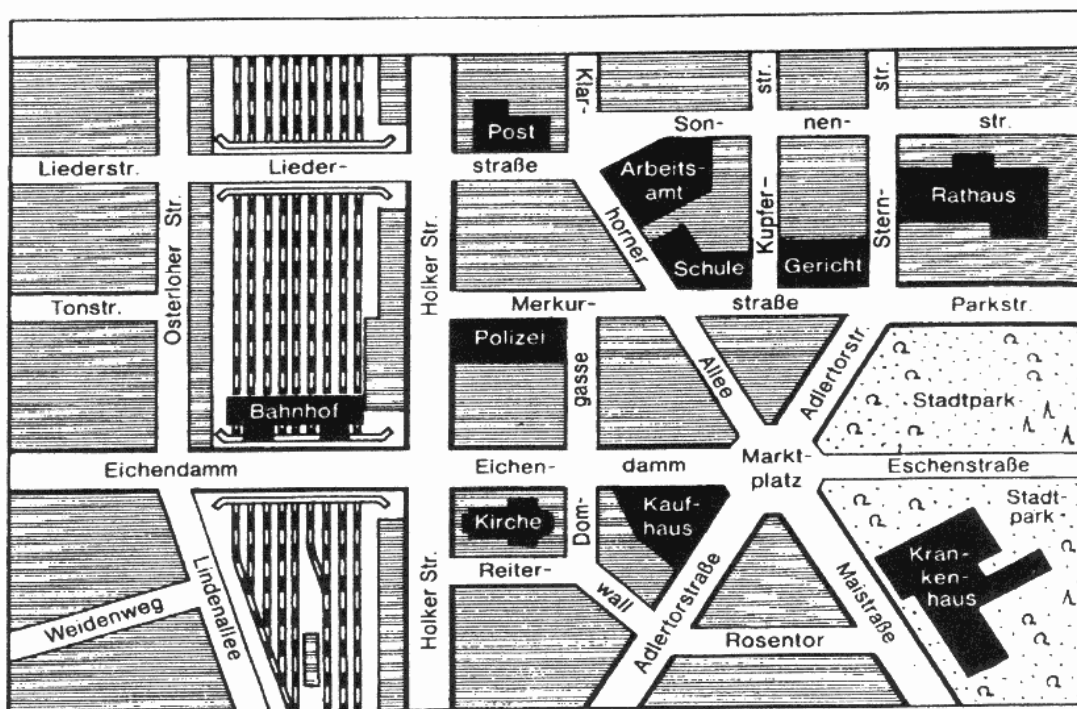


Gesamtstichprobe (N = 2.104)

¹ Stadtplan siehe folgende Seite

Die Items diskriminieren also in der hier untersuchten Gruppe am besten im Bereich der vergleichsweise guten Leistungen. Gleichwohl lassen sich die Ergebnisse für diesen Untertest im vorherigen Kontext durchaus sinnvoll interpretieren. Beispielsweise gilt, dass eine Aufgabenschwierigkeit aus dem Bereich Informationsentnahme, die dem mittleren Wert 100 (und damit zugleich der Schwierigkeit des mittleren der drei Beispielitems) entspricht, von 57,2 Prozent der Schülerinnen und Schüler mit der gebotenen Erfolgswahrscheinlichkeit gemeistert wurde. Es handelte sich hier beispielhaft um eine auf einen Stadtplan bezogene Aufgabe. In dem Stadtplan sind öffentliche Gebäude besonders gekennzeichnet, und die Schülerinnen und Schüler sollten das von der Post am weitesten entfernte öffentliche Gebäude (Krankenhaus, Bahnhof, Rathaus oder Kaufhaus) angeben.

Eine schwierigere Testaufgabe mit dem Rasch-Skalenwert 107 erforderte insoweit eine abstrahierende Leistung, als die räumliche Beziehung zwischen zwei Gebäuden bestimmt werden sollte. Immerhin 42,6 Prozent der untersuchten Schülerinnen und Schüler konnten mit der festgelegten Mindestwahrscheinlichkeit von $p = 0,65$ angeben, in welcher Himmelsrichtung das fragliche Gebäude, vom Referenzpunkt aus gesehen, liegt.



Bei diesem Untertest ist die Zweigipfligkeit, die „Bimodalität“, der Leistungsverteilung besonders ausgeprägt. Diese stark von dem gewohnten Muster der Normalverteilung abweichende Beobachtung überrascht insoweit, als die Befähigung zum Umgang mit diskontinuierlichen Texten keineswegs von jeher zu den unumstrittenen Kernbestandteilen schulischer Curricula gehört. Somit stellt sich erneut die Frage, wieso es an den Schulen für Lernbehinderte zu derart auffällig erwartungswidrigen Leistungsverteilungen kommt. Offenkundig liegt hier eine Mischverteilung zweier unterschiedlicher Gruppen vor, deren eine mit den Anforderungen des Untertests *Informationsentnahme* gut zurecht kommt, während die andere große Schwierigkeiten damit hat. Leistungsprofile über die vier Untertests *Leseverständnis*, *Wortkenntnis*, *Informationsentnahme*, *Mathematik* hinweg zeigen nun, dass 61,3 Prozent der Schülerinnen und Schüler, die den Untertest *Informationsentnahme* gut bewältigen, auch in den drei übrigen Untertests zum eher erfolgreichen Teil der Leistungsverteilung⁴ gehören. Dies wirft die weitere Frage auf, wie die entsprechende Kompetenz auf einer Schule für Lernbehinderte, bei Vorliegen partieller Lernschwächen, optimal gefördert werden kann.

3.1.4 Mathematik

Der Untertest *Mathematik* spiegelt ausgewählte mathematische Teilgebiete wider, die im Unterricht der Schülerinnen und Schüler eine Rolle gespielt haben. Im Bereich *Zahlenverständnis* werden Kenntnisse der Ziffern- und Stellenwerte, Kenntnisse des Zusammenhangs zwischen Zahlwort und Zahlendarstellung, Grundkenntnisse der Rechenoperationen und Fähigkeiten im Umgang mit Balkendiagrammen und Zahlenstrahl gefordert. Im Bereich *Größen* werden Kenntnisse der Maßeinheiten überprüft, und der Bereich *Rechnen* umfasst Aufgaben mit Sachverhalten zur Prüfung der Rechenfertigkeit.

Bei der Skalierung erwiesen sich drei der insgesamt 30 eingesetzten Mathematikitems unter teststatistischen Gesichtspunkten als ungeeignet und blieben aufgrund ihrer zu hohen Schwierigkeit und zu niedrigen Trennschärfe bei den weiteren Analysen unberücksichtigt. Die in LABEL 8-10 geltende Skala Mathematik umfasst somit 27 Items.

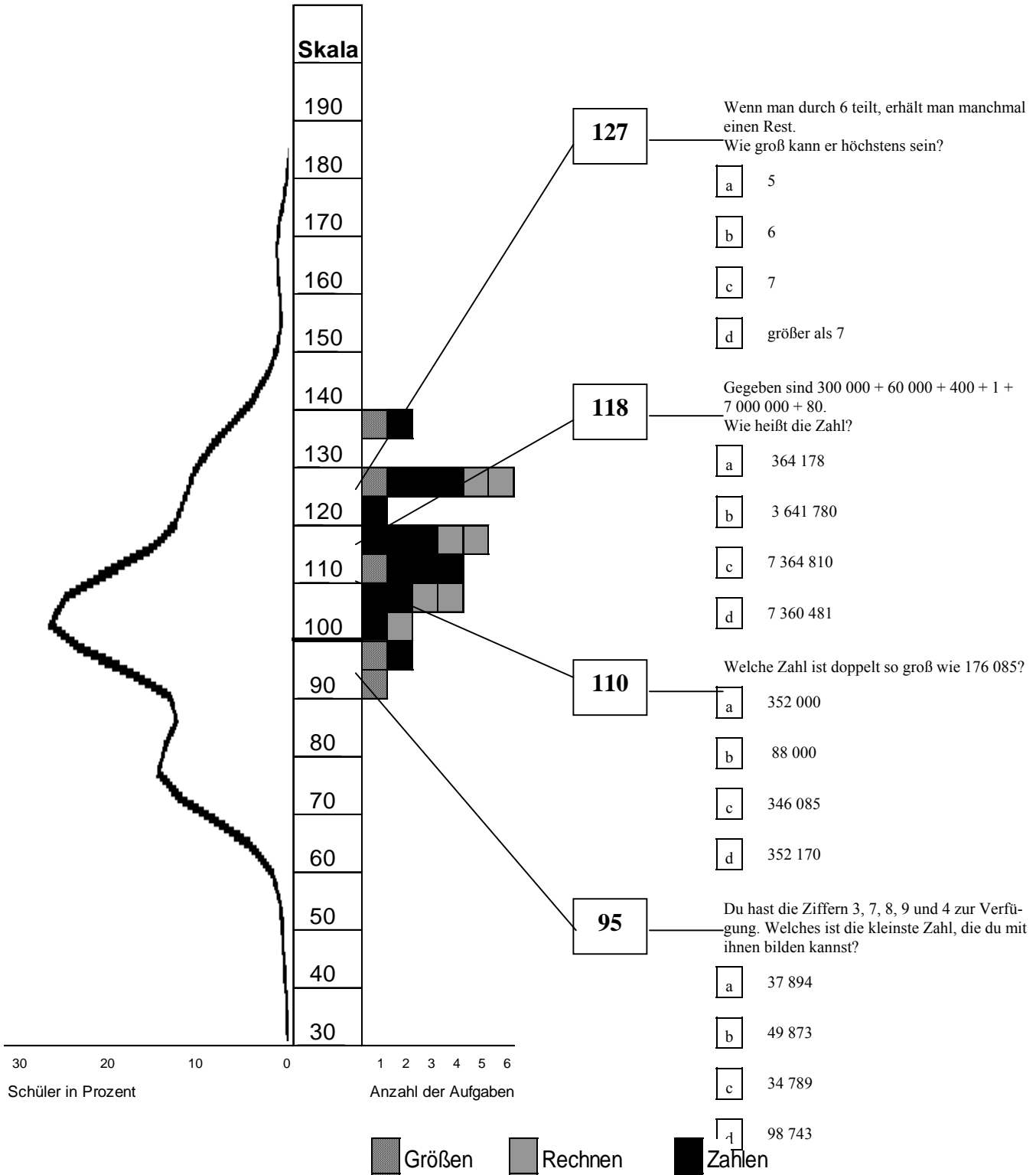
Abbildung 5 Verteilung der Schülerleistungen in Mathematik im Vergleich mit den Schwierigkeiten der Testaufgaben im Untertest Mathematik (nach Themenbereichen)

⁴ Als Kriterien wurden die Wende- oder Sattelpunkte gewählt, die die Abweichungen von der Normalverteilung markieren: 90 Punkte bei den Untertests *Leseverständnis* und *Mathematik*, 100 Punkte beim Untertest *Wortkenntnis*.

Beispielaufgaben aus dem Bereich Zahlenverständnis

Verteilung der Schülerleistungen

Testaufgaben



Gesamtstichprobe (N = 2.091)

In der Abbildung 5 sind auf der rechten Seite der Skala die 27 Testaufgaben nach ihrem statistisch bestimmten Schwierigkeitsgrad geordnet eingetragen. Die Schwierigkeiten der Aufgaben sind wiederum insgesamt und auch innerhalb der Themenbereiche eher im oberen Leistungsbereich der untersuchten Population angesiedelt; dies lässt sich leicht in der Überrepräsentanz von Aufgaben oberhalb des Mittelwertes 100 und der in den unteren Skalenbereich verschobenen Verteilung der Schülerleistungen ablesen. Dass der Test für eine bestimmte Gruppe von Schülerinnen und Schülern etwas zu schwierig war, hat zur Folge, dass er im internen Leistungsbereich nicht in dem wünschenswerten Maße diskriminiert. Vor allem aber ist angesichts der Bimodalität der Verteilung der Mathematikleistungen neuerlich keineswegs von einer Normalverteilung auszugehen. Vielmehr liegt hier wieder eine Mischverteilung vor, was sich damit belegen lässt, dass 50,3 Prozent der Schülerinnen und Schüler im hohen Leistungsbereich auch in den drei übrigen Fachleistungstests relativ hohe Leistungen gezeigt haben, während diejenigen, die zur leistungsschwachen Gruppe zählen, in keinem der drei übrigen Untertests besonders erfolgreich waren. Dies deutet darauf hin, dass die an den Schulen für Lernbehinderte angetroffenen Lernschwierigkeiten nicht primär als sektorale Defizite zu beschreiben sind, sondern als generalisierte Lernprobleme auftreten.

Das leichteste Beispielitem stammt aus dem Themenbereich Zahlenverständnis und erfasst das Vorhandensein grundlegender Prinzipien des Zahlenaufbaus; sein Rasch-Skalenwert beträgt 95. 36,2 Prozent der Probanden vermochten Aufgaben dieses Kompetenzniveaus nicht mit der spezifizierten Erfolgswahrscheinlichkeit von $p \geq 0,65$ zu bearbeiten. Die Schwierigkeit von Additionen mit sechsstelligen Zahlen (zweite Beispielaufgabe, Skalenwert 110) überschreitet die Kompetenz von 70,5 Prozent der untersuchten Zielgruppe. Die dritte Beispielaufgabe (ungeordnete Zerlegung einer siebenstelligen Zahl in Zehnerpotenzen; Rasch-Skalenwert 118) hat 84,7 Prozent überfordert, und der Schluss auf den größtmöglichen ganzzahligen Rest bei Divisionen durch 6 (Rasch-Skalenwert 127) konnte nur noch von 8,0 Prozent der Schülerinnen und Schüler in der Zielgruppe mit der gebotenen Sicherheit gezogen werden.

3.1.5 Gesamtergebnis in den Fachleistungen

Jeder der vier eingesetzten Untertests *Leseverständnis*, *Wortkenntnis*, *Informationsentnahme* und *Mathematik* kann einzeln und für sich interpretiert werden. Es ist darüber hinaus möglich und sinnvoll, alle Ergebnisse zu einem ein-

zigen Testwert zusammenzuziehen⁵. Testanalysen zeigen nämlich, dass die Untertests verhältnismäßig hoch miteinander korrelieren, wobei der Gesamtwert insbesondere durch den sprachbasierten Untertest *Leseverständnis* bestimmt ist (vgl. Tabelle 5).

Tabelle 5 Interkorrelationen der Ergebnisse im LABEL 8-10-Fachleistungstest

	Leseverständnis	Wortkenntnis	Informationsentnahme	Mathematik
Wortkenntnis	.67			
Informationsentnahme	.59	.57		
Mathematik	.52	.53	.59	
Gesamttest	.86	.80	.78	.79

Die Untertests und die Einzelaufgaben können deshalb nicht nur als gute Indikatoren für eine spezielle Fähigkeit (z. B. das *Leseverständnis*) gelten, sondern sie sind zugleich charakteristisch für die *allgemeine* Fachleistung. Somit ist es empirisch gut begründet, eine Gesamtskala zu bilden, deren Mittelwert wiederum auf 100 (Standardabweichung: 20) gesetzt ist. Die Gesamtskala umfasst 73 Aufgaben, von denen 25, also reichlich ein Drittel, aus dem Untertest *Leseverständnis* stammt. Dies allein erklärt allerdings kaum die gegenüber den drei übrigen Untertests so deutlich erhöhte Korrelation mit dem Gesamtergebnis. Vielmehr ist davon auszugehen, dass das Leseverständnis in der Tat konstitutiv für die meisten schulischen Lernprozesse ist, so wie sich dies auch in anderen Untersuchungen gezeigt hat⁶.

Der Abbildung 6 ist zu entnehmen, dass sich die Items aus den vier Untertests ebenso wie in den konstituierenden Untertests auf den oberen Teil der Gesamtskala verteilen, also insgesamt besser zwischen den vergleichsweise leistungsfähigen Schülerinnen und Schülern unterscheiden. Umgekehrt scharren sich die Ergebnisse der Schülerinnen und Schüler definitionsgemäß um den Durchschnittswert von 100 Punkten, allerdings asymmetrisch mit einem

⁵ Das Ergebnis für den Gesamttest beruht auf einer Neuberechnung sämtlicher Items. Diese Berechnung geht über die bloße Addition der Unterskalen hinaus. Im Kontext aller Aufgaben haben sich unter teststatistischen Gesichtspunkten drei der ursprünglichen 27 Mathematikitems als ungeeignet erwiesen. Entsprechend umfasst die Gesamtskala anstatt 76 Aufgaben aus den vier Untertests noch 73 Aufgaben.

⁶ Zur fundamentalen Bedeutung des Leseverständnisses vgl. DEUTSCHES PISA-KONSORTIUM

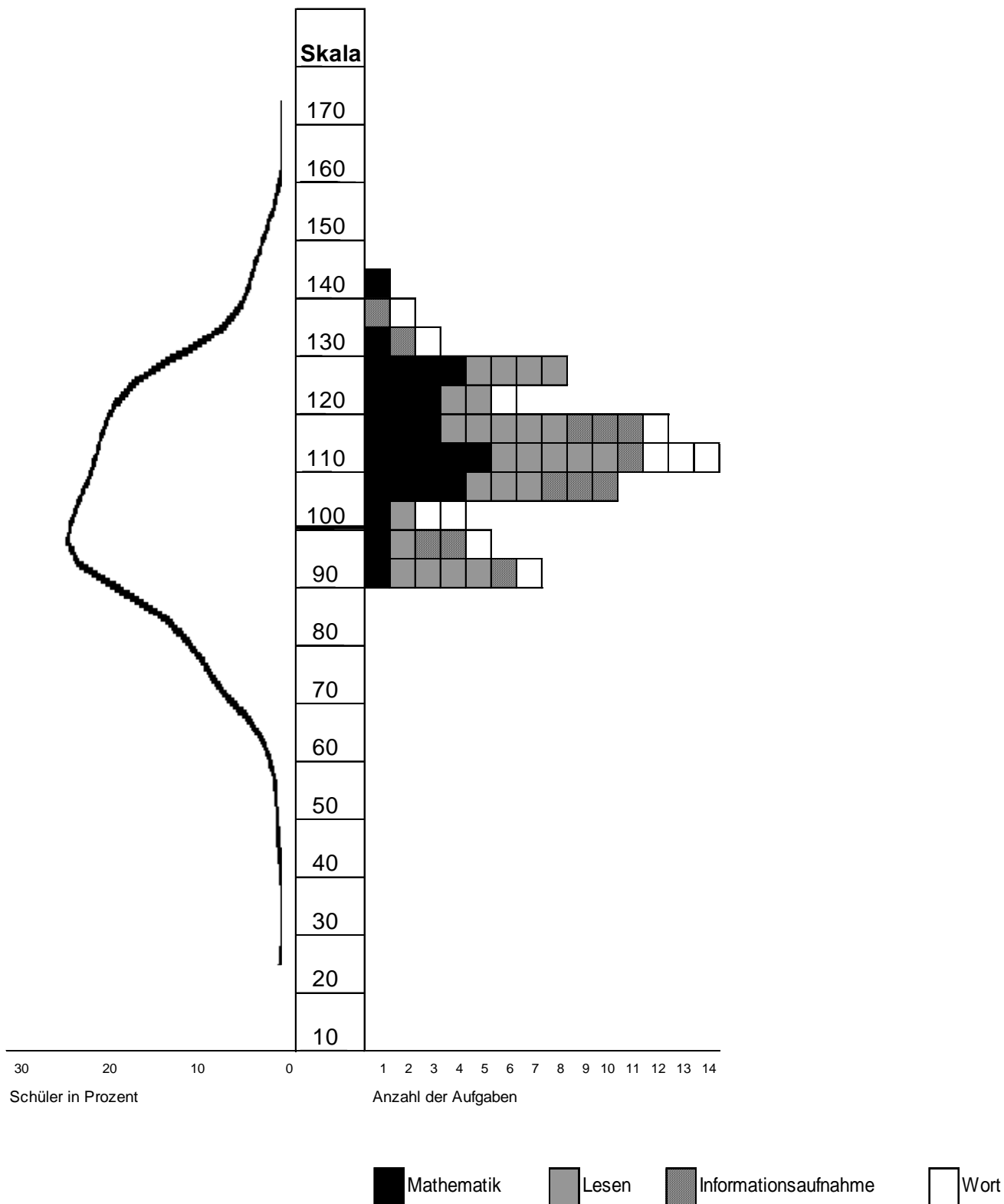
Maximum im Bereich der leichtesten Testaufgaben. Selbst unter Berücksichtigung der Messungenauigkeit im unteren Leistungsbereich ist daraus der Schluss zu ziehen, dass sich an den untersuchten Schulen ein erheblicher Schüleranteil befindet, deren Fähigkeiten gerade dem untersten Anforderungsniveau des eingesetzten Tests entsprechen.

Der Gesamtwert, den eine Schülerin bzw. ein Schüler beim Fachleistungstest erzielt hat, darf angesichts der testanalytischen Befunde mit Ausnahme der wenigen ausgeschlossenen Items als ein recht guter Indikator für die allgemeine Schulleistung angesehen werden. Der Einwand, dass im einen oder anderen Falle eine schlechte „Tagesform“ oder glückliches Raten zu einer Fehlbeurteilung geführt haben könnte, ist für die weiteren Analysen deswegen von geringem Gewicht, weil es hier nicht um Individualdiagnostik geht, sondern stets um Gruppenvergleiche und eine Beurteilung der Gesamtsituation. Schon auf der niedrigsten Ebene der Analysen, der Ebene der Schulklassen, stehen mögliche Über- und Unterschätzungen nebeneinander und gleichen sich so weitgehend aus.

Abbildung 6 Verteilung der Schülerleistungen im LABEL 8-10-Gesamttest im Vergleich mit den Schwierigkeiten der Testaufgaben (nach Untertests)

Verteilung der Schülerleistungen

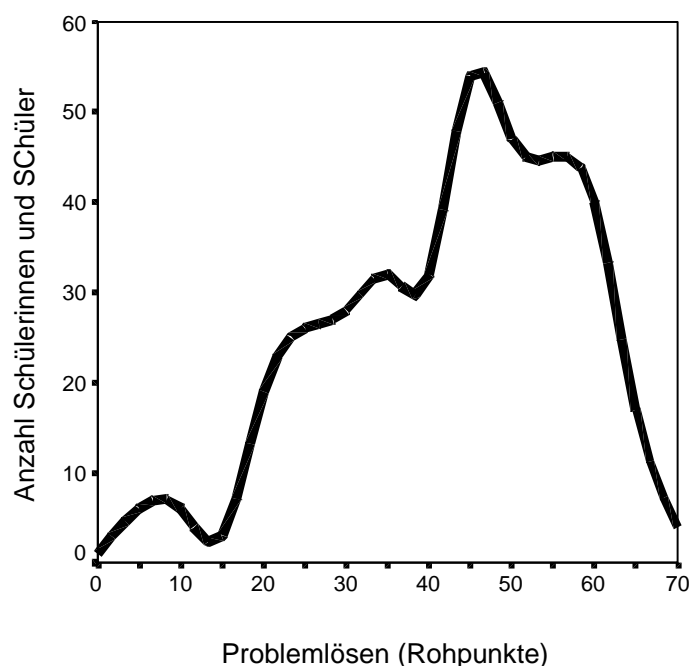
Testaufgaben



3.2 Problemlösen

Die drei Aufgabengruppen zur fächerübergreifenden (vielleicht besser: schulfachunabhängigen) Kompetenz des Problemlösens enthalten 73 Items, die sich mit konventionellen Methoden zu einer Skala hoher Reliabilität kombinieren lassen; Cronbach's α beträgt 0,94⁷. Die Form der Verteilung der Testwerte ist wieder asymmetrisch (Abbildung 7).

Abbildung 7: Verteilung der Schülerleistungen im schulfachunabhängigen Untertest „Problemlösen“ im Rahmen der Untersuchung LABEL 8-10



Teilstichprobe (N=337)

Es bestätigt sich somit nochmals das Bild einer Mischverteilung an den Schulen für Lernbehinderte. Vergleicht man die Zusammensetzung der Gruppe mit Testwerten unterhalb des ‚Sattelpunktes‘ der Verteilung (40 Testpunkte) mit derjenigen im Hinblick auf die allgemeine Fachleistung (gleichfalls dichotomisiert, und zwar nach Maßgabe des Medians von 100 Testpunkten), so zeigt sich ein interessanter Zusammenhang (vgl. Tabelle 6).

⁷ Angesichts der hier nicht gegebenen lokalen stochastischen Unabhängigkeit kommt eine probabilistische Skalierung nicht in Betracht. Auch sind nur minimale Erträge bei dem Bemühen möglich, durch die sukzessive Elimination der wenigen weniger geeigneten Items die interne Konsistenz der Skala marginal weiter zu erhöhen.

Tabelle 6: Erfolgsniveau im Fachleistungstest nach Erfolg im Problemlösetest (dichotomisiert), in Prozent; $N = 344$

Problemlösen	Allgemeine Fachleistung		Σ
	< 100 Punkte	\geq 100 Punkte	
< 40 Punkte	51,7	20,5	36,6
\geq 40 Punkte	48,3	79,5	63,4
Gesamt	100,0	100,0	100,0

Hier wird sichtbar, dass unter denjenigen Schülerinnen und Schülern, die mit dem Fachleistungstest offenkundig Schwierigkeiten hatten, unerwartet viele, nämlich 51,7 Prozent gegenüber 36,6 Prozent in der Randverteilung, unterdurchschnittlich Leistungen im Problemlösetest aufwiesen. Umgekehrt haben die weitaus meisten Schülerinnen und Schüler mit überdurchschnittlichen Fachleistungen (79,5 Prozent) auch den Problemlösetest vergleichsweise erfolgreich bearbeitet. Man kann dies als Indiz (wenn auch nicht als Beweis) dafür deuten, dass die hier gemessene allgemeine Fachleistung – und dabei nicht zuletzt das Leseverständnis – eine notwendige Voraussetzung für die Bearbeitung der vorgelegten ‚Probleme‘ darstellt. Für die entgegengesetzte Annahme lassen sich dagegen in den vorliegenden Daten kaum plausible Gründe finden. Die Korrelationen zwischen den Ergebnissen im Problemlösetest und jenen im Fachleistungstest unterstützen diese Einschätzung (Tabelle 7).

Tabelle 7: Korrelationen der Ergebnisse im Problemlösetest mit denen im Fachleistungstest und seinen Untertests

Leseverständnis	Wortkenntnis	Informationsentnahme	Mathematik	Allgemeine Fachleistung
0,39	0,31	0,36	0,32	0,42

3.3 Einstellungen, Einschätzungen und Werthaltungen

Im Schülerfragebogen sollten die Schülerinnen und Schüler Angaben zu ihrem *fachleistungsbezogenen Selbstkonzept*, zu ihrer Neigung zu *external attributierender Erfolgdeutung*, zu ihrer Selbsteinschätzung hinsichtlich ihrer *Arbeitstugenden* sowie ihrer *Flexibilität und Belastbarkeit* und schließlich zu ihrer *Einstellung gegenüber Technik* machen.

- **Fachleistungsbezogenes Selbstkonzept**

Die erste Dimension der erfragten Einstellungen bzw. Selbsteinschätzungen der Schülerinnen und Schüler bezieht sich auf die folgenden vier Aussagen⁸: „Meine Leistungen in Mathematik reichen für eine Ausbildung aus.“, „Meine Leistungen im Technisch-Naturwissenschaftlichen Unterricht reichen für eine Ausbildung aus.“, „Meine Leistungen in Arbeitslehre reichen für eine Ausbildung aus.“ und „Meine Leistungen in Lesen und Schreiben reichen für eine Ausbildung aus“. Die Antworten lassen sich nach Maßgabe der probabilistischen Messtheorie gut zu einer Skala mit definitorisch normiertem Mittelwert 10 und Standardabweichung 2 verbinden.⁹ Wenn nicht anders erwähnt, sind im Folgenden die entsprechenden Skalenwerte Grundlage der Beschreibungen und Analysen.

Gleichwohl ist auch die Betrachtung der Einzelitems durchaus aufschlussreich. So ist das Fachleistungsbezogene Selbstkonzept der Befragten hinsichtlich Mathematik, Technik-Naturwissenschaft und Arbeitslehre keineswegs auffallend schwach ausgeprägt; die entsprechenden Durchschnittswerte befinden sich mit 2,5 bis 2,6 im Bereich der theoretischen Mittelwerte. Auffällig ist indessen mit einem Mittelwert von 2,9 die eher zuversichtliche Selbsteinschätzung der Schülerinnen und Schüler bezüglich des Lesens und Schreibens, eine Auffassung, die durch die gemessenen Kompetenzen schwerlich gerechtfertigt wird.

Das fachleistungsbezogene Selbstkonzept hat generell wenig mit der tatsächlich feststellbaren allgemeinen Fachleistung zu tun ($r = 0,12$), hinsichtlich der erfassten Teilleistungen am ehesten noch mit den Ergebnissen im Untertest *Mathematik* ($r = 0,14$).

⁸ Die Einschätzungen der Schülerinnen und Schüler gründen sich auf einen Wertebereich, der bei den Einzelaussagen von 1 (für positiv formulierte Items: „nein“) bis 4 (für positiv formulierte Items: „ja“) reicht.

⁹ Die interne Konsistenz der Skala *Fachleistungsbezogenes Selbstkonzept*, berechnet nach dem ‚klassischen‘, von Cronbach entwickelten Verfahren liegt bei $\alpha = 0,72$.

• External attribuierende Erfolgsdeutung

Die zweite Dimension der erfragten Einstellungen drückt sich in der Bewertung von 8 Aussagen aus, in denen Erfolg external begründet wird, bei denen eigener Erfolg also eher auf andere Menschen oder auf äußere Umstände zurückgeführt und weniger dem Respondenten selbst zugeschrieben wird.¹⁰ Wiederum ist wegen des lediglich definitorisch bestimmten Mittelwerts bei probabilistischer Skalierung (die für die Untersuchung korrelativer Zusammenhänge hier vorgezogen wird) der Rekurs auf die Mittelwerte der Einzeli-tems hilfreich, wenn es um die Deutung der allgemeinen Antworttendenzen geht.

Dabei stellt sich heraus, dass für die befragten Schülerinnen und Schüler von Sonderschulen die motivational ungünstige Neigung zu externaler Erfolgs-
attribution dominiert: Typisch sind in diesem Bereich Mittelwerte um 2,9 Rohpunkte, also eine vergleichsweise deutliche Tendenz, Erfolge gerade *nicht* der eigenen Fähigkeit zuzuschreiben.

Im Übrigen bestätigt der negative korrelative Zusammenhang der Skala mit der allgemeinen Fachleistung ($r = - 0,23$) den auch aus anderen Studien bekannten Sachverhalt, dass die Leistungsschwächsten in besonderem Maße zu solchen ungünstigen Deutungsmustern neigen. Bemerkenswert ist indessen, dass dieser Zusammenhang noch in der relativ homogen leistungsschwachen Gruppe der Sonderschüler nachweisbar ist: Selbst hier gilt, dass die Befragten um so eher die Ursachen der eigenen Lernschwächen bei unkontrollierbaren Faktoren suchen, je offenkundiger diese Schwächen sind.

• Arbeitstugenden

¹⁰ Der Skala *External attribuierende Erfolgsdeutung* liegen Einschätzungen folgender Aussagen zugrunde: („Das ist wichtig, um Erfolg zu haben: ...“)

- „Man muss gut aussehen.“
- „Man muss der richtige Typ sein“
- „Man muss die richtigen Beziehungen haben.“
- „Man muss das richtige Geschlecht haben.“
- „Man muss die richtigen Leute kennen.“
- „Man muss die richtige Ausstrahlung haben.“
- „Die Eltern und Verwandten müssen helfen können.“
- „Man muss Glück haben.“

Die interne Konsistenz der Skala liegt bei $\alpha = 0,74$.

Eine Skala, die die Selbsteinschätzungen der Schülerinnen und Schüler zu ihren eigenen *Arbeitstugenden* zusammenfasst, geht auf zwölf Einzelaussagen zurück, die den probabilistischen Kriterien der Eindimensionalität genügen, aber auch nach konventionellen Maßstäben bei einer internen Konsistenz von $\alpha = 0,71$ gut zueinander passen¹¹.

Die unerwartet positive Selbsteinschätzung der Probanden ist mit einer durchschnittlichen Zustimmung zu diesen Aussagen in Höhe von 3,2 Rohpunkten sehr auffällig. Unabhängig davon, ob diese Selbsteinschätzung der Realität sehr nahe kommt bzw. ob hier eher ‚sozial erwünschte‘ Antwortmuster auftreten, wird damit doch belegt, dass die Befragten hinsichtlich der Bedeutung der sog. „Sekundärtugenden“ wenig Zweifel hegen. Denkbar ist, dass sich darin erzieherische Intentionen widerspiegeln, die in den untersuchten Sonderschulen bewusst und nicht ohne Erfolg zur Geltung gebracht werden. Für diese Deutung spricht zum Einen die geringe Varianz zwischen den Schülern im Hinblick auf dieses Merkmal, zum Anderen der Umstand, dass die Einschätzung der eigenen Arbeitstugenden – abgesehen von der sogleich zu behandelnden Skala *Flexibilität und Belastbarkeit* ($r = 0,40$) – mit keinem der anderen untersuchten Schülermerkmale sonderlich hoch korreliert ist.

• **Flexibilität und Belastbarkeit**

Wie gerade angedeutet, wurden die Schülerinnen und Schüler auch danach befragt, wie sie selbst ihre Belastbarkeit bzw. ihre Flexibilität einschätzen. Insgesamt 12 Items¹² im Schülerfragebogen bezogen sich auf diesen Aspekt. Sie konnten zu einer messtheoretisch befriedigenden, probabilistisch begründeten Skala zusammengefasst werden, deren nach dem konventionellen Verfahren bestimmte interne Konsistenz $\alpha = 0,68$ beträgt.

¹¹ Die Items lauten u. a.:

- „Ich kann gut Ordnung halten.“
- „Auf mich kann man sich verlassen.“
- „Andere halten mich für fleißig.“
- „Pünktlichkeit gehört zu meinen Stärken.“
- „Ich kann gut früh aufstehen.“
- „Ich finde es wichtig, dass Absprachen eingehalten werden.“

¹² Eine Auswahl der Items lautet:

- „Ich bin körperlich belastbar und gesund.“
- „Ich kann viel aushalten.“
- „Für eine Arbeit würde ich in eine andere Stadt ziehen.“
- „Ich mag es, wenn sich Arbeitsaufgaben öfter verändern.“

Anders als die eigenen Arbeitstugenden werden die eigene *Flexibilität und Belastbarkeit* von den Befragten als eher schwach entwickelt eingeschätzt (Mittelwert: 2,3 Rohpunkte, also als einziges der hier diskutierten Merkmale unterhalb des theoretischen Mittels). Nicht ganz unerwartet ist es, dass sich die Jungen für belastbarer halten als die Mädchen (Effektstärke $d = 0,54$, was etwas mehr als einer halben Standardabweichung entspricht).

• **Technikfreundlichkeit**

Ein umfangreicher Fragebogen mit insgesamt 50 Items wurde eingesetzt, um grundsätzliche Einschätzungen der Schülerinnen und Schüler gegenüber der Technik zu erfragen¹³. Abweichend von den anderen Einstellungsmerkmalen kam in diesem Falle probabilistische Skalierung nicht in Betracht. Nach konventionellen Maßstäben erwies sich indessen eine Auswahl von 23 dieser Items als effizient, wobei die Reliabilität der so gebildeten Rohwertskala *Technikfreundlichkeit* eine beachtliche interne Konsistenz von $\alpha = 0,84$ erreicht. Hohe Werte indizieren eine technikfreundliche, niedrige Werte eine distanzierte Haltung. Die leistungsstärkeren Befragten sind etwas technikfreundlicher als die weniger leistungsstarken ($r = 0,14$). Nicht unerwartet haben sich die Jungen wesentlich technikfreundlicher geäußert als die Mädchen, und zwar mit einem Unterschied von mehr als einer Standardabweichung (Effektstärke $d = 1,22$)! Tabelle 8 fasst die wichtigsten Befunde noch einmal zusammen, indem die Interkorrelationen der die fünf Einstellungsskalen und der beiden kognitiven Hauptkriterien aufgeführt werden. Über weitere Zusammenhänge, insbesondere mit den wichtigsten Gruppierungsmerkmalen innerhalb der Schülerschaft, gibt das folgende Kapitel Auskunft.

¹³ Die Skala besteht aus den Angaben der Schülerinnen und Schülern zu Aussagen wie z. B.:

- „Maschinen sind interessant.“
- „Es langweilt mich, ein Fahrrad selbst auseinander zu montieren.“ (umgepolt)
- „Technik gehört zum heutigen Leben.“
- „Mit Werkzeugen und rohem Material etwas herzustellen ist schön.“
- „Ich bin froh, dass ich so viele technische Dinge benutzen kann.“
- „Ich möchte mehr über Technik erfahren.“

Tabelle 8 : Interkorrelationen der Einstellungsskalen und kognitiven Kriterien

	Selbst konzept	Erfolgs attributi- on	Arbeits tugenden	Belast barkeit	Technik	Fach leistung
Externale Erfolgs attribution	0,12					
Arbeits tugenden	0,19	0,21				
Flexibilität und Belastbarkeit	0,25	0,18	0,40			
Technik freundlich- keit	0,09	0,06	0,08	0,36		
Allgemeine Fachleistung	0,12	-0,23	-0,04	0,07	0,14	
Problem lösen	0,06	-0,11	0,10	0,10	0,17	0,42

4 Differenzierungen in den kognitiven Leistungen

4.1 Unterschiede in den Leistungen zwischen Klassenstufen

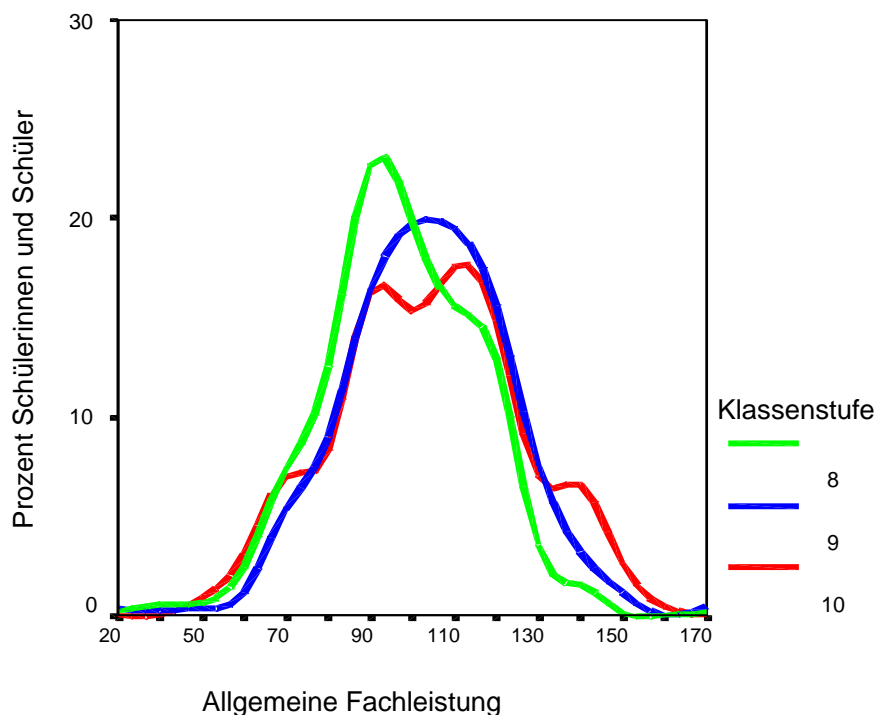
Bislang wurden die in die Untersuchung einbezogenen Schülerinnen und Schüler als eine Gruppe gesehen, ohne dabei Differenzierungen nach der aktuell besuchten Klassenstufe vorzunehmen. In Tabelle 9 sind die entsprechenden Werte für die Klassenstufen 8, 9 und 10 aufgeführt.

Tabelle 9 Fachleistung insgesamt zu Beginn der Klassenstufen 8, 9 und 10: Mittelwerte, Standardabweichungen und Effektstärken

	Mittelwert	Standard- abweichung	Effektstärke d	N
Jahrgangsstufe 8	97,15	18,5	-0,16	988
Jahrgangsstufe 9	103,14	19,7	0,14	878
Jahrgangsstufe 10	103,77	22,4	0,17	228

Die Durchschnittsleistungen der Schülerinnen und Schüler aus der Klassenstufe 8 unterscheiden sich von denen in den Klassenstufe 9 statistisch signifikant ($p < 0,05$). Die beiden Mittelwerte differieren um 6 Skalenpunkte, d.h. 30 Prozent der Standardabweichung aller Werte *in der hier untersuchten Zielpopulation*. Anders ausgedrückt: Der mutmaßliche Lernzuwachs von Klassenstufe acht nach 9 entspricht (allerdings nur bezogen auf die hier untersuchte Zielgruppe) etwa dem, was in Allgemeinen Schulen in Deutschland üblich ist. Hingegen unterscheiden sich die gemittelten Fachleistungen zwischen den Neunt- und Zehntklässlern praktisch nicht hinsichtlich des Mittelwerts, wohl aber hinsichtlich der Form der Verteilung. Die folgende Abbildung 8 zeigt in graphischer Form die Verteilungen im LABEL 8-10-Fachleistungstest in den verschiedenen Klassenstufen. Darin nun ist in der Jahrgangsstufe 10 deutlich eine Eigentümlichkeit (eine sog. „Bimodalität“) der Werteverteilung zu erkennen, die anzeigt, dass hier eine Mischung von zwei Schülergruppen vorliegt. Diese Beobachtung verweist auf die Unterscheidung zwischen den Bildungsgängen BO 10 und BESO 10, auf die weiter unten noch näher eingegangen wird.

Abbildung 8 Allgemeine Fachleistung. Verteilung der Lernstände, nach Klassenstufe



Im Folgenden werden die Ergebnisse für die drei eingesetzten Untertests aufgezeigt.

Abbildung 9 zeigt in graphischer, Tabelle 10 in numerischer Form die Unterschiede zwischen den Klassenstufen im Leseverständnis. Wie hinsichtlich der allgemeinen Fachleistung unterscheiden sich hier die Durchschnittswerte der Schülerinnen und Schüler aus der Klassenstufe 8 von den in den Klassenstufen 9 und 10 beobachteten statistisch signifikant ($p < 0,05$). Erneut entspricht der Zuwachs in diesem Untertest in der Klassenstufe 8, wenn auch wieder nur bezogen auf die Zielpopulation und auf niedrigem Anfangsniveau, nahezu den jährlichen Zuwachsraten der Allgemeinen Schulen. Auch die Differenz zwischen den Klassenstufen 9 und 10 ist – unbeschadet der Unterscheidung zwischen BO 10 und BESO 10 – mit 0,10 Standardabweichungen immer noch beträchtlich.

Im Bereich Lesen zeigen sich auf der Ebene der Klassenstufen die markantesten Unterschiede. Bei keinem anderen Untertest sind so deutliche Mittelwertunterschiede beobachtbar wie hier: 7,3 Skaleneinheiten oder 0,37 Standardabweichungen zwischen Klassenstufe 8 und 10.

Abbildung 9 Leseverständnis. Verteilung der Lernstände, nach Klassenstufe

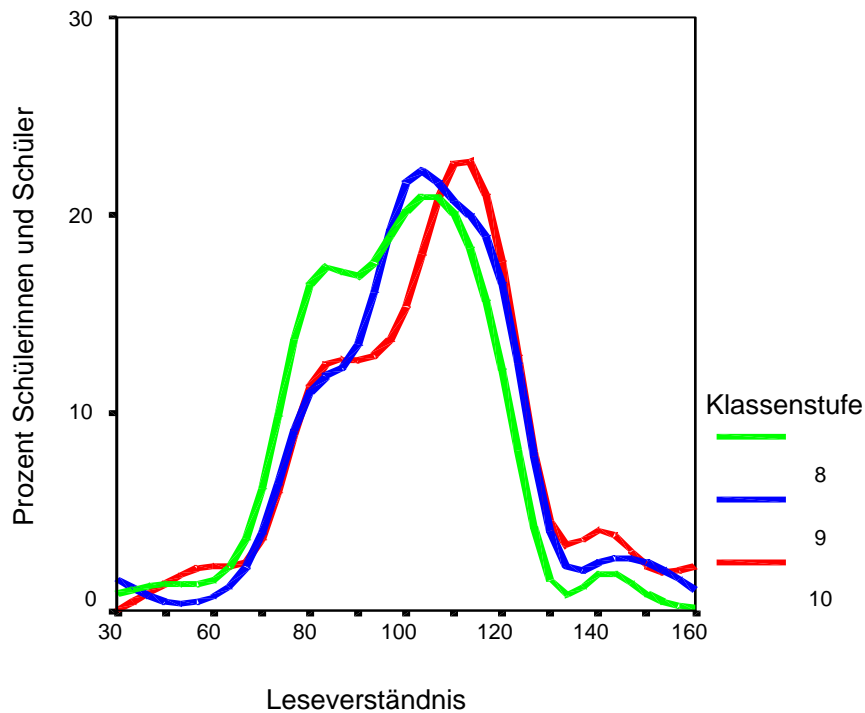


Tabelle 10 Leseverständnis zu Beginn der Klassenstufen 8, 9 und 10: Mittelwerte, Standardabweichungen und Effektstärken

	Mittelwert	Standardabweichung	Effektstärke d	N
Klassenstufe 8	97,4	18,5	-0,15	976
Klassenstufe 9	102,7	20,3	0,12	869
Klassenstufe 10	104,7	21,5	0,22	221

Bezogen auf die Kompetenzen im Umgang mit diskontinuierlichen Texten seien zunächst die maßgeblichen Verteilungsparameter einander nach Klassenstufe in numerischer Form gegenübergestellt (Tabelle 11).

Tabelle 11 Leistung im Bereich Informationsentnahme aus Karten/ Tabellen/ Diagrammen zu Beginn der Klassenstufen 8, 9 und 10: Mittelwerte, Standardabweichungen und Effektstärken

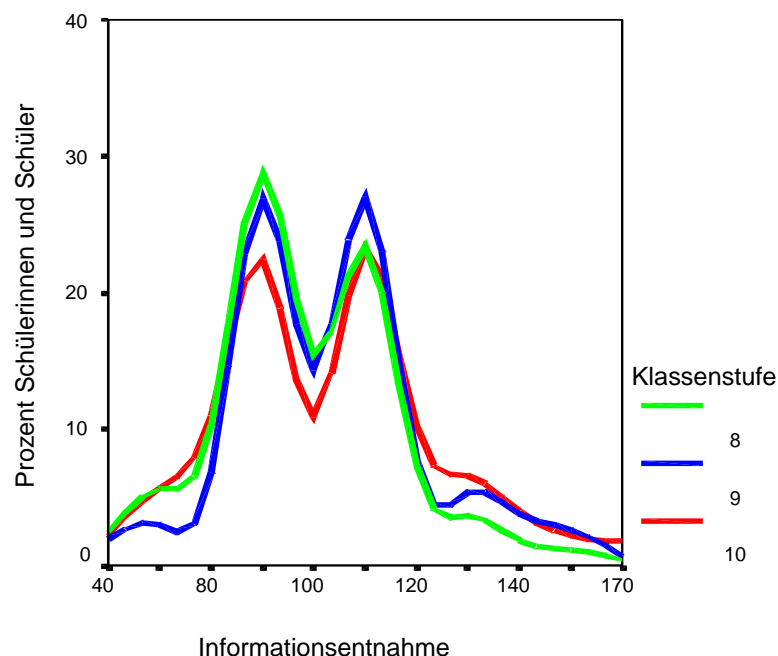
	Mittelwert	Standardabweichung	Effektstärke d	N
Klassenstufe 8	97,7	19,2	-0,12	988
Klassenstufe 9	102,5	19,9	0,11	878
Klassenstufe 10	102,0	22,8	0,09	228

Die Durchschnittsleistungen der Schülerinnen und Schüler aus der Klassenstufe 8 weichen von denen in den Klassenstufen 9 und 10 statistisch signifi-

kant ab ($p < 0,05$). Insofern sich die Mittelwerte allerdings nur um 4,8 bzw. 4,3 Skalenpunkte unterscheiden, sind die Differenzen etwas geringer, als man vielleicht erwartet hätte.

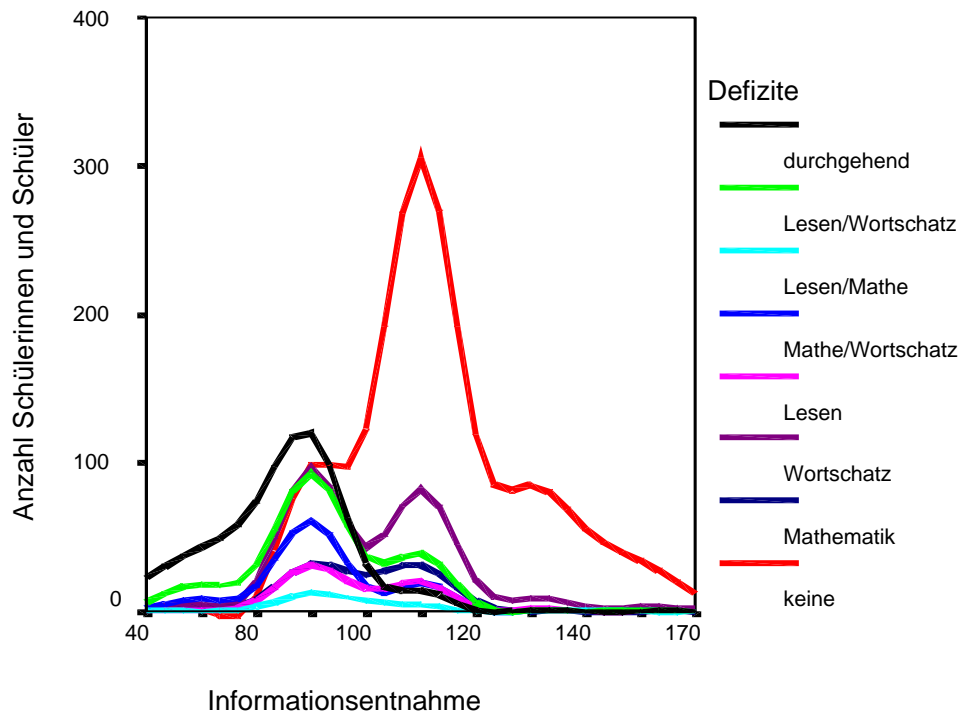
Wie Abbildung 10 zeigt, ist jedoch hinsichtlich der Informationsentnahme aus Karten, Tabellen und Diagrammen die bimodale Verteilungsstruktur besonders ausgeprägt und über die Klassenstufen hinweg von bemerkenswerter Stabilität, was Anlass gibt, nach Hintergründen dafür zu suchen, dass sich diese so auffällige Polarisierung der Ergebnisse um den Skalenpunkt 100 in allen drei Klassenstufen in so ungewöhnlicher Gleichförmigkeit wiederholt.

Abbildung 10 Informationsentnahme aus Karten / Tabellen / Diagrammen bzw. Umgang mit diskontinuierlichen Texten: Verteilung der Lernstände, nach Klassenstufe



Mit den gängigen multivariaten Analyseverfahren, die den Erfolg im Untertest *Informationsentnahme* zu anderen Ergebnissen der Untersuchung in Beziehung setzen, erhält man recht gute Erklärungsansätze: Die Untertests *Mathematik*, *Leseverständnis*, *Wortschatz* sowie der *CFT 20* und der Migrationsstatus ‚erklären‘ mit 49,6 Prozent knapp die Hälfte der individuellen Leistungsvarianz im Umgang mit diskontinuierlichen Texten. Gleichzeitig erlaubt diese Kombination von Prädiktoren bei einer Diskriminanzanalyse in mehr als 76 Prozent aller Fälle eine korrekte Vorhersage darüber, ob sich ein Schüler oder eine Schülerin im oberen oder unteren Teil der Verteilung befindet. Noch aussichtsreicher erscheint es aber, die Zugehörigkeit der Schüler und Schülerinnen zu einer der beiden Teilgruppen in Beziehung zu Besonderheiten des sonstigen Leistungsprofils zu setzen (vgl. Abbildung 11).

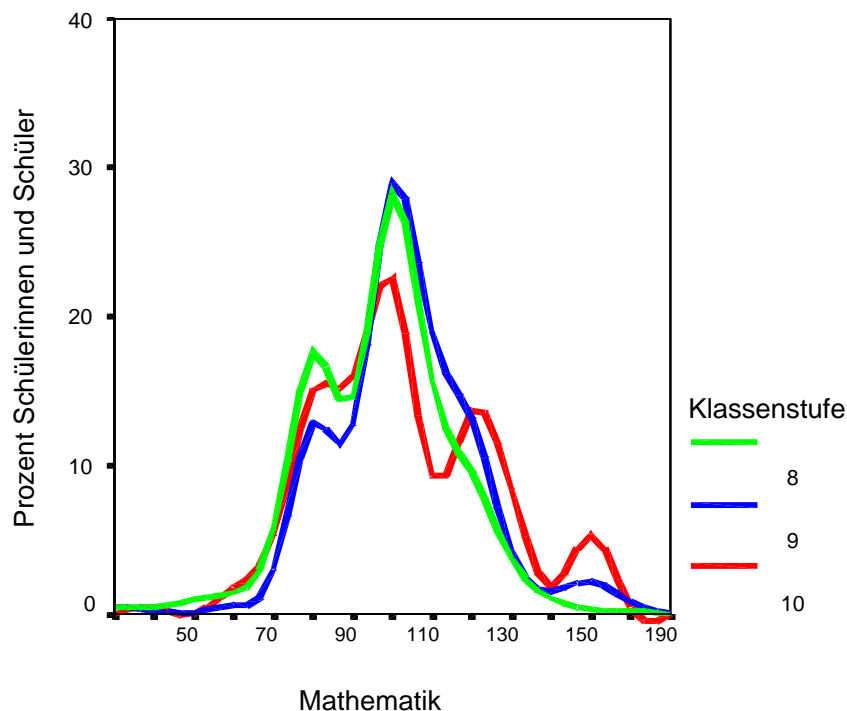
Abbildung 11: Erfolg im Untertest *Informationsentnahme* nach Leistungsprofil



In dieser Abbildung werden die Schülerinnen und Schüler danach unterschieden, in welchen Untertests bzw. Leistungsbereichen sie – ausweislich der Zugehörigkeit zur unteren Hälfte des Leistungsspektrums – Defizite gezeigt haben. Dabei wird sichtbar, dass die erfolgreichen Bearbeiter des Untertests *Informationsentnahme* fast ausschließlich diejenigen sind, die in keinem einzigen der anderen Untertests Schwächen demonstrieren: zumindest ein ausreichendes Leseverständnis und ein Mindestmaß mathematischer Kompetenz scheinen gleichermaßen notwendige Voraussetzungen für den erfolgreichen Umgang mit diskontinuierlichen Texten zu sein. Der gleiche Schluss liegt nahe, wenn man die Gruppe derer betrachtet, die einzig im Bereich „Wortschatz“ Schwierigkeiten haben, also in den Bereichen „Leseverständnis“ und „Mathematik“ unauffällig sind; denn in der vorliegenden Darstellung weist diese Gruppe als einzige die ausgeprägt bimodale Struktur des Untertests *Informationsentnahme* auf.

Die Leistungsverteilung im Untertest *Mathematik* wird, differenziert nach Klassenstufen, in Abbildung 12 illustriert. Auch hier deuten sich, sofern man den Kohortenvergleich als Entwicklung deuten darf, Verbesserungen über die Klassenstufe 8 hinaus an. Dies zeigt sich nicht zuletzt darin, dass das auf der Klassenstufe 8 noch sehr deutliche sekundäre Maximum beim Skalenwert 80 auf der Klassenstufe 9 deutlich geringer ausgeprägt ist, während in der Klassenstufe 10 oberhalb der Mittelkategorie (Skalenwert 100) zwei Nebenmaxima erkennbar sind.

Abbildung 12 Mathematik: Verteilung der Lernstände, nach Klassenstufe



In der numerischen Darstellung (Tabelle 12) artikulieren sich diese Befunde in der deutlichen Zunahme des Mittelwerts zwischen den Klassenstufen 8 und 9 um 0,29 Standardabweichungen sowie durch eine Steigerung der Streuung (der Standardabweichung) um 2,7 Skalenpunkte zwischen den Klassenstufen 9 und 10. Wäre die Steigerung des Mittelwerts lediglich eine Folge von vorzeitigen Schulabgängen schwächerer Schülerinnen und Schüler, müsste man an dieser Stelle eine Abnahme der Streuung erwarten. Dass eine solche, an einer verminderten Standardabweichung erkennbare, Homogenisierung nicht festzustellen ist, spricht dafür, dass wirkliche Lerngewinne haben erreicht werden können.

Tabelle 12 Fachleistung Mathematik zu Beginn der Klassenstufen 8, 9 und 10: Mittelwerte, Standardabweichungen und Effektstärken

	Mittelwert	Standardabweichung	Effektstärke d	N
Jahrgangsstufe 8	97,2	18,8	-0,14	962
Jahrgangsstufe 9	103,0	19,6	0,15	857
Jahrgangsstufe 10	102,8	22,3	0,14	226

Mit (statistisch signifikanten) Differenzen gegenüber dem durchschnittlichen Lernstand auf Klassenstufe 8 von 5,8 bzw. 5,6 Skalenpunkten bzw. fast einem Drittel der Standardabweichung dieser Zielgruppe ist wieder eine Zuwachsrate zu beobachten, die jener der Regelschulen fast entspricht.

Leistungsstände auf der Klassenstufe 10: Lehrgangprofil BO 10 vs. BESO 10

Detailliertere Analysen zeigen, dass in der Klassenstufe 10 erhebliche Unterschiede zwischen den berücksichtigten Lehrgangprofilen auftreten. Die Schülerinnen und Schüler, die hier zusammengefasst als Zehntklässler bezeichnet sind, können zwei recht unterschiedlichen Gruppen zugeordnet werden, dem Lehrgangprofil BO 10 und dem Lehrgangprofil BESO 10. Getestet wurden 141 Schülerinnen und Schüler aus dem Bildungsgang BO 10 und 69 Schülerinnen und Schüler aus dem Bildungsgang BESO 10. Die beiden Gruppen unterscheiden sich in ihrem CFT-Wert deutlich: Die Schülerinnen und Schüler aus BO 10 haben einen Mittelwert von 25,2 ($s = 7,0$) Rohpunkten, diejenigen aus BESO 10 den wesentlich geringeren Mittelwert von 18,7 ($s = 6,4$) erreicht.

In Übereinstimmung damit haben die Schülerinnen und Schüler aus dem Bildungsgang BO 10 erheblich höhere Fachleistungen gezeigt als die aus dem Bildungsgang BESO 10 (vgl. Tabelle 13).

Tabelle 13 Fachleistungen in der Klassenstufe 10 nach Bildungsgang: Mittelwerte (M) und Standardabweichungen (s)

	Leseverständnis		Informationsentnahme		Mathematik		Gesamttest	
	Mittelwert	s	Mittelwert	s	Mittelwert	s	Mittelwert	s
BO10	113,0	18,4	109,5	21,4	111,1	20,7	113,4	18,9
BESO10	88,4	17,2	86,6	19,2	86,8	15,7	85,2	16,4

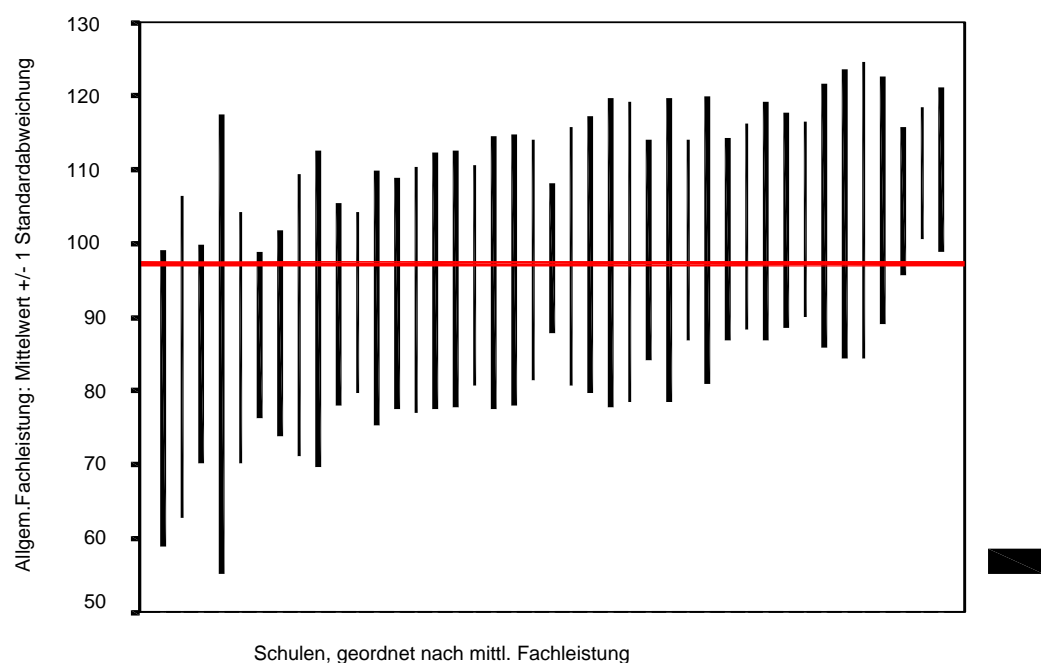
Bei der Aufspaltung der Schülerschaft in die beiden Bildungsgänge der Klassenstufe 10 kommt es offenbar zu einer Negativauslese im Bildungsgang BESO 10, die neben den niedrigeren Mittelwerten an den Standardabweichungen erkennbar ist. Ob dieser Effekt den vergleichsweise leistungsstärkeren Schülern im Lehrgangprofil BO 10 zu Gute kommt oder mit überwiegend problematischen Auswirkungen für diejenigen verbunden ist, die auf vermutlich dem niedrigeren Anspruchsniveau der BESO 10 unterrichtet werden, lässt sich mit dem vorliegenden Querschnittvergleich nicht klären.

4.2 Unterschiede in den Leistungen zwischen Schulen

4.2.1 Unterschiede in den Leistungen zwischen Schulen allgemein

Nach Vorerfahrungen aus Schulleistungsstudien im allgemeinbildenden Bereich (vgl. z. B. LEHMANN & PEEK 1997; LEHMANN ET AL. 1999, 2001), die diesen Befund für Hamburg nachgewiesen haben bzw. die das Phänomen für das Bundesland Brandenburg dokumentieren (LEHMANN ET AL. 2000, S. 46ff. und 59ff.) war zu erwarten, dass sich die allgemeinen Fachleistungsstände von Schule zu Schule erheblich unterscheiden können: So gibt es in Berlin Schulen für Lernbehinderte, in denen in der Oberstufe die besten Fachleistungen kaum an die schwächsten Lernergebnisse in denjenigen heranreichen, wo auf vergleichsweise viel höherem Niveau gearbeitet werden konnte (vgl. Abbildung 13).

Abbildung 13 Verteilung der Ergebnisse im LABEL 8-10-Fachleistungstest (Klassenstufe 8), nach Schulen (Durchschnittswerte plus/minus eine Standardabweichung)



In der Abbildung sind die untersuchten 41 Schulen nach der erreichten Durchschnittsleistung auf der Klassenstufe 8 aufsteigend geordnet, wobei die Länge jeden „Balkens“ den Bereich des Mittelwerts plus/minus eine Standardabweichung markiert, also eine Streuweite, innerhalb derer etwa zwei Drittel der Schülerinnen und Schüler der betreffenden Institution zu finden sind. Leistungshomogenere Schulen sind somit durch kürzere Balken gekennzeichnet, solche mit stärker streuenden Fachleistungen durch längere. Die horizontale Linie bezeichnet den Mittelwert über alle Schulen hinweg.

Wie diese Abbildung zeigt, haben an der leistungsstärksten Schule in der Klassenstufe 8 fast alle Schülerinnen und Schüler den entsprechenden Gesamtmittelwert übertroffen, während an der leistungsschwächsten Schule nur sehr wenige in diesem Sinne überdurchschnittliche Resultate erzielt haben. Der Anteil der Leistungsvarianz, der sich mit der Einzelschule und ihrem jeweiligen Umfeld verbindet, ist also nicht unbeträchtlich.

Ferner weist die Abbildung Schulen mit vergleichsweise homogenem Fachleistungsspektrum gegenüber Schulen mit einer deutlich heterogeneren Schülerschaft aus. Auch diesen Unterschieden, die ebenfalls mit Differenzen im außerschulischen Umfeld und/oder mit unterschiedlichen Rekrutierungsmustern zusammenhängen können, wird noch nachzugehen sein.

Eine Aggregation der Leistungen für jede Schule ohne Unterscheidung der Klassenstufen 8, 9 und 10 wäre indessen schon deshalb problematisch, weil nicht in allen in die Studie einbezogenen Schulen Zehntklässler untersucht werden konnten. Um jedoch zumindest für die Klassenstufe 8 auch Differenzierungen sichtbar zu machen, die sich mit dem jeweils untersuchten Leistungsaspekt verknüpfen, sei der mit dem Merkmal „Schulzugehörigkeit“ verbundene Varianzanteil wenigstens numerisch ausgewiesen (vgl. Tabelle 14).

Tabelle 14: Varianzanteile, die mit der Zugehörigkeit zu einer bestimmten Einzelschule verbunden sind nach Untertest (nur Klassenstufe 8)

Test bzw. <i>Untertest</i>	<i>Eta²</i>	<i>N</i>
Allgemeine Fachleistung (HST 4/5)	0,156	988
<i>Leseverständnis</i>	0,152	976
<i>Wortkenntnis</i>	0,129	974
<i>Informationsentnahme</i>	0,097	988
<i>Mathematik</i>	0,115	962
Problemlösen	0,119	210
CFT 20	0,086	958

Die Unterschiede auf Schulebene sind am deutlichsten im Bereich *Leseverständnis* und am wenigsten durch den Untertest *Informationsentnahme* bestimmt. In der Allgemeinen Fachleistung sind diese Differenzen zwischen den Schulen gewissermaßen kumulativ zusammengefasst, wobei interessanterweise praktisch keine weitere Differenzierung der Schulen über die Unterschiede im Leseverständnis hinaus mehr erfolgt.

Besondere Bedeutung hat angesichts dieser Zahlen die Frage, inwieweit an den Unterschieden zwischen den Einzelschulen institutionelle und pädagogi-

sche Faktoren beteiligt sind oder inwieweit sich darin lediglich unterschiedliche Rekrutierungsmuster spiegeln. Einige Hinweise zur Beantwortung dieser Frage lassen sich der nachstehenden Tabelle 15 entnehmen.

Tabelle 15: Interkorrelationen von durchschnittlicher Fachleistung und gemessener Intelligenz auf Schulebene, innerhalb und zwischen den Klassenstufen (Klassen 8, 9: $N = 41$; Klasse 10: $N = 16$)

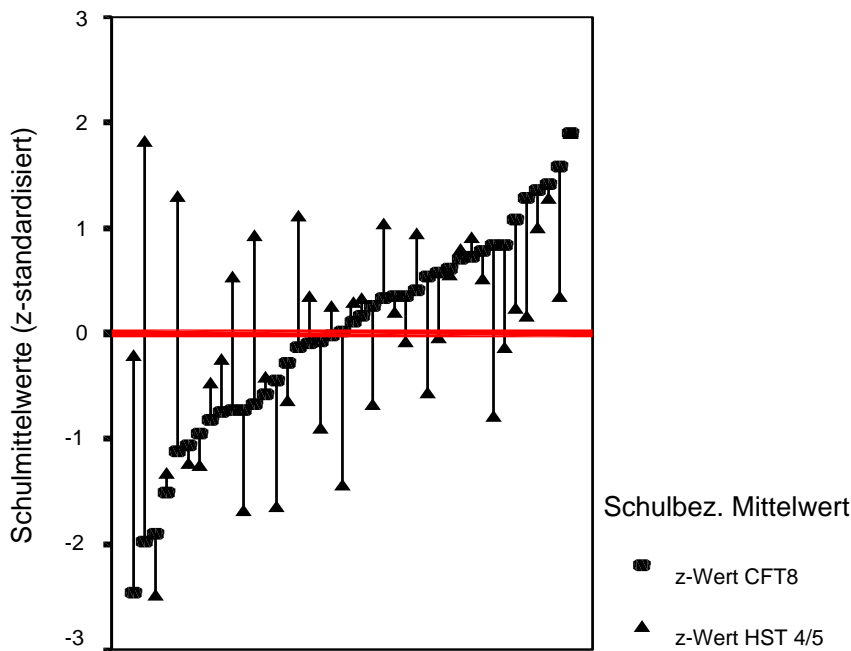
	HST 4/5			CFT 20		
	Klasse 8	Klasse 9	Klasse 10	Klasse 8	Klasse 9	
HST 4/5 Klasse 9	0,64					
Klasse 10	0,52					0,34
CFT 20 Klasse 8	0,40	0,19	0,36			
Klasse 9	0,26	0,45	0,20			0,00
Klasse 10	0,21	0,28	0,81			0,10

Wie die Teilmatrix rechts unten, die sich auf den CFT 20 bezieht, mit dem Median von $r = 0,10$ zeigt, sind die Einzelschulen in ihrer Schülerschaft über die Jahrgänge hinweg keineswegs auf bestimmte kognitive Eingangsvoraussetzungen festgelegt: Die niedrigen Korrelationen lassen offenbar erheblichen Raum für Variationen zwischen den Jahrgängen. Hinsichtlich der erzielten Fachleistungen (vgl. die Teilmatrix oben links) ist jedoch angesichts einer mittleren Korrelation von $r = 0,52$ (Median) mit weitaus mehr Berechtigung von einem schultypischen Niveau zu sprechen. Dass die Fachleistung innerhalb eines Jahrgangs gleichzeitig wesentlich durch die jeweiligen kognitiven Eingangsvoraussetzungen bestimmt ist, wird an den Diagonalelementen der linken unteren Teilmatrix erkennbar, die die Zusammenhänge zwischen dem Fachleistungstest HST 4/5 und dem nonverbalen Intelligenztest CFT 20 repräsentiert.

Um die komplexen Zusammenhänge zwischen kognitiven Eingangsvoraussetzungen, so wie sie mit dem CFT 20 erhoben wurden, und erzielter Fachleistung gemäß HST 4/5 auch optisch sichtbar zu machen, sind in Abbildung 14 die z-standardisierten Schulmittelwerte dargestellt, und zwar so, dass die Schulen nach ihren CFT-Mittelwerten geordnet sind und die jeweils im Durchschnitt erreichten Fachleistungen erkennbar werden. Aus Gründen der Übersichtlichkeit beschränkt sich die Abbildung auf die Verhältnisse in der Klassenstufe 8. Die waagerechte Referenzlinie bezeichnet den allgemeinen Durchschnitt für beide Variablen.

Abbildung 14 Zusammenhang zwischen kognitiven Eingangsvoraussetzungen

(CFT 20) und Allgemeiner Fachleistung (HST 4/5) auf Schulebene (z-standard. Schulmittelwerte, nur Klassenstufe 8)



Schulen, geordnet nach CFT 20

Wäre die mit dem CFT 20 gemessene durchschnittliche kognitive Eingangsvoraussetzung die einzige außerschulische Determinante des an einer Schule erreichten Lernstands, so könnte man die entsprechenden Werte gleichsam als ‚Erwartungswerte‘ für eine Schule interpretieren und die Differenzen gegenüber erzielten Testwerten als Indikator für den Ertrag der pädagogischen Arbeit deuten. Nach dem oben Gesagten wäre diese Sichtweise indessen so nicht gerechtfertigt. Allenfalls lässt sich sagen, dass die im Mittel erzielten Lernstände in einigen Schulen mit besonders ungünstigen Eingangsvoraussetzungen auffällig höher sind, als dies unter Berücksichtigung der CFT-Ergebnisse zu erwarten gewesen wäre¹.

4.2.2 Unterschiede in den Leistungen zwischen Schulen nach ihrer Teilnahme am *Netzwerk Berliner Schülerfirmen*

Von besonderem Belang sind in diesem Zusammenhang die Ergebnisse der Schulen, die am „Netzwerk Berliner Schülerfirmen“ beteiligt sind.

¹ Über eine analoge Analyse zu den Verhältnissen auf der Klassenstufe 9 konnte geprüft werden, ob insbesondere die auffälligen positiven Abweichungen zwischen ‚erwarteten‘ und erzielten Testwerten jahrgangübergreifend schultypisch sind, was u.U. den Verdacht auf Unregelmäßigkeiten bei der Testbearbeitung hätte entstehen lassen können. Es findet sich bei diesem Vergleich jedoch kein Anhaltspunkt für ein solches Artefakt.

Tabelle 16: Testergebnisse der Schulen nach Klassenstufe und Teilnahme am *Netzwerk Berliner Schülerfirmen*: Mittelwerte und Effektstärken

Klassenstufe 8	Netzwerk Berliner Schülerfirmen		Effektstärke <i>d</i>	<i>N</i>
	nein	ja		
Allg. Fachleistung (HST 4/5)	97,3	94,7	- 0,14	988
<i>Leseverständnis</i>	97,4	97,3	- 0,00	976
<i>Wortkenntnis</i>	97,8	95,3	- 0,13	974
<i>Informationsentnahme</i>	97,9	96,3	- 0,08	988
<i>Mathematik</i>	97,7	91,7	- 0,32	962
Problemlösen	40,2	40,5	+ 0,02	210
CFT 20	22,6	22,3	- 0,04	958
<i>N</i>	915	73		
Klassenstufe 9				
Allg. Fachleistung (HST 4/5)	104,1	99,1	- 0,26	878
<i>Leseverständnis</i>	103,7	98,6	- 0,25	869
<i>Wortkenntnis</i>	103,5	97,9	- 0,28	867
<i>Informationsentnahme</i>	103,2	99,4	- 0,19	878
<i>Mathematik</i>	103,7	100,44	- 0,17	857
Problemlösen	43,1	51,1	+ 0,61	90
CFT 20	23,2	22,5	- 0,11	856
<i>N</i>	705	173		
Klassenstufe 10				
Allg. Fachleistung (HST 4/5)	104,2	103,6	- 0,03	228
<i>Leseverständnis</i>	105,0	104,5	- 0,02	221
<i>Wortkenntnis</i>	105,6	102,9	- 0,14	221
<i>Informationsentnahme</i>	103,5	101,1	- 0,10	228
<i>Mathematik</i>	103,1	102,7	- 0,02	226
Problemlösen	53,6	47,0	- 0,46	60
CFT 20	23,2	22,3	- 0,11	226
<i>N</i>	80	146		

Die Tabelle 16 zeigt, dass zum gegenwärtigen Zeitpunkt nicht durchgängig von einer Überlegenheit der NBS-Schulen bezüglich der Fachleistungen ausgegangen werden kann. Eine Differenzierung nach Klassenstufen und Leistungsaspekten sowie die Berücksichtigung der kognitiven Lernvoraussetzungen (CFT 20) bringt jedoch aufschlussreiche Details ans Licht.

Die Mittelwerte der noch nicht dem Netzwerk angeschlossenen Schulen liegen zumeist etwas höher als die der Netzwerkschulen. Dies ist jedoch auch

in Zusammenhang damit zu sehen, dass auf den Klassenstufen 9 und 10 die kognitiven Lernvoraussetzungen an den Netzwerkschulen etwas ungünstiger sind ($d = - 0,11$ ggü. dem unbedeutenden Effekt $d = - 0,04$ auf Klassenstufe 8). Auf Klassenstufe 8 ist einzig die vergleichsweise geringere Mathematikleistung der Befragten auffällig, die Erfahrungen mit Schülerfirmen haben ($d = - 0,32$). Gerade deshalb ist es von Bedeutung, dass ein derart deutliches Defizit der Netzwerkschüler auf Klassenstufe 9 im Fach Mathematik nicht mehr nachweisbar ist ($d = - 0,17$). Leider entziehen sich die Befunde zu den Problemlösungsaufgaben wegen der geringen Fallzahlen weitgehend der Interpretation; denn auf die Ergebnisse von 23 bzw. 21 bzw. 40 Schülerinnen und Schülern, je nach Klassenstufe, wird man kaum weitreichende Schlussfolgerungen gründen können.

Aus diesen heterogenen Befunden Bewertungen hinsichtlich der Wirksamkeit der Arbeit mit Schülerfirmen abzuleiten, wäre jedoch verfrüht. Immerhin kann man mit exploratorischer Zielsetzung eine Reihe von multiplen Regressionsanalysen rechnen, um wenigstens Vermutungen über die Auswirkungen auf die Testleistung zu generieren (vgl. Tabelle 17).

Tabelle 17: Vier multiple Regressionsanalysen zum Einfluss der Arbeit mit Schülerfirmen auf die Testleistung (HST 4/5 bzw. Problemlösen) (R^2 und β -Koeffizienten)²

	1.	2.	3.	4.
R^2 HST 4/5	0,32**	0,32**		
R^2 Problemlösen			0,13**	0,13**
Klassenstufe	0,14**	0,14**	0,13**	0,13**
CFT 20	0,42**	0,42**	0,21**	0,21**
Geschlecht (weibl.)	- 0,11**	- 0,11**	- 0,17**	- 0,18**
Migrant	- 0,24**	- 0,24**	- 0,04ns	- 0,03ns
Netzwerkschule	- 0,06**		0,04ns	
Anzahl der Beteiligungen		- 0,05*		0,08ns

² Signifikanzprüfungen im strengen Sinne waren hier nicht möglich. Um gleichwohl Anhaltspunkte für die statistische Bedeutsamkeit der Effekte zu geben, werden unter der kontrafaktischen Annahme einer echten Zufallstichprobe von Schülern ‚Signifikanzen‘ berichtet: $p \geq 0,05$: ‚ns‘; $0,01 < p < 0,05$: ‚*‘; $p < 0,01$: ‚**‘

Hiernach lassen die vorhandenen Daten vorläufig allenfalls zwei vage Tendenzen erkennen:

- Es gibt keine Indizien für positive Auswirkungen der Arbeit mit Schülerfirmen auf die allgemeine Fachleistung; hingegen können solche Effekte im Bereich des Problemlösens nicht unbedingt ausgeschlossen werden.
- Maßgeblich dafür ist offenbar nicht die Existenz (mindestens) einer Schülerfirma an einer Schule, sondern die Intensität, mit der sich die Befragten in diesem Bereich engagiert haben, gemessen an der Anzahl der eigenen Beteiligungen.

Die Effekte, die sich mit der Klassenstufe, den kognitiven Lernvoraussetzungen, dem Geschlecht und dem Migrantenstatus verbinden und die hier zur Absicherung der Befunde zu den Schülerfirmen herangezogen werden mussten, werden in diesem Bericht an anderer Stelle diskutiert. Im Übrigen wäre es gewiss sinnvoll, diejenigen unter den ‚Netzwerkschulen‘ zu identifizieren, an denen unter Berücksichtigung der genannten Faktoren besonders gute Fachleistungen, aber auch besonders gute Leistungen im Problemlösetest beobachtet wurden, die sich also als „*benchmarks*“ für andere Schulen anbieten. Ein solcher Ansatz überschreitet indessen die gegenwärtigen Möglichkeiten und muss ggf. späteren Analysen vorbehalten bleiben.

4.3 Unterschiede in den Leistungen zwischen Schulklassen

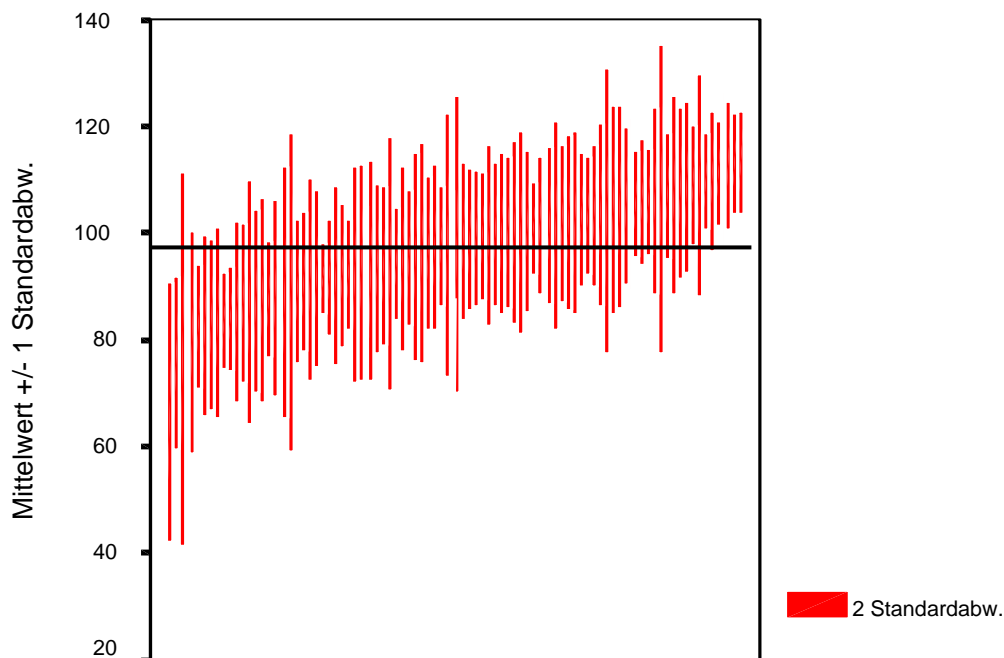
4.3.1 Unterschiede in den Leistungen zwischen Schulklassen allgemein

Im vorigen Abschnitt 4.2 waren Unterschiede in den Fachleistungen zwischen den einzelnen Schulen dargestellt worden, und zwar exemplarisch für die Klassenstufe 8. Die Leistungsunterschiede zwischen den untersuchten Schul-

klassen sind zwangsläufig höher als die Differenzen zwischen den Schulen: Die Parallelklassen in einer Schule werden sich so gut wie niemals auf exakt gleichem Leistungsstand befinden, und so wird die Gesamtvarianz der individuellen Testergebnisse in höherem Maße durch die Klassenzugehörigkeit als durch die Schulzugehörigkeit bestimmt.

Grafisch zeigt dies die Abbildung 16, die analog zur Abbildung 8 aufgebaut ist. In dieser Abbildung sind die Klassen ohne Rücksicht auf die Zugehörigkeit zu einer bestimmten Schule nach ihrer durchschnittlichen Fachleistung angeordnet.

Abbildung 16 Verteilung der Ergebnisse im LABEL 8-10-Fachleistungstest (Klassenstufe 8), nach Klassen (Mittelwerte plus/minus eine Standardabweichung)



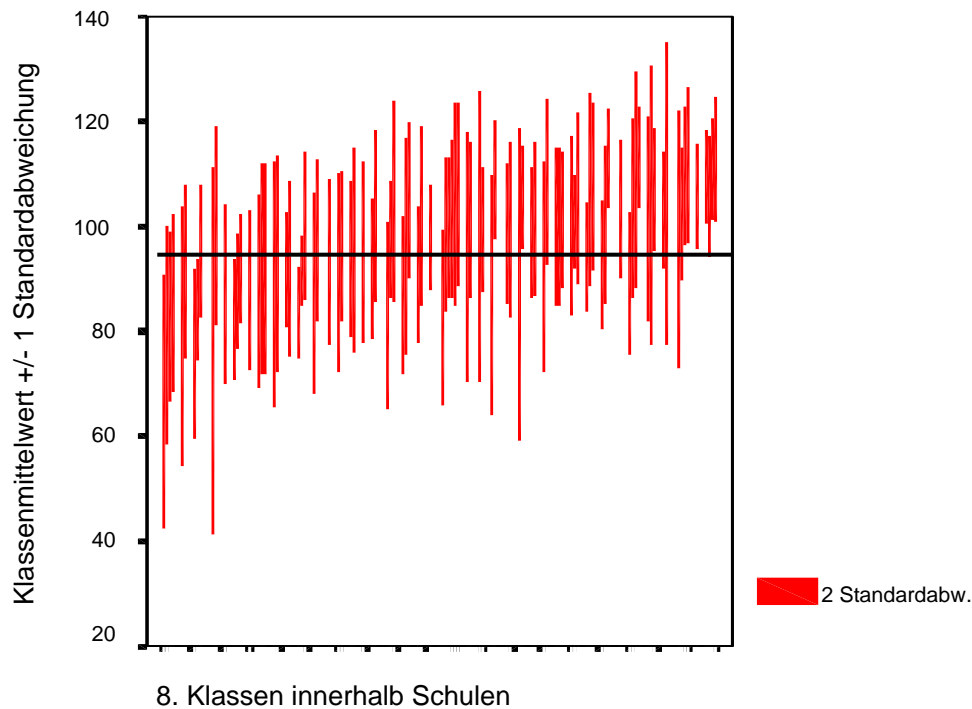
8. Klassen, geordnet nach mittlerer allg. Fachleistung

In der üblichen Darstellung einer Häufigkeitsverteilung sind die Klassenmittelwerte mit einer Standardabweichung von 9,4 Skalenpunkten glockenförmig um ihren Mittelwert von 96,5 verteilt. In der Abbildung 16 ist dieser Mittelwert durch die waagerechte Linie markiert. Wären hier die Klassenmittelwerte miteinander verbunden, so ergäbe sich eine am linken Rand leicht nach unten und am rechten Rand leicht nach oben abgebogene Kurve. Es können jedoch die großen Unterschiede zwischen den Schulklassen, die sich in einem Bestimmtheitsmaß von η^2 ausdrücken, nicht umstandslos auf die Effektivität pädagogischer Maßnahmen zurückgeführt werden. So gehen allein mehr als zwei Fünftel der Mittelwertunterschiede zwischen den Schulklassen (41,3 Prozent) auf Unterschiede im CFT-Klassenmittelwert und im Migrantenanteil zurück. Hier wäre im Einzelfall – und auch im Vergleich der Bezirke – zu untersuchen, welche unterschiedlichen Praktiken der Zuweisung von Schülerinnen und Schülern zu den Sonderschulen hinter solchen Zusammenhängen stehen.

Dass jedoch Lernerfolge ihre Wurzeln auch auf der Ebene des Fachunterrichts im Klassenverband haben, legt die nachstehende Abbildung 17 nahe. Hier sind die Ergebnisse der Schulen so angeordnet, dass sie in die Einzelergebnisse aus den zugehörigen Parallelklassen in der achten Jahrgangsstufe aufgelöst sind. Die Mitte eines Balkens bezeichnet wieder den Klassenmittelwert, seine Länge die doppelte Standardabweichung als Maß für die Streuung. Klassen verschiedener Schulen sind durch eine Lücke voneinander getrennt.

Nimmt man nun an, dass die Rekrutierungsmuster und –standards für alle Klassen einer Schule annähernd gleich sind, so verlangen die beobachteten Unterschiede zwischen ihnen in Durchschnittsleistung und Streuung nach Erklärungen, die prinzipiell auch die Effektivität des Unterrichts betreffen können. Im einfachen Querschnitt und in Abwesenheit von Indikatoren für die Unterrichtsqualität sind solche Erklärungsansätze hier nicht möglich. Auch bliebe noch zu prüfen, ob sich die vorgefundenen Effekte in der Tat mit dem Fachunterricht verbinden lassen oder ob sie über die Grenzen des Fachunterrichts hinweg stabil sind.

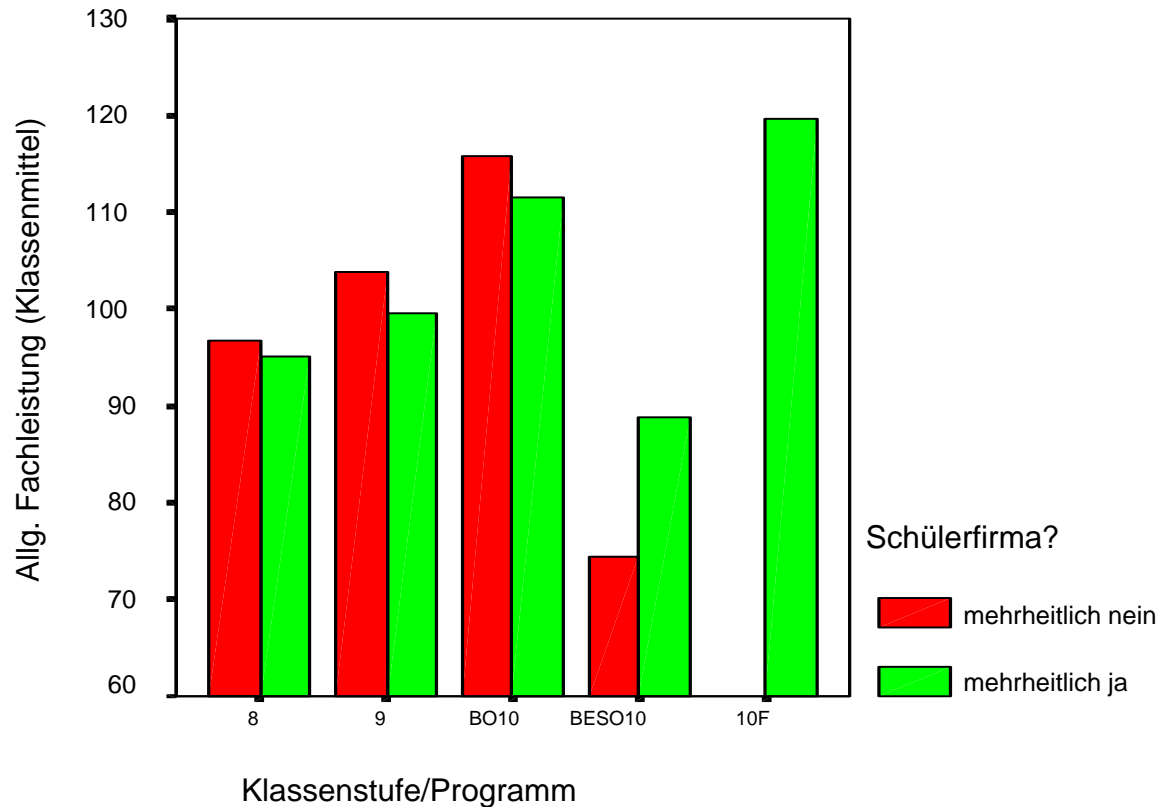
Abbildung 17: Verteilung der Ergebnisse im LABEL 8-10-Fachleistungstest, nach Klassen der achten Jahrgangsstufe innerhalb von Schulen (Durchschnittswerte plus/minus eine Standardabweichung)



4.3.2 Unterschiede in den Leistungen zwischen Schulklassen nach ihrer Teilnahme am *Netzwerk Berliner Schülerfirmen*

Fragt man danach, ob auf Klassenebene Effekte der Schülerfirmen zu verzeichnen sind, so ergeben sich interessante Differenzierungen nach Klassenstufe (vgl. Abbildung 18). Die Abbildung zeigt zwischen den Klassen, deren Schülerschaft mehrheitlich in Schülerfirmen mitarbeitet, und jenen, in denen dies nicht der Fall ist, in den Klassenstufen 8 und 9 sowie im Bildungsgang BO 10 annähernd gleiche Leistungsstände, mit geringen Vorteilen zugunsten der überwiegend konventionell unterrichteten Lerngruppen. Im leistungsmäßig eher problematischen Bildungsgang BESO 10 hingegen, der in der Stichprobe immerhin acht Schulklassen umfasst, sind in den Schulklassen mit einem mehrheitlich akzeptierten Angebot von Schülerfirmen recht beachtliche Leistungen zu beobachten, und erst recht gilt dies für die (einzige) Schulklasse im Bildungsgang 10 F.

Abbildung 18: Allgemeine Fachleistung (LABEL 8-10-Fachleistungstest), nach Klassenstufe und mehrheitlicher Zugehörigkeit der Schülerschaft zu einer Schülerfirma



Eine angemessene Würdigung der relativen Leistungsposition der Klassen, in denen die Schülerschaft überwiegend in Schülerfirmen eingebunden ist, erfordert allerdings die Berücksichtigung von Hintergrundmerkmalen, die schulisch kaum oder gar nicht zu beeinflussen sind:

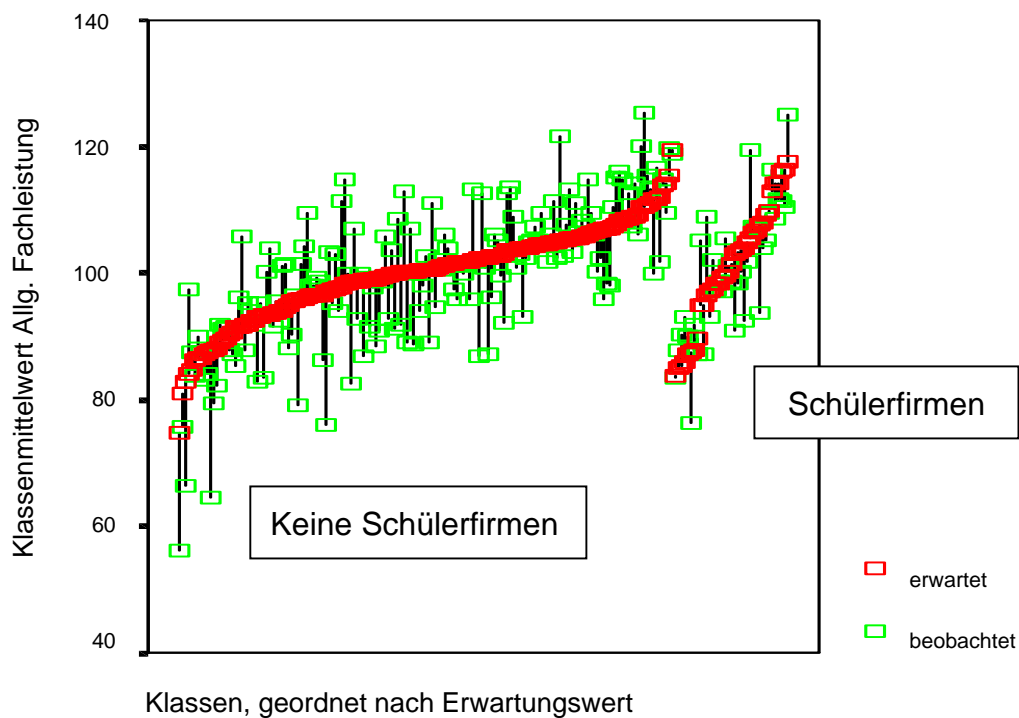
- Klassenstufe und Bildungsgang ($0,232 < \beta < 0,345$)
- Kognitive Lernvoraussetzungen (CFT 20; $\beta = 0,374$)
- Ausländerstatus ($\beta = - 0,214$)
- (weibl.) Geschlecht ($\beta = - 0,117$)
- Selbstwirksamkeitsüberzeugung ($\beta = 0,058$)
- Tendenz zur externalen Kausalattribution ($\beta = - 0,127$).

Unter Berücksichtigung dieser Merkmale lassen sich für die einzelnen Schulklassen ‚Erwartungswerte‘ berechnen, mittlere Lernstände also, die unter Berliner Verhältnissen bei bekannten Lernvoraussetzungen zu erwarten sind. Der

hierzu eingesetzte Algorithmus, eine „multiple Regressionsanalyse“, liefert auf Individualebene mit $R^2 = 0,363$ akzeptable Vorhersagegenauigkeiten.

Diese Erwartungswerte können nun mit den tatsächlich gemessenen Fachleistungen in den einzelnen Schulklassen verglichen werden. In Abbildung 19 geschieht dies so, dass die Schulklassen in aufsteigender Reihenfolge nach ihren Erwartungswerten (im Diagramm repräsentiert durch rote Punkte) angeordnet sind, welche ihrerseits mit den tatsächlich beobachteten Werten (grüne Punkte) mit einer schwarzen Geraden verbunden sind. Obwohl für Schulklassen mit und ohne (nennenswerte) Schülerfirmenbeteiligung eine einheitliche Gleichung zur Berechnung der Erwartungswerte verwendet worden ist, sind diese beiden Gruppen optisch getrennt: nur die Klassen im rechten Teil der Grafik sind im Bereich Schülerfirmen engagiert.

Abbildung 19 Allgemeine Fachleistung (LABEL 8-10-Fachleistungstest), nach mehrheitlicher Zugehörigkeit der Schülerschaft zu einer Schülerfirma: erwartete und gemessene Lernstände



In dieser Darstellung wird sichtbar, dass von den 38 Schulklassen, in denen mehr als die Hälfte der Schüler und Schülerinnen an Schülerfirmen beteiligt ist, nur acht in ihrer tatsächlichen Fachleistung nennenswert von den erwarteten Mittelwerten abweichen, und zwar je vier mit erwartungswidrig hohen bzw. niedrigen Werten. Im Durchschnitt liegen die Abweichungen weniger als einen Skalenpunkt unter den Erwartungswerten. Aus analytischen, rein

innermathematischen Gründen ist der Durchschnitt für die übrigen Schulklassen nahezu identisch mit dem Gesamterwartungswert, doch die (quadrierten) Abweichungen sind hier mit einer so genannten Residualvarianz von $s^2 = 68,9$ erheblich größer als bei den Schülerfirmen ($s^2 = 40,5$). Es lässt sich also aus dieser Analyse keinerlei Argument gegen das Engagement in Schülerfirmen ableiten. Im Gegenteil scheint mit diesem Ansatz eine Stabilisierung erwartbarer Leistungen verbunden zu sein, insbesondere eine Vermeidung unerwartet geringer Leistungen in bestimmten Schulklassen. Umgekehrt können Schulklassen, die – gemessen an ihren Eingangsvoraussetzungen – erwartungswidrig günstige Testergebnisse erzielt haben, als Maßstäbe dafür dienen, was unter realistischen Bedingungen erreichbar ist, und zwar auch unter den besonderen Bedingungen der Arbeit mit Schülerfirmen. Damit wird belegt, dass auch hier, auf Klassenebene, das sog. „*Benchmarking*“ einen realistischen Ansatz darstellt.

4.4 Leistungsunterschiede zwischen Mädchen und Jungen

Dass Jungen bzw. männliche Jugendliche an den Sonderschulen mit dem Schwerpunkt Lernen überrepräsentiert sind, ist eine keineswegs auf Berlin beschränkte Erscheinung, die auch oben, im Zusammenhang mit der Stichprobenzusammensetzung bereits kurz angesprochen wurde. Dabei war allerdings die Frage offen geblieben, ob die männlichen Jugendlichen auf Grund mangelnder Fachleistung – gleichsam zu Recht – in größerer Anzahl an die Sonderschulen relegiert werden oder ob sich die Lehrerschaft der Allgemeinen Schulen durch die Maßnahme der eigentlich ungerechtfertigten Überweisung an Sonderschulen vor allem bei Jungen einer unbequemen Klientel entledigt hat.

In diesem Zusammenhang ist zunächst die Vorfrage zu klären, ob sich *innerhalb der Schülerschaft der Sonderschulen* Jungen und Mädchen nach ihren intellektuellen Ressourcen unterscheiden; denn zu den Basisannahmen des Systems der getrennten Regel- und Sonderschulen gehört es, dass Schülerinnen und Schüler gleicher Fähigkeit auch dieselbe Schulform besuchen können müssen. Die Befunde in dieser Hinsicht sind eindeutig (vgl. Tabelle 18).

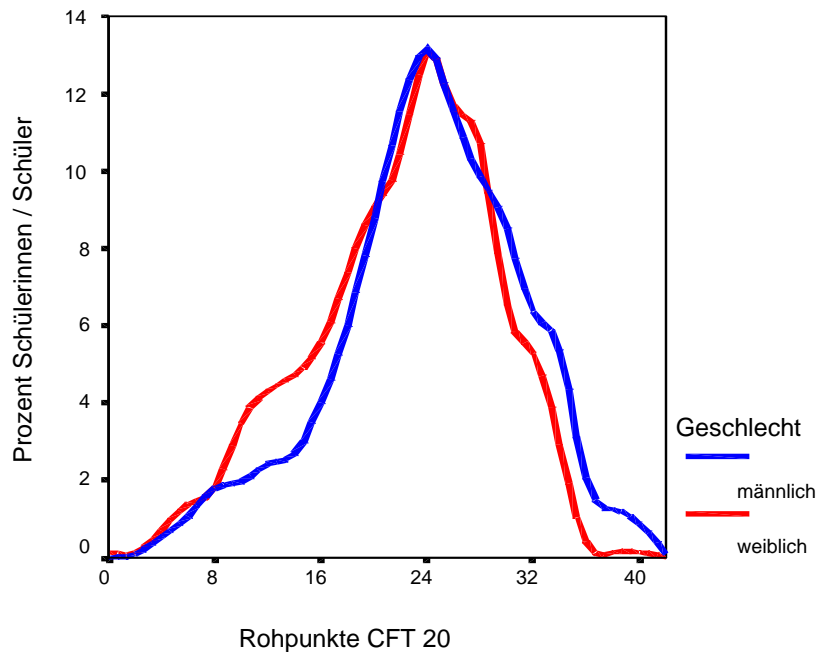
Tabelle 18: Ergebnisse zweier Intelligenztests nach Geschlecht

Intelligenztest	Jungen	Mädchen	Effektstärke d	N
CFT 20	23,6	21,7	0,27	1.836
HAWIK-R	73,4	70,5	0,29	591

Im CFT 20, der in der vorliegenden Untersuchung als Maß für die intellektuellen Ressourcen der Jugendlichen verwendet worden ist, beträgt der Geschlechterunterschied mehr als eine Viertel Standardabweichung zu Gunsten der Jungen. Es ist also davon auszugehen, dass hier seitens der überweisenden Institutionen eine ziemlich grobe Ungerechtigkeit auf ihre Kosten stattgefunden hat. Angesichts der Brisanz dieses Befundes wurden zur Kontrolle weitere Intelligenztestergebnisse ausgewertet, die aus den Schülerakten ersichtlich waren, und zwar diejenigen, die sich auf den zumeist verwendeten HAWIK-R bezogen. Das Resultat bestätigt und verschärft die Befunde der CFT-Auswertung: Diejenigen Intelligenztestergebnisse, die am häufigsten zur Begründung der Sonderschulüberweisung herangezogen wurden, wiesen nicht etwa auf eine schwächere Benachteiligung der Jungen hin, sondern sogar auf einen – freilich geringfügig – verstärkten Effekt.

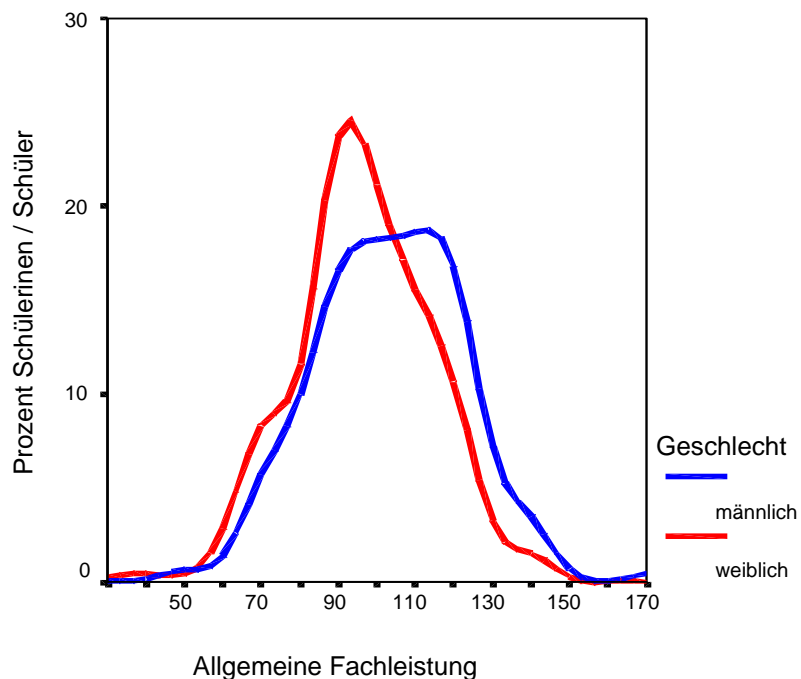
Abbildung 20 zeigt noch einmal graphisch, welche Verteilungseigenschaften hinter diesen Mittelwertunterschieden stehen.

Abbildung 20 Fähigkeit zum schlussfolgernden Denken (CFT 20) nach Geschlecht



Hier wird deutlich, dass an den Sonderschulen mit dem Schwerpunkt Lernen Jugendliche beiderlei Geschlechts anzutreffen sind, die man dort nach ihren Ergebnissen im nonverbalen Intelligenztest CFT nicht vermuten würde, etwa bei mehr als 28 Rohpunkten. Bei den männlichen Jugendlichen ist dieser Fall indessen mit 23,1 Prozent wesentlich häufiger als bei den weiblichen mit 14,6 Prozent.

Abbildung 21 Allgemeine Fachleistung nach Geschlecht



In Abbildung 21 wird nun geprüft, ob die allgemeine Leistungsverteilung der Intelligenzverteilung folgt oder ob in einem der beiden Geschlechter erwartungswidrige zu verzeichnen sind. Zwar entspricht die Effektstärke für den mit dem Geschlecht verbundenen Mittelwertunterschied in der allgemeinen Fachleistung mit 0,27 Standardabweichungen genau den Vorgaben der Intelligenzmessung (vgl. Tabelle 12), doch die Verteilungsform weicht im Falle der Jungen charakteristisch von derjenigen für die CFT-Rohwerte ab. Überraschenderweise ist die Erscheinung des Overachievement, also das Erbringen von höheren Fachleistungen, als nach der gemessenen Intelligenz zu erwarten sind, bei Jungen häufiger als bei Mädchen; denn 46 Prozent der Jungen, aber nur 30 Prozent der Mädchen haben mehr als 105 Punkte auf der Fachleistungsskala erzielt, d.h. mehr als den Skalenwert des Schnittpunkts der geschlechtsspezifischen Verteilungen. Dies wird auch an der flachen Verteilungsform für die Jungen erkennbar. Man kann also im Falle der Jungen nicht eine zu erwartende Form des Underachievement als Begründung für die benachteiligende Überweisung an die Sonderschulen mit dem Schwerpunkt Lernen anführen. Die numerischen Unterschiede zwischen den Geschlechtern in den Fachleistungen, ergänzt um den fächerübergreifenden Problemlösetest, und die entsprechenden Effektstärken sind in Tabelle 19 zusammengefasst.

Tabelle 19 Fachleistung nach Geschlecht: Mittelwerte und Effektstärken

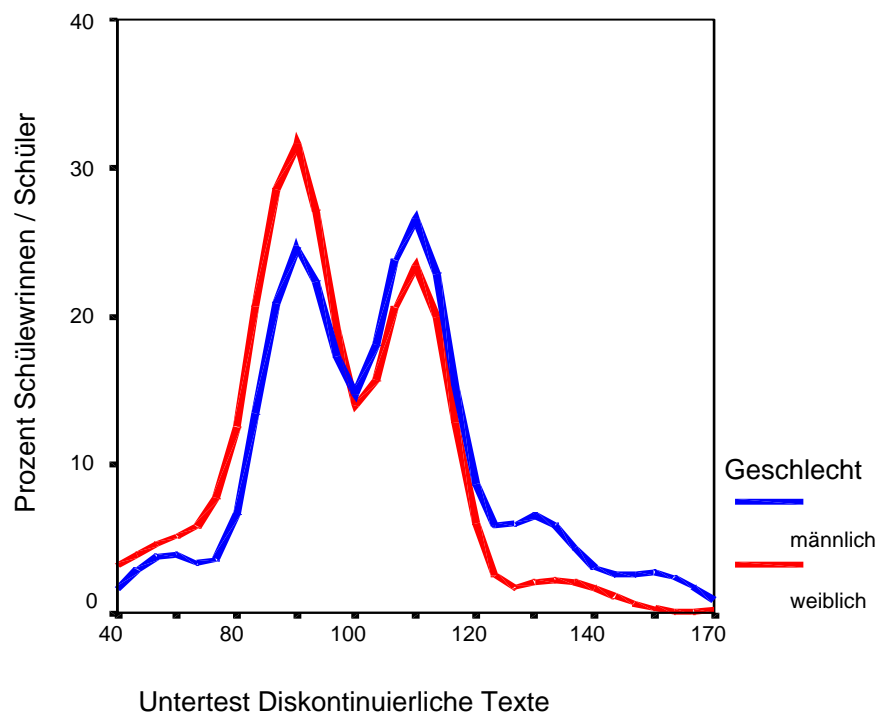
	Geschlecht		Effektstärke <i>d</i>	N
	männlich	weiblich		
Allg. Fachleistung (HST 4/5)	102,8	96,4	- 0,33	1.878
<i>Leseverständnis</i>	100,7	99,4	- 0,06	1.852
<i>Wortkenntnis</i>	103,0	95,8	- 0,36	1.848
<i>Informationsentnahme</i>	103,1	95,3	- 0,39	1.878
<i>Mathematik</i>	103,9	94,7	- 0,47	1.839
Problemlösen	44,7	39,5	- 0,36	341

Berücksichtigt man die grundlegende Leistungsdifferenz, die nicht zuletzt durch die unterschiedlichen kognitiven Grundfähigkeiten bedingt ist, so folgen die hier beobachteten Geschlechterunterschiede weitgehend den bekannten Mustern. Im Bereich des Leseverständnisses – einer ‚weiblichen Domäne‘ – können die Mädchen die Basisdifferenz fast kompensieren, während in der Mathematik – der ‚männlichen Domäne‘ der Geschlechtereffekt noch verstärkt ist. In den übrigen Untertests wie auch im Problemlösen sind die Unterschiede etwas stärker ausgeprägt, als es die CFT-Werte (und die HAWIK-R-Werte) erwarten lassen. Anders als in den Allgemeinen Schulen bilden also an

den Sonderschulen mit dem Schwerpunkt Lernen die weiblichen Jugendlichen, die sich in der klaren Minderheit befinden und paradoxerweise durch die mädchenfreundliche Überweisungspraxis eine kognitiv besonders stark negativ ausgelesene Gruppe darstellen, eine Risikogruppe dar.

Die Sonderstellung der Mädchen an diesen Schulen drückt sich allgemein so aus, dass die relativ höheren Kompetenzstufen überproportional mit Jungen besetzt sind. Am Beispiel zweier Untertest seien jedoch markante Unterschiede in der Verteilungsform auch grafisch demonstriert. Hinsichtlich des Umgangs mit diskontinuierlichen Texten (Karten, Tabellen, Diagrammen) ist bereits die ausgeprägt bimodale Verteilungsform angesprochen worden. Im gegenwärtigen Kontext ist es interessant zu prüfen, ob diese Erscheinung in irgendeinem Zusammenhang mit den geschlechtsspezifischen Leistungsprofilen steht (vgl. Abbildung 22).

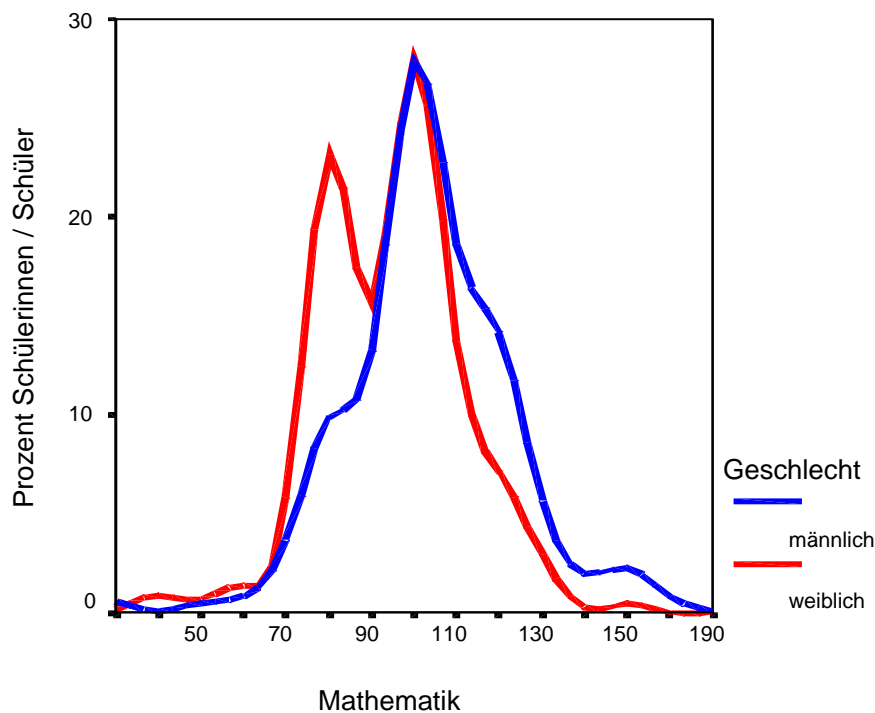
Abbildung 22 Testergebnisse im Umgang mit diskontinuierlichen Texten (Karten, Tabellen, Diagrammen) nach Geschlecht



Es ist bemerkenswert, dass die Polarisierung (die „Bimodalität“) der Verteilung für *beide* Geschlechter gilt, wobei lediglich verschiedene Akzente gesetzt sind: Bei den weiblichen Jugendlichen sind die ausgesprochen niedrigen Kompetenzbereiche deutlich häufiger besetzt, während bei den männlichen in mehr als der Hälfte der Fälle vergleichsweise hohe Kompetenzen festzustellen sind. Angesichts dessen, was differenziell über die kognitiven Lernvoraussetzungen und die sonstigen Fachleistungen gefunden wurde, war ein solcher Unterschied wohl zu erwarten.

Im Bereich Mathematik hingegen sind die Formen der Leistungsverteilung von Jungen und Mädchen deutlich unterschiedlich (vgl. Abbildung 23).

Abbildung 23 Testergebnisse in Mathematik nach Geschlecht



Hier wird erkennbar, dass im Bereich Mathematik nur die Leistungsverteilung der Mädchen polarisiert ist: Obwohl von den kognitiven Voraussetzungen her annähernd eine Normalverteilung zu erwarten gewesen wäre, gibt es unter den weiblichen Jugendlichen an den Sonderschulen offenbar eine relativ große Gruppe von Underachievern. Es wäre viel gewonnen, wenn erstens diese Mädchen diagnostisch identifiziert werden könnten, wenn in diesen Fällen zweitens die Gründe für die erwartungswidrig niedrigen Leistungen in Erfahrung gebracht und auf dieser Basis drittens spezifische Fördermaßnahmen entwickelt und umgesetzt würden.

4.5 Leistungen von Jugendlichen mit Migrationshintergrund

Von den untersuchten Jugendlichen an den Sonderschulen mit dem Förderschwerpunkt Lernen besitzen 15,5 Prozent nicht die deutsche Staatsbürgerschaft; 18,4 Prozent haben als Muttersprache eine andere Sprache als Deutsch angegeben. Beide Werte belegen – im Gegensatz zu häufig geäußerten Vermutungen – keine Überrepräsentanz von Schülern mit Migrationshintergrund in dieser Schulform. Wegen der Mehrdeutigkeit des Kriteriums der Staatsbür-

gerschaft (verdeckte Migrationsgeschichte bei Aussiedlern und Eingebürgerten, abgesehen von der mehrfachen Staatsbürgerschaft) basieren die nachstehenden Analysen auf Hinweise der Klassenlehrkräfte aus den Schülerakten zur ‚Muttersprache‘. Danach haben 39,9 Prozent der Jugendlichen fremder Muttersprache Türkisch oder Kurdisch (bei der Befragung nicht unterschieden) als Erstsprache; Arabisch folgt mit 22,0 Prozent und die Sprachen des ehemaligen Jugoslawien mit 15,3 Prozent. Russisch und Polnisch sind mit zusammen 10,6 Prozent als Muttersprache unter den Jugendlichen mit Migrationshintergrund an dieser Schulform vergleichsweise selten.

In Tabelle 20 werden die Testergebnisse zusammengefasst – im Falle des Schulleistungstests HST 4/5 auch differenziert nach Untertests.

Tabelle 20 Testergebnisse nach Muttersprache: Mittelwerte und Effektstärken

	Muttersprache		Effektstärke <i>d</i>	<i>N</i>
	deutsch	nicht deutsch		
CFT 20	23,4	19,9	- 0,51	1.938
Allg. Fachleistung (HST 4/5)	103,6	86,2	- 0,88	1.983
<i>Leseverständnis</i>	102,9	88,6	- 0,72	1.956
<i>Wortkenntnis</i>	103,4	85,6	- 0,89	1.952
<i>Informationsentnahme</i>	102,8	88,8	- 0,70	1.983
<i>Mathematik</i>	102,5	90,6	- 0,60	1.943
Problemlösen	44,5	39,8	- 0,33	333

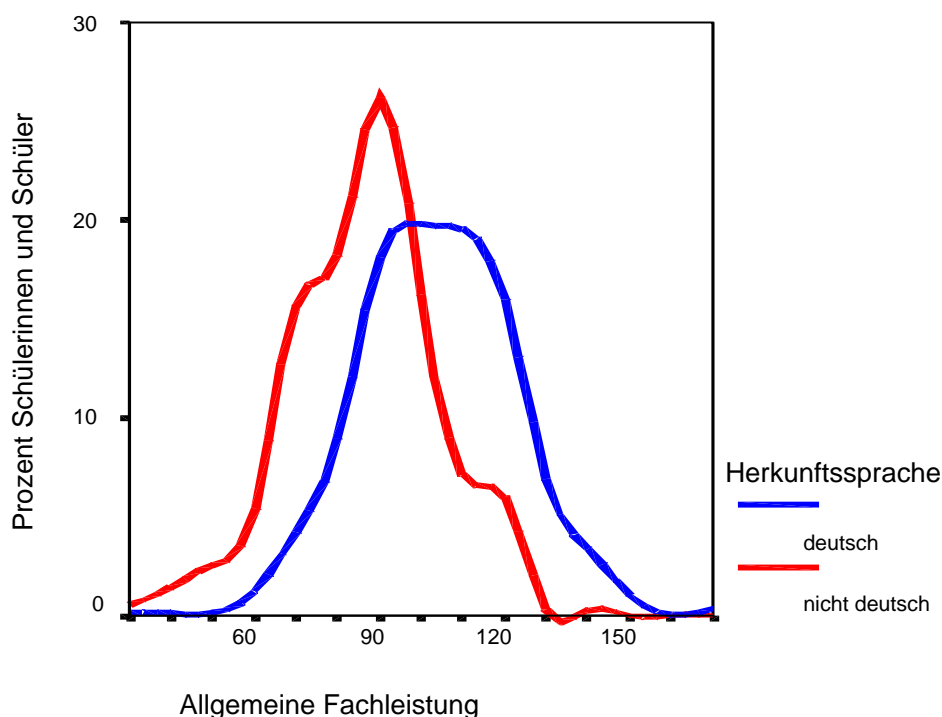
Die Ergebnisse aus der Messung der nonverbalen Intelligenz (CFT 20) sind hier mit aufgeführt, weil sie eine wichtige Kontrollvariable darstellen. Würden für Kinder und Jugendliche deutscher und nicht deutscher Herkunftssprache gleiche Maßstäbe gelten und wären zudem die kognitiven Grundfähigkeiten gleich verteilt, so dürfte hier keine Differenz auftreten. Der beobachtete Effekt ist aber mit mehr als einer halben Standardabweichung nicht zu vernachlässigen. Wie immer sich die hier nicht zu untersuchenden Verhältnisse in der Berliner Gesamtschülerschaft darstellen, so belegt doch der beobachtete Unterschied in den CFT-Ergebnissen zwischen Schülern deutscher und nicht deutscher Herkunftssprache, dass bei der Überweisung an die Sonderschule mit dem Förderschwerpunkt Lernen offenbar unterschiedliche Maßstäbe angelegt werden, und zwar im Durchschnitt zu Gunsten der Kinder mit Migrationshintergrund; denn für die Kinder deutscher Herkunftssprache liegt die kritische Intelligenzschwelle offenkundig höher. Diese Beobachtung kann verschiedene Hintergründe haben: Denkbar wäre es, dass Lehrkräfte angesichts manifester Sprachdefizite tiefer liegende kognitive Probleme verkennen.

Gleichfalls plausibel (und durchaus mit der eben erwogenen Erklärung zu vereinbaren) wäre es anzunehmen, dass angesichts der sozialräumlich sehr ungleichen Verteilung der Migranten in Berlin die Praxis der Überweisung an Sonderschulen in Gegenden mit hohen Ausländeranteilen mit niedrigeren Standards verbunden ist.

Auffällig ist es allerdings, dass die festgestellten Fachleistungsunterschiede noch wesentlich stärker ausgeprägt sind als Differenzen in den Lernvoraussetzungen. Das Argument, dass für Kinder und Jugendliche fremder Herkunftssprache an den Regelschulen offenbar verminderte Anforderungen gelten, gilt hier also in nochmals verstärkter Form. Interessanterweise – und nicht ganz unerwartet – sind die stärksten Unterschiede im Untertest *Wortkenntnis* sowie in den kumulativ berechneten Gesamtergebnissen im HST 4/5 zu beobachten. Lediglich im Problemlösetest zeigen sich geringere Unterschiede als in den gemessenen kognitiven Lernvoraussetzungen.

Abbildung 24 zeigt die Form und relative Position der Leistungsverteilungen für Jugendliche deutscher und nicht deutscher Herkunftssprache noch einmal in graphischer Form.

Abbildung 24 Allgemeine Fachleistung, nach Herkunftssprache



Hier wird klar erkennbar, dass die Jugendlichen nicht deutscher Herkunftssprache in mehrfacher Hinsicht eine ‚eigene Population‘ darstellen, die ganz besondere pädagogische Herausforderungen bieten. Nicht unerwähnt bleiben sollte in diesem Zusammenhang, dass allem Anschein nach die Arbeit mit

Schülerfirmen für diese Gruppe wirkliche Vorteile bietet (vgl. Tabelle 21). Hierauf verweist ja auch der Umstand, dass deren Anteil an der Schülerschaft wohl durch vorzeitigen Abgang nach der Pflichtschulzeit von 20,3 Prozent (Klassenstufe 8) über 17,7 Prozent (Klassenstufe 9) auf 12,4 Prozent (Klassenstufe 10) sinkt.

Tabelle 21 Allgemeine Fachleistung bei Jugendlichen nicht deutscher Herkunftssprache zu Beginn der Klassenstufen 8, 9 und 10 nach Teilnahme an der Arbeit mit Schülerfirmen: Mittelwerte und Effektstärken

	Schülerfirmen		Effektstärke <i>d</i>
	nein	ja	
Jahrgangsstufe 8	83,2 (<i>N</i> = 160)	90,9 (<i>N</i> = 22)	0,45
Jahrgangsstufe 9	89,7 (<i>N</i> = 123)	85,3 (<i>N</i> = 19)	- 0,25
Jahrgangsstufe 10	76,0 (<i>N</i> = 9)	89,9 (<i>N</i> = 18)	0,60

Abschließend mag es von Interesse sein, welche Unterschiede sich zwischen den Ethnien zeigen, die in nennenswerter Stärke an den Sonderschulen mit dem Schwerpunkt Lernen angetroffen wurden (vgl. Tabelle 22).

Tabelle 22 Allgemeine Fachleistung nach sprachlich-ethnischer Zugehörigkeit: Mittelwerte und Effektstärken

Sprachlich definierte Ethnie	Mittelwert	Effektstärke <i>d</i>	<i>N</i>
Deutsch	103,6	0,16	1.632
Sonstige	94,7	-0,29	156
Türkisch /kurdisch	85,5	- 0,76	142
Arabisch	88,6	- 0,60	77
Ehem. jugoslawisch	82,0	- 0,93	49
Polnisch / Russisch	90,9	- 0,48	38

Hiernach sind, abgesehen von der Residualkategorie („Sonstige“), zu der EU-Ausländer, aber auch Iraner und andere relativ bildungsnahe Gruppen gehören, vor allem die vergleichsweise hohen Lernstände unter den Jugendlichen mit Polnisch oder Russisch als Herkunftssprache bemerkenswert. Am anderen Ende des Leistungsspektrums fallen die extrem niedrigen Lernstände der Sonderschüler mit einer jugoslawischen Herkunftssprache, gefolgt von jenen der türkisch-kurdischen Gruppen auf. Auch diese Erscheinungen könnten mit dem jeweils typischen sozial-topographischen Umfeld und seinen spezifischen Überweisungsstandards zusammenhängen. In jedem Falle fehlen einst-

weilen Referenzwerte, die sich auf mögliche ethnische Differenzierungen in der gesamten Berliner Schülerschaft, also nicht nur an den Schulen mit dem Förderschwerpunkt Lernen, beziehen.

4.6 Sozialräumlich bedingte Leistungsunterschiede

Im Unterschied zu anderen Städten weist Berlin eine zweifache Binnengliederung auf: zum Einen die trotz der Verwaltungsreform von 2002 immer noch weitgehend erhaltene Unterscheidung zwischen Ost- und West-Berliner Bezirken (zwei Ausnahmen: die neuen Großbezirke Friedrichshain-Kreuzberg und Mitte mit Anteilen am früheren Ost- und West-Berlin), zum Anderen die sozialräumliche Gliederung mit Unterschieden im Sozialstatus der Bewohner eines Quartiers. Zu erwarten ist, dass die Praxis der Überweisung von Kindern und Jugendlichen an die Schulen mit dem Förderschwerpunkt Lernen in der Stadt keineswegs nach einheitlichen Kriterien erfolgt.

Wären solche Unterschiede nicht gegeben, so müsste man für die Hauptkriterien der Untersuchung in guter Näherung gleiche Durchschnittswerte für alle Stadtregionen erwarten. Schon die Prüfung dieser Hypothese nach der einfachen Unterscheidung zwischen Ost- und West-Berlin widerlegt diese Vermutung (vgl. Tabelle 23).

Tabelle 23 Testergebnisse nach Stadtregion: Mittelwerte und Effektstärken

	Ost-Berlin	West-Berlin	Effektstärke d	N
CFT 20	23,6	21,8	0,26	2.040
Allg. Fachleistung (HST 4/5)	103,7	96,3	0,38	2.094
<i>Leseverständnis</i>	102,9	97,4	0,28	2.066
<i>Wortkenntnis</i>	103,0	97,0	0,30	2.062
<i>Informationsentnahme</i>	102,8	96,9	0,29	2.094
<i>Mathematik</i>	103,3	96,5	0,34	2.045
Problemlösen	44,7	41,3	0,24	360

Hier sind generell höhere Testwerte für die östlichen Stadtbezirke zu beobachten, und zwar in der Größenordnung eines Viertels einer Standardabweichung für die allgemeine Fähigkeit zum schlussfolgernden Denken (CFT 20) und die fächerübergreifende Fähigkeit zum Lösen sequenzieller Aufgaben („Problemlösen“) gegenüber den durchgehend etwas höheren Werten für den Fachleistungstest HST 4/5 und seine Unterskalen. Will man nicht davon aus-

gehen, dass es generelle Intelligenzunterschiede dieser Größenordnung zwischen Ost- und West-Berlin gibt, so bleibt nur die Möglichkeit anzunehmen, dass Ost-Berliner Kinder und Jugendliche bereits mit höheren kognitiven Fähigkeiten an die Sonderschulen mit dem Schwerpunkt Lernen relegiert werden. Dafür spricht auch der disproportional höhere Anteil Ost-Berliner Jugendlicher ($N_{\text{ost}} = 1.050$ gegenüber $N_{\text{west}} = 944$). Darüber hinaus lässt der nochmals verstärkte Effekt für den Schulleistungstest vermuten, dass die West-Berliner Zurückhaltung in der Überweisungspraxis jedenfalls nicht automatisch und gleichzeitig zu besonders günstigen Lernentwicklungen unter den Sonderschülern geführt hat. Die mit dieser Zurückhaltung im Westen zwangsläufig verbundene extrem negative Auslese – und d.h. auch: die Seltenheit vergleichsweise leistungsstarker Mitschüler an den dortigen Schulen mit dem Förderschwerpunkt Lernen – ist im Gegenteil begleitet von besonders schwachen Fachleistungen in dieser Schulform.

Obwohl ein Vergleich der einzelnen Bezirke wegen der kleinen Fallzahlen nur mit größter Vorsicht zu interpretieren ist, seien die entsprechenden Zahlen hier genannt (vgl. Tabelle 24).

Tabelle 24 Testergebnisse (Allgemeine Fachleistung und CFT 20) nach Bezirken

Bezirk	Testpunkte	N	CFT 20
Treptow-Köpenick	106,52	157	23,80
Pankow	106,20	220	22,85
Marzahn-Hellersdorf	103,40	408	23,88
Steglitz-Zehlendorf	103,27	88	23,40
Lichtenberg	102,08	287	23,84
Spandau	98,79	57	24,50
Neukölln	98,76	259	21,61
Reinickendorf	98,41	128	22,06
Charlottenburg-Wilmersdorf	96,13	88	20,30
Tempelhof-Schöneberg	95,40	127	21,22
Mitte	94,03	185	22,18
Friedrichshain-Kreuzberg	86,26	90	21,13
Total	100,38	2094	22,80

Hier wird nochmals deutlich, mit welcher Konsistenz in den Bezirken des ehemaligen Ost-Berlin auch solche Kinder an die Schulen für Lernbehinderte überwiesen worden sind, die in den Westbezirken an Allgemeinen Schulen – vermutlich in der Mehrzahl Hauptschulen – hätten bleiben können. Die beiden Ausnahmen von dieser Tendenz sind leicht aufzuklären: Steglitz-

Zehlendorf weist vor allem deshalb günstige Werte auf, weil der Altbezirk Zehlendorf mit einem Mittelwert von 107,02 Testpunkten herausragt, erklärbar vielleicht als Konsequenz der in diesem Teil der Stadt zu vermutenden Verschiebung des allgemeinen Leistungsspektrums in Richtung auf günstige Werte. Umgekehrt ist die Schlussposition von Friedrichshain-Kreuzberg ganz wesentlich auf die extrem niedrigen Werte im Altbezirk Kreuzberg (Durchschnitt: 80,20 Punkte) zurückzuführen, auch wenn der Altbezirk Friedrichshain mit einem Mittelwert von 93,20 Punkten die Werte der Altbezirke Wedding (91,83 Punkte) und Willmersdorf (91,72 Punkte) nur knapp übertrifft. Offenkundig ist jedenfalls, dass die Praxis der Überweisung von Kindern und Jugendlichen an die Sonderschulen mit dem Förderschwerpunkt Lernen in hohem Maße von umgebenden Sozialmilieu abhängt.

5 Differenzierungen in den Selbsteinschätzungen der Jugendlichen

Im Folgenden sollen fünf charakteristische Formen der Selbsteinschätzung, die bei den Jugendlichen untersucht werden konnten, darauf hin analysiert werden, ob sie Gruppenunterschiede sichtbar machen, die ggf. pädagogisch relevant sind. Im Einzelnen handelt es sich um die folgenden, nach Vorgaben der probabilistischen Testtheorie konstruierten Skalen:

- Fachleistungsbezogenes Selbstkonzept
- Selbstwirksamkeit („externale Erfolgsattribution“)
- Eigene Arbeitstugenden
- Flexibilität und Belastbarkeit
- Technikfreundlichkeit.

5.1 Fachleistungsbezogenes Selbstkonzept

Ein relativ positives Selbstbild wird üblicherweise bei denjenigen Schülerinnen und Schülern angetroffen, die regelmäßig die Erfahrung machen, die ihnen gestellten Aufgaben lösen zu können, also bei den leistungsstärkeren. Man hätte also vermuten können, dass in der Schülerschaft von Schulen mit dem Förderschwerpunkt Lernen generell ein negatives Selbstkonzept dominiert. Es hatte sich jedoch gezeigt, dass dies durchaus nicht durchgängig der Fall ist (vgl. oben, S. 34).

Auch die Vermutung, Jugendliche mit Migrationshintergrund könnten angesichts ihrer erheblich geringeren Fachleistungen im Vergleichsrahmen der Sonderschulen, vielleicht auch angesichts einer exponierten sozialen Situation mit bedrohten Geltungsansprüchen, ein geringeres Selbstvertrauen aufweisen, wird, wie sogleich zu zeigen ist, durch die Daten nicht bestätigt.

Weiterhin ist aus anderen Untersuchungen bekannt, dass insbesondere leistungsschwächere Mädchen in dieser Hinsicht über ein geringeres Selbstvertrauen verfügen (vgl. HORSTKÄMPER 1987). Man ist vielleicht versucht, daraus die Hypothese abzuleiten, dass dieser Effekt in Anbetracht der besonders ungünstigen Leistungsposition der Mädchen in der untersuchten Klientel ungewöhnlich stark zur Geltung kommt. In Tabelle 25 sind die entsprechenden Befunde in simultaner Analyse aufgeführt.

Tabelle 25 Fachleistungsbezogenes Selbstkonzept nach Migrationsstatus und Geschlecht

	Geschlecht	Mittelwert	Standard- abweichung	N
Ohne Migrations- hintergrund	männlich	10,22	1,98	875
	weiblich	9,70	1,88	554
	Gesamt	10,02	1,96	1.429
Mit Migrations- hintergrund	männlich	9,84	2,03	177
	weiblich	9,91	2,30	136
	Gesamt	9,87	2,15	313
Gesamt	männlich	10,16	1,99	1.052
	weiblich	9,74	1,97	690
	Gesamt	9,99	1,99	1.742

Wie diese Daten bestätigen, gibt es zwischen den Jugendlichen mit und ohne Migrationshintergrund hinsichtlich ihres fachleistungsbezogenen Selbstkonzepts nur geringfügige Unterschiede, die keinerlei Signifikanzprüfung standhalten ($p = 0,22$). Die in der Literatur beschriebene Differenz zwischen den geschlechtsspezifischen Ausprägungen des Selbstvertrauens ist in der untersuchten Schülerschaft auf die Jugendlichen ohne Migrationshintergrund beschränkt ($d = 0,27$) und dort auch selbst dann signifikant, wenn regressions- oder kovarianzanalytisch der intervenierende Einfluss der Leistungspositionen berücksichtigt wird ($p = 0,00$).

Bedeutsame Unterschiede zwischen den Ost- und Westberliner Bezirken wurden nicht beobachtet. Bei den Schülerinnen und Schülern, die an einer größeren Zahl Schülerfirmen beteiligt waren, konnte eine leichte und angesichts der übrigen Befunde, insbesondere der Leistungsdaten, keineswegs selbstverständliche Tendenz in Richtung auf ein etwas günstiger ausgeprägtes Leistungsselbstbild festgestellt werden ($p = 0,06$). Mit steigender Klassenstufe steigt in der untersuchten Gruppe offenbar auch das fachleistungsbezogene Selbstkonzept ($r = 0,06$; $p = 0,01$).

5.2 External attribuierende Erfolgsdeutung

Die Neigung, Ursachen des eigenen Erfolges oder Misserfolges nicht bei sich selbst, sondern in den äußeren Umständen zu suchen, stellt bekanntlich keine günstige Voraussetzung für erfolgreiches Lernen dar: Im Erfolgsfalle „Glück“ (statt die eigene Fähigkeit), im Misserfolgss Falle „Pech“ und „Ungerechtigkeit“ (statt mangelnde Anstrengung) für den Lernerfolg verantwortlich zu machen ist eine Strategie, die insbesondere bei den weniger Erfolgreichen das Selbstbild schont. Die Befunde der vorliegenden Untersuchung betätigen auch für die besondere Gruppe der Schülerinnen und Schüler von Sonderschulen mit dem Förderschwerpunkt Lernen diese Erfahrung.

So ist diese Neigung in der untersuchten Schülerschaft durchgängig bei den weniger Leistungsstarken anzutreffen: Sie korreliert mit der Gesamtestleistung in Höhe von $r = -0,23$, mit den Ergebnissen für die einzelnen Untertests in Höhe von $r = -0,19$ (*Mathematik, Leseverständnis diskontinuierlicher Texte*) bzw. $r = -0,21$ (*Leseverständnis, Wortkenntnis*). Nachweisbar ist diese Beziehung auch für die Problemlösekompetenz ($r = -0,11$) und die gemessene Fähigkeit zum schlussfolgernden Denken ($r = -0,14$). Dass diese Neigung in den ehemals Westberliner Bezirken ($r = -0,17$) sowie unter den Jugendlichen mit Migrationshintergrund ($r = 0,20$) verbreiteter ist als im ehemaligen Ostberlin bzw. unter den Jugendlichen ohne Migrationshintergrund, hängt vermutlich mit der relativ schwachen Leistungsposition dieser Gruppen zusammen; auch sind die beiden Befunde nicht unabhängig voneinander. Erwartungswidrig ist hingegen, dass an den untersuchten Schulen die Jungen etwas stärker zu externaler Erfolgsattribuierung tendieren als die Mädchen ($r = -0,05$; zu entgegengesetzten Befunden aus Regelschulen vgl. HORSTKÄMPER 1987). Die Einbeziehung in die Arbeit mit Schülerfirmen des Netzwerkes scheint die Neigung zu externaler Erfolgsattribuierung nicht unbedingt zu verringern ($r = 0,05$), und zwar auch dann nicht, wenn man deren Konzentration auf Jugendliche mit eher geringeren Schulleistungen berücksichtigt (partielles $r = 0,06$).

5.3 Eigene Arbeitstugenden

Von der unerwartet positiven Selbsteinschätzung der Schülerinnen und Schüler im Hinblick auf ihre Arbeitstugenden wurde bereits berichtet (vgl. oben, Abschnitt 3). Bemerkenswert ist darüber hinaus, dass die schwachen Korrelationen mit den kognitiven Ausgangsvoraussetzungen (CFT 20) und den Leistungsvariablen selbst (mit Ausnahme des Problemlösens) *negativ* sind: Je ungünstiger das Leistungsniveau, desto positiver – tendenziell – das Selbstbild der Befragten unter der Frage nach Ordnung, Fleiß, Pünktlichkeit u.s.f..

Trotz der generell niedrigen Korrelationen sind die so ermittelten Befunde aufschlussreich. So steigt die Betonung der eigenen Arbeitstugenden in dem Maße, wie das Ende der formalen Schulzeit näher rückt ($r = 0,06$; $p = 0,004$). Schülerinnen und Schüler mit Migrationshintergrund betonen ihre Arbeitstugenden etwas stärker als diejenigen, die nicht aus zugewanderten Familien stammen ($r = 0,09$; $p = 0,00$). Erwartungskonform sehen Mädchen hier eher als Jungen ihre besondere Stärke ($r = 0,10$; $p = 0,00$). Westberliner Jugendliche in ihrer extremeren Leistungsschwäche rekurrieren stärker als die Ostberliner auf diese Merkmale ($r = 0,13$; $p = 0,00$). Die Erfahrung mit Schülerfirmen geht ebenfalls, wenn auch noch schwächer ausgeprägt, mit einer besonderen Betonung der Arbeitstugenden einher ($r = 0,06$; $p = 0,01$). Die Zusammenhänge mit den Testergebnissen sind niedrig und eher negativ; nur in den Untertests *Leseverständnis diskontinuierlicher Texte* und *Mathematik* kann überhaupt von signifikanten Befunden gesprochen werden ($r = -0,06$; $p = 0,01$). Es ist denkbar, bedürfte aber eingehenderer Untersuchungen, dass die Betonung der eigenen Arbeitstugenden in der Selbstdarstellung Jugendlichen kompensatorische Funktion hat angesichts der in der Zugehörigkeit zur Sonderschule auch nach außen signalisierten Lernschwächen.

5.4 Flexibilität und Belastbarkeit

Eine solche Kompensationsthese legt vielleicht die Vermutung nahe, dass auch die Einschätzung der eigenen Flexibilität und Belastbarkeit durch die Jugendlichen einem ähnlichen Muster folgt. Dies ist indessen nicht der Fall. Für flexibel und belastbar halten sich diejenigen Jugendlichen, deren Lernvoraussetzungen ausweislich des CFT 20 noch vergleichsweise günstig sind ($r = 0,09$; $p = 0,00$) und die auch im Schulleistungstest eher bessere Ergebnisse erzielt haben ($r = 0,07$; $p = 0,002$). Die Einschätzung der eigenen Belastbarkeit steigt anscheinend in den letzten Klassenstufen der Sonderschulen ($r = 0,05$; $p = 0,04$); sie ist jedoch offenbar unabhängig von der Zugehörigkeit zu einer der beiden Stadtregionen und auch der Erfahrung mit Schülerfirmen. Relativ eng hängt sie hingegen mit dem Geschlecht der Befragten zusammen und mittelbar dann auch mit dem Migrationsstatus (vgl. Tabelle 26). Man erkennt, dass der Geschlechterunterschied unter der Perspektive der selbst eingeschätzten Flexibilität und Belastbarkeit unter den Schülern ohne Migrationshintergrund mit einer Effektstärke von $d = 0,59$ noch stärker ausgeprägt ist als unter den Schülerinnen und Schülern aus zugewanderten Familien ($d = 0,47$).

Tabelle 26 Einschätzung der eigenen Flexibilität und Belastbarkeit nach Migrationsstatus und Geschlecht

	Geschlecht	Mittelwert	Standard- abweichung	<i>N</i>
Ohne Migrations- hintergrund	männlich	10,47	2,122	882
	weiblich	9,26	1,690	555
	Gesamt	10,00	2,051	1.437
Mit Migrations- hintergrund	männlich	10,31	1,908	178
	weiblich	9,49	1,384	136
	Gesamt	9,96	1,747	314
Gesamt	männlich	10,44	2,088	1.060
	weiblich	9,31	1,636	691
	Gesamt	9,99	2,000	1.751

5.5 Technikfreundlichkeit

Der Umstand, dass für die Schülerschaft der Schulen mit dem Förderschwerpunkt Lernen Berufsperspektiven im Bereich manuell-praktischer Tätigkeiten eine wichtige Rolle spielen, weist der Frage nach der Affinität dieser Jugendlichen gegenüber der modernen Technik eine besondere Bedeutung zu. In diesem Zusammenhang wurde bereits oben davon berichtet, dass sich die Lernstärkeren etwas aufgeschlossener gegenüber der modernen Technik geäußert haben als die langsameren Lerner und dass insbesondere starke Geschlechterdifferenzen beobachtet wurden. Es wird wenig überraschen, dass vor allem der Untertest *Mathematik* wie auch der Untertest *Problemlösen* (beide $r = 0,17$) einen noch deutlicheren Zusammenhang mit der Technikfreundlichkeit aufweist als der Gesamttest ($r = 0,14$; ebenso Leseverständnis diskontinuierlicher Texte und Wortkenntnis; deutlich abgeschwächt das Leseverständnis kontinuierlicher Texte mit $r = 0,08$). In den ehemaligen Westberliner Bezirken ist die Technikfreundlichkeit anscheinend etwas stärker ausgeprägt als im

Osten Berlins ($r = 0,07$, $p = 0,002$), was angesichts der unterschiedlichen Mechanismen der Übergangsauslese und vielleicht auch angesichts der starken Betonung von Naturwissenschaft und Technik in der DDR überraschen mag. Es bleibt aber vor allem der Bereich der Geschlechterdifferenzen im Umgang mit moderner Technik, der in dieser Schülerschaft Anlass für besondere Besorgnis liefert und in jedem Falle zusätzliche Analysen erfordert.

6 Zusammenfassung

Die vorliegende Studie ist unter anderem durch die Frage nach der Wirksamkeit einer bestimmten Konzeption für die Berliner Schulen mit dem Förderschwerpunkt Lernen veranlasst worden. Es geht um den künftigen Ertrag des Angebots der Arbeit in sog. „Schülerfirmen“, wie sie im „Netzwerk Berliner Schülerfirmen“ organisatorisch zusammengefasst sind. Um diese Frage schlüssig beantworten zu können, war es notwendig, zur Ermittlung der Ausgangslage dieser Zielgruppe, eine Vollerhebung an allen Berliner Schulen dieses Typs durchzuführen. Dabei wurden folgende zentralen Aspekte der schulischen Leistung berücksichtigt:

- Leseverständnis
- Wortkenntnis
- Informationsentnahme aus Tabellen, Karten und Diagrammen
- Mathematik
- Problemlösen als fächerübergreifende Kompetenz.

Um die Lern- und Verhaltensmerkmale der Zielgruppe in besonderer Weise zu berücksichtigen, kamen aus dem Bereich der Selbsteinschätzungen und Werthaltungen hinzu:

- Fachleistungsbezogenes Selbstkonzept
- External attribuierende Erfolgsdeutung
- Eigene Arbeitstugenden
- Flexibilität und Belastbarkeit
- Technikfreundlichkeit.

Daten zu außerschulischen Einflussfaktoren und Informationen aus den Schülerakten vervollständigten die Datengrundlage für die 2239 in die Untersuchung einbezogenen Schülerinnen und Schüler.

Die Verfügbarkeit einer so breiten Datenbasis ermöglichte es, die Reichweite der ursprünglich gestellten Fragen wesentlich zu erhöhen. So konnten auch allgemeinere Probleme wissenschaftlichen Analysen zugänglich gemacht werden. Zu diesen zählen u.a. etwa solche der kognitiven Lernzuwächse in den Abschlussklassen der einschlägigen Sonderschulen, der differenziellen Behandlung bei der Zuweisung zu solchen Schulen und in gleicher Weise

auch die jeweils spezifischen Probleme von Jugendlichen aus einheimischen und zugewanderten Familien.

Die Ergebnisse dokumentieren die Lernausgangslage der gesamten Population. Aussagen zur Wirksamkeit von Schülerfirmen können daraus in diesem Anfangsstadium nur begrenzt abgeleitet werden.

Da sich die Befunde auf den bisher kaum repräsentativ untersuchten Bereich der Schulen mit dem Förderschwerpunkt Lernen (anderenorts: „Schulen für Lernbehinderte“, „Allgemeine Förderschulen“ oder „Förderschulen mit dem Schwerpunkt Lernen“) beziehen, sind sie aber zweifellos als Charakterisierungen der Arbeit an dieser Schulform von besonderer Bedeutung.

Einige wenige Punkte seien herausgehoben:

- (1) Schulen mit dem Förderschwerpunkt Lernen verhalten sich in vieler Hinsicht zu einander wie Allgemeine Schulen: Sie weichen in ihrem Leistungsniveau und ihrem Leistungsspektrum erheblich voneinander ab, und zwar nicht zuletzt in Reaktion auf die Gegebenheiten des außerschulischen Umfeldes. In Berlin spielen dabei die Unterschiede bei der Feststellung des sonderpädagogischem Förderbedarfs in den früheren Ost- bzw. Westbezirken eine besondere Rolle: Allem Anschein nach werden in den Allgemeinen Schulen in den westlichen Bezirken von Berlin die Kinder und Jugendlichen relativ seltener an Schulen mit dem Förderschwerpunkt Lernen überwiesen. Folge dieses Vorgehens ist eine Differenzierung der Leistungsstände an den Sonderschulen in den unterschiedlichen Regionen.
- (2) In den Abschlussklassen der Schulen mit dem Förderschwerpunkt Lernen (Klassenstufen 8 bis 10) sind in allen untersuchten Unterrichtsbereichen durchaus erhebliche Lernfortschritte bei den Jugendlichen zu verzeichnen. In der Größenordnung entsprechen diese mit etwa dem Viertel einer Standardabweichung pro Jahr denen, die an Allgemeinen Schulen beobachtet werden, auch wenn dabei in Rechnung zu stellen ist, dass die Schülerschaft der Sonderschulen durch negative Ausleseprozesse homogener sein dürfte als die Schülerschaft der Allgemeinen Schulen.

Im Übrigen geht ein Teil dieser Lernzuwächse darauf zurück, dass offenbar vor allem in der Abschlussklasse 10 ein Teil der Schülerschaft, vielleicht motiviert durch die Antizipation des nahen Über-

tritts in das Berufsleben, bemerkenswert gute Leistungen zeigt und so für eine sekundäre Auffächerung des Leistungsspektrums sorgt.

- (3) Für den Verbleib an Allgemeinen Schulen gelten für männliche Kinder und Jugendliche offenbar strengere Maßstäbe als für weibliche. Dies folgt aus den generell höheren Testleistungen der Jungen aus Sonderschulen.
- (4) Jugendliche aus Familien mit einer Migrationsbiographie werden in Berlin bei gleicher Fachleistung seltener an eine Schule mit dem Förderschwerpunkt Lernen überwiesen als solche aus Familien, die seit Generationen in Deutschland ansässig sind. Auch prozentual liegt ihr Anteil an diesem Typ der Sonderschule unter der Quote, die nach der Verteilung des Merkmals „Migrationshintergrund“ zu erwarten wäre. In dieser Hinsicht kann also entgegen manchen Befürchtungen nicht von einer Benachteiligung ausländischer Schüler die Rede sein. Es besteht offenbar eine vergleichsweise größere Bereitschaft, lernschwachen Kindern und Jugendlichen aus Migrantenfamilien die Integration in Allgemeinen Schulen offen zu halten. Daran dürfte die extrem ungleiche sozialräumliche Verteilung der Migrantenfamilien in der Stadt beteiligt sein: In Stadtregionen mit hohen Migrantenanteilen ist zu vermuten, dass Schulen schon aus Eigeninteresse weniger dazu neigen, schwächere Lerner an Sonderschulen abzugeben.

In mancher Hinsicht ist die vorliegende Untersuchung eher als eine Machbarkeitsstudie für die spezielle Zielgruppe der Schülerinnen und Schüler mit dem sonderpädagogischen Förderschwerpunkt Lernen zu verstehen, die abschließende Urteile noch nicht zulässt. Es konnte gezeigt werden, dass mit Mitteln der empirischen Bildungsforschung, die für Untersuchungen an Allgemeinen Schulen entwickelt wurden, neue Einsichten auch in Prozesse gewonnen werden können, die sich an Sonderschulen und in deren Umfeld vollziehen. Einige der Befunde weisen auf die Notwendigkeit einer intensiveren Erforschung des Phänomens hin. So ist etwa in einigen Regionen die Zuweisung von Kindern und Jugendlichen zu den Sonderschulen eigentümlichen Regelmäßigkeiten unterworfen die schwerlich rational genannt werden können. Auch wurde immer wieder erkennbar, mit welchem hohem Differenzierungsgrad trotz aller institutionalisierter Bemühung um Homogenisierung auch an den Schulen mit dem Förderschwerpunkt Lernen zu rechnen ist.

Die Befunde dieser Untersuchung zur Lernausgangslage arbeitsrelevanter Basiskompetenzen von Berliner Schülerinnen und Schülern im Förderschwer-

punkt Lernen – Klassenstufe 8 bis 10 – bilden eine gute Grundlage für Folgeuntersuchungen. Erst mit einer weiteren Erhebung lässt sich die Grundfrage nach dem durch das komplexe Lernarrangement Schülerfirma bewirkten Lernfortschritten genauer beantworten. Besonders begrüßenswert ist es, dass sich ein Projekt, das aufwändig mit Mitteln des Europäischen Sozialfonds unterstützt wird, einer solchen objektiven empirischen Evaluation stellt.

Den Schulen liefern die Ergebnisse der Untersuchung zweifellos zahlreiche Impulse für interessante Fachdiskussionen, aus denen sich vielfältige Anregungen für die Qualitätsentwicklung von Unterricht und Erziehung ergeben werden.

Literaturverzeichnis

- BAUMERT, J. ET AL. (2001). Programme for International Student Assessment. Zielsetzung, theoretische Konzeption und Entwicklung von Messverfahren. In: WEINERT, F. (Hg.). Leistungsmessung in Schulen. Weinheim und Basel (Beltz)
- BOS, W. ET AL. (2003). Erste Ergebnisse aus IGLU. Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. Münster (Waxmann)
- DEUTSCHES PISA-KONSORTIUM (Hrsg.) (2001). PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen (Leske & Budrich)
- DUISMANN, G.H. (2002). Arbeitsorientierte Basiskompetenzen. Ein neues Konzept zur Qualitätssicherung. In: Unterricht – Arbeit und Technik 4(2002)15, 50 - 52
- DUISMANN, G.H. / ESSER, R. / HASEMANN, K. / HÜBNER, P. / MESCHENMOSER, H. (2001). Qualitätsentwicklung beim Übergang Schule in den Beruf. In: Sonderpädagogik in Berlin. Ausgabe 4, 2001, S. 26-45
- DUISMANN, G.H. / MESCHENMOSER, H. (2003). Zum Umgang mit den LABEL-Rückmeldungen. Senatsverwaltung für Schule, Jugend und Sport Berlin (Hrsg.)
- HORSTKÄMPER, M. (1987). Schule, Geschlecht und Selbstvertrauen. Eine Längsschnittstudie über Mädchensozialisation in der Schule. Weinheim und München
- KLIEME, E. ET AL. (2003). Expertise zur Entwicklung nationaler Bildungsstandards. Bundesministerium für Bildung und Forschung (Hrsg.)
- LEHMANN, R. ET AL. (1995). Leseverständnis und Lesegewohnheiten deutscher Schüler und Schülerinnen. Weinheim und Basel (Beltz)
- LEHMANN, R. & PEEK, R. (1997). Aspekte der Lernausgangslage von Schülerinnen und Schülern der fünften Klassen an Hamburger Schulen. Bericht über die Untersuchung im September 1996. Herausgegeben von der Behörde für Bildung und Sport. Hamburg
- LEHMANN, R. ET AL. (1999). Aspekte der Lernausgangslage und der Lernentwicklung von Schülerinnen und Schülern an Hamburger Schulen – Klassenstufe 7. Herausgegeben von der Behörde für Bildung und Sport. Hamburg
- LEHMANN, R. ET AL. (2002). Aspekte der Lernausgangslage und der Lernentwicklung – Klasse 9. Herausgegeben von der Behörde für Schule, Jugend und Berufsbildung. Hamburg
- MIETZEL, G. & WILLENBERG, H. (2000). Hamburger Schulleistungstest für 4. und 5. Klassen – HST 4/5 . Göttingen (Hogrefe)
- OECD (1995). Grundqualifikationen, Wirtschaft und Gesellschaft. Ergebnisse der ersten internationalen Untersuchung von Grundqualifikationen Erwachsener. Statistics Canada
- ROST, J. (1996). Lehrbuch Testtheorien – Testkonstruktion. Bern (Hans Huber)
- STRYCK, T. & LEHMANN, R. (2003). Erhebungen zum Lese- und Mathematikverständnis-Entwicklung in den Jahrgangsstufen 4-6 in Berlin (ELEMENT).
- WEISS, R. H. (1998): Culture Fair Intelligence Test – CTF 20

Glossar

Alpha

Alpha (genauer: „Cronbachs Alpha“) ist eine Kennzahl für die interne Konsistenz einer \Rightarrow Skala. Werden inhaltlich zusammenpassende Aufgaben oder Fragen zu einer Skala zusammengefasst, z. B. 10 Fragen aus dem Schülerfragebogen zur „Schulzufriedenheit“, so gibt Alpha Auskunft darüber, wie gut sich die Antworten zu einer Gesamttendenz zusammenfügen. Alpha kann maximal den Wert 1 annehmen. Als Faustregel gilt, dass bei Skalen ab $\alpha = 0,75$ eine befriedigende und ab $\alpha = 0,85$ eine gute Skalenqualität vorliegt.

arithmetisches Mittel, arithmetischer Mittelwert, Durchschnittswert

\Rightarrow Mittelwert.

Beta(-Gewicht)

\Rightarrow Regressionsanalyse.

d, Effektstärke *d*, standardisierte Effektstärke *d*

ist ein standardisiertes Maß für Merkmalsunterschiede zwischen zwei Gruppen. Sie wird berechnet, indem die Differenz der \Rightarrow Mittelwerte der Gruppen durch die gemeinsame \Rightarrow Standardabweichung dividiert wird.

erklärte Varianz \Rightarrow Kovarianzanalyse,

\Rightarrow Regressionsanalyse, \Rightarrow Varianzanalyse

*Eta*², Bestimmtheitsmaß *Eta*²

ist die Maßzahl für die erklärte Varianz bei der \Rightarrow Kovarianzanalyse und der \Rightarrow Varianzanalyse. *Eta*² bezeichnet den Varianzanteil, den man erhält, wenn man die Einzelwerte durch den jeweiligen Gruppenmittelwert ersetzt, daraus die Varianz berechnet und durch die ursprüngliche Varianz teilt; es variiert zwischen 0 und 1. Z. B. gibt das mit der Schulform verbundene *Eta*² für ein Leistungsmerkmal Auskunft darüber, welchen Anteil Schulformunterschiede (Mittelwertdifferenzen) an den Leistungsunterschieden zwischen den Schülerinnen und Schülern überhaupt haben.

interne Konsistenz

\Rightarrow Alpha.

intervallskalierte Variable

Wenn Gleichheit der Abstände zwischen zwei benachbarten Werten einer Variable angenommen werden darf, spricht man von einer intervallskalierten Variable. Ist dies nicht der Fall, liegt eine nominalskalierte (keine Rangfolge der Werte) oder eine ordinalskalierte (Rangfolge der Werte, aber keine Abstandsgleichheit) Variable vor.

Item

⇒ Variable

Koeffizient

Ein Koeffizient ist ein statistischer, ein mathematischer Kennwert. ⇒ Pearsons r ist z.B. ein Korrelationskoeffizient, d. h. ein statistisches Zusammenhangsmaß.

Korrelation, korrelieren

Zusammenhang zwischen zwei ⇒ Variablen (Merkmalen). Für ⇒ intervallskalierte Daten ist das Korrelationsmaß der Pearsonsche Produkt-Moment-Korrelationskoeffizient r (kurz „Pearsons r “ oder nur „ r “). Er variiert zwischen minus eins und plus eins. Ein hohes negatives r besagt: Je höher das eine Merkmal ausgeprägt ist, desto niedriger das andere Merkmal, und je niedriger das eine Merkmal, desto höher das andere Merkmal. Ein hohes positives r besagt sinngemäß entsprechend: Je höhere Werte das eine Merkmal annimmt, desto höhere auch das andere (bzw. je niedriger, desto niedriger). Ein r nahe Null sagt aus, dass zwischen den beiden Merkmalen kein Zusammenhang besteht. r^2 gibt direkt die so ⇒ erklärte Varianz an.

Kovarianzanalyse

Die Kovarianzanalyse ist eine Mischform aus ⇒ Regressionsanalyse und ⇒ Varianzanalyse. Die abhängige Variable wird durch unabhängige Variablen erklärt, von denen mindestens eine wie bei der Varianzanalyse nominalskaliert ist (der sog. „Faktor“), andere dagegen wie bei der Regressionsanalyse intervallskaliert (Kovariaten). Die Maßzahl für die insgesamt erklärte Varianz ist ⇒ R^2 ; für den Faktor allein kann auch ein ⇒ Eta^2 bestimmt werden.

Kriterium(s-Variable)

abhängige Variable, ⇒ Regressionsanalyse.

Lösungswahrscheinlichkeit

Die Lösungswahrscheinlichkeit einer Aufgabe gibt an, wie groß die Wahrscheinlichkeit ist, dass irgend ein Schüler bzw. irgendeine Schülerin diese Aufgabe löst. Die Lösungswahrscheinlichkeit wird mit dem Wert p (vom englischen „probability“) angegeben und liegt zwischen 0 und 1. Eine Lösungswahrscheinlichkeit von $p = 0,47$ beispielsweise besagt, dass „auf lange Sicht“ 47 Prozent der Schülerinnen und Schüler einer definierten Gruppe diese Aufgabe lösen.

Mittelwert

Kurzbezeichnung für den arithmetischen Mittelwert. Der Mittelwert darf berechnet werden, wenn mindestens \Rightarrow intervallskalierte Daten vorliegen. Er ist die Summe der Einzelwerte aller Fälle dividiert durch die Fallzahl.

multiple Regressionsanalyse

\Rightarrow Regressionsanalyse.

multiple R , multiple R^2

\Rightarrow Regressionsanalyse.

Objektivität

Objektivität ist ein Gütekriterium für sozialwissenschaftliche Messungen. Objektivität in Schulleistungsuntersuchungen ist gegeben, wenn für alle Schülerinnen und Schüler gleiche Aufgabenstellungen, Bearbeitungszeiten, Erläuterungen der Aufgaben, Arbeitsmaterialien u. ä. gelten; die Auswertung und die Interpretation erfolgen nach klaren Kriterien, die unabhängig von der Person des Auswerters sind.

Prädiktor(variable)

unabhängige Variable, \Rightarrow Regressionsanalyse.

r Abkürzung für Pearsons r , \Rightarrow Korrelation.

Regressionsanalyse

Die (multiple) Regressionsanalyse ist ein Analyseverfahren, das den Zusammenhang zwischen einer \Rightarrow intervallskalierten abhängigen (zu erklärenden) Variable (dem sogenannten Kriterium) und mehreren, ebenfalls intervallskalierten unabhängigen (erklärenden) Variablen (den sogenannten Prädiktoren) aufdeckt. Bei der Berechnung der Regressionsgleichung werden die \Rightarrow Korrelationen der Prädiktoren untereinander berücksichtigt. Die Maßzahl für den Zusammenhang zwischen allen Prädiktoren ei-

nerseits und dem Kriterium andererseits ist das „multiple R“. Es kann wie \Rightarrow Pearsons r interpretiert werden, kann allerdings keine negativen Werte annehmen. Das quadrierte multiple R entspricht der erklärten Varianz. Der eigenständige Beitrag jedes einzelnen Prädiktors (bei Konstanthaltung der anderen Prädiktoren) zur Aufklärung der Unterschiede im Kriterium wird mit den \Rightarrow Beta-Gewichten angegeben.

Reliabilität

Reliabilität ist ein Gütekriterium für sozialwissenschaftliche Messungen. Sie entspricht der Zuverlässigkeit einer Messung. Reliabel ist ein Test oder eine Skala, wenn nur geringe Messfehler auftreten. Es gibt verschiedene Methoden, die Reliabilität zu überprüfen, z. B. \Rightarrow Cronbachs Alpha.

Skala

1. Kurzbezeichnung für die Ausprägungen einer Einschätzskala (Ratingskala). Bei der Vorgabe der Antwortmöglichkeiten von 1 = „trifft völlig zu“ bis 4 = „trifft überhaupt nicht zu“ im Schülerfragebogen spricht man z. B. von einer vierstufigen Skala.
2. Inhaltlich zusammenfassende Einzelitems können, z. B. durch Aufsummieren oder Mittelwertbildung, zu einer Skala zusammengefasst werden. Ein Beispiel ist die Skala „Schulzufriedenheit“, die auf dem Wege der individuellen Mittelwertbildung über 19 Fragen aus dem Fragebogen für Schülerinnen und Schüler ermittelt wurde.

Standardabweichung, SD

Die Standardabweichung ist ein sogenanntes Streuungsmaß, das für intervallskalierte Daten Auskunft darüber gibt, wie homogen oder heterogen eine Merkmalsverteilung ist. Je kleiner die Standardabweichung ist, desto enger gruppieren sich die Werte der einzelnen Fälle um den Mittelwert, je größer sie ist, desto weiter streuen sie um den Mittelwert.

Liegt eine Normalverteilung vor, so lässt sich über die Verteilung folgendes sagen: Im Bereich Mittelwert plus/minus eine Standardabweichung liegen etwa 68 Prozent der Fälle; im Bereich Mittelwert plus / minus zwei Standardabweichungen liegen etwa 95 Prozent der Fälle.

Streuung

\Rightarrow Standardabweichung.

Validität

Validität ist ein Gütekriterium für sozialwissenschaftliche Messungen. Validität gibt die Gültigkeit eines Messinstruments, z. B. eines Tests, an.

Ein Test ist valide, wenn er das misst, was er zu messen vorgibt. Validität lässt sich an Außenkriterien überprüfen; z. B. lassen sich die Aufgaben der fachbezogenen Leistungstests mit den Curricula der jeweiligen Schulform und Klassenstufe validieren (curriculare Validität).

Variable

Ein erhobenes Merkmal, das mehrere Ausprägungen haben kann, wird als Variable bezeichnet, z. B. Alter oder erreichter Rasch-Skalenwert im Mathematiktest. Die Ausprägungen nennt man auch Merkmalsausprägungen.

Varianz

Die Varianz ist das Quadrat der \Rightarrow Standardabweichung. Mathematisch ist die Varianz der Durchschnitt aus den quadrierten Abweichungen aller Einzelwerte vom Gesamtmittelwert.

Varianzanalyse

Dieses Analyseverfahren ist mit der \Rightarrow Regressionsanalyse verwandt. Auch hier geht es um die Aufklärung individueller Unterschiede in einer abhängigen Variable durch eine oder mehrere unabhängige Variablen. Der Hauptunterschied zur Regressionsanalyse besteht darin, dass die unabhängigen Variablen nicht \Rightarrow intervallskaliert zu sein braucht. Die Maßzahl für die erklärte Varianz heißt \Rightarrow Eta^2 .