

Was messen internationale Schulleistungsstudien?

Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz?

Heiner Rindermann

Zusammenfassung. Über Aufgabenanalysen und die Analyse empirischer Beziehungen lässt sich bestimmen, was Tests wie die der internationalen Schulleistungsstudien (PISA, IGLU, TIMSS u. a.) messen. Inhaltliche Aufgabenanalysen belegen, dass Aufgaben unterschiedlicher Skalenzuordnung einander sehr ähnlich sind: Leseaufgaben enthalten Tabellen und Grafiken, Naturwissenschaftsaufgaben umfangreiche Lesetexte, Mathematikitems mit viel Text versehene alltagsnahe Denkaufgaben. Problemlösen ist konzeptuell am geringsten von Intelligenz zu unterscheiden. Zur Lösung der Aufgaben sind in unterschiedlichem Maße innerhalb der Bereiche Denken, Allgemeinwissen und Schulwissen notwendig. Empirische Beziehungen in Form von Korrelationen der Skalen untereinander, mit Intelligenztests und von Faktorenanalysen lassen sich in der Literatur nur schwierig oder gar nicht finden. Skaleninterkorrelationen bei Individualdaten liegen weit über den gängigen Skaleninterkorrelationen bei Intelligenztests. Skaleninterkorrelationen auf Ebene von Staaten sind extrem hoch; Korrelationen zwischen verschiedenen Studien und mit Intelligenztestergebnissen auf Ebene von Staaten sind hoch. Es finden sich selten Belege für diskriminante Validität. Alles spricht für die Messung eines *g*-Faktors kognitiver Fähigkeiten. Auf makrosozialer Ebene ist eine Unterscheidung von Intelligenz empirisch nicht möglich.
Schlüsselwörter: Schulleistungsstudien, Schulleistungstests, schulnahe kognitive Fähigkeiten, Wissen, Intelligenz, OECD

What do international student assessment studies measure? School performance, student abilities, cognitive abilities, knowledge or general intelligence?

Abstract. What tests measure, like tests of international student assessments (PISA, PIRLS, TIMSS), is possible to describe by analysis of tasks and analysis of empirical relationships. Results of content-based task analysis show that items of different scales are very similar: Reading literacy items contain tables and graphs, science items extensive texts, math items common cognitive tasks with long texts. Conceptually most difficult to separate from intelligence is problem solving. Thinking, general knowledge, and school knowledge are necessary in different extent in each domain for solutions. Empirical relationships in form of correlations between used scales, with intelligence tests and results of factor analysis are difficult or even not possible to find in published reports. Correlations between scales on level of individual data are higher than known correlations between scales of intelligence tests. Correlations between scales on level of state data are extremely high; correlations between studies and with results of intelligence tests on level of state data are high. Discriminant validity is rarely found. All results indicate a strong *g*-factor of cognitive abilities. On level of macrosocial data a separation from intelligence is empirically not possible.

Key words: International student assessment, educational achievement, school-near cognitive abilities, knowledge, intelligence, OECD

Die OECD- und IEA-Schulleistungsstudien der letzten Jahre – PISA, IGLU und TIMSS – haben für großes Aufsehen gesorgt. Insbesondere in Deutschland, das in den 70er und 80er-Jahren kaum an internationalen Studien teilnahm, führten die Vergleiche mit anderen Staaten und zwischen Bundesländern zu öffentlicher Aufmerksamkeit. Spätestens PISA 2000 hat in der Bildungspolitik und in der Forschung einen Mentalitätswandel hervorgerufen. Schulsysteme müssen sich nun auch hinsichtlich ihrer Ergebnisse beurteilen lassen und wissenschaftlich gelten „Schulleistungen“ inhaltlich wie Forschungsorganisatorisch als ein höchst relevantes Thema (Weinert, 2001 a).

Dabei ist das, was in diesen Studien im Zentrum des Interesses steht, die Messung von „Schulleistungen“,

begrifflich, inhaltlich, psychometrisch und hinsichtlich empirischer Dimensionalität und Validität unklar. Wenn man unter „Schulleistungen“ Leistungen der Schule verstehen würde, wäre der Begriff auf jeden Fall falsch gewählt, da Leistungstestergebnisse von Schülern nicht nur von der Schule, sondern auch von vielen weiteren Faktoren abhängen, etwa von Ernährung, von Erziehung durch die Eltern und von privater Lernzeit (z. B. Armor, 2003). Interindividuelle Unterschiede sind vornehmlich auf familiäre Faktoren umweltlichen oder (im Detail unbekannt) genetischen Ursprungs zurückführbar. Ernährung, Gesundheit, Schule, Erziehung und Lernzeit wiederum hängen von allgemeinen politischen, wirtschaftlichen und gesellschaftlichen Hintergrundfaktoren ab, nicht zuletzt in der Form weltanschaulich-religiös-kultureller Orientierung.

gen (z. B. Max Weber, 1988/1920; Lenhart, 1998). Zudem wird nie die Leistung der Schule gemessen, sondern immer die von Schülern; wenn von Leistung, dann müsste man von *Schülerleistung* sprechen (Weinert, 2001 b; „student performance“, OECD, 2002, S. 3). Doch ein Leistungsergebnis ist für sich stehend zunächst wissenschaftlich wenig interessant und wird erst in einer Interpretation als Indikator für eine inhaltlich näher bestimmte *Fähigkeit* bedeutsam.

Leistung als empirisches Datum stellt zunächst das Ergebnis eines Leistungstests oder in einer lebensweltlich bedeutsamen Aufgabe dar. Solche Ergebnisse stehen für Fähigkeiten. Doch für welche? Naheliegend ist bei Tests, die (auch) schulische Themen beinhalten, von *schulischen Fähigkeiten* zu sprechen. Doch auch *Intelligenztests*, wie etwa der von Heller und Perleth (2000) in Anlehnung an den CogAT von Thorndike entwickelte Kognitive Fähigkeitstest, enthalten durch die Schule vermittelte Themen und werden vereinzelt sogar – wie von Ditton und Kreckler (1995, S. 511) – als „Schulleistungstests“ verwendet. Intelligenzforscher wiederum nutzen mitunter „Schulleistungstests“ als Intelligenztests, etwa den SAT.

Verschiedene Autoren zweifeln kategoriale Unterschiede zwischen Schulleistungstests und Intelligenztests in Inhalten und empirisch bestimmbaren Validitäten an. Aus der Perspektive schullaffiner Forscher wie Ceci (1991, S. 703 u. 708) oder Cole (1995) wird in Frage gezogen, ob Intelligenztests etwas anderes als schulisch-kognitive Fähigkeiten messen, was sich gut durch ihre gemeinsame Abhängigkeit vom Schulbesuch und durch die Ähnlichkeit der Messverfahren belegen lässt. Aus der Perspektive intelligenzaffiner, insbesondere g-naher Forscher wie Armor (2003, S. 18 ff.), Gottfredson (2003), Herrnstein und Murray (1994), Jensen (1989), Lehl (2005), Rost (2005) und – im Ausdruck in dieser Sache etwas drastisch – Weiss (2002) wird unter Bezug auf die hohen Korrelationen von einer Identität oder zumindest hohen Überlappung mit Intelligenz gesprochen. Linda Gottfredson zieht für den Beleg einer Äquivalenz von Intelligenz und schulnahen kognitiven Fähigkeiten das US-amerikanische National Adult Literacy Survey (NALS) heran, das ähnlich wie PISA Leseverständnis erhebt. NALS messe ihrer Meinung nach *g* oder verbales *g*, da es den Umgang mit komplexer verbaler Information, schlussfolgerndes Denken und Informationsverarbeitung erhebe, seine verschiedenen Unterdimensionen untereinander hoch korreliert seien und zudem die prognostische Validität ähnlich die der Intelligenz sei (Berufe, Einkommen).

Das Erkennen einer möglichen Ähnlichkeit oder gar Identität der Konzepte wurde bislang aber durch unterschiedliche Fächertraditionen (einschließlich Methoden, Zeitschriften und Tagungen) erschwert, umso erstaunlicher, dass kontextualistische Entwicklungspsychologie und psychometrische Persönlichkeitspsychologie zu gleichen Einschätzungen gelangten.

Allerdings ist nicht alles, was *g* ist, also auf einem Generalfaktor lädt, meist der erste unrotierte Faktor einer Fak-

torenanalyse, mit Intelligenz gleichzusetzen. *g* ist nur ein rein statistisches Ergebnis irgendwelcher Daten und wäre z. B. auch in einem Sportleistungstest mit Zeitmessungen verschiedener Distanzen in Laufen, Schwimmen und Fahrradfahren erzielbar. Eine inhaltliche Definition von Intelligenz ist somit unverzichtbar. Unter (kognitiver) *Intelligenz* soll nun verstanden werden: die komplexe Fähigkeit zum *denkgestützten Lösen von Aufgaben und Problemen* in Situationen, die für die Person *neu* und nicht allein durch Wissensabruf erfolgreich bearbeitbar sind, die Fähigkeit zum induktiv und deduktiv-logisch *schlussfolgernden Denken*, die Fähigkeit zum *abstrakten Denken* und die Fähigkeit zu *Verständnis* und *Einsicht* – zum *Erkennen* und zur *Herstellung* von Strukturen, Beziehungen, Sinnzusammenhängen und Bedeutungen (s. a. Rindermann, 2004). Kurz: *Intelligenz ist die Fähigkeit zum Denken*.

Fragestellung

Untersucht werden soll, was international vergleichende Schulleistungstudien messen. Hierzu werden zunächst die Konstruktbeschreibungen der Studien analysiert und anschließend Aufgaben inhaltsanalytisch ausgewertet. Über Korrelationen auf verschiedenen Datenebenen (Individualdaten, Staaten, deutsche Bundesländer) zwischen Dimensionen und Subdimensionen einer Studie und makrosozial zwischen Dimensionen und Gesamtwerten verschiedener Studien und im Vergleich mit Intelligenztests und Bildungsindikatoren wird die These aufgestellt, dass eine empirisch einheitliche und bildungsabhängige kognitive Fähigkeit erfasst wird.

Konstrukte der internationalen Schulleistungstudien

In den internationalen Studien werden drei explizit schulbezogene Dimensionen unterschieden: *Lesekompetenz* (PISA 2000 und 2003, IGLU 2001, IEA-Lesen 1990-1991), *mathematische Grundbildung, Fachleistung* oder *Kompetenz* (PISA 2000 und 2003 und TIMSS 1994-1995) und *naturwissenschaftliche Grundbildung, Fachleistung, Verständnis* oder *Kompetenz* (PISA 2000 und 2003 und TIMSS 1994-1995), in PISA 2003 kam noch die eher als schuldistant betrachtete Skala *Problemlösen* hinzu, in IGLU 2001 innerhalb Deutschlands *Orthografie*. Im Englischen steht hinter den Skalen jeweils der Begriff „literacy“, manchmal wird auch von „knowledge“, „skills for life“ oder „literacy skills“ gesprochen. „Literacy“ beschreibt funktionalistisch verstandene „Basiskompetenzen, die in modernen Gesellschaften für eine befriedigende Lebensführung in persönlicher und wirtschaftlicher Hinsicht sowie für eine aktive Teilnahme am gesellschaftlichen Leben notwendig sind.“ (DPK, 2001, S. 78) Die Konzepte verschiedener Studien ähneln einander, im Gegensatz zu TIMSS sind PISA und IGLU aber nicht curriculumsbezogen und entsprechen in ihrer Zielsetzung nicht klassischen, curricular validen Schulleistungstests. Die Beschreibungen nach PISA werden fokussiert.

Unter *Lesekompetenz* wird Folgendes verstanden: „Verstehender Umgang mit Texten“ (DPK, 2001, S. 79), Verständnis verschiedener Formen von Text, von Tabellen über Anzeigen bis zu Erzählungen; Subskalen: Informationsermittlung, textbezogene Interpretation sowie Reflexion-Bewertung. Nach OECD (2003 a, S. 37): „Reading literacy is defined in PISA as the ability to understand, use and reflect on written texts in order to achieve one’s goals, to develop one’s knowledge and potential, and to participate effectively in society.“

Mathematische Grundbildung: „Mathematical Literacy wird in aller Knappheit als die Fähigkeit definiert, die Rolle, die Mathematik in der Welt spielt, zu erkennen und zu verstehen, begründete mathematische Urteile abzugeben und sich auf eine Weise mit der Mathematik zu befassen, die den Anforderungen des gegenwärtigen und künftigen Lebens einer Person als eines konstruktiven, engagierten und reflektierenden Bürgers entspricht“ (DPK, 2001, S. 141, auf Englisch OECD, 2004 a, S. 37). PISA 2000 unterschied in Deutschland die Inhaltsbereiche begriffliches Modellieren, rechnerisches Modellieren und technische Fertigkeiten, PISA 2003 die Inhalte Quantität, Veränderung und Beziehung, Raum und Form sowie Unsicherheit; als Kompetenzen Reproduktion, Verbindung und Reflexion. Bei TIMSS gab es Subskalen unterschieden nach mathematischen Bereichen (Zahlenverständnis, Geometrie, Algebra, Wahrscheinlichkeit, Messen, Proportionalität), und nach Anforderungen (Wissen, Routineverfahren, Denken, Kommunizieren).

Naturwissenschaftliche Grundbildung: „Naturwissenschaftliche Grundbildung (Scientific Literacy) ist die Fähigkeit, naturwissenschaftliches Wissen anzuwenden, naturwissenschaftliche Fragen zu erkennen und aus Belegen Schlussfolgerungen zu ziehen, um Entscheidungen zu verstehen und zu treffen, welche die natürliche Welt und die durch menschliches Handeln an ihr vorgenommenen Veränderungen betreffen“ (DPK, 2004, S. 112, Englisch: OECD, 2003 b, S. 15 u. 133). Drei oder vier Subskalen werden in verschiedenen Varianten bei PISA unterschieden: Konzepte (Begriffe, Wissen, Verständnis), Prozesse (Untersuchungsmethoden und Denkweisen) und Situationen (Vorstellungen über die Besonderheit der Naturwissenschaft und Vorstellungen über die Beziehungen zwischen Naturwissenschaften, Technik und Gesellschaft); auch werden genannt: Biologie, Chemie, Physik; Verständnis naturwissenschaftlicher Untersuchungen, Umgang mit Evidenz, Kommunikation naturwissenschaftlicher Beschreibungen und Argumente, Verstehen naturwissenschaftlicher Konzepte; Grafiken, mentale Modelle, konvergentes Denken, divergentes Denken, Verbalisieren, Zahlen, Bewerten. Bei TIMSS gab es Subskalen nach Fächern unterschieden (Biologie, Chemie, Physik) und nach Anforderungen (Verstehen, Kommunizieren).

Problemlösen: „Fähigkeit einer Person, kognitive Prozesse zu nutzen, um sich mit solchen realen, fächerübergreifenden Problemstellungen auseinanderzusetzen und sie zu lösen, bei denen der Lösungsweg nicht unmittelbar erkennbar ist und die zur Lösung nutzbaren Wissensbe-

reiche nicht einem einzelnen Fachgebiet der Mathematik, der Naturwissenschaften oder des Lesens entstammen“ (DPK, 2004, S. 148, Englisch: OECD, 2004 b, S. 26). Unterschieden werden nach Problemstellungen: Entscheidungen treffen, Systeme analysieren und entwerfen, Fehlersuche; nach Kontexten Privat-Persönliches, Arbeit und Freizeit, Gesellschaft. In Deutschland wurde in Ergänzung des international mittels Papier-Bleistift-Tests gemessenen „analytischen Problemlösens“ noch über Computersimulationen „dynamisches Problemlösen“ im Sinne Dörners erhoben (DPK, 2004, S. 162 ff.).

Bei der Beschreibung der Dimensionen fallen verschiedene Aspekte auf, die Skepsis hinsichtlich einer Differenzierbarkeit der genannten Dimensionen aufkommen lassen:

1. Die Definitionen der Konstrukte wirken im Vergleich zu sonst in der Diagnostik üblichen Definitionen global und diffus („to achieve one’s goals, to develop one’s potential, to participate effectively in society, die Rolle, die Mathematik in der Welt spielt, zu erkennen, konstruktive, engagierte und reflektierende Bürger, Kommunikation naturwissenschaftlicher Beschreibungen“; s. o.). Begriffe schwanken (Grundbildung, Fachleistung, Verständnis, Kompetenz), die Aufteilung in Subskalen variiert. Die Beschreibungen wirken manchmal wie das Ergebnis eines politischen Aushandlungsprozesses.

2. Bei den Definitionen wird Ähnlichkeit zwischen verschiedenen Inhaltsdimensionen und zur Intelligenz offensichtlich („verstehen, reflect, erkennen, verstehen, Reflexion, erkennen, Schlussfolgerungen ziehen, verstehen, Verständnis“).

3. Sonst übliche psychometrische, statistische und testbezogene Informationen sind unzureichend dargestellt, d. h. sie fehlen, sind schwer zu finden, sind erst über Nachfragen oder auch dann nicht erhältlich. Z. B. in den deutschen Bänden zu PISA 2000, TIMSS und IGLU die Korrelationen zwischen den Skalen; Reliabilitäten; es fehlen manifeste Korrelationen; es gibt keine Darstellung von Skalenhomogenitäten und Korrelationen zwischen allen Subskalen und Items; Reliabilitätsmessungen sind unklar oder fehlen; bei Pfadmodellen fehlen die Angaben standardisiert/unstandardisiert, der Fehler und von Korrelationen; wird Intelligenz gemessen („kognitive Grundfähigkeit“, DPK, 2001, S. 129 f.), sind Testverfahren nicht angegeben; vor allem fehlen Faktorenanalysen aller Skalen zusammen (insbesondere Varianzaufklärung des ersten unrotierten Faktors). Zwar wären die Analysen auf Grund des bei jedem Schüler nur teilweise eingesetzten Aufgabensatzes schwierig (Multimatrixdesign), über Schätzungen oder vor allem Unterskalen als Variablen wären aber Berechnungen möglich. Aus wissenschaftlichen Gründen hätten in Teilstichproben alle Aufgaben eingesetzt werden können, zumal die Gesamtbearbeitungszeit von 6,5 Zeitstunden – PISA 2000 (DPK, 2001, S. 49) – noch auf zwei Testtage verteilbar ist und Vorstudien mit etwas umfangreicheren Aufgabensätzen durchgeführt wurden (OECD, 2002, S. 109). Das Nächstliegende, die

Berechnung einer Faktorenanalyse, fehlt. So sind die Tests nur schwierig bewertbar.¹

Wie lässt sich bestimmen, was Tests messen?

Zwei Strategien liegen nahe zur Feststellung dessen, was Tests kognitiver Fähigkeiten messen:

1. *Aufgabenanalyse*: Man untersucht die Aufgaben eines Tests und schließt über deren Inhalte, die Anforderungen, die sie stellen, und über die für die Lösung notwendigen kognitiven Prozesse auf den Messgegenstand (z. B. Carpenter, Just & Shell, 1990).

2. *Analyse empirischer Beziehungen*: Was mit dem Testergebnis in einer aussagekräftigen Stichprobe korreliert, misst der Test oder misst er mit. Innerhalb eines Tests lässt sich durch Faktorenanalysen die Dimensionalität überprüfen.

Eine bloße Analyse empirischer Beziehungen (Korrelationen) reicht zur Bestimmung eines Messkonstruktes nicht aus, sonst ließe sich etwa behaupten, über die Messung der Haarlänge inhaltlich das Geschlecht oder gar räumliches Vorstellungsvermögen zu erfassen. Auf das Geschlecht wird aber nur *schlussfolgert*, gemessen wird ausschließlich die Haarlänge. So messen Tests kognitiver Fähigkeiten *indirekt* auch Persönlichkeitsmerkmale (Fleiß, Anstrengungsbereitschaft), Bildungsorientierung der Eltern, Bildungsnähe des Getesteten, Länge und Qualität der Schulbildung, Sozialstatus, Berufstätigkeit, ethnische und kulturelle Zugehörigkeit usw., selbst ein reiner Wissenstest könnte auch Intelligenz messen. Empirische Validitätsbestimmungen sind stärker von spezifischen Stichprobenzusammensetzungen abhängig. Zudem ist bei korrelativen Beziehungen oft die Art der Merkmal-Indikator-Beziehung unklar (warum gibt es eine Korrelation?). Empi-

¹ Es ist prinzipiell nicht möglich nachzuweisen, dass etwas nicht existiert. Die gesuchten Informationen waren bei üblicher wissenschaftlicher Informationssuche und mehrtausendseitiger Rezeption deutscher und englischer Texte nicht findbar, oder, falls sie auffindbar waren, waren sie in Endnoten oder in Unterkapiteln mit nicht auf diese Information verweisender Überschrift. Im deutschen 2003er Band zu PISA sind Korrelationen in Tabellen angegeben, die Reliabilitäten fehlen aber. In den IGLU-Bänden finden sich keine Reliabilitäten, Korrelationen und Faktorenanalysen für die vier Skalen (Faktorenanalyse: nicht für alle Subskalen zusammen). Bei Nachfragen wurden dem Autor meist Informationen, falls vorhanden, zugesandt. Vielen Dank! Auf diesen Nachfragen basieren im Wesentlichen die Informationen zu Individualdaten in PISA 2000 und TIMSS. Faktorenanalysen für die verschiedenen Bereiche zusammen fehlen, es werden aber des öfteren Faktorenanalysen innerhalb einzelner Leistungsskalen und vor allem für Unterrichts- und Persönlichkeitsskalen berichtet, ebenso eine MDS (z. B. DPK, 2004, S. 168). Falls Faktorenanalysen benutzt wurden, fehlen Informationen zu Abbruchkriterien, zur Varianzaufklärung der Faktoren und zu numerischen Ladungsmustern; konfirmatorische Faktorenanalysen, die naheliegen, sucht man vergebens (Ausnahme: in IGLU nur innerhalb Lesen). Faktorenanalysen sollten nicht auf Itemebene, sondern auf Ebene von Subskalen gerechnet werden. Bei Nachfragen wurde geantwortet, dass Faktorenanalysen nicht einfach (Multimatrixdesign) und nicht sinnvoll seien; nur: für Lesekompetenz etwa gibt es solche (bei IGLU explorativ und konfirmatorisch; s. u.).

rische Analysen müssen immer durch *inhaltliche* Aufgabenanalysen ergänzt werden.

Aufgabenanalysen

Methodisches Vorgehen

In den Publikationen und auf den Netzseiten der an PISA, IGLU und TIMSS beteiligten Institutionen werden Aufgaben vorgestellt. Eine eingehende Aufgabenanalyse müsste den Umfang einer Monografie annehmen (und auch die empirische Untersuchung der Lösungswege von getesteten Schülern einschließlich von Protokollen lauten Denkens beinhalten; vgl. Waldmann & Weinert, 1990). Aus Platzgründen wird nur eine Aufgabe je Fähigkeitsdimension näher behandelt, weitere werden kurz skizziert. Aufgaben beziehen sich, falls nicht anders angegeben, auf PISA 2000. Aufgaben werden inhaltlich beschrieben, anschließend deren Anforderungen skizziert und auf kognitive Fähigkeiten bezogen.

Lesen

Bekannt ist die Aufgabe „Tschadsee“ (s. a. Schnotz & Dutke, 2004): Hier ist eine Verlaufsgraphik mit Wasserständen zwischen ca. 11000 v. Chr. und 1300 n. Chr. abgebildet, separat eine Zeittafel, auf der erkennbar ist, wann welche Tiere auf Felsmalereien auftauchen (z. B. Nilpferd vs. Kamel). Gefragt wurde nach der heutigen Tiefe des Tschadsees, nach dem zeitlichen Beginn auf dem Diagramm, warum gerade zu dieser Zeit mit der Darstellung begonnen wurde, warum die Zeittafel mit den Tieren beigegeben wurde und in welchem Zusammenhang Wasserstand und Tiervorkommen stünden.

Diese Aufgabe hat mit Lesen nicht spezifischer zu tun als viele Aufgaben, die unter Mathematik oder Naturwissenschaften aufgelistet werden (Informationen aus Grafiken, Tabellen und Texten entnehmen). Inhaltlich ist sie eher den Naturwissenschaften zuordenbar. Alle Aufgaben bei „Tschadsee“ sind mit den gegebenen Informationen lösbar, zur Lösung müssen Abbildungen verstanden, einem Text Informationen entnommen und Schlussfolgerungen gezogen werden und wie bei allen anderen Aufgaben der verschiedenen Skalen ist die Beherrschung der Testsprache notwendig.

Andere Aufgaben unter Lesekompetenz sind ähnlich, es gibt aber auch Aufgaben, in denen spezifisch sprachliche Aspekte wie eine Unterscheidung von Inhalt und Stil thematisiert werden („Grippe“, „Graffiti“). Zum Teil kann Vorwissen wie bei „Grippe“ eine richtige Lösung behindern, weil sich die Lösung ausschließlich auf die gegebenen Informationen stützen soll. In der Aufgabe „Struktur der erwerbstätigen Bevölkerung“ sind in einem Baumdiagramm absolute Zahlen und Prozentsatzangaben mit jeweils unterschiedlicher Prozentuierungsgrundlage angegeben. Dies sind mathematische Inhalte, es muss aber ähnlich wie bei Aufgaben mit Zahlen in Tabellen nicht gerechnet, sondern nur interpretiert werden. In weiteren

Leseaufgaben ist deutlich mehr Text enthalten (dafür keine Abbildungen und Tabellen, z. B. in „Geschenk“), noch expliziter als bei „Tschadsee“ müssen in „Turnschuhe“ und „Geschenk“ Motivationen von Akteuren (der Text- oder Aufgabenautoren, von Personen in Texten) erschlossen werden. In einer Aufgabe („Amanda“) sind Informationen aus einem Text in eine räumliche Repräsentation umzusetzen.

Ähnliche Aufgaben (aber deutlich kürzer) finden sich in IGLU-Lesen: Texte sind zu lesen, Abbildungen (mit hier für die 4. Klassenstufe eher illustrativ-motivierender Absicht) zu betrachten, dazu werden Fragen gestellt, die aus dem Text beantwortbar sind, teilweise müssen Motivationen und Gefühle von Akteuren erschlossen werden.

Zur erfolgreichen Bearbeitung der Aufgaben ist kaum gelerntes spezifisches Vorwissen notwendig, wenn, dann eher allgemeines Weltwissen aus den Naturwissenschaften. Kognitive Anforderungen sind Informationen finden, Verbindungen herstellen, schlussfolgerndes Denken, Passung/Konsistenz/Kohärenz vs. Widerspruch bestimmen; hierin ähneln sie sehr stark Anforderungen von Intelligenztests verschiedenen Inhalts (z. B. Carpenter et al., 1990, für die figuralen Raven-Matrizentests); hinzu kommen Kenntnisse der Testsprache und manchmal kognitive Aspekte von Empathie.

Mathematik

In der Aufgabe „Geschwindigkeit eines Rennwagens“ ist in einer Verlaufskurve dessen Geschwindigkeit abgetragen. In Fragen müssen Längen der Strecken bestimmt (der Abbildung entnehmbar), Geschwindigkeitskurve und Streckenkilometer in Beziehung gesetzt (der Abbildung entnehmbar), die Verlaufskurve interpretiert sowie ein plausibler Streckenverlauf aus verschiedenen Vorgaben herausgesucht werden.

Die Anforderungen ähneln nicht nur optisch stark der Tschadsee-Aufgabe unter Lesekompetenz. Alle Fragen bei „Rennwagen“ sind mit den gegebenen Informationen lösbar, zur Lösung müssen Abbildungen verstanden, einem kurzen Text Informationen entnommen und Schlussfolgerungen gezogen werden und wie bei allen anderen Aufgaben der verschiedenen Skalen ist die Beherrschung der Testsprache notwendig.

Es gibt aber unter Mathematik auch weit mathematikspezifischere Aufgaben, so sind etwa in „Äpfel“ eindeutig auch Rechnungen und Formelanwendung notwendig; in anderen zusätzlich ein Umgang mit Plausibilitäten und allgemein-schlussfolgerndes Denken. In der Aufgabe „Kontinent“ ist die Fläche der Antarktis plausibel zu schätzen, Voraussetzung dafür ist das Wissen, dass Länge mit Breite multipliziert Flächeninhalt ergibt. Die Aufgaben „Dreieck“ und „Bauernhof“ sind ohne schulspezifisches Geometrie-Wissen nicht lösbar (Winkel, Flächenberechnung).

Ähnliches lässt sich unter PISA 2003 (Schwerpunkt Mathematik) beobachten: In der Aufgabe „Größer werden“ ist eine Verlaufskurve der Körpergrößen bei Jun-

gen und Mädchen im Alter von 10 bis 20 Jahren angegeben; beantwortet werden muss, wann wer größer ist. Die Aufgabe ist der Tschadseeaufgabe unter Lesen 2000 sehr ähnlich. Unter „Wechselkurs“ sind Währungen umzurechnen, in „Auto“ sind verschiedene Qualitätsmerkmale nach vorgegebener Gewichtung rechnerisch zu kombinieren (was auch im Kopf geht) oder selbst eine Gewichtung zu entwickeln. In der Aufgabe „Spielwürfel“ ist ein Würfel als Flächennetz dargestellt, dieses muss mental gefaltet werden. Ähnliche Aufgaben finden sich auch in Intelligenztests, dort unter räumlichem Vorstellungsvermögen oder figuralen und nonverbalen Fähigkeiten (LPS Unterskala 8 Faltaufgaben, KFT-N3 Faltaufgaben, BIS-T4 Abwicklungen).

Die Mathematik-Aufgaben in TIMSS (7./8. Klasse) sind deutlich kürzer als die in PISA, es gibt kaum Text, manche sind völlig ohne. In ihrer Kürze und Schulnähe sind sie denjenigen in KFT-Q1 vergleichbar. Spezifisches Schulmathematikwissen ist für viele Aufgaben notwendig, allerdings sind durch einen gebildeten Erwachsenen beinahe alle noch lösbar (was für den auch für 12.-Klässler eines Gymnasiums gedachten KFT-Q1 nicht mehr zutrifft). In mehreren TIMSS-Aufgaben sind Grafiken (Temperatur auf Thermometer, Gewicht auf Waage, Balkendiagramm, Bremskurve, Landkarte, Baumgröße, Bleistiftlänge) und Tabellen zu interpretieren, in einer ein Würfel zu drehen (mentales Rotieren), Geometrieaufgaben (Winkel) kommen auch vor. TIMSS ist schulnäher als PISA, enthält aber auch Aufgaben, die auf allgemeines Weltwissen rekurrieren und über dieses lösbar sind.

In den PISA-Aufgaben variiert die Schwierigkeit zwischen allgemeinem mathematischem Wissen und speziellem Formelwissen, das ohne kontinuierlichen Kontakt mit Schulunterricht nicht vorhanden ist. Oft ist über Plausibilitäten zu agieren, Texten und Grafiken sind Informationen zu entnehmen und schlussfolgernd auf Antworten zu übertragen. Der spezifische Mathematik-Wissensanteil ist eher gering, beträgt schätzungsweise 30% (aus Gründen des Testschutzes werden nur Teile der Aufgabensätze veröffentlicht). Die TIMSS-Aufgaben sind in Inhalt und Form schulnäher.

Naturwissenschaft

In der Aufgabe „Sammelweis“ wird die Geschichte der Entdeckung des Zusammenhangs zwischen Leichensezierung, ausbleibendem Waschen der Hände durch die Ärzte, Entbindung durch diese Ärzte, Fieber der Wöchnerinnen und ihrem anschließenden Tod geschildert. In Antworten auf Fragen muss argumentiert werden, warum Erdbeben als Ursache für das Kindbettfieber nicht in Frage kommen kann; die Idee von Semmelweis ist wiederzugeben; darüber hinausgehend ist zu begründen, warum hohe Temperatur beim Waschen das Fiebrisiko minimiert, und anzugeben, was Ursache für die heute nachlassende Wirkung von Antibiotika sein könnte. Informationen zu den beiden letzten Fragen finden sich nicht im Text.

Teile der Aufgabe wären auch unter Lesefähigkeit subsumierbar, die beiden ersten Teilfragen sind durch Infor-

mationsentnahme (Idee von Semmelweis) und schlussfolgerndes Denken (Erdbeben) lösbar, die beiden letzten aber nicht. Für letztere ist Wissen notwendig, das wahrscheinlich weniger dem Biologieunterricht entstammt, sondern der Allgemeinbildung.

In der Aufgabe „Ozon“ ist ebenso ein längerer Text zu lesen. Aus dem Allgemeinwissen ist abzuleiten, dass geringes Ozon in der Stratosphäre über verstärkte UV-Strahlung Hautkrebs verursachen kann. Schließlich ist eine wissenschaftliche Fragestellung von einer ethisch-politischen zu unterscheiden. Letzteres ähnelt der Unterscheidung zwischen Inhalt und Stil unter Lesekompetenz. Die „Gewitter-Akustik“-Aufgabe gehört eigentlich zur Mathematik (ein Donner ist 9 Sekunden nach einem Blitz zu hören, das Gewitter ist 3 km entfernt, wie hoch ist die Schallgeschwindigkeit?). Dass ein Knall vom Mond nicht hörbar sein kann (weil die den Schall transportierende Luft fehlt), ist eine reine Wissensfrage.

Die Naturwissenschaftsaufgaben in TIMSS (7./8. Klasse) sind stark wissensbezogen (warum geht eine Kerze unter einer Glocke aus? warum Stahlbrücken streichen? wieso löscht Kohlendioxid Feuer? ist Salz ein Gemisch? fossile Brennstoffe entstanden aus? wie lange braucht Licht von einem anderen Stern als der Sonne zur Erde?). Das meiste davon dürfte zwischen allgemeinem und schulbezogenem Wissen stehen. Im Vergleich zu PISA ist weit weniger zu lesen und zu denken, oft sind Antworten – auch weil in der Regel zwischen verschiedenen Vorgaben zu wählen ist – in Sekundenbruchteilen möglich, es ist nur ein Wissensabruf notwendig. Fehlt dieses Wissen, ist die Aufgabe nicht lösbar. Grafiken sind eher selten (z. B. Thermometer, Talskizze, Aquarium). Wie in TIMSS-Mathematik sind die Aufgaben weit kürzer als in PISA.

In den Naturwissenschaftsaufgaben von PISA sind ungefähr die Hälfte bis zwei Drittel der Fragen durch Informationsentnahme aus den gegebenen Materialien und durch schlussfolgerndes Denken lösbar, für den Rest sind allgemeines Weltwissen und/oder Schulwissen notwendig. Spezifisch naturwissenschaftliche Aspekte in Inhalt, Aufgabenstellung und Anforderungen sind eher in der Minderzahl, große Überlappungen bestehen nicht nur zu den Leseaufgaben, sondern auch zur Mathematik.

Problemlösen (nur PISA 2003)

In der Aufgabe „Bibliothekssystem“ sind die Ausleihmodi zweier Schulbibliotheken (Gymnasien) beschrieben. Für die erste Schule ist das Ausleihverfahren (Lehrer vs. Schüler) verbal und per Entscheidungsbaum dargestellt, für die zweite Schule (Lehrer vs. Schüler, vorbestellt vs. nicht, Bücher vs. Zeitschriften, Nichtsäumige vs. Säumige) nur verbal. In der ersten Frage muss die Ausleihfrist für ein nicht vorbestelltes Buch durch einen nichtsäumigen Schüler angegeben werden. In der zweiten ist das deutlich kompliziertere Ausleihverfahren für Schule 2 ähnlich wie für Schule 1 als Entscheidungsbaum darzustellen.

Die Frage wäre auch gut unter Lesekompetenz einordenbar, Informationen müssen entnommen und verstan-

den werden; deduktiv-schlussfolgerndes Denken (Anwendung von Regeln, nicht deren Findung) ist notwendig. Das Aufmalen eines Entscheidungsbaums bedeutet die Transformation einer verbalen Information in eine grafische gemäß Vorlage. Vermutlich sind an der Lösung räumlich-vorstellende Fähigkeiten beteiligt.

Unter „Ferienlager“ sind Kinder gemäß vorgegebener Regeln (ähnlich wie in „Bibliothekssystem“; Jungen vs. Mädchen, Kinder vs. Erwachsene, Geschlecht der Erwachsenen gleich wie der Kinder, Gesamtkinderzahl und Bettenzahl je Zimmer) auf Zimmer zu verteilen. Eine Aufgabe mit mathematischem Inhalt. Regeln müssen beachtet und angewendet werden. Der Text ist kurz. Stärker figurale Inhalte sind die Aufgaben „Bewässerungskanäle“ und „Anschlusszüge“. In letzterer ist die zeitlich kürzeste Strecke mit Wartezeiten an Haltestationen und Umsteigepunkten zu finden. Diese Aufgabe wäre auch als mathematische oder naturwissenschaftliche Aufgabe denkbar, nur zählen und addieren können sind notwendig, zusätzlich wie in anderen Aufgaben Informationsentnahme.

Bei *dynamischem Problemlösen* ist in an Rechnern zu bearbeitenden Aufgaben das Zusammenwirken mehrerer Determinanten auf Effekte hin zu prüfen (z. B. zwei Parabeln, welche Regler bzw. welche Variablen und Gewichte sind für die horizontale und vertikale Lage oder Steigung und Öffnungsrichtung der zwei Kurven verantwortlich).

Problemlösen ähnelt (noch) mehr als die bisherigen drei Konstrukte der Intelligenz. Die kognitiven Anforderungen sind schwer von Intelligenz zu unterscheiden. Nach den Autoren selbst messen sie schlussfolgerndes Denken (den definitorischen Kern von Intelligenz): „The key skill needed to solve problems is analytical reasoning“ (OECD, 2004 b, S. 51). Weitere Beispiele aus der Beschreibung kognitiver Anforderungen beim Problemlösen: (OECD, 2004 b, S. 29 u. S. 51): „choosing among alternatives under constraints“, „understanding a situation“, „identify causally related variables“, „communicating the solution“, „checking and evaluating the decision“. „In most items, students are required to organise and analyse information, and derive solutions satisfying given constraints.“ (s. a. DPK, 2004, S. 148 ff.). Spezifisches, den Aufgaben nicht entnehmbares Wissen ist nicht notwendig.

Schlussfolgerungen

Die Aufgaben innerhalb von PISA und innerhalb von TIMSS ähneln einander in Inhalten, Formaten und Anforderungen. In TIMSS sind die meisten Fragen wissensabhängiger, in IGLU-Lesen ähnlich wenig(er) wie in PISA. Schwierigkeit entsteht durch spezifische Wissensabhängigkeit, durch Anforderungen an Konzentration, Lesegenauigkeit und Denken (Aufnahme, Integration und Vergleich verschiedener Informationen, Beziehungsbildung, Schlussfolgerungen ziehen und anwenden). Aufgaben in Intelligenztests sind im Vergleich deutlich kürzer, kognitive Prozesse und oft Wissen werden elementarer erfasst. In PISA sind die Aufgaben lang, Relevantes ist zu suchen,

zur Lösung ist eine Zusammenarbeit verschiedener kognitiver Prozesse notwendig.

Nur selten ist spezifisch-schulisches Wissen notwendig, am ehesten noch in Mathematik. Bei Vergleich mit dem KFT fällt in dessen verbalen Skalen die weit größere Wissensbeteiligung auf (Kenntnis von seltenen Fremdwörtern und abstrakten Begriffen), KFT-Quantitativ ist in vielen Aufgaben (Q1, z. B. „ 3^{-2} , $(3^2)^2$, $3^{1/2}$, \log_{10} “, Geometriaufgaben, spezielle mathematische Termini) schulnäher als PISA-Mathematik. Hier wäre die Frage zu stellen, ob der KFT für einen Intelligenztest nicht zu schulnah ist. Im Vergleich zu LPS und BIS-T4, die zwar nur die Kenntnis elementarer Rechenregeln voraussetzen, aber auch Fremdwortwissen (LPS) und Umgang mit Sprache (BIS-T4), ist der KFT eindeutig ein schulnaher Test. Weit reiner werden Denkfähigkeiten über figurale Aufgaben (wie in KFT-, LPS- und BIS-Unterskalen, im CFT oder in Matrixtests) erfasst. Auffällig ist aber, dass Kovariationen der figuralen KFT-Aufgaben oder anderer nonverbaler Intelligenztests mit Umwelt- und Persönlichkeitsmerkmalen nur wenig anders als die von sprachlichen und mathematischen Testteilen sind (vgl. Rindermann, in Vorb.). Weitgehend irrelevant dürfte auch der Geschwindigkeitsfaktor in der Durchführung sein, viele Intelligenztests sind nur unter enormem Zeitdruck vollständig bearbeitbar, was zu Operieren mit Plausibilitäten oder gar zum Raten zwingt.

Die Messkonstrukte der Schulleistungstudien – insbesondere von PISA und IGLU – sind denjenigen aus Intelligenztests weit ähnlicher als bislang angenommen. Darauf weisen nicht nur ähnliche Aufgabenmerkmale und Anforderungen hin, sondern auch die Beschreibungen der Autoren. So spricht das Deutsche PISA-Konsortium (DPK, 2001, S. 21 ff.) von „fächerübergreifenden Schlüsselqualifikationen“, „Modellierungsfähigkeit“, „Lernfähigkeit“, „Reflexion“, „kognitiver Rationalität“, „Verstehen“, „Erkennen“, „Problemlösen“, „intellektuellen Fähigkeiten“, „Kompetenzen“ und „Verständnis“. Unter Kompetenzen und Fähigkeiten werden kognitive gemeint, aber die Wortkombination „kognitive Fähigkeit“ ist nicht findbar und es wird von vielem gesprochen, oft in einer etwas unklaren Begrifflichkeit, aber nie von dem nahe liegenden Begriff Intelligenz. Hier dürfen allerdings zwei Probleme nicht übersehen werden: 1. Intelligenz wird in Wissenschaft und außerhalb dieser oft äußerst überdehnt für alles Erwünschte verwendet (vgl. die Kritik von Weber & Westmeyer, 2001). Fasst man Intelligenz aber präzise als kognitive Fähigkeit auf – neue kognitive Probleme lösen, Schlussfolgern, Abstraktion, Verstehen und Einsicht – dann ist kein relevanter Unterschied zu den genannten PISA-Begriffen erkennbar. Bildung wird als funktionale verstanden, d. h. anwendbare Kompetenz, im Wesentlichen Denken. 2. Varianzaufklärung in verhaltensgenetischen Studien wird mit Unveränderbarkeit von Intelligenz verwechselt. Genetische Faktoren müssen Umwelteinflüsse aber nicht ausschließen (z. B. Körpergröße). Ein prinzipieller Unterschied besteht aber zu einem (momentan) simplen *Wissensabruf*, über den sowohl einige Aufgaben in Intelligenztests als auch in Schulleistungstests lösbar sind. *Die Grenze Denken-Wissen verläuft aber nicht*

zwischen Intelligenztests und Schulleistungstests, sondern innerhalb ihrer. Wissen ist generell ein schwieriger Begriff: Die Abgrenzung schulischen Wissens von allgemeinem Wissen oder Kenntnissen der Alltagswirklichkeit der Moderne und schließlich von prozeduralem Wissen (Lösungsstrategien) ist von der spezifischen Wertschätzung und dem alltäglichen Gebrauch von Wissensinhalten abhängig, kompliziertere Mathematik gehört nicht dazu.

Empirische Beziehungen auf Ebene der Individuen

Methodisches Vorgehen

Soweit in der Literatur auffindbar werden latente und manifeste Korrelationen zwischen Dimensionen und Subdimensionen innerhalb der verschiedenen Studien sowie Ergebnisse aus dimensional Analysen beschrieben. Wenn nur latente Korrelationen berichtet werden, wird über Minderungskorrektur versucht, manifeste zu berechnen. Ein Vergleich mit Informationen aus Intelligenztests als Referenz soll die Interpretation der erzielten Koeffizienten erleichtern. Gesamtwerte aus Dimensionen einer Studie wurden durch einfache arithmetische Mittelung errechnet, die Reliabilität dieses Maßes durch die Spearman-Brown-Testverlängerungsformel berechnet.

In den international vergleichenden Schulleistungstudien arbeitet man mit Plausible Values. An der Stelle eines Wertes einer Dimension eines Schülers gibt es jeweils fünf Plausible Values. Hier für PISA 2000 näher ausgeführt: Diese Schätzwerte individueller Wertverteilungen für die gemessenen Dimensionen und Subdimensionen wurden verwendet, da Schüler aus Zeitgründen nur Teile des Gesamtaufgabensatzes bearbeiteten (Multi-matrixdesign); international bearbeiteten die meisten Schüler sogar nur Aufgaben aus dem Bereich Lesen und aus einem der beiden Bereiche Mathematik und Naturwissenschaften, nur zwei der neun Testhefte (Nr. 8 und 9) enthielten Aufgaben aus allen drei Bereichen (DPK, 2001, S. 50; OECD, 2002, S. 23). Die KMK hat den deutschen PISA-Datensatz von 2000 ins Netz gestellt. Mit dessen Daten wurden durch den Autor Reanalysen vorgenommen (Korrelationen zwischen Dimensionen, Variablenkürzel: *pv1read*, *pv1math*, *pv1scie*; Faktorenanalysen mit Subdimensionen, Variablenkürzel: *pv1read1*, *pv1read2*, *pv1read3*, *pv1math*, *pv1bio*, *pv1che*, *pv1phy* oder *pv1rcont*, *pv1rnonc*, *pv1math*, *pv1bio*, *pv1che*, *pv1phy*). Für Mathematik gibt es im Datensatz anders als in Publikationen angegeben keine Subdimensionen.

Anhand der Plausible Values bzw. der Mittel von 5 Plausible Values (Korrelation dieser innerhalb einer Dimension zwischen $r = .93$ bis $.99$, $\alpha = .99$ bis 1) können deskriptive Auswertungen (Korrelation, Faktorenanalyse) mit konventionellen Programmen vorgenommen werden (Mislevy et al., 1992, S. 148), nur für inferenzstatistische Prüfungen (Problem: Standardfehler), deren Nutzen bei den hier üblichen Stichprobengrößen zwischen 5000

und 200000 aber zweifelhaft ist, wären Programme, die die Fehlerbehaftetheit der Plausible Values als Verteilungswerte berücksichtigen (z. B. ConQuest und WesVar), notwendig. Plausible Values wären nur dann für Korrelationen und Faktorenanalysen problematisch, wenn zu ihrer Schätzung auch Items außerhalb ihrer Dimensionen herangezogen werden würden *und* wenn diese die Beziehung zwischen den Dimensionen verändern würden, Ersteres ist der Fall, fehlende Werte werden auch aus den Schülerfragebögen geschätzt (wahrscheinlich auch mit Hilfe von Informationen zur Bildung der Eltern), Letzteres soll aber nicht der Fall sein (eine präzise Beschreibung dieser Prozedur ist aber leider nicht zu finden; OECD, 2002).

In allen nationalen, erweiterten Testheften (Kunter et al., 2002, S. 79 und 96) gab es Aufgaben der Mathematik und Naturwissenschaften. Hier beträgt die Stichprobe in Faktorenanalysen mit Subdimensionen um $N = 17000$, auf Dimensionsebene um 28000. Nur Faktorenanalysen mit den internationalen Testheften Nr. 8 und 9, die Aufgaben aller drei Dimensionen enthalten, sind nicht möglich, da diese Information im KMK-Datensatz fehlt. Jeder Plausible Value der 5 verschiedenen ist ein erwartungstreuer Schätzer des Populationsmittels, alle fünf zusammen (nicht als Mittel) auch der Varianz; Plausible Values schätzen eine Verteilung. Empfehlenswert ist eine jeweils fünffache Rechnung und der anschließende Vergleich der Ergebnisse (OECD, 2002, S. 107). Gängig und durch die hohen Korrelationen der Plausible Values untereinander gut begründbar ist auch die Rechnung mit nur einem (meist dem ersten) Plausible Value, insbesondere in großen Stichproben und in multivariaten Verfahren (DPK, 2001, S. 521). Programme wie SPSS oder SAS sind hier nutzbar. Bei 6 oder 7 Subskalen, wie hier für Faktorenanalysen verwendet (2 oder 3 für Lesen, 1 für Mathematik, 3 für Naturwissenschaften), wären bei jeweils 5 Plausible Values 5^6 (15625) oder 5^7 (78125) verschiedene Faktorenanalysen denkbar, deren Ergebnisse zu mitteln wären. Aus Zeitgründen wurden nur zwei gerechnet (mit dem ersten Plausible Value oder mit dem Mittelwert aller 5). Nähere Informationen finden sich bei den einzelnen Ergebnissen.

Ergebnisse

Korrelationen zwischen den Skalen der Schulleistungstests sind in den Publikationen mit Ausnahme PISA 2003 schwierig zu finden. Nirgendwo gibt es Faktorenanalysen aller Skalen gemeinsam, für PISA 2003 jedoch eine Multidimensionale Skalierung (DPK, 2004, S. 168).

Bei PISA 2000 korrelieren die Skalen untereinander im Schnitt latent (l) zu $r_l = .85$ (.87, .83 und .84), manifest (m) zu $r_m = .78$ (berechnet über Minderungskorrektur, Reliabilitäten .90, .91 und .94; DPK, 2001, S. 221, 245, u. OECD, 2002, S. 153 f.; s. a. Tabelle 1; zur Minderungskorrektur DPK, 2001, S. 520 u. 245), selbst über den PISA-Datensatz mit allen fünf Plausible Values errechnet und anschließend Koeffizienten gemittelt $r_m = .80$, nur für Gymnasiasten $r_m = .60$. Errechnet man aus den drei hoch miteinander korrelierenden Skalen einen PISA-Gesamtwert, dann würde

dieser eine Homogenität von $\alpha_m = .91$ aufweisen (berechnet über Spearman-Brown-Formel; Homogenität mittels Cronbach- α selbst über den PISA-Datensatz errechnet $\alpha_m = .92$). Nach Artelt und Schlagmüller (2004, S. 171 f.), die den deutschen PISA-Datensatz 2000 analysierten, betragen die Korrelationen $r_l = .80$ und $r_m = .65$ (nicht über Minderungskorrektur, sondern über weighted maximum likelihood estimates, hier $\alpha_m = .85$ Gesamtwert). Die beiden TIMSS-Skalen (Baumert & Lehmann, 1997) korrelieren untereinander zu $r_m = .61$ (Individualdaten, nicht latent, E-Mail Rainer Watermann, 8-1-4; $\alpha_m = .76$ des TIMSS-Gesamtwertes). Die Korrelation ist wohl auf Grund der deutlichen Wissensbasis geringer. Die vier internationalen Skalen von PISA 2003 korrelieren untereinander im Schnitt zu $r_l = .82$ (DPK, 2004; S. 167), Gesamtwert $\alpha_l = .95$ (vorläufige Skalenreliabilitäten nach Mail vom 14-4-5 zwischen $r = .84$ und .92; über Minderungskorrektur $r_m = .71$ und Gesamtwert $\alpha_m = .91$). Für IGLU fehlen jegliche Angaben zu Skaleninterkorrelationen (auch bei Nachfrage).

Es kann nicht von empirisch unabhängigen Dimensionen ausgegangen werden. Zur Einordnung: Die mittlere Korrelation der drei KFT-Skalen verbal, quantitativ und nonverbal beträgt laut Manual $r_m = .44$ (Heller & Perleth, 2000; eigene Daten Gymnasium mit drei Dimensionen: $r_m = .40$, $N = 2041$ Schülerrohdaten, $r_m = .53$ mit $N = 526$ Schülermittelwerten über verschiedene Jahre, Rindermann, in Vorb.), die der drei BIS-Inhalts-Skalenwerte $r_m = .55$ (Jäger, Süß & Beauducel, 1997), die von Schulnoten zwischen fünf Fächergruppen $r_m = .55$ (Sprachen, Mathematik-Physik, Naturwissenschaften, Geisteswissenschaften, Kunst-Musik; $N = 1969-2092$, Rindermann, in Vorb.). Im Vergleich zu gängigen Intelligenztests und selbst Schulnoten sind die PISA- und TIMSS-Skalen auch auf Individualdatenebene deutlich homogener! Solch hohe Interkorrelationen, auch latente, findet man sonst nie in der Diagnostik.

Diese *ungewöhnlich hohe Homogenität* zeigt sich auch bei Betrachtung von Interkorrelationen innerhalb der Skalen (was nicht problematisch, sondern erwünscht ist): Die Subskalen Mathematik PISA 2003 korrelieren zwischen $r_l = .82$ und .95 (mathematisches Arbeiten: latent, arithmetisch gemittelt: $r_l = .91$, manifest mit Minderungskorrektur $r_m = .79$; Stoffgebiete: latent, arithmetisch gemittelt: $r_l = .88$, manifest mit Minderungskorrektur $r_m = .72$; DPK, 2004, S. 62 f.). Die durchschnittliche latente Korrelation zwischen verschiedenen Subskalen ($r_l = .91$ und $r_l = .88$) ist größer als die durchschnittliche Reliabilität der Subskalen ($r_{\text{Relia}} = .87$ und .82), die manifeste nur knapp darunter ($r_m = .79$ und .72). Diese Homogenität wird aber nicht als Homogenität, sondern als Differenziertheit interpretiert:

„Die latenten Korrelationen zwischen den Subskalen der Typen mathematischen Arbeitens und den Subskalen der Stoffgebiete weisen auf sinnvolle Differenzierungen hin. In Bezug auf die Typen mathematischen Arbeitens hebt sich vornehmlich die Skala der technischen Fertigkeiten mit Korrelationen von jeweils $r = .89$ deutlich von den Skalen zur Modellierung [mit $r = .95$] ab“ (DPK, 2004, S. 62).

Es ist völlig ungewöhnlich, dass eine Korrelation von $r_l = .89$ (manifest $r_m = .78$) als Nachweis von Verschiedenartigkeit gilt.

Tabelle 1. Korrelationen zwischen PISA-Skalen und mit anderen Maßen kognitiver Fähigkeiten

PISA 2000						
	Lesekompetenz	Mathematik	Naturwissenschaft			
Individualdaten (links unten latent, rechts oben manifest, in Diagonale Reliabilität)						
Lesekompetenz	.94	.78	.80			
Mathematik	.84	.91	.75			
Naturwissenschaft	.87	.83	.90			
Staaten, PISA 2000, $N = 41$						
Lesekompetenz	1					
Mathematik	.95	1				
Naturwissenschaft	.97	.97	1			
Deutsche Länder, PISA 2000, $N = 14$						
Lesekompetenz	1					
Mathematik	.93	1				
Naturwissenschaft	.95	.97	1			
PISA 2003						
Individualdaten (nur latente Ebene, in Klammern deutsche Daten)						
Lesekompetenz	1					
Mathematik	.77 (.80)	1				
Naturwissenschaft	.83 (.87)	.83 (.87)	1			
Problemlösen, anal.	.82 (.81)	.89 (.90)	.80 (.85)			
Problemlösen, dynam.	(.61)	(.69)	(.65)			
KFT-N2, figurale Ana.	(.64)	(.74)	(.68)			
Staaten, PISA 2003, $N = 40$						
Lesekompetenz	1					
Mathematik	.94	1				
Naturwissenschaft	.93	.97	1			
Problemlösen	.94	.99	.97			
PISA 2003 mit PISA 2000, Staatenebene, $N = 33$						
Lesekompetenz 03	.92	.89	.88			
Mathematik 03	.90	.96	.90			
Naturwissenschaft 03	.89	.93	.92			
Problemlösen 03	.89	.96	.91			
PISA 2000–2003, Staaten, $N = 40$, in Gesamtwerten: $r = .95$						
PISA 2003 mit KF-Gesamtwert, IQ Lynn & Vanhanen, IEA-Lesen, TIMSS8, TIMSS4, IGLU-Lesen (alle unkorrigiert)						
	KF-Gesamt	IQ-L&V	IEA-Lesen	TIMSS8	TIMSS4	IGLU
Lesekompetenz 03	.85	.78	.82	.30	.60	.55
Mathematik 03	.87	.85	.74	.54	.72	.52
Naturwissenschaft 03	.90	.84	.81	.61	.78	.69
Problemlösen 03	.89	.86	.77	.59	.74	.59
PISA 2003 Gesamt	.89	.85	.80	.54	.74	.61
N	39	38	22	26	19	18

Anmerkungen: Individualdaten-Korrelationen PISA 2003 s. OECD (2004b, S. 55). Korrelationen in Deutschland mit Intelligenztest (Figuren-analogien des KFT, N2; $N = 535$ Schüler, DPK, 2004, S. 167): analytisches Problemlösen $r = .72$, dynamisches Problemlösen $r = .63$, die beiden letzteren untereinander $r = .68$. PISA 2000 und IQ-L&V unkorrigiert. Manifeste Korrelation über Minderungskorrektur berechnet. KF-Gesamtwert aus IEA-Lesen, TIMSS4+8, PISA 2000, IGLU 2001 und Lynn-&Vanhanen-Intelligenzsammlung (nicht PISA 2003).

Zwischen den drei Leseskalen von PISA-2000 beträgt die mittlere Interkorrelation $r_1 = .93$ (bei Artelt & Schlagmüller, 2004: $r_1 = .91$ und $r_m = .70$), zwischen Lesen, Mathematik und Naturwissenschaften $r_1 = .85$ (OECD, 2002,

S. 153). Die diskriminante Validität auf Ebene von Individualdaten ist gering (Differenz $r = .08$ oder Fishers- $z = .40$). Ähnliche Ergebnisse werden innerhalb der Naturwissenschaften PISA 2000 und 2003 berichtet: Interkorrelation

von fünf kognitiven Teilkompetenzen 2000 zwischen $r = .64$ und $.87$ (vermutlich latent, arithmetisch gemittelt: $r = .74$, DPK, 2001, S. 226); Interkorrelation der Teilkompetenzen 2003 zwischen $r_1 = .58$ und $.93$ (latent, arithmetisch gemittelt: $r_1 = .77$, DPK, 2004, S. 129, $r_{\text{Relia}} = .74$, manifest $r_m = .57$); PISA 2000 Aufgaben Biologie-Chemie-Physik im Schnitt $r_m = .80$ (selbst berechnet). Senkbeil, Rost, Carsensen und Walter (2005, S. 166) bewerten dies angemessen: „Ähnlich wie in PISA 2000 korrelieren die Fächer-scores [$r_1 = .75$ – $.83$] außerordentlich hoch miteinander.“ Oder Interkorrelation der Teilkompetenzen innerhalb Lesen bei IGLU zwischen $r_1 = .84$ und $.90$ (latent, arithmetisch gemittelt: $r_1 = .88$, Bos et al., 2003, S. 80f., Reliabilitäten nicht dokumentiert). Ein „Generalfaktormodell“ nur für IGLU-Lesekompetenz hat einen guten Fit:

„Die Vorstellung, dass ein Generalfaktor einer derart komplexen Fähigkeit wie der Lesekompetenz zu Grunde liegt, entspricht nicht den oben dargelegten differenzierten theoretischen Ansätzen. Dennoch gelingt es mit Hilfe dieses Modells erstaunlich gut, einen Teil der in den Daten enthaltenen Variation zu erklären. Eine explorative Hauptkomponentenanalyse ergab ebenfalls eine einfaktorische Lösung mit einem erklärbaren Varianzanteil von 45 Prozent“ (Bos et al., 2003, S. 81). Bei einem Modell mit zwei Faktoren korrelieren diese untereinander zu $r = .89$ oder $.88$ (ähnliche Ergebnisse für Leseverständnis schon Rost, 1985).

Die Korrelationen innerhalb der Dimensionen zwischen Unterskalen sind ähnlich hoch oder etwas geringer als diejenigen zwischen den Dimensionen. Entscheidend wäre nun eine Faktorenanalyse auf Ebene aller Unterskalen. Lassen sich drei oder vier Oberskalen oder nur ein gemeinsamer g -Faktor finden? Wie groß ist die Varianzaufklärung eines ersten unrotierten Faktors? Faktorenanalysen fehlen, aber in PISA 2003 wurde das Ergebnis einer MDS dargestellt (DPK, 2004, S. 168): Lesen, Mathematik, Naturwissenschaften und analytisches Problemlösen bilden eine Fassade, eine andere dynamisches Problemlösen, die dritte eine figurale Intelligenztestskala des KFT (N2, figurale Analogien). KFT-N2 korreliert im Schnitt zu $r_1 = .70$ mit den vier internationalen PISA-Skalen (manifest ca. $r_m = .64$; s. a. Tabelle 1). Das Ergebnis spricht für eine Methoden- statt einer Inhaltsaufteilung (sonst müssten die beiden Problemlösemethoden in einer Fassade liegen), Inhaltsfaktoren wären erst mit zusätzlichen verbalen und quantitativen Intelligenztestaufgaben nachweisbar. Eine MDS ersetzt nicht eine Faktorenanalyse auf Ebene der Subskalen. Nach Herausnahme der KFT-N2-Skala bleiben Korrelationen zwischen den anderen Skalen übrig, die aber nicht für eine eigenständige kognitive Fähigkeit sprechen müssen (Methodenfaktor?), wichtig wäre, g herauszunehmen.

Für Homogenität spricht auch, dass in Kausalmodellen Bedingungsfaktoren und Ladungen auf Pfadkoeffizienten für unterschiedliche Dimensionen nahezu identisch sind: für Lesekompetenz (L) und Mathematik (M), wahrscheinlich standardisierte Beta, Muttersprache $\beta_L = .18$ und $\beta_M = .17$, Schulform $\beta_L = .39$ und $\beta_M = .41$, kulturelle Praxis $\beta_L = .22$ und $\beta_M = .25$, abweichende Orientierung $\beta_L = -.13$ und $\beta_M = -.15$. Einzig im Geschlecht, das nur für Mathematik angegeben ist, dürfte es Unterschiede geben (DPK, 2001, S. 501 ff.).

Die veröffentlichten Daten von PISA 2000 (www.kmk.org/schul/home.htm) ermöglichen Dritten, Faktorenanalysen zu rechnen. Hierfür können drei oder zwei Subskalen von Lesen, eine von Mathematik (im Datensatz nur eine Gesamtskala, in der Dokumentation dazu aber sieben Subskalen; Kunter et al., 2002) und drei Subskalen der Naturwissenschaften (im Datensatz nur Biologie, Chemie, Physik, in der Dokumentation dazu noch fünf kognitive Komponenten) herangezogen werden. Kognitive Fähigkeiten (KFT-Subskalen, Arbeitsgedächtniskapazität, Problemlösen), die in der Dokumentation der Variablen des Datensatzes beschrieben werden, fehlen leider im veröffentlichten Datensatz und sind für Dritte kaum zugänglich.² Auf Itemebene können, da Schüler jeweils nur Teile der Itemsätze bearbeiteten (Multimatrixdesign), keine Faktorenanalysen mit dem Gesamtdatensatz gerechnet werden. Würde man die fehlenden regressionsanalytisch aus den vorhandenen schätzen und dann mit den Einzelitems rechnen, wäre der dimensionsspezifische g -Faktor wahrscheinlich überhöht. In allen mit den gegebenen Subdimensionen des Datensatzes zu rechnen möglichen Varianten (Gruppierungen innerhalb der Skalen, erster Plausible Value, $N = 17843$, vs. Mittel von 5, $N = 29567$) ergeben sich immer eindeutig einfaktorische Lösungen: Der erste unrotierte Faktor bindet zwischen 80, 82 und 84 % der Varianz, der zweite Eigenwert liegt unter 0,5 (nur Gymnasium: 65, 67 und 70 %, 2. Eigenwert unter 0,9, $N = 5106$ oder 8833; zum Vergleich Gymnasium: bei KFT mit sechs Subdimensionen, alte Skalen V1, V4, Q2, Q3, N1, N2, Faktor 1 unrotiert 46,2 %, 2. Eigenwert 0,96, bei Schulnoten 59,9 %, 2. Eigenwert 1,1; Rindermann, in Vorb.). Die Rechnung mit Mitteln aller fünf Plausible Values statt nur mit dem ersten erhöht die Varianzaufklärung des ersten Faktors um maximal 2 % (alle Schüler) oder 5 % (nur Gymnasiasten). Bei gängigen Abbruchkriterien (Eigenwerte ≥ 1 , Scree test) würde immer eine einfaktorische Lösung empfohlen werden. Allerdings ließe sich mit vorgegebenen drei Faktoren ungefähr die vorgegebene Struktur replizieren (Probleme: Biologie auf allen drei Skalen, Physik eher bei Mathematik als bei Naturwissenschaften, Mathematik auch über .50 auf Lesen)³.

Schlussfolgerungen

Die verschiedenen Dimensionen innerhalb der internationalen Schulleistungsstudien sind untereinander weit höher korreliert als Dimensionen innerhalb von Intelligenztests. Alles – auch das Ergebnis von Faktorenanalysen –

² Eine Veröffentlichung der Intelligenzdaten wurde von der KMK untersagt, sie wären nur über Sonderantrag mit Begründung des Forschungsanliegens zugänglich, um – wie unter der Hand mitgeteilt wird – „Unsinn“ zu vermeiden.

³ Faktorenanalysen auf Itemebene (Items aus Testheften 8 und 9; Wittmann, 2005) ergeben gemäß Scree test auch eher Hinweise auf Eindimensionalität: der erste unrotierte Faktor bindet 33,0 oder 35,7 %, der zweite 5,5 oder 5,8 %, der dritte 5,1 oder 4,9 % usw. Nach dem Eigenwertkriterium werden drei Faktoren mit zum Teil beträchtlichen Fremdladungen extrahiert. Zum Vergleich: Faktorenanalyse mit Items des KFT, 7. Klasse Gymnasium, der erste unrotierte Faktor bindet 14,7 %, der zweite 8,7 %; 9. Klasse Gymnasium, erster unrotierter Faktor 9,1 %, zweiter 8,4 %; 11. Klasse Gymnasium, erster unrotierter Faktor 10,1 %, zweiter 8,3 %.

spricht für einen sehr starken g -Faktor. D. h. interindividuelle Unterschiede in verschiedenen Dimensionen kovariieren sehr stark und sind wahrscheinlich auf ähnliche oder identische Ursachen zurückführbar. Die Korrelationen mit figuralen Intelligenztestresultaten unterstreichen die Nähe zur Intelligenz, Unterschiede scheinen auf Erhebungsmethoden zurückführbar zu sein. Die Varianzaufklärung des ersten Faktors dürfte nicht nur auf Grund der Ähnlichkeit der PISA-Aufgaben aus verschiedenen Dimensionen so hoch sein (s. a. Artelt & Schlagmüller, 2004, S. 171), sondern auch wegen der vergleichsweise sehr fehlerarmen Messungs- und Auswertungsorganisation (vgl. OECD, 2002), der höheren Varianz im Gesamtdatensatz als oft in Intelligenzteststudien und – was psychologisch relevant ist – vor allem wegen der grundlegenden Einheitlichkeit kognitiver Prozesse und ihrer Entwicklungsdeterminanten.

Empirische Beziehungen auf Ebene von Staaten und Ländern

Methodisches Vorgehen, Quellen

Auf internationaler Ebene ist man bei vergleichenden Analysen mit vielerlei Problemen konfrontiert: Ist eine Repräsentativität der Daten gegeben, wie sahen die Erfassungsmethoden aus, sind Schätzungen aussagekräftig usw.? So sind etwa in Albanien nur noch 43 % der 15-Jährigen noch Schüler, in Mexiko 52 %, in Brasilien 53 % und in Peru 66 %, PISA bezieht sich nur auf diese. Möchte man aber Aussagen für alle 15-Jährigen eines Landes vornehmen, sind Korrekturen notwendig, sicherlich nach unten. Lynn und Vanhanen (2002) haben für 81 Staaten veröffentlichte Intelligenztestresultate gesammelt, für 104 haben sie durch Mittel aus den Nachbarländern mit ähnlicher ethnischer Zusammensetzung die Daten geschätzt. Da aber nicht zufällig bestimmte Staaten wie Afghanistan fehlen, dürften Schätzungen als Mittel zu hoch gegriffen sein. Weiteres: In der IEA Reading-Literacy-Study 1990-1991 waren in Simbabwe die „14-Jährigen“ 15,5 Jahre alt, in Venezuela die „9-Jährigen“ 10,7. Die Teilnahmequote variierte zwischen 59 % (Nigeria) und 100 % (z. B. Singapur; Elley, 1992). Auch hier müssen Korrekturen vorgenommen werden. Bei den nationalen Intelligenztests der Bundeswehr (nur nicht früh verweigernde Männer; Ebenrett, Hansen & Puzicha, 2002) sind Testverfahren nicht völlig klar, die Repräsentativität ist eingeschränkt. In Analysen zeigt sich aber immer wieder, dass die internationalen und nationalen Testergebnisse trotzdem aussagekräftiger sind als die je nach Land von unterschiedlichen Definitionen betroffenen (Weede, 2003) oder sogar von absichtlichen Fälschungen befallenen Bildungsmaße (Barro & Lee, 1993 a, S. 366f.).

Datengrundlage sind Daten auf Ebene der Staaten oder Länder (für Deutschland z. B. 14 oder 16 Werte der teilnehmenden Bundesländer). Gesamtwerte innerhalb der Studien (TIMSS, PISA) wurden aus deren Dimensionen durch einfache Mittelung berechnet, was empirisch auf Grund extrem hoher Korrelationen innerhalb der Studien

gerechtfertigt ist. Sowohl auf Ebene der Staaten als auch auf Ebene der deutschen Länder wurde auch ein Gesamtwert kognitiver Fähigkeiten berechnet, was durch hohe Korrelationen und Eindimensionalität zwischen Studien empirisch unterstützt wird. KF-Gesamt besteht aus allen Studien der 90er-Jahre bis IGLU 2001 (nicht PISA 2003). Da Studien fast immer in verschiedenen Staaten- und Länderstichproben durchgeführt wurden, waren vor einer Mittelung Standardisierungen (z -Werte) an den jeweils gleichen Staaten- und Länderstichproben notwendig. International wurden zuvor noch, soweit Informationen vorliegen, Korrekturen vorgenommen, um zu für die jeweilige Bevölkerung repräsentativen Werten zu gelangen (für Lynn-Vanhanen-Intelligenz: Abzug bei den geschätzten Länderdaten; bei PISA-2000: Abzug um Prozentsätze der 15-Jährigen, die nicht im Schulsystem sind; bei IEA Reading-Literacy-Study 1990-1991 um geringe Teilnahmequote und Überalterung; nähere Informationen s. Rindermann, in Vorb.). Die Korrekturen an kognitiven Fähigkeiten nehmen nur geringen Einfluss auf die Höhe der Korrelationen (s. Tabellen 2 und 4).

Folgende Studien wurden berücksichtigt: PISA 2003 (OECD, 2004 a), IGLU/PIRLS 2001 (Bos et al., 2003; Mullis et al., 2003), Intelligenztest (Lynn & Vanhanen, 2002), PISA 2000 (DPK, 2002; OECD, 2003 a), TIMSS 1994-1995 für Klassenstufen 8 und 4 (Baumert & Lehmann, 1997; Gonzales, 2000) sowie IEA Reading-Literacy-Study 1990-1991 (Elley, 1992). Innerhalb Deutschlands wurden Ergebnisse von Intelligenzmessungen bei Musterungen herangezogen (Ebenrett, Hansen & Puzicha, 2002). Leider fehlen bislang systematisch aufbereitete Datensätze zu kulturvergleichenden Piagetstudien. Diese wären zukünftig zusätzlich zu berücksichtigen.

Auf internationaler Ebene wurden korrelativ Vergleiche mit wichtigen gesellschaftlichen Indikatoren vorgenommen: Für das *nationale Bildungsniveau* wurden drei Werte zusammengeführt: 1. Nichtalphabetenanteil unter Erwachsenen 1991 nach Vanhanen (1997) in 172 Staaten, 2. Prozentsatz der Menschen im Alter von 12-19 Jahren von 1960 bis 1985, die über Sekundarschulbildung verfügen, $N = 117$ (Mankiw, Romer & Weil, 1992), 3. Anzahl der Schuljahre insgesamt 1990, 1995 und 2000 von Personen im Alter über 25 Jahren nach Barro und Lee (2000; $N = 107$); Gesamtbildungsindex der Bevölkerung (Mittel nach Standardisierung) aus diesen dreien $\alpha = .94$, $N = 173$. *Religionen* wurden in Anlehnung an Max Weber (1988/1920) und Werner Sombart (1913) gewichtet nach ihrer affirmativen Einstellung zu Bildung, Denken, Rationalität und Leistung (Reihenfolge in der Stärke des Gewichts: Protestantismus, Judentum, ostasiatischer Konfuzianismus, Katholizismus, orthodoxes Christentum, Buddhismus, Hinduismus und Islam, Animismus). Religionsdaten fußen auf Internetinformationen des Auswärtigen Amtes und einem Länderlexikon (Jahrbuch, 2004) mit der Fokussierung auf die traditionell in einem Land vorherrschende Religion. Das kaufkraftbereinigte *Bruttonationalprodukt* 1998 stammt von Lynn und Vanhanen (2002), $N = 185$. *Rechtsstaat* 1970-2000 mit Betonung von Eigentumschutz nach Gwartney und Lawson (2003) und Knack und Keefer (1995), $\alpha = .90$, $N = 131$. *Lebenstempo* 1992-1995

Tabelle 2. Korrelationen zwischen kognitiven Fähigkeiten (international)

Skala	IQ	Iqk	IQD	PTG	PG	PGk	P-L	P-M	P-N	T8G	T8M	T8N	EL	ELk	EL9	EL14	I-L	T4G	T4M	T4N
IQ-L&V	1	.96	.30	.75	.78	.79	.72	.80	.78	.83	.83	.78	.84	.84	.61	.86	.82	.87	.83	.86
IQ-L&V-k	.96**	1	.55	.78	.80	.80	.77	.79	.80	.85	.84	.81	.84	.81	.62	.86	.76	.88	.84	.87
IQ-Daten (1 = da)	.30**	.55**	1	.34	.33	.32	.38	.28	.34	.36	.34	.37	.38	.25	.44	.39	.28	.48	.45	.49
PISA-TIMSS8-G	.75**	.78**	.34*	1	.97	.95	.94	.96	.96	.96	.94	.92	.68	.72	.58	.43	.73	.89	.90	.82
PISA00-Gesamt	.78**	.80**	.33*	.97**	1	.97	.98	.99	.99	.45	.49	.31	.84	.84	.79	.59	.65	.68	.56	.76
PISA00-G-k	.79**	.80**	.32*	.95**	.97**	1	.97	.95	.96	.37	.40	.25	.87	.89	.84	.62	.75	.61	.47	.71
PISA00-Lesen	.72**	.77**	.38*	.94**	.98**	.97**	1	.95	.97	.29	.32	.20	.86	.87	.84	.55	.63	.57	.42	.69
PISA00-Mathe	.80**	.79**	.28	.96**	.99**	.95**	.95**	1	.97	.46	.55	.26	.78	.76	.69	.56	.62	.67	.57	.73
PISA00-Natwiss.	.78**	.80**	.34*	.96**	.99**	.96**	.97**	.97**	1	.53	.53	.43	.83	.84	.80	.59	.69	.74	.63	.80
TIMSS8-Gesamt	.83**	.85**	.36*	.96**	.45*	.37	.29	.46*	.53**	1	.98	.97	.31	.39	.26	.30	.74	.92	.94	.83
TIMSS8-Mathe	.83**	.84**	.34*	.94**	.49**	.40*	.32	.55**	.53**	.98**	1	.90	.29	.34	.23	.28	.72	.90	.94	.78
TIMSS8-Natwis.	.78**	.81**	.37*	.92**	.31	.25	.20	.26	.43*	.97**	.90**	1	.30	.40	.25	.29	.71	.89	.88	.84
Elley-Lesen	.84**	.84**	.38*	.68**	.84**	.87**	.86**	.78**	.83**	.31	.29	.30	1	.92	.97	.99	.30	.45	.28	.61
Elley-Lesen-k	.84**	.81**	.25	.72**	.84**	.89**	.87**	.76**	.84**	.39	.34	.40	.92**	1	.88	.90	.45	.56	.43	.64
Elley-Lesen-9j	.61**	.62**	.44*	.58**	.79**	.84**	.84**	.69**	.80**	.26	.23	.25	.97**	.88**	1	.83	.24	.26	.11	.44
Elley-Lesen-14j	.86**	.86**	.39*	.43**	.59**	.62**	.55**	.56**	.59**	.30	.28	.29	.99**	.90**	.83**	1	.31	.46	.32	.57
IGLU-Lesen	.82**	.76**	.28	.73**	.65**	.75**	.63**	.62**	.69**	.74**	.72**	.71**	.30	.45	.24	.31	1	.87	.77	.91
TIMSS4-Gesamt	.87**	.88**	.48*	.89**	.68**	.61**	.57**	.67**	.74**	.92**	.90**	.89**	.45	.56*	.26	.46	.87**	1	.97	.96
TIMSS4-Mathe	.83**	.84**	.45*	.90**	.56*	.47*	.42	.57**	.63**	.94**	.94**	.88**	.28	.43	.11	.32	.77**	.97**	1	.87
TIMSS4-Natwis.	.86**	.87**	.49*	.82**	.76**	.71**	.69**	.73**	.80**	.83**	.78**	.84**	.61*	.64**	.44	.57*	.91**	.96**	.87**	1
Korrelationsmittel	.82	.83	.37	.88	.85	.82	.80	.81	.85	.75	.73	.68	.77	.75	.65	.66	.69	.80	.75	.79
r mit KF-Gesamt	.95	.99	.53	.91	.92	.95	.91	.90	.92	.93	.91	.91	.92	.92	.81	.92	.93	.97	.92	.96
N	185	185	185	52	41	41	41	41	41	38	38	38	31	31	26	30	33	26	26	26
IQ-L&V Mess	1	1	-	.77	.78	.79	.73	.81	.78	.82	.82	.78	.84	.86	.47	.88	.76	.84	.80	.80
IQ-L&V NMess	1	1	-	.56	.79	.86	.79	.76	.79	.98	.96	.98	.82	.64	.52	.82	.93	.93	.82	.99

Anmerkungen: Korrelationsmittel der 19 Dimensionen (ohne „IQ-Daten“) ohne die Korrelation mit sich selbst berechnet, über Fishers-Z-Werte, Gesamtmittel $r = .77$ (79 bei Gesamtskalen, 73 bei Unterskalen). Signifikanz hier nicht sinnvoll, nur als indirektes Maß des Stichprobenumfangs. N zwischen 51 (PISA-TIMSS-Gesamt mit IQ-L&V), 27 (PISA mit TIMSS) und 16 (Elley mit TIMSS4). IQ-Daten: Daten für Staaten von Lynn & Vanhanen aus Testergebnissen ermittelt (1) oder aus Nachbarstaaten und ethnischer Zusammensetzung geschätzt (0). „k“ jeweils von HR korrigierte Daten. „IQ-L&V Mess“: nur diejenigen Staaten, für die gemessene Werte vorliegen (hier: IQ-L&V-k=IQ-L&V; $N = 81$, $N = 22-42$ mit anderen Skalen). „IQ-L&V NMess“: nur diejenigen Staaten, für die nicht gemessene Werte vorliegen ($N = 104$, $N = 9-4$ mit anderen Skalen). PISA-TIMSS8-G: Gesamtwert aus PISA 2000 und TIMSS. PISA00-G-k: korrigierter Wert für PISA 2000 (um die im Schulsystem fehlenden Schüler). TIMSS8: 8. Klasse, TIMSS4: 4. Klasse 1994-1995. Elley-Lesen: IEA Reading Literacy Study 1990-1991. IGLU-Lesen 2001. KF-Gesamt: Gesamtwert aus allen Studien 90er-Jahre bis 2001.

Tabelle 3. Faktorenanalyse der kognitiven Fähigkeiten (Unterskalen international)

Skalen	unrotierte Lösung	zwei Faktoren vorgegeben	
	erster Faktor	Jugendliche (?)	Kinder (?)
IQ Lynn & Vanhanen	.85	.79	.35
PISA-2000-Lesekompetenz	.83	.88	.18
PISA-2000-Mathematik	.91	.88	.32
PISA-2000-Naturwissenschaften	.91	.85	.37
TIMSS8-Mathematik	.93	.73	.58
TIMSS8-Naturwissenschaften	.88	.78	.40
Elley Lesen-9-Jährige	.86	.83	.30
Elley Lesen-14-Jährige	.85	.88	.21
IGLU-Lesen	.80	.49	.72
TIMSS4-Mathematik	.66	.15	.96
TIMSS4-Naturwissenschaften	.79	.35	.90
Varianzaufklärung	71,5%	53,5%	29,7%

Anmerkungen: Eigenwerteverlauf: 7,86, 1,29, 0,87, 0,49 usw. $N = 186$, fehlende Werte ersetzt. Ersetzung nach Regionen, Sprachen, Religionen, Welt (1./2./3.). Ohne Ersetzungen: Über 23 Staaten nur mit PISA-2000-Unterskalen, TIMSS8-Unterskalen und IQ-Lynn-Vanhanen resultiert eine einfaktorielle Lösung (Faktor 1 Eigenwert 4,2, 70%, Faktor 2 0,9, 15%); zwingt man zu zwei Faktoren, werden PISA und TIMSS unterschieden (Methoden- statt Inhaltsfaktoren).

nach Levine (2002/1997), $\alpha = .72$, $N = 31$. *Bücherzahl* von Eltern und Kindern nach IGLU-Studie 2001 (Mullis et al., 2003), $\alpha = .88$, $N = 34$. *Demokratie* 1950-2002 nach Vanhanen (1997) und Marshall und Jagers (2000), $\alpha = .94$, $N = 183$. *Zwischenmenschliches Vertrauen* in den 90er-Jahren nach Inglehart (1997), $N = 41$. *Wirtschaftswachstum* zwischen 1950 und 1990 nach Lynn und Vanhanen (2002), $N = 185$. *Morde* (Tötungsdelikte je 100.000 Einwohner) 1995/2002 ($\alpha = .41$) und *Aufklärung Morde* 1995/2002 ($\alpha = .77$) nach Interpol (2004), $N = 138$ und 122. *Krieg* nach Sala-i-Martin (1997; 1960-90, $N = 118$) ergänzt und korrigiert durch mich auf den Zeitraum zwischen 1960 und 2000, $N = 186$, die Variable gibt nicht nur Kriegsteilnahme, sondern auch Intensität von Krieg und Kriegsschäden im eigenen Land wieder. *HIV-Infektionsrate* 2001 und 2003 ($\alpha = .99$) nach UNAIDS/WHO (2003), $N = 165$. *Staatsquote* 1980-89 nach Barro und Lee (1993 b), $N = 138$. *Einkommensungleichheit* (Differenz zwischen arm und reich, Gini-Koeffizient) nach Fukuda-Parr (2004) und Deininger und Squire (1996), 1951-2002 zusammengefasst (Schwerpunkt 70er bis 90er-Jahre), $\alpha = .89$, $N = 148$. *Kinderzahl* je Frau 1960-1984 nach Barro und Lee (1993 b), $N = 130$. *Ökonomische Freiheit* 1960-2000 nach Gwartney und Lawson (2003), O'Driscoll, Holmes und O'Grady (2002) und Sala-i-Martin (1997), $\alpha = .89$, $N = 165$. *Nichtkorruption* 1980-2003 nach Transparency International (2004), $N = 132$. *Bürokratieschnelligkeit* zwischen 1972 und 1995 nach Knack und Keefer (1995), $N = 60$.

Nur „Rationalität“ und „gesellschaftliche Modernität“ sind ausschließlich für diese Studie ermittelte Variablen, $N = 186$. „Rationalität“: Probleme denkend unter Zuhilfenahme von empirisch gewonnenen Informationen lösen, Orientierung an Regeln des Denkens, nicht oder nicht überwiegend an Astrologie, Hexerei, Tradition oder

Macht. „Gesellschaftliche Modernität“: Menschenrechte, Demokratie, Freiheit, Gleichberechtigung. Zu deren Einschätzung wurden vier Akademiker gebeten: Herkunft Deutschland und Türkei, ein Protestant, zwei Katholiken und ein Muslim, zusammen Kenntnis von 11 verschiedenen Fremdsprachen, Arbeitsaufenthalte im Ausland einschließlich Dritter Welt.

Die Zuverlässigkeit der Daten ist oft gering und gewinnt auch nicht immer durch Mittelbildung aus verschiedenen Jahren und Studien, die zeitliche Stabilität ist aber relativ hoch (ausführliche Diskussion in Rindermann, in Vorb.).

Ergebnisse

Die drei PISA-Skalen 2000 korrelieren untereinander zu $r = .96$ (41 Staaten; Mittel aus diesen drei Skalen: $\alpha = .98$) oder zu $r = .95$ (14 deutsche Länder; $\alpha = .98$), die vier PISA-Skalen 2003 korrelieren untereinander zu $r = .96$ (40 Staaten; $\alpha = .99$), die beiden TIMSS-8-Skalen zu $r = .90$ (38 Staaten; $\alpha = .94$ des Gesamtwertes), die beiden TIMSS-4-Skalen zu $r = .87$ (26 Staaten; $\alpha = .93$ des Gesamtwertes; s. Tabellen 1, 2 und 4). Die Korrelationen sind extrem hoch, es kann auf keiner Datenebene von empirisch innerhalb der Studien unabhängigen Dimensionen ausgegangen werden.

Bei Vergleich zwischen den Studien IEA-Lesen, TIMSS, PISA, IGLU und der IQ-Sammlung von Lynn und Vanhanen oder den Intelligenzmesswerten der Bundeswehr (s. Tabellen 1, 2 und 4) fallen sehr hohe Korrelationen auf. International und national erheben die verschiedenen Skalen der Schulleistungsstudien in unterschiedlichen Altersstufen empirisch weitgehend das Gleiche und

Tabelle 4. Korrelationen zwischen kognitiven Fähigkeiten (national)

Skala	IQ98	IQ92	IQM	P-G	P-L	P-M	P-N	I-G	I-L	I-M	I-N	I-O
IQ Bundeswehr 1998	1	.68	.87	.87	.89	.84	.85	.92	.87	.82	.94	.88
IQ Bundeswehr 1992	.68**	1	.96	.53	.65	.41	.48	.70	.44	.70	.78	.80
IQ Bundeswehr Mittel	.87**	.96**	1	.71	.80	.62	.67	.85	.67	.81	.92	.90
PISA00-Gesamt	.87**	.53	.71**	1	.98	.98	.99	.93	.90	.87	.90	.98
PISA-Lesen	.89**	.65*	.80**	.98**	1	.93	.95	.92	.88	.83	.87	.98
PISA-Mathematik	.84**	.41	.62*	.98**	.93**	1	.97	.93	.93	.89	.93	.98
PISA-Natwiss.	.85**	.48	.67**	.99**	.95**	.97**	1	.92	.88	.86	.89	.95
IGLU-Gesamt	.92**	.70	.85**	.93**	.92**	.93**	.92*	1	.95	.98	.98	.97
IGLU-Lesen	.87*	.44	.67	.90*	.88*	.93**	.88*	.95**	1	.96	.96	.95
IGLU-Mathematik	.82*	.70	.81*	.87	.83	.89*	.86	.98**	.96**	1	.96	.92
IGLU-Natwiss.	.94**	.78	.92**	.90*	.87	.93*	.89*	.98**	.96**	.96**	1	.91
IGLU-Orthographie	.88*	.80	.90*	.98**	.98**	.98**	.95*	.97**	.95**	.92**	.91*	1
Korrelationsmittel	.87	.69	.83	.93	.91	.92	.91	.94	.90	.90	.93	.95
r mit KF-Gesamt	.97	.61	.80	.96	.96	.93	.94	.97	.91	.90	.96	.96
N	16	16	16	14	14	14	14	7	7	6	6	6

Anmerkungen: Korrelationsmittel ohne die Korrelation mit sich selbst berechnet, über Fishers-Z-Werte, Gesamtmittel $r = .89$ (.90 bei Gesamtskalen, .89 bei Unterskalen). Signifikanzen hier nicht sinnvoll, nur als indirektes Maß des Stichprobenumfangs. N zwischen 14 (Intelligenz-PISA) und 5 (IGLU mit PISA). Keine Korrekturen oder Ergänzungen fehlender Werte. Weitere Studien mit hohen Korrelationen weggelassen (Studienleistungen in Medizin, Studienauswahltests, Berufseignungstests, Orthografie-Bundeswehr, Studien vor 1990, PISA 2003).

Tabelle 5. Faktorenanalyse der kognitiven Fähigkeiten (Unterskalen national)

Skalen	Werte ohne Ergänzung, $N = 5$		Werte mit Ergänzung, $N = 16$	
	unrotiert erster Faktor		unrotiert erster Faktor	zwei Faktoren vorgegeben, rotiert
	Kognitive Fähigkeit		Kognitive Fähigkeit	Schul- studien
		IQ 92		IQ 92
IQ Bundeswehr 1998	.95		.82/.84	.54
IQ Bundeswehr 1992	.94		-.36	-.08
PISA-Lesen	.97		.90/.92	.67
PISA-Mathematik	.99		.95/.95	.83
PISA-Natwiss.	.97		.94/.95	.80
IGLU-Lesen	.96		.95/.94	.97
IGLU-Mathematik	.94		.93/.92	.97
IGLU-Natwiss.	.96		.95/.93	.96
IGLU-Orthographie	.99		.95/.93	.96
Varianzaufklärung	92,9 %		85,3 %/77,0 %	64,5 %
		28,7 %		

Anmerkungen: Eigenwerteverlauf: 8,37, 0,43, 0,16, 0,05 usw.; oder 6,82, 0,80, 0,23, 0,07 usw. (rechts, links vom Strich); oder 6,93, 1,45, 0,36, 0,13 usw. (rechts, rechts vom Strich). Links vom Strich ohne IQ 1992, rechts vom Strich mit IQ 1992. Bei Faktorenanalyse der drei Summenskalen (IQ 1998, PISA-G, IGLU-G) sind die Ergebnisse selbstverständlich einfaktoriell mit Ladungen $1 > .90$. Ergänzung fehlender Werte nach Gruppierung durch Schulpolitik (nach Parteizugehörigkeit des Regierungschefs) und Ostwest.

das Gleiche wie Intelligenztests. Deshalb lässt sich aus den verschiedenen Studien auch ein Gesamtwert kognitiver Fähigkeiten „KF-Gesamt“ berechnen. Der große Vorteil eines solchen Gesamtmaßes ist, Analysen für praktisch alle Staaten der Welt durchführen zu können, nicht nur für die der Ersten Welt. OECD-Mitgliedschaft ist kein wissenschaftliches Kriterium.

Faktorenanalysen (Tabellen 3 und 5) bestätigen auch auf internationaler und nationaler Ebene einfaktorielle

Datenstrukturen: Staaten, deren Schüler gut in PISA abschneiden, erzielten auch gute TIMSS-Ergebnisse und Intelligenztestresultate (meist auch Personen jüngerer Alters). Es gibt keinerlei inhaltlich deutbare Faktorenstrukturen. Allerdings scheinen gewisse Evidenzen für Methoden- oder Alterseffekte vorzuliegen. Innerhalb Deutschlands fallen die 1992er Ergebnisse der Bundeswehr-Intelligenzmessungen aus dem Rahmen. Vielleicht waren damals die in der Bundesrepublik stabilen sozio-kulturell-schul-

politischen Milieus in Ostdeutschland noch nicht etabliert. Die sonst extrem hohen Korrelationen sind um so beachtlicher, als innerhalb Deutschlands zum Teil nur fünf Länder in die Berechnungen eingehen (PISA-2000-Mathematik 15-Jährige mit IGLU-Orthographie 2001 Viertklässler $r = .98$, $N = 5$ Länder). Die Schulleistungsstudien und die Intelligenzmessungen (international wie national) scheinen extrem reliabel zu sein. Ein Ergebnis zweifellos sehr akkurater Arbeit.

Bei Faktorenanalysen ausschließlich mit den Skalen Lesen, Mathematik und Naturwissenschaft aus PISA 2000 und PISA 2003 resultieren auf Staatenebene ($N = 33$) wie Länderebene ($N = 14$) eindeutig einfaktorische Lösungen (erster unrotierter Faktor: 93 % oder 91 % Varianzaufklärung). Zwingt man zu einer dreifaktoriellen Lösung, laden die inhaltlich korrespondierenden Dimensionen nicht paarweise (Lesen vs. Mathematik vs. Naturwissenschaft), sondern Dimensionen laden jährlich (2000 vs. 2003), der dritte Faktor ist uninterpretierbar.

In seltenen Fällen gibt es Indizien für aussagekräftige differenzielle Validitäten. Diese beziehen sich auf Inhalte, nicht auf Schulleistung vs. Intelligenz: Israelische Schüler erreichten in PISA 2000 höhere Lesekompetenz (452) als mathematische (433) oder naturwissenschaftliche (434), ein ähnliches sprachaffines Muster berichten Herrnstein und Murray (1994, S. 275) aus Intelligenztests für amerikanische Juden. Ostasiaten (Japan, Südkorea, Hongkong, Macao, Singapur, Taiwan) sind in Mathematik und Natur-

wissenschaften der Schulleistungsstudien jeweils besser als in Lesen – dies korrespondiert zu Intelligenztestresultaten von Ostasiaten in den USA (Oesterdiekhoff, 2000, S. 103, IQ 97 vs. 110), zu geringen Bücherzahlen daheim und in der Schule (und wenn naturwissenschaftliche), zu hoch ausgeprägten naturwissenschaftlichen Interessen, zu Unterricht in einer Fremdsprache (Hongkong und Singapur) und zu der oft „schweigsamen Haltung der Schüler“ im Unterricht (Hesse, 2005, S. 45).

Bildung und Intelligenz

Testresultate (alle Maße kognitiver Fähigkeiten zusammengefasst in „KF-Gesamt“, $N = 186$ Staaten) korrelieren sehr hoch mit internationalen Maßen durchschnittlicher Bildung ($r = .78$, s. Abbildung 1; vgl. Barber, 2005). Diese Korrelation sinkt durch die Herausnahme des Bruttosozialprodukts nur auf $r_p = .64$.

Die Korrelationen kultureller, gesellschaftlicher, politischer und wirtschaftlicher Merkmale mit den zusammengefassten Ergebnissen in Intelligenz- und Schulleistungstests eines Staates auf Staatenebene sind von erstaunlicher Höhe. Wahrscheinlich werden die Zusammenhänge sogar noch unterschätzt, weil Daten nicht zufällig aus bestimmten Regionen und Ländern fehlen und zudem mit einer Vielzahl von Fehlern belastet sind. Kognitive Fähigkeiten sind ein sensibler Indikator für gesellschaftliche Zustände und für gesellschaftliche Entwicklungen. Individuell gemessene und untersuchte und als rein personenbezogen verstandene Merkmale sind in Entstehung und Folge mit makrosozialen Bedingungen verwoben. Intelligenz ist kein rein individuell verstehbares Merkmal, sondern auch ein kulturelles und gesellschaftliches (vgl. Oesterdiekhoff, 2000). Die Fähigkeit zum Denken scheint in einem Netz von Merkmalen, das für eine *moderne und funktionierende bürgerliche Gesellschaft* steht, eine zentrale Position einzunehmen. Intelligenz ist aber nicht nur *Indikator*, sondern auch vermittelnder *Kausalfaktor* dieser Entwicklungen.

Verrechnet man nun in einer separaten Faktorenanalyse die drei Größen der Bildung gemeinsam mit Ergebnissen der Intelligenztests und Schulleistungsstudien (mit Lynn-Vanhanen-Intelligenz, PISA 2000 und PISA 2003, $N = 25$ Staaten oder mit Lynn-Vanhanen-Intelligenz, PISA 2000 und TIMSS, $N = 21$), ergeben sich nach dem Eigenwertkriterium ein- oder zweifaktorielle Lösungen, auf Faktor 1 die Bildungsmaße, auf Faktor 2 Tests, der erste unrotierte Faktor klärt aber zwischen 79,9 % und 59,3 % der Varianz auf und PISA

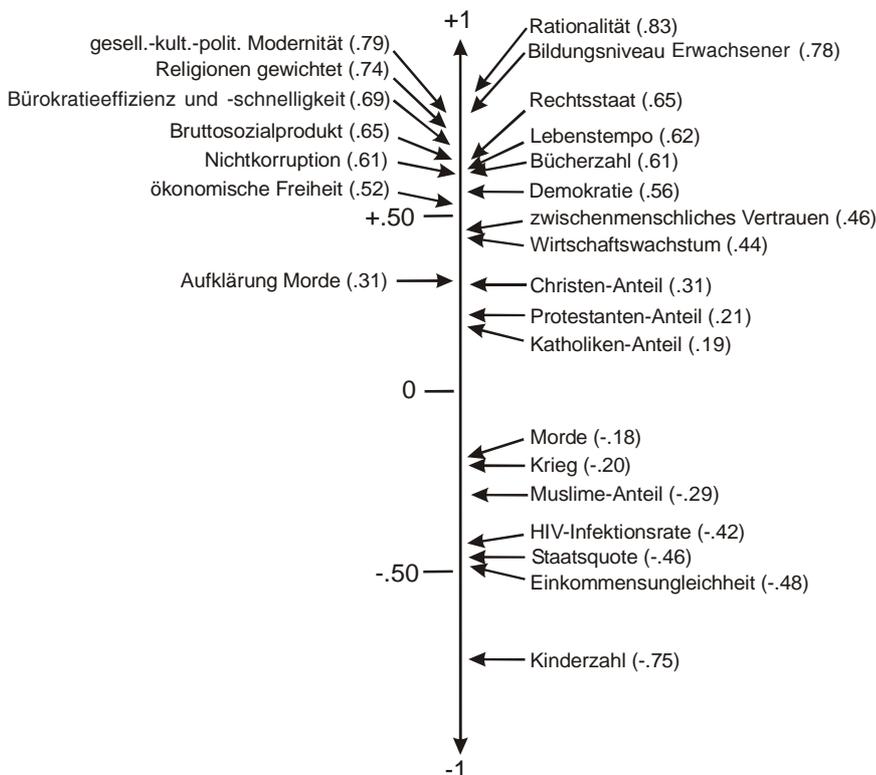


Abbildung 1. Korrelationen zwischen makrosozialen Merkmalen und Gesamtwert kognitiver Fähigkeiten (Intelligenz und Wissen), $N = 31$ bis 186 Staaten je nach Merkmal, Staatenebene (in Klammern Korrelationen numerisch).

2000 bzw. PISA 2000 und 2003 laden auch in der rotierten Lösung immer mit über $\lambda = .60$ auf dem Bildungsfaktor (und nicht etwa die viel schulwissensnähere TIMS-Studie). Kognitive Fähigkeiten sind international schwierig von Bildung separierbar ($r = .78$ der Gesamtwerte), die Korrelation ist höher als die zu gesellschaftlichem Reichtum (mit Bruttosozialprodukt $r = .65$). Damit wird die von Ceci (1991) behauptete Ähnlichkeit von Schulbildung und Intelligenz unterstrichen und der Zusammenhang ist ein weiteres sehr wichtiges Indiz dafür, dass kognitive Fähigkeiten bildungsabhängig sind, nicht nur individuelle, sondern auch gesellschaftliche (vgl. Lurija, 1987/1974).

Konsequenzen

Ergebnisse der Schulleistungsstudien sind auf allen Datenebenen in sich höchst homogen, sie korrelieren hoch mit anderen Schulleistungsstudien und mit Intelligenztestresultaten. Die Ergebnisse sind in extremem Maße – weit mehr als aus Intelligenztests selbst bekannt – auf einen gemeinsamen g -Faktor zurückführbar. Inhaltsanalytisch untersucht sind sich verschiedene Konstrukte und Aufgaben sehr ähnlich. Die diskriminanten Validitäten sind gering. Die Korrelationen sind auch zwischen deutlich unterschiedlichen Konstrukten und Erfassungsmethoden noch hoch. Dies zieht viele Konsequenzen nach sich:

Je ähnlicher sich zwei Konstrukte sind, ähnlich in begrifflicher Definition, in ihrer Messung oder in ihrer Entwicklung bei Individuen und Gesellschaften (und in Determinanten dieser), desto geringer ist ihr gegenseitiger Erklärungswert. Es ist nicht sinnvoll, das schnelle Laufen eines 100m-Läufers mit dessen schnellem Rennen zu erklären. Gutes Abschneiden in kognitiven Schulleistungstests mit hoher Intelligenz zu erklären, hat kaum Erkenntniswert, genauso wenig wie mathematische Kompetenz durch Lesekompetenz oder sprachliche Intelligenz durch mathematische (vgl. DPK, 2001, S. 129, 184).

Wenn Intelligenz wie „Schulleistung“ von Bildung abhängig ist, dann muss *Intelligenz als eine plastische Eigenschaft* aufgefasst werden, Intelligenz im Lebenslauf, interindividuelle Unterschiede und makrosoziale Unterschiede in Intelligenz.

Interindividuelle Unterschiede sind in verhaltensgenetischen Studien zu einem beträchtlichen Teil durch bislang unbekannte genetische Faktoren aufklärbar, dies gilt auf Grund der hohen Korrelationen auch für Schulleistung. Gene könnten auch an Fähigkeitsunterschieden auf makrosozialer Ebene beteiligt sein (so die These von Lynn & Vanhanen, 2002). Genetische Erklärungen sind nicht unumstritten, wissenschaftlich problematisch ist das Ausbleiben eines Nachweises des Kausalweges zwischen Genen und Intelligenz: Welches Gen oder welches Genom oder welche Genkombination ist für welche Ausprägung von Intelligenz auf welchem Weg verantwortlich? Weit klarer belegbar ist ein Kausalweg von Bildung auf kognitive Fähigkeiten (z. B. Armor, 2003; Ceci, 1991), Bildung

wiederum ist eingebettet und beeinflusst von Kultur. Bildung ist nicht nur institutionell, sondern auch familiär, schließt Erziehung ein. Bildungseffekte müssen genetische Faktoren nicht unbedingt ausschließen und umgekehrt.

Wenn „Intelligenz“ und „Schulleistung“ einander so ähnlich sind, sollte ein gemeinsamer Begriff für beide gefunden werden. Wie soll man diese kognitive Fähigkeit benennen? Allgemeine kognitive Fähigkeiten („general mental ability“ bei Schmidt & Hunter, 1998, oder „cognitive ability“ bei Herrnstein & Murray, 1994), schulnahe kognitive Fähigkeiten, Denkfähigkeit, Intelligenz? TIMSS ist am schulnahesten nach Schulnoten in kognitiven Fächern selbst, dann wissensabhängige Fragen in PISA und in Intelligenztests, schließlich figurale Tests. Alle Varianten sind aber abhängig von Bildung (vgl. Lurija, 1987/1974; Ceci, 1991; Oesterdiekhoff, 2000). Komplexe kognitive Fähigkeiten sind nie schulfremd. Deshalb ist auch wenig sinnvoll, Sprache als Sprache (nicht als spezifische Sprache) oder Kultur als Kultur aus Tests kognitiver Fähigkeiten zu entfernen. Sprache ist ein wichtiges Instrument und Medium des Denkens und bestimmte kulturelle Bedingungen fördern über Zwischenglieder wie Bildungsinstitutionen und Erziehung besser kognitive Fähigkeiten als andere (vgl. Max Weber, 1988/1920; Lurija, 1987/1974; Lenhart, 1998). Die Begriffe *allgemeine kognitive Fähigkeiten*, *Denkfähigkeit* und *Intelligenz* sind einander ähnlich. Intelligenz ist aber ein schärferer Begriff, kognitive Fähigkeiten schließen etwa auch basale kognitive Kompetenzen wie Informationsverarbeitungsgeschwindigkeit mit ein (Carroll, 1993), Intelligenz als Fähigkeit zum Denken definiert (schlussfolgerndes Denken, Abstraktion, Verstehen) jedoch nicht. Allerdings besteht hier das Problem der fehlenden Berücksichtigung von *Wissen*, wenn Aufgaben durch einfachen Wissensabruf lösbar sind. Bei einem Überwiegen von Wissensfragen ist nicht mehr von Denkfähigkeit zu sprechen, dann müsste ein neuer Gesamtbegriff für Denken und Wissen gefunden werden, etwa „*Denkwissensfähigkeit*“. Zwischen über figurale Testaufgaben ermittelter Intelligenz und reinem, aber bedeutsamem und richtigem Wissen besteht ein Kontinuum der Zunahme spezifischer Einflüsse von Persönlichkeit und Umwelt.

Warum sind aber gemessene Intelligenz und Schulleistung oder Denkfähigkeit, Wissen und Bildung einander so ähnlich? Warum besteht eine so enge Beziehung zwischen Denken und Wissen? Mögliche Ursachen:

1. *Ähnliche Aufgaben und Inhalte*: Viele Intelligenztests beinhalten schulnahe Aufgaben, vor allem in den quantitativen (numerischen, mathematischen) Teilen und etwas schwächer in den sprachlichen Teilen. Die Aufgaben beinhalten eine gemeinsame Wissensbasis (z. B. Kenntnis von Rechenregeln oder Wortbedeutungen).

2. *Leistungssituation Test und Schule* (mündliche wie schriftliche Prüfungen): Für erfolgreiche Leistungen in kognitiven Anforderungssituationen sind bestimmte Persönlichkeitsmerkmale förderlich, z. B. Anstrengungsbereitschaft, Fleiß, mäßige Angst sowie Interessen und Ein-

stellungen. Die Anpassungsfähigkeit an diese Leistungssituationen oder die erfolgreiche Sozialisation in solchen fließt in die Ausprägung beider Merkmale und in deren Messung ein.

3. *Ähnliche Denkanforderungen*: Leistungen in Intelligenztests und Schulprüfungen basieren auf gemeinsamen Denkkomponenten (Schlussfolgern, Abstraktion, Verständnis). Diese finden darüber hinaus eine gemeinsame Grundlage in kognitiven Basiskompetenzen (Informationsverarbeitungsgeschwindigkeit, Aufmerksamkeit, Arbeitsgedächtniskapazität).

4. *Schulfähigkeiten sind Ergebnis investierter Intelligenz*: Intelligenz fördert Lernen und so kognitive Schulfähigkeiten. Der Aufbau „kristalliner Intelligenz“ benötigt „fluide Intelligenz“ (Cattell, 1987/1971) oder im Sinne der (genetisch orientierten) Informationsverarbeitungstheorie geht der Schulleistung Intelligenz voraus und dieser Informationsverarbeitungsgeschwindigkeit (Jensen, 1989; Rindermann & Neubauer, 2004). Kristalline wie fluide Intelligenz sind aber bildungsabhängig, je wissensnäher oder komplexer Fähigkeiten, desto stärker (vgl. Merz, Remer & Ehlers, 1985), und Informationsverarbeitungsgeschwindigkeit und Intelligenz korrelieren nur zu $r = .30$.

5. *Wissen und Denkfähigkeit sind Ergebnis einer vorangegangenen Praxis von Denken und Bildung*: Wissen – wie vor allem in TIMSS erhoben – als Rückgriff auf vergangenes Denken und Bildung ermöglicht momentane Denkanforderungen zu reduzieren, Denken wiederum operiert meist mit Wissensinhalten.

6. *Schule und Unterricht fördern Intelligenz wie Schulfähigkeiten*: Quantität und Qualität von Schulunterricht fördern Schulfähigkeiten und Intelligenz. In der Schule ist aber das Lösen figuraler Intelligenztestaufgaben nie Thema.

7. *Erziehung und Bildungsnähe fördern Intelligenz wie kognitive Schulfähigkeiten*: Kompetente Erziehung (Anregungen, Fragen, Induktivität, dosierte Überforderung, Regelkontrolle usw.) fördert beide Varianten kognitiver Fähigkeiten, genauso Bildungsnähe (Gespräche, Fragen, Lesen, selektives Fernsehen etc.).

Dank der internationalen Schulleistungsstudien ist dies klarer geworden. Kognitive Fähigkeiten und ihre Beziehung zu Bildung rücken in das Zentrum des Interesses und werden in ihrer Bedeutung als Indikator, Folge und Kausalfaktor ökonomischer, politischer, gesellschaftlicher und kultureller Prozesse deutlich. Das Beziehungsgeflecht wäre aufzuklären.

Literatur

- Armor, D. J. (2003). *Maximizing intelligence*. New Brunswick: Transaction Publishers.
- Artelt, C. & Schlagmüller, M. (2004). Der Umgang mit literarischen Texten als Teilkompetenz im Lesen. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz* (S. 169–196). Wiesbaden: VS.
- Barber, N. (2005). Educational and ecological correlates of IQ: A cross-national investigation. *Intelligence*, 33 (3), 273–284.
- Barro, R. J. & Lee, J.-W. (1993 a). International comparisons of educational attainment. *Journal of Monetary Economics*, 32, 363–394.
- Barro, R. J. & Lee, J.-W. (1993 b). *Barro-Lee data set. International measures of schooling years and schooling quality. Distributed by the World Bank Group*. Retrieved from: <http://www.worldbank.org/research/growth/ddbarle2.htm>. Aktualisierung 1994 [14.07.2004].
- Barro, R. J. & Lee, J.-W. (2000). *Barro-Lee data set. International data on educational attainment: updates and implications*. Retrieved from: <http://www2.cid.harvard.edu/ciddata/barrolee/readme.htm> [18.11.2004].
- Baumert, J. & Lehmann, R. (1997). *TIMSS – Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich*. Opladen: Leske & Budrich.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Walther, G. & Valtin, R. (Hrsg.) (2003). *Erste Ergebnisse aus IGLU*. Münster: Waxmann.
- Carpenter, P. A., Just, M. A. & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review*, 97 (3), 404–431.
- Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge: Cambridge University Press.
- Cattell, R. B. (1987/1971). *Intelligence: Its structure, growth and action*. Amsterdam: Elsevier.
- Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, 27 (5), 703–722.
- Cole, P. G. (1995). The bell curve: Should intelligence be used as the pivotal explanatory concept of student achievement? *Issues In Educational Research*, 5 (1), 11–22.
- Deininger, K. & Squire, L. (1996). *Measuring income inequality: a new database*. Retrieved from: <http://www.worldbank.org/research/growth/dddeisqu.htm> [20.09.2004].
- Ditton, H. & Krecker, L. (1995). Qualität von Schule und Unterricht. *Zeitschrift für Pädagogik*, 41 (4), 507–529.
- DPK (Deutsches PISA-Konsortium; Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann K.-J. & Weiß, M.) (Hrsg.) (2001). *Pisa 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- DPK (Baumert, J., Artelt, C., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Schümer, G., Stanat, P., Tillmann K.-J. & Weiß, M.) (Hrsg.) (2002). *Pisa 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.
- DPK (Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J. & Schiefele, U.) (Hrsg.) (2004). *PISA 2003. Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Ebenrett, H. J., Hansen, D. & Puzicha, K. J. (2002). „Brain drain“ in deutschen Regionen. Effekte von Arbeitslosigkeit und innerdeutscher Migration. Bonn: Arbeitsberichte des Psychologischen Dienstes der Bundeswehr (1/02).
- Elley, W. B. (1992). *How in the world do students read? The Hague: The International Association for the Evaluation of Educational Achievement (IEA)*.
- Fukuda-Parr, S. (2004). *Human development report 2004*. Retrieved from: http://hdr.undp.org/statistics/data/indic/indic_138_1_1.html [20.09.2004].
- Gonzales, P. (2000). *Mathematics and science in the eighth grade*. Washington: National Center for Education Statistics.
- Gottfredson, L. (2003). g, jobs and life. In H. Nyborg (Ed.), *The scientific study of general intelligence. Tribute to Arthur R. Jensen* (pp. 293–342). Oxford: Pergamon.
- Gwartney, J. & Lawson, R. (2003). *Economic freedom of the world. 2003 annual report*. Vancouver: Fraser Institute.
- Heller, K. A. & Perleth, Ch. (2000). *Kognitiver Fähigkeits-Test (KFT 4-12+R)*. Weinheim: Beltz.

- Herrnstein, R. J. & Murray, Ch. (1994). *The bell curve. Intelligence and class structure in american life*. New York: Free Press.
- Hesse, H.-G. (2005). Lernen innerhalb und außerhalb der Schule aus interkultureller Perspektive. In G. Trommsdorff & H.-J. Kornadt (Hrsg.), *Enzyklopädie der Psychologie. C/VII/2 Kulturvergleichende Psychologie* (Manuskript: S. 1–95). Göttingen: Hogrefe (Manuskript in Vorbereitung).
- Inglehart, R. (1997). *Modernization and postmodernization. Cultural, economic, and political change in 43 societies*. Princeton: Princeton University Press.
- Interpol (2004). *International Crime Statistics*. Retrieved from: [http://www.interpol.int/Public/Statistics/ICS/download List.asp](http://www.interpol.int/Public/Statistics/ICS/download>List.asp) [22.10.2004].
- Jäger, A. O., Süß, H.-M. & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test (BIS-T4)*. Göttingen: Hogrefe.
- Jahrbuch (2004). *Jahrbuch*. München: DTV.
- Jensen, A. R. (1989). The relationship between learning and intelligence. *Learning and Individual Differences, 1* (1), 37–62.
- Knack, S. & Keefer, Ph. (1995). Institutions and economic performance: crosscountry tests using alternative institutional measures. *Economics and Politics, 7* (3), 207–227.
- Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann K.-J. & Weiß, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente*. Berlin: MPI.
- Lehrl, S. (2005). PISA – ein weltweiter Intelligenz-Test. *Geistig Fit, 1*, 3–6.
- Lenhart, V. (1998). *Protestantische Pädagogik und der „Geist“ des Kapitalismus*. Frankfurt: Lang.
- Levine, R. (2002/1997). *Eine Landkarte der Zeit. Wie Kulturen mit Zeit umgehen*. München: Piper.
- Lurija, A. R. (1987/1974). *Die historische Bedingtheit individueller Erkenntnisprozesse*. Berlin: Deutscher Verlag der Wissenschaften.
- Lynn, R. & Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport: Praeger.
- Mankiw, N. G., Romer, D. & Weil, D. N. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics, 107* (2), 407–437.
- Marshall, M. G. & Jaggers, K. (2000). *Polity IV Project. Political regime characteristics and transitions, 1800–1999*. Retrieved from: <http://www.bsos.umd.edu/cidcm/inscr/polity> [05.10.2004].
- Merz, F., Remer, H. & Ehlers, Th. (1985). Der Einfluß des Schulbesuchs auf Intelligenztestleistungen im Grundschulalter. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 17* (3), 223–241.
- Mislevy, R. J., Beaton, A. E., Kaplan, B. & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement, 29* (2), 133–161.
- Mullis, I., Martin, M. O., Gonzales, E. J. & Kennedy, A. M. (2003). *PIRLS 2001 international report*. Chestnut Hill: IEA.
- O'Driscoll, G. P., Holmes, K. R. & O'Grady, M. A. (2002). *2002 index of economic freedom*. Washington: Heritage Foundation & Wall Street Journal.
- OECD (Adams, R. & Wu, M.) (2002). *PISA 2000 technical report*. Paris: OECD.
- OECD (2003a). *Literacy skills for the world of tomorrow. Further results from PISA 2000*. Paris: OECD.
- OECD (2003b). *PISA 2003 assessment framework Mathematics, reading, science and problem solving. Knowledge and skills*. Paris: OECD.
- OECD (2004a). *Learning for tomorrow's world. First results from PISA 2003*. Paris: OECD.
- OECD (2004b). *Problem solving for tomorrow's world*. Paris: OECD.
- Oesterdiekhoff, G. W. (2000). *Zivilisation und Strukturgenese*. Frankfurt: Suhrkamp.
- Rindermann, H. (2004). Intelligenz. In H. S. Friedman & M. W. Schustack (Hrsg.), *Persönlichkeitspsychologie und Differentielle Psychologie* (S. 372–389). München: Pearson.
- Rindermann, H. & Neubauer, A. (2004). Processing speed, intelligence, creativity and school performance: Testing of causal hypotheses using structural equation models. *Intelligence, 32* (6), 573–589.
- Rindermann, H. *Intelligenz, Bildung und Gesellschaft: Entwicklung, Förderung und Auswirkung kognitiver Fähigkeiten auf individueller und makrosozialer Ebene*. Manuskript in Vorbereitung.
- Rost, D. H. (1985). *Dimensionen des Leseverständnisses*. Braunschweig: Agentur Pdersen.
- Rost, D. H. (2005). *Interpretation und Bewertung pädagogisch-psychologischer Studien*. Weinheim: Beltz UTB.
- Sala-i-Martin, X. (1997). I just ran two million regressions. *American Economic Review, 87* (2), 178–183.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124* (2), 262–274.
- Schnotz, W. & Dutke, S. (2004). Kognitionspsychologische Grundlagen der Lesekompetenz. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz* (S. 61–99). Wiesbaden: VS.
- Senkbeil, M., Rost, J., Carstensen, C. H. & Walter, O. (2005). Der nationale Naturwissenschaftstest PISA 2003. Entwicklung und empirische Überprüfung eines zweidimensionalen Facettendesigns. *Empirische Pädagogik, 19* (2), 166–189.
- Sombart, W. (1913). *Der Bourgeois*. München: Duncker & Humblot.
- Transparency International (2004). *Global corruption report 2004*. London: Pluto Press.
- UNAIDS/WHO (2003). *AIDS epidemic update*. Genf: WHO.
- Vanhanen, T. (1997). *Prospects of democracy*. London: Routledge.
- Waldmann, M. & Weinert, F. E. (1990). *Intelligenz und Denken*. Göttingen: Hogrefe.
- Weber, H. & Westmeyer, H. (2001). Die Inflation der Intelligenzen. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 251–266). Lengerich: Pabst.
- Weber, M. (1988/1920). *Gesammelte Aufsätze zur Religionssoziologie I (Protestantische Ethik, Konfuzianismus und Taoismus)*. Tübingen: J. C. B. Mohr.
- Weede, E. (2003). Intelligenztests, Humankapital und Wirtschaftswachstum: eine international vergleichende Studie. *List Forum für Wirtschafts- und Finanzpolitik, 29* (4), 390–406.
- Weinert, F. E. (2001a). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17–31). Weinheim: Beltz.
- Weinert, F. E. (2001b). Schulleistungen – Leistungen der Schule oder der Schüler? In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 73–86). Weinheim: Beltz.
- Weiss, V. (2002). Bevölkerung hat nicht nur eine Quantität, sondern auch eine Qualität. Ein kritischer Beitrag zur politischen Wertung der PISA-Studie. *Veröffentlichungen der Gesellschaft für Freie Publizistik, 18*, 31–59.
- Wittmann, W. W. (2005). Group differences in intelligence and related measures. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 223–238). Thousand Oaks: Sage.

Dr. habil. Heiner Rindermann

Institut für Psychologie
 Otto-von-Guericke-Universität Magdeburg
 Postfach 4120
 39016 Magdeburg
 E-Mail: heiner.rindermann@gse-w.uni-magdeburg.de