

---

Werner Blum/Gabriele Kaiser/David Burghes/Nigel Green

## **Entwicklung und Erprobung eines Tests zur „mathematischen Leistungsfähigkeit“ deutscher und englischer Lernender in der Sekundarstufe I**

Im Rahmen von empirischen Untersuchungen zum Lehren und Lernen von Mathematik haben wir einen Test entwickelt ("Potentialtest"), der die "mathematische Leistungsfähigkeit" von 13/14-jährigen Jugendlichen in England und Deutschland für Vergleichszwecke messen soll. Im vorliegenden Beitrag beschreiben wir die Entstehung des Tests sowie Resultate der Durchführung des Tests bei 1036 englischen und deutschen Lernenden. Die Resultate werden unter Berücksichtigung von - aus unseren früheren Fallstudien bekannten - Charakteristika des Mathematikunterrichts in beiden Ländern interpretiert.

In the context of empirical investigations into the learning and teaching of mathematics, we have developed a test ("Potential Test") which is to measure the "mathematical potential" of 13/14-year-olds in England and Germany for the purpose of comparisons. In this paper we describe the construction of the test as well as results of testing it with 1036 English and German pupils. The results will be interpreted with regard to characteristic features of mathematics teaching in both countries, known from our previous case studies.

### **(1) FRAGESTELLUNG**

Wie schon in unserem Beitrag Kaiser-Meßmer/Blum (1993) dargestellt, vollzieht sich in England seit Mitte der achtziger Jahre eine grundlegende Umgestaltung des Bildungssystems, z.B. die Einführung eines "National Curriculum" mit verpflichtenden Kernfächern oder die verstärkte Einführung von Leistungskontrollen (vgl. etwa Brown, 1993). Die Begründungen für diese Veränderungen beziehen sich in wesentlichen Teilen auch auf Leistungen und Erfolge des deutschen Bildungssystems. Auch gibt es in England ein großes Interesse an vergleichenden empirischen Untersuchungen, insbesondere bezogen auf den Mathematikunterricht, u.a. bedingt durch das schlechte Abschneiden englischer Lernender in internationalen Mathematiktests (z.B. in der Second International Mathematics Study [SIMS] von 1981, siehe Cresswell/Gubb, 1987 und Robitaille/Garden, 1989, sowie beim Second International Assessment of Educational Progress [IAEP] von 1990, siehe Educational Testing Service, 1992) bzw. durch Veröffentlichungen, in denen auf die im Vergleich zu Deutschland deutlich schlechteren mathematischen Leistungen in England hingewiesen wird (siehe Prais/Wagner, 1986).

Wir führen seit 1990 in einem gemeinsamen Projekt der Universitäten von Kassel und Exeter **vergleichende empirische Untersuchungen** zum Lehren und Lernen von Mathematik, insbe-

sondere eingebettet in Realitätsbezüge, zwischen England und Deutschland durch (für erste Ergebnisse siehe Blum et al., 1992 und Kaiser-Meißner/Blum, 1993). Die Ergebnisse unserer ersten Fallstudien haben u.a. darauf hingewiesen, daß es, um zu breiteren und gesicherteren Erkenntnissen über Unterschiede und Gemeinsamkeiten zu kommen, sinnvoll ist, noch längerfristige Vergleichsuntersuchungen durchzuführen. Dabei ist es natürlich wichtig, die **Vergleichbarkeit** der in beiden Ländern beteiligten Lerngruppen abzusichern.

Wir haben 1993 begonnen, die Entwicklung der mathematischen Leistungen einer größeren Anzahl von Lernenden (ca. 900 in jedem Land) über einen Zeitraum von 2 bzw. 3 Jahren (von Anfang Klasse 8 bis Ende der Sek.I) zu verfolgen und zu vergleichen. Um sicherzustellen, daß die englischen und deutschen Lernenden über ein vergleichbares mathematisches "Ausgangsniveau" verfügen, haben wir in 1992/93 einen **Test zur mathematischen Leistungsfähigkeit** (Mathematical Potential Test) entwickelt, den alle beteiligten Lernenden zu Anfang unserer Langzeit-Untersuchung (Beginn Schuljahr 1993/94) zu absolvieren hatten. Die vorangehenden Erprobungen dieses Tests in mehreren deutschen und englischen Schulen haben zu Ergebnissen geführt, die auch für sich genommen interessant erscheinen und die wir daher hier zur Diskussion stellen möchten<sup>1</sup>.

Die eben erwähnte **Verbindung verschiedener Methoden** (eher quantitative im Rahmen der begonnenen Langzeitstudie und eher qualitative in den Fallstudien) ist bewußt geschehen. Die Forderung nach einer solchen Verbindung - vor allem bei umfassenderen Fragestellungen - wird innerhalb der einschlägigen mathematikdidaktischen Diskussion zunehmend erhoben<sup>2</sup>; siehe etwa Beck/Maier (1993). Damit ergeben sich neben den im Gesamtprojekt in der Zusammenschau quantitativer und qualitativer Methoden erzielten Ergebnissen natürlich auch Resultate aus **einzelnen** Projektkomponenten, die auch allein in sich stimmig und von Interesse sind. Um solche Resultate handelt es sich in diesem Beitrag.

Ein kurzer **Überblick**: In Abschnitt 2 beschreiben wir die Entwicklung des Tests, und in Abschnitt 3 geben wir Informationen über seine Erprobung. In beiden Abschnitten thematisieren wir bewußt auch einige dabei aufgetretene methodologische Fragen. Die Ergebnisse der Testerprobung werden dann in Abschnitt 4, dem Hauptteil dieses Beitrags, erläutert und interpretiert, wobei vor allem die aufgetretenen Unterschiede bei einzelnen mathematischen Fähigkeiten interessant und aufschlußreich sind.

---

<sup>1</sup> Ein Teil dieser Ergebnisse ist auch in Blum et al. (1993) enthalten.

<sup>2</sup> Diese Methodenvielfalt erscheint gerade unserem Untersuchungsgegenstand als besonders angemessen, da bisher noch kein für die ausschließliche Verwendung quantitativer Methoden nötiger vorab entwickelte Theorierahmen, der die Datenerhebung und Datenauswertung steuert, existiert. Vielmehr sind qualitative Methoden, die zu einer empirisch begründeten Theoriebildung geeignet sind und auf einer offenen Datenerhebung sowie einer interpretativen Datenauswertung beruhen, ebenfalls nötig.

## (2) ENTWICKLUNG DES TEST

Zunächst zu unseren **Grundannahmen**: Wir gehen davon aus, daß es sinnvoll ist, von der **mathematischen Leistungsfähigkeit** (dem mathematischen "**Potential**") von etwa 13/14jährigen Jugendlichen in England und Deutschland zu sprechen. Dieses theoretische **Konstrukt** umfaßt sämtliche mathematischen **Kenntnisse, Fertigkeiten und Fähigkeiten**, wie auch immer sie zustande gekommen sind (angeboren, durch Unterricht ausgebildet, im Alltag erworben o.a.). Des weiteren gehen wir davon aus, daß dieses Potential (gemessen etwa als Gesamtpunktzahl für Aufgabenlösungen in einem geeignet konzipierten Test) bei den deutschen und bei den englischen 13/14jährigen im wesentlichen **identisch verteilt** ist. Bei **einzelnen Komponenten** der mathematischen Leistungsfähigkeit können natürlich relevante **Unterschiede** zwischen beiden Ländern auftreten (etwa indem bei einem solchen Test die beiden jeweiligen Mittelwerte über alle 13/14jährigen signifikant voneinander verschieden sind). Unsere eben genannte Annahme besagt aber, daß sich diese Einzel-Unterschiede in der Summe über alle Komponenten i.w. aufheben.

Eine weitere Ausgangs-Hypothese ist, daß solche Einzel-Unterschiede **instruktionsbedingt** sind, d.h. durch die Unterschiedlichkeit des jeweils vorangegangenen Mathematikunterrichts verursacht worden sind. Hiermit wird natürlich keineswegs behauptet, es gäbe nicht noch andere Variable, die für die Ausformung mathematischer Fähigkeiten relevant sind, so etwa die organisatorischen und sozialen Rahmenbedingungen des schulischen Lehrens und Lernens. Unsere Hypothese bezieht sich nur auf die Erklärung von **Unterschieden**.

Unser Ziel war es, das Konstrukt "Potential" mittels eines geeigneten Tests **für unsere Zwecke** zu **operationalisieren**. Und zwar ging es uns, wie in Abschnitt 1 gesagt, um die Identifikation vergleichbarer (Teil-)Populationen für unsere geplante Langzeitstudie. Im Sinne unserer Ausgangs-Annahmen sollte unser Test so konstruiert werden, daß er - bei Anwendung auf repräsentative Stichproben in beiden Ländern - sowohl

- genügend "**trennscharf**" ist, d.h. in beiden Ländern eine Unterscheidung nach verschiedenen Fähigkeitsniveaus gestattet, als auch
- zu beiden Ländern "**fair**" ist, d.h. i.w. identische Verteilungen der Gesamtergebnisse liefert (genauer: es dürfen und sollen sich sehr wohl bei einzelnen Fähigkeitskomponenten instruktionsbedingte Unterschiede zwischen beiden Ländern zeigen, diese sollen sich aber insgesamt herausmitteln).

In einem solchen Test können natürlich nur ausgewählte Fähigkeitskomponenten durch Aufgaben repräsentiert sein. Nach welchen Gesichtspunkten haben wir die für uns relevanten Komponenten und zugehörige Testaufgaben ausgewählt?

Eine recht restriktive Ausgangsvoraussetzung war, daß aus schulorganisatorischen Gründen ein

Test zu entwickeln war, der in **einer Schulstunde** (d.h. in 40 Minuten) durchgeführt werden kann. Daher war es nicht sinnvoll zu versuchen, sehr viele verschiedene (oder gar "alle", was immer dies heißen mag) Fähigkeitskomponenten abzudecken. Insbesondere schloß dieser Zeitrahmen das Testen "höherer" Fähigkeiten wie Definieren, Mathematisieren oder Beweisen ebenso aus wie das Bearbeiten umfangreicherer neuartiger Situationen im Stile der in England üblichen "Investigations". Weiter war unser Ziel nicht die Entwicklung eines universellen, alters- und kulturunabhängigen Fähigkeitstests, sondern wie gesagt eines Tests für einen definierten Zweck. Deshalb haben wir uns nicht an der vorhandenen allgemeinen Literatur zum Themenkomplex "mathematische Fähigkeiten" orientiert, sondern sind von vorneherein eigene Wege gegangen.

Grob gesprochen sollte unser Test **zwei Typen** von Fähigkeitskomponenten enthalten:

Zum ersten sollten Komponenten berücksichtigt werden, die möglichst wenig "instruktionsabhängig" sind, d.h. die nicht direkt spezifischen Themen des Mathematikunterrichts zugeordnet werden können, etwa entsprechend solchen Fähigkeiten, wie sie bei **Intelligenztests** abgeprüft werden. Hierzu haben wir uns an gängigen Intelligenztests (z.B. Mannheimer Intelligenztest [MIT], Kognitiver Fähigkeitstest [KFT] oder Leistungsprüfsystem [L-P-S]) orientiert.

Andererseits ist klar, daß direkte Bezüge zum Mathematikunterricht unvermeidbar, ja sinnvoll und notwendig sind; z.B. gehören einfache Rechenfertigkeiten sicherlich auch zum "mathematischen Potential". Deshalb sollten zum zweiten auch erkennbar **instruktionsabhängige** Komponenten hinzugezogen werden. Hierbei sollten die jeweiligen Schwerpunkte in den **Mathematik-Curricula** in England und Deutschland angemessen (im Sinne der intendierten Fairness und Trennschärfe, s.o.) berücksichtigt werden (z.B.: "Dreisatz" als eher deutscher oder "Graphen-Interpretation" als eher englischer Schwerpunkt). Hierzu haben wir die Sek.I-Curricula in beiden Ländern eingehend analysiert, z.T. auch mit Hilfe der Analysen bei Howson (1991) (beide Länder), bei DES (1991) bzgl. England und bei Fanghänel/Stamm/Weber (1992) bzgl. Deutschland. Um nicht nur die intendierten, sondern auch die faktisch implementierten Curricula zu berücksichtigen, haben wir zusätzlich mit zahlreichen Lehrenden diskutiert. Wir möchten behaupten, daß diese explizite und konstitutive Berücksichtigung der Curricula gerade eine Stärke unseres Tests ist, verglichen mit Tests in globaleren internationalen Studien.

Insgesamt haben wir **zwölf Fähigkeitskomponenten** ausgewählt:

- |  |   |
|--|---|
| [1] Einfaches Zahlenrechnen                  | [7] Räumliches Vorstellungsvermögen             |
| [2] Erkennen von Strukturen von Zahlenfolgen | [8] Interpretieren realer Graphen               |
| [3] Angewandtes algebraisches Rechnen        | [9] Vorstellungen v. Wahrscheinlichkeitsbegriff |
| [4] Umgehen mit proportionalen Beziehungen   | [10] Kombinatorisches Schließen                 |
| [5] Zahlensymbolrätsel                       | [11] Logisches Schließen                        |
| [6] Umgehen mit ebenen Figuren               | [12] Erkennen von Strukturen                    |

Diesen Komponenten haben wir (i.a. mehrere) **Testaufgaben** zugeordnet, orientiert an Aufgaben

aus Intelligenztests bzw. an Mathematik-Aufgaben bekannten Typs, im Sinne der eben beschriebenen Leitlinien. Das Validierungsverfahren für unsere Testaufgaben orientierte sich am Konzept der inhaltlichen Validierung, das ja bekanntlich auf der Annahme basiert, daß die Testaufgaben selbst das bestmögliche Kriterium für das zu untersuchende Persönlichkeitsmerkmal darstellen ("logische Validität"). Dementsprechend wurde mehrere kompetente Beurteiler in beiden Ländern konsultiert.

**Unser Konzept** von "mathematischer Leistungsfähigkeit" wird also durch die Summe all der Fähigkeiten definiert, die in unseren Testaufgaben enthalten sind. Natürlich erfaßt diese pragmatische Arbeitsdefinition nur einen kleinen Teil der Vielfalt, die im theoretischen Konstrukt enthalten ist. Trotzdem behalten wir hierfür denselben Begriff bei.

Wir haben die ersten Testversionen in England und in Deutschland mehrfach ausprobiert und revidiert, um Unklarheiten in den Aufgabenstellungen zu beseitigen oder um einzelne Aufgaben auszutauschen oder zu eliminieren. Dabei haben wir auch darauf geachtet, daß die deutschen und englischen Fassungen der Testaufgaben sprachlich äquivalent sind, u.a. durch Prüfung der Übersetzungen durch einen bilingualen Mathematiklehrer<sup>3</sup>. Hierdurch ist der im Anhang in seiner deutschen Version wiedergegebene **Test mit 28 Aufgaben** entstanden; die Aufgaben entsprechen folgenden Fähigkeitskomponenten (s.o.):

[1]: Aufgaben 2, 5	[7]: Aufgaben 7, 21
[2]: Aufgaben 1, 17, 19, 22	[8]: Aufgaben 6, 23
[3]: Aufgaben 8, 10, 13, 24	[9]: Aufgabe 9
[4]: Aufgaben 11, 15, 20, 26	[10]: Aufgaben 14, 27
[5]: Aufgabe 28	[11]: Aufgabe 25
[6]: Aufgaben 3, 12, 16	[12]: Aufgabe 4, 18

### (3) ERPROBUNG DES TESTS

Wir wollten unseren Test in beiden Ländern an **repräsentativen** Stichproben erproben um zu überprüfen, ob er für unsere Zwecke brauchbar ist (siehe Abschnitt 2), speziell auch um auftretende Mittelwert-Unterschiede bei einzelnen Fähigkeitskomponenten (das wohl Interessanteste dabei!) genügend plausibel und verläßlich als instruktionsbedingt interpretieren zu können und nicht etwa als Folge anderer Einflußfaktoren wie etwa durchschnittliche allgemeine "Intelligenz", mittleres Lebensalter oder eben durchschnittliche "mathematische Leistungsfähigkeit". Zur Reprä-

<sup>3</sup> Wie wichtig bzw. schwierig die Herstellung wirklich äquivalenter Fragen ist, insbesondere wenn die gewählten Formulierungen den in den jeweiligen Unterrichtsmaterialien verwendeten entsprechen sollen, zeigt die Untersuchung von Hanna (1993) an der Second International Mathematics Study. Die englische Fassung des Test ist in Blum et.al. (1993) wiedergegeben.

sentativität einer Stichprobe ist also u.a. vergleichbares mathematisches Potential notwendig, was wiederum nur mittels "repräsentativer" Stichproben überprüfbar ist. Dieser *Circulus vitiosus* beschreibt ein grundsätzliches methodologisches Problem bei empirischen Untersuchungen, das prinzipiell als kaum befriedigend lösbar anzusehen ist; insbesondere ist dieses Problem auch nicht durch größere Stichproben (wie bei den großen internationalen Studien; vgl. etwa Robitaille/Donn, 1993) in den Griff zu bekommen.

Wir haben dieses Problem pragmatisch gelöst, indem wir in beiden Ländern Stichproben ausgewählt haben, die in ihrer Zusammensetzung das jeweilige staatliche Schulwesen widerspiegeln. Unsere Arbeitshypothese lautete dann, daß die so ausgewählten Gruppen in für unsere Zwecke ausreichendem Maße repräsentativ sind. Insgesamt wurde der Test in 1992/93 mit 496 deutschen und 540 englischen Lernenden durchgeführt, und zwar jeweils anonym.

In **Deutschland** wurden sechs Schulen in Stadt und Landkreis Kassel ausgewählt, zwei Gymnasien, eine Haupt- und Realschule, zwei additive und eine integrierte Gesamtschule(n). Die beteiligten Lernenden besuchten die Jahrgangsklassen 6 (9%), 7 und 8, ihr Durchschnittsalter lag - wie von uns gewünscht - bei etwa 13 1/2 Jahren. Die Verteilung der Lernenden auf die entsprechenden Schultypen ist mit etwa 46% Gymnasialschülern, 34% Realschülern und 20% Hauptschülern als nahezu repräsentativ für die Situation der Staatsschulen in Kassel und Umgebung anzusehen (vgl. Hessisches Kultusministerium, 1992, S.32 bzw. Daten des Staatlichen Schulamts Kassel für 1993).

Bei den beteiligten Schulen in **England** handelt es sich um fünf (integrierte) Gesamtschulen (Comprehensive Schools)<sup>4</sup> aus unterschiedlichen Gegenden Englands (aus London und Umgebung, dem Norden sowie dem Westen Englands), und zwar sowohl aus innerstädtischen wie ländlichen Bezirken. Die beteiligten Lernenden besuchten die Jahrgangsklassen 7 (17%), 8 und 9 (dies entspricht i.w. den deutschen Jahrgängen 6 bis 8), ihr Durchschnittsalter war mit etwa 13 Jahren geringer als das der deutschen Kinder. Andererseits waren die englischen Kinder bereits jeweils ein Jahr länger zur Schule gegangen<sup>5</sup>. Da sich nachträglich zeigte, daß die Testresultate in England mit wachsendem Alter im Mittel nur ganz geringfügig besser geworden sind, erweist sich die Altersfrage für England ohnehin als kaum bedeutsam. Alle beteiligten Schulen besitzen im Fach Mathematik ein volles Differenzierungsangebot für Kinder aller Fähigkeitsniveaus, wobei 3 bis 5 Niveaus unterschieden werden. Unsere Stichprobe kann somit als repräsentativ für die Situation an englischen Gesamtschulen angesehen werden. Allerdings muß offen bleiben, inwie-

---

<sup>4</sup> Comprehensive Schools sind die in England mit Abstand dominierende Schulform, sie werden von über 90% der Lernenden besucht.

<sup>5</sup> Ob man sich bei der Auswahl der Stichprobe auf das Lebensalter oder das Schulalter der Kinder bezieht, ist ein prinzipielles Problem aller internationaler Vergleichsuntersuchungen, das sich in der Regel nicht befriedigend lösen läßt.

weit das gesamte Fähigkeitsspektrum bei englischen Kindern durch die Staatsschulen adäquat erfaßt wird, da bei den Gegenden, aus denen unsere Schulen stammen, davon auszugehen ist, daß ein nicht vernachlässigbarer Anteil höher begabter Kinder Privatschulen (häufig selective Grammar Schools) besucht. Diese Frage stellt sich allerdings auch für Deutschland; der entsprechende Prozentsatz ist jedoch (noch) etwas geringer als in England.

#### (4) TESTERGEBNISSE UND INTERPRETATIONEN

Die **Ergebnisse** der Testerprobung lassen sich in verschiedener Hinsicht auswerten, z.B. bzgl. der Anteile richtiger Lösungen bei den einzelnen Testaufgaben (diese Daten sind im Anhang wiedergegeben)<sup>6</sup>; auf den Vergleich dieser Anteile konzentrieren wir uns im folgenden. Im Sinne der beschriebenen Anlage unserer Untersuchung werden wir dabei kleinere, statistisch nicht signifikante Unterschiede zwischen den beiden Gruppen nicht besonders betonen, insbesondere da bei solchen Untersuchungen natürlich auch mit zufälligen Unterschieden gerechnet werden muß. Bei **signifikanten** Unterschieden soll, wie vorhin gesagt, stets versucht werden, ihre Ursache auf der Grundlage unserer Kenntnisse über den Mathematikunterricht in Deutschland und in England (siehe Kaiser-Meßmer/Blum, 1993) zu **interpretieren** und zu erklären. Ziel sind dabei eher globale Aussagen über ganze Aufgaben-Klassen und nicht Detail-Aussagen über einzelne Aufgaben.

Zunächst ist zu sagen, daß Aufgabe 9 nicht in die Auswertung eingehen kann, da aufgrund eines Abstimmungsfehlers in England eine etwas anspruchsvollere Version (mit vier statt zwei Beuteln) verwendet worden ist.

Global gesehen ergeben sich bei etwa zwei Dritteln der Aufgaben signifikante Unterschiede in den Resultaten der englischen und der deutschen Lernenden, die unter Benutzung des Chi-Quadrat-Tests auf dem 1%- (bzw. 2,5%/5%-)Niveau abgesichert werden können. Dabei sind in etwa 60% der Aufgaben, in denen signifikante Unterschiede festgestellt werden können, diese zugunsten der deutschen und in etwa 40% zugunsten der englischen Lernenden.

Der **Mittelwert** der richtigen Antworten aller Lernenden bezogen auf den gesamten Test (stets ohne Aufgabe 9) liegt mit 13,34 bei den deutschen Lernenden deutlich höher als mit 12,79 bei den englischen Lernenden. Bei der zugehörigen Streuung zeigt sich ein Unterschied insofern, als die englische Bandbreite etwas größer ist (Standardabweichungen 4,63 bzw. 4,96). Die **Gesamtzahl** der richtigen Antworten (aus den Daten im Anhang nicht zu entnehmen) ist in beiden Ländern in sehr befriedigender Näherung normalverteilt, so wie erwünscht. Daher ist auch der übliche Test auf Signifikanz des Unterschieds zwischen den Mittelwerten anwendbar. Er zeigt, daß dieser Unterschied auf dem 5%-Niveau signifikant ist. Das bedeutet: Entweder war unser Test nicht

---

<sup>6</sup> Die vollständigen Ergebnisse sind bei den Autoren erhältlich.

"fair" konstruiert, oder die Stichproben waren nicht genügend repräsentativ, oder beides ist der Fall. Wir kommen sogleich auf diese Frage zurück.

Im folgenden nun die **Ergebnisse im Detail**:

1. Beim Themengebiet **Arithmetik/Algebra** (Fähigkeitskomponenten [1] bis [5]) erzielten die deutschen Lernenden insgesamt signifikant bessere Resultate. So hatten sie bessere Resultate in allen sechs Aufgaben zum **Zahlenrechnen** (Nr. 2, 5) und zu **proportionalen Beziehungen** (Nr. 11, 15, 20, 26), wobei die Unterschiede bei vier der sechs Aufgaben auf dem 1%-Niveau signifikant sind (bei Aufgabe 2, 5, 11, 15). Auch bei zwei der vier Fragen zum **angewandten algebraischen Rechnen** erzielten die deutschen Lernenden signifikant bessere Resultate (Nr. 8 und 10). Die englischen Lernenden erreichten signifikant bessere Resultate in den beiden anderen Aufgaben (Nr. 13 und 24). Dabei verlangt Aufgabe Nr. 24 (die insgesamt von nur wenigen richtig gelöst wurde) komplexere algebraische Vorgehensweisen oder (angemessener) Versuch-und-Irrtum-Methoden. Eine mögliche Erklärung für die besseren Resultate der englischen Lernenden könnte u.a. darin liegen, daß diese mit solchen Methoden besser vertraut sind, da solche Vorgehensweisen beim entdeckenden Lernen bzw. im individualisierten Unterricht, wie er in England weit verbreitet ist, stärker gefördert und geübt werden.

In den Aufgaben zur **Struktur von Zahlenfolgen** (einer der Intelligenztest-Komponenten) erzielten die deutschen Lernenden in drei von vier Aufgaben bessere Ergebnisse (Nr. 1, 19 und 22), wobei die Unterschiede aber nur in Aufgabe 19 auf dem 1%-Niveau signifikant sind. Aufgabe Nr. 17, bei der die englischen Lernenden bessere Resultate erreichten (auf dem 2,5%-Niveau signifikant), benutzt Quadratzahlen, die den Lernenden einer der beteiligten deutschen Schulen nach Aussagen der beteiligten Lehrpersonen bis dato noch nicht explizit begegnet waren. Berücksichtigt man nur die Lernenden der anderen deutschen Schulen, so ergeben sich nur noch unerhebliche Unterschiede (22% zu 21% richtige).

In der einzigen Frage zu **Zahlensymbolrätseln** (Nr. 28) erreichten die deutschen Lernenden ebenfalls auf dem 1%-Niveau signifikant bessere Ergebnisse.

Eine naheliegende Erklärung für die besseren Resultate der deutschen Lernenden insbesondere im Bereich einfacher Zahlenrechnungen, Prozentrechnung und proportionaler Zuordnungen ist die hohe Bedeutung, die diesen Themen im deutschen Mathematikunterricht zukommt, und die dementsprechend hohe Unterrichtszeit, die ihnen in der Sek.I gewidmet wird. Der deutsche Mathematikunterricht ist - wie bereits an anderer Stelle ausgeführt (siehe Kaiser-Meßmer/Blum, 1993) - u.a. charakterisiert durch seine Orientierung an der präzisen Ausführung von Rechenalgorithmen, insbesondere in den oben genannten Bereichen. Dagegen weist der englische Mathematikunterricht der Übung von Rechenalgorithmen eine weit geringere Bedeutung zu und betont demgegenüber mehr ein entdeckendes Lernen (sog. Investigations), dessen Effekte deutlich schwerer durch einen Test meßbar sind.

2. Im Bereich **Geometrie** (Fähigkeitskomponenten [6] und [7]) erreichten die englischen Lernenden insgesamt signifikant bessere Ergebnisse. In den Fragen zu **ebenen Figuren** (Nr. 3, 12, 16) waren die englischen Lernenden in zwei von drei Aufgaben signifikant besser. Bei der Aufgabe, die Figuren in einem Gitternetz beinhaltet (Nr. 12), zeigten sich die stärksten Unterschiede zugunsten einer Gruppe innerhalb des gesamten Tests. Eine Erklärung hierfür ist darin zu sehen, daß im Mathematikunterricht an deutschen Schulen solche Fragestellungen in Gitternetzen eher wenig behandelt werden, während sie in England absolut üblich sind. Bei der dritten Aufgabe mit ebenen Figuren (Nr. 16) erzielten die deutschen Lernenden auf dem 5%-Niveau signifikant bessere Ergebnisse; allerdings beinhaltet diese Aufgabe nur zum Teil geometrische Aspekte (Drehungen) und betont noch stärker strukturelle und logische Aspekte.

Bei den Aufgaben zu **räumlichen Figuren** erreichten in der einfacheren Aufgabe (Nr. 7) die deutschen Lernenden bessere Resultate (allerdings nicht signifikant), währenddessen in der komplexeren Aufgabe (Nr. 21) die englischen Lernenden etwas besser waren (ebenfalls nicht signifikant).

3. Beim Themenbereich **Graphen/Funktionen** (Fähigkeitskomponente [8]), wo es um die Interpretation von Graphen geht, die Realsituationen beschreiben (Nr. 6, 23), erzielten die englischen Lernenden bei einer Aufgabe (Nr. 6) signifikant bessere Ergebnisse, bei der anderen Aufgabe zeigten sich keine Unterschiede.

Die Ergebnisse zu Nr. 6 sind durch die im Vergleich zu England deutlich geringere Bedeutung, die nach unseren Beobachtungen solchen Interpretationen von Graphen im deutschen Mathematikunterricht üblicherweise zukommt, leicht erklärbar. Die Ergebnisse zu Nr. 23 müssen dann aber überraschen.

4. In den Themengebieten **Kombinatorik/Wahrscheinlichkeit/Logik** (Fähigkeitskomponenten [9] bis [12]) ergaben sich in drei von fünf Aufgaben (Nr. 9 ist stets ausgeschlossen) keine wesentlichen Unterschiede.

Bei den zwei Aufgaben zur **Kombinatorik** (Nr. 14 und 27) zeigten sich kaum Unterschiede zwischen den englischen und den deutschen Lernenden. In der anspruchsvolleren Aufgabe (Nr. 27), die allgemeine kombinatorische Fähigkeiten abtestet, versagten nahezu alle Lernenden. Die Ergebnisse der zwei intelligenztestartigen Aufgaben (Nr. 4 und 18) zu **Strukturen**, die ebenfalls gewisse kombinatorische Aspekte beinhalten, waren nicht einheitlich; so ergaben sich kaum Unterschiede in der einfacheren Aufgabe (Nr. 4), in der komplexeren Aufgabe (Nr. 18) erzielten die englischen Lernenden signifikant bessere Resultate.

In der Aufgabe zur **Logik** (Nr. 25) waren die deutschen Lernenden signifikant besser. Folgende Erklärung für dieses Ergebnis erscheint plausibel: Obwohl Logik üblicherweise nicht explizit in den deutschen Lehrplänen verankert ist, betont der deutsche Mathematikunterricht, insbesondere am Gymnasium, viel stärker logische Überlegungen und Schlußfolgerungen als der englische, wodurch auch logische Fähigkeiten gefördert werden. Es ist allerdings nicht auszuschließen, daß bei

dieser Aufgabe, die besonders stark sprachlich geprägt ist, die beiden Fassungen sprachlich doch nicht äquivalent sind.

So viel zu den Ergebnissen bei einzelnen Aufgaben(-Klassen). Wir haben noch weitere Analysen vorgenommen, z.B. eine getrennte Auswertung in Deutschland nach den drei Schulformen (wobei sich zeigt, daß auch für die Gesamtpunktzahlen in diesen drei Untergruppen approximativ Normalverteilungen resultieren und daß unser Test sehr deutlich zwischen den drei Schulformen trennt) oder eine Auswertung in beiden Ländern getrennt nach Jahrgangsstufen (wobei, wie schon kurz erwähnt, in England die Mittelwerte überraschenderweise kaum besser wurden, was sich auch als Folge der dort verwendeten Materialien interpretieren läßt, die weit weniger fachsystematisch aufgebaut sind als in Deutschland üblich). Auch eine genauere Analyse der Verteilungen zeigt weitere interessante Resultate (z.B. daß die größere englische Streuung auch von der deutlich höheren Zahl von englischen Lernenden mit alles richtig bzw. nur 1 bis 2 falschen Antworten herrührt<sup>7</sup>). Auf all dies wollen wir nicht weiter eingehen. Weitere statistische Verfahren (z.B. multivariate Regressionsmodelle; vgl. etwa Knoche, 1990) haben wir nicht verwandt.

Die obige detaillierte Analyse und Interpretation der Ergebnisse des Tests macht nochmals deutlich, daß die aufgetretenen Einzel-Unterschiede wohl tatsächlich - ganz im Sinne unserer Hypothese - im wesentlichen vom vorangegangenen Unterricht determiniert sind. Der Unterschied im Mittelwert zugunsten der deutschen Lernenden ist dann weitgehend erklärbar durch die stärkere Berücksichtigung arithmetisch-algebraischer Aufgaben im Vergleich mit geometrischen. Insofern war unser Test offenbar **nicht** genügend "fair" konzipiert.

Insbesondere die Ergebnisse zu den Aufgaben Nr. 16 (deutscher Vorsprung trotz nicht erkennbarem Unterrichtsbezug) und Nr. 23 (kein Unterschied trotz Schwerpunktsetzung im englischen Unterricht) könnten allerdings auch so interpretiert werden, daß die englische Gruppe im Mittel doch ein wenig "schwächer" war als die deutsche (gemessen an der allgemeinen mathematischen Fähigkeit der Lernenden; vgl. unsere Bemerkungen in Abschnitt 3). Dem widerspricht jedoch das Ergebnis zu Aufgabe Nr. 18. Da solche Einzel-Ergebnisse auch durch Zufallsschwankungen zustande gekommen sein können, ist es nicht sinnvoll, sie weiter zu diskutieren. Jedenfalls beeinflußt dies unsere instruktionsbezogenen Interpretationen von (signifikanten) Unterschieden bei einzelnen Aufgaben(klassen) offensichtlich in keiner Weise.

Davon unabhängig ist aber die vielleicht wichtigste Erkenntnis aus unserer Untersuchung die folgende: Verallgemeinernde Aussagen über Unterschiede in den mathematischen Leistungen zwischen deutschen und englischen Lernenden, insbesondere über global deutlich schlechtere Lei-

---

<sup>7</sup> Hierin könnte sich eine in England vorhandene Tendenz zur Eliteförderung (durch frühe leistungsmäßige Differenzierung auch innerhalb des Gesamtschulwesens) ausdrücken.

stungen der englischen, wie sie in anderen Studien (u.a. von Prais/Wagner, 1986) formuliert werden, sind nach unseren Kenntnissen mit gewissen **Einschränkungen** und **Spezifizierungen** zu versehen. Natürlich wird ein Fähigkeitstest mit anderen Zielen und anderen Schwerpunktsetzungen wohl andere Ergebnisse hervorbringen als unser Test. Insbesondere wird ein Test, der in noch stärkerem Maße arithmetisch-algebraisch-logische Fähigkeiten anspricht, das Resultat erbringen, das deutsche Kinder unvergleichlich höhere mathematische Fähigkeiten haben als gleichaltrige englische. All dies ist eine Frage der Gewichtung verschiedener Fähigkeitskomponenten.

### (5) AUSBLICK

Der hier diskutierte Test zur mathematischen Leistungsfähigkeit war - wie mehrfach betont - konzipiert worden, um vergleichbare Schülergruppen in England und Deutschland identifizieren zu können, die an der jetzt laufenden Langzeit-Untersuchung beteiligt sind. Es war für uns daher konsequent, bei der Konstruktion der endgültigen, zu Beginn des Schuljahres 1993/94 eingesetzten Version des Tests gewisse Fragen aus der hier vorgestellten Fassung herauszunehmen bzw. auszutauschen, mit dem Ziel, den Test wirklich "fair" im Hinblick auf die in England und Deutschland resultierenden Verteilungen zu machen. Hierüber wird - im Zusammenhang mit der Langzeitstudie - zu gegebener Zeit zu berichten sein.

### LITERATUR

BECK, C.; MAIER, H. (1993). Das Interview in der mathematikdidaktischen Forschung. In: *Journal für Mathematik-Didaktik*, 14, 147-179.

BLUM, W.; BURGHESE, D.; GREEN, N.; KAISER-MESSMER, G. (1992). Teaching and Learning of Mathematics and its Applications: First Results from a Comparative Empirical Study in England and Germany. In: *Teaching Mathematics and its Applications*, 11, 3, 112-123.

BLUM, W.; BURGHESE, D.; GREEN, N.; KAISER-MESSMER, G. (1993). British/German Comparative Project: Some Preliminary Results. In: *Teaching Mathematics and its Applications*, 12, 1, 13-21.

BROWN, M. (1993). Assessment in Mathematics Education: Developments in Philosophy and Practice in United Kingdom. In: M. Niss (Ed.), *Cases of Assessment in Mathematics Education*. Dordrecht: Kluwer, 71-84.

CRESSWELL, M.; GUBB, J. (1987) *The Second International Mathematics Study in England and Wales*. Windsor: NFER-Nelson.

DEPARTMENT OF EDUCATION AND SCIENCE (1991). *Mathematics in the National*

Curriculum. London: DES.

EDUCATIONAL TESTING SERVICE (1992). *Learning Mathematics*. Princeton: Educational Testing Service.

FANGHÄNEL, G.; STAMM, R.; WEBER, K. (1992). *Mathematikunterricht in Ländern der Bundesrepublik Deutschland*. Berlin: Technische Universität Berlin, Fachbereich Mathematik, Preprint Nr.4.

HANNA, G. (1993). *The Validity of International Performance Comparisons*. In: M. Niss (Ed.), *Investigations into Assessment in Mathematics Education*. Dordrecht: Kluwer, 245-252.

HESSISCHES KULTUSMINISTERIUM (1992) (Hrsg.). *Hessische Bildungspolitik in Zahlen - Daten aus dem Schulbereich 1991*. Wiesbaden: Hessisches Kultusministerium.

HOWSON, G. (1991). *National Curricula in Mathematics*. Leicester: The Mathematical Association.

KAISER-MESSMER, G.; BLUM, W. (1993). *Einige Ergebnisse von vergleichenden empirischen Untersuchungen (Deutschland/England) zum Lehren und Lernen von realitätsbezogener Mathematik*. In *Journal für Mathematik-Didaktik*, 14, 3/4, 269-305.

KNOCHE, N. (1990). *Modelle der empirischen Pädagogik*. Mannheim: BI.

PRAIS, S.J.; WAGNER, K. (1986). *Schooling Standards in England and Germany - Some Summary Comparisons Bearing on Economic Performance*. In: *Compare*, 16, 1, 5-35.

ROBITAILLE, D.; DONN, S. (1993). *TIMSS: The Third International Mathematics and Science Study*. In M. Niss (Ed.), *Investigations into Assessment in Mathematics Education*. Dordrecht: Kluwer, 229-243.

ROBITAILLE, D.; GARDEN, R. (Eds.) (1988). *The IEA Study of Mathematics II*. Oxford: Pergamon Press.

Werner Blum und Gabriele Kaiser-Meißner  
Universität Gesamthochschule Kassel  
Fachbereich Mathematik/Informatik  
D-34109 Kassel

David Burghes  
University of Exeter  
Centre for Innovation in Mathematics Teaching  
St. Luke's, Exeter EX1 2LU  
England

Nigel Green  
Evesham High School  
Mathematics Department  
Four Pools Road, Evesham WR11 6DQ  
England

## ANHANG

## ERGEBNISSE ZU DEN EINZELNEN AUFGABEN

Aufgaben- nummer	Prozentsatz richtiger Antworten		Chi-Quadrat-Wert $\chi^2$	Signifikanzniveau
	England n=540	Deutschland n=496		
1	98	<b>99</b>	0,38	
2	79	<b>90</b>	<b>21,70</b>	p<0,01
3	<b>79</b>	71	<b>9,11</b>	p<0,01
4	<b>68</b>	67	0,08	
5	90	<b>98</b>	<b>26,63</b>	p<0,01
6	<b>42</b>	34	<b>6,29</b>	p<0,025
7	65	<b>71</b>	4,75	
8	33	<b>43</b>	<b>11,40</b>	p<0,01
9	64	<b>90</b>	Nicht anwendbar	
10	63	<b>75</b>	<b>18,05</b>	p<0,01
11	33	<b>45</b>	<b>16,20</b>	p<0,01
12	<b>39</b>	24	<b>26,62</b>	p<0,01
13	<b>65</b>	59	<b>5,23</b>	p<0,025
14	66	<b>68</b>	0,47	
15	26	<b>37</b>	<b>13,99</b>	p<0,01
16	72	<b>77</b>	<b>4,22</b>	p<0,05
17	<b>22</b>	17	<b>5,26</b>	p<0,025
18	<b>42</b>	35	<b>5,38</b>	p<0,025
19	44	<b>57</b>	<b>15,93</b>	p<0,01
20	17	<b>19</b>	0,65	
21	<b>40</b>	37	0,82	
22	78	<b>79</b>	0,12	
23	34	34	0,04	
24	<b>13</b>	7	<b>13,56</b>	p<0,01
25	29	<b>39</b>	<b>10,32</b>	p<0,01
26	22	<b>26</b>	2,45	
27	4	<b>5</b>	0,55	
28	16	<b>25</b>	<b>12,59</b>	p<0,01
Mittelwert (ohne Frage 9)	12,79	<b>13,34</b>		
Standard- abweichung	<b>4,96</b>	4,63		

Schulform:

Bearbeitungszeit: 40 Minuten

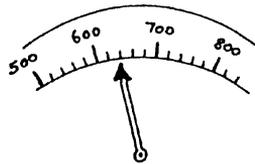
Schreibe die Antworten in die dafür vorgesehenen Kästen auf der rechten Seite.

Du kannst die Nebenrechnungen schriftlich auf Extrapapier durchführen. Taschenrechner dürfen nicht benutzt werden.

1. Wie lautet die nächste Zahl in dieser Folge?

25 20 15 10 ...

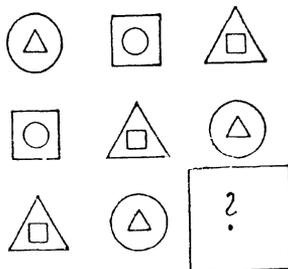
2. Auf welche Zahl deutet der Zeiger?




3. Gib die Nummern aller zusammengesetzter Figuren an, die aus den beiden gegebenen Stücken hergestellt werden können (jedes Stück darf nur einmal benutzt werden).

gegebene Stücke	
zusammengesetzte Figuren	

4. Gib die Nummer derjenigen Figur an, die das Muster vervollständigt; wähle aus den abgebildeten Figuren.

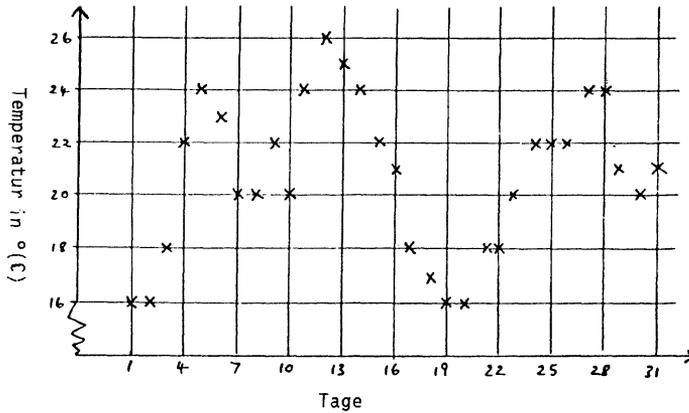


- 1 2 3 4 5 6

5. Eine Frau besitzt 100 DM. Sie verdient weitere 40 DM und gibt 70 DM aus. Wieviel besitzt sie nun?

6. Das Diagramm zeigt die Tageshöchsttemperaturen eines Ferienorts während des Monats Juli.

An wieviel Tagen betrug die Höchsttemperatur zumindest 20°C ?



7. Wieviele Flächen hat dieser Körper?




8. Eine Gärtnerin pflanzt Samen in Schalen. Sie hat 20 Schalen, und sie benötigt 10 Minuten für jede Schale. Sie beginnt ihre Arbeit morgens um 7.30. Um wieviel Uhr wird sie drei Viertel der Schalen bepflanzt haben?

9. Wir haben zwei Beutel. Jeder Beutel enthält schwarze und weiße Kugeln:

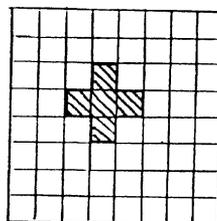
Beutel A: 12 schwarze und 4 weiße  
 Beutel B: 20 schwarze und 20 weiße

Du sollst aus einem der Beutel eine Kugel herausnehmen (mit geschlossenen Augen). Bei welchem Beutel hast Du die besten Chancen, eine schwarze Kugel zu bekommen?

10. Ich denke mir eine Zahl. Ich verdopple sie und nehme 13 weg. Die Antwort ist 45. Wie heißt die Zahl, die ich mir ursprünglich gedacht habe?

11. Zwanzig Mandarinen kosten 4,20 DM. Wieviel kosten wohl 12 Mandarinen?

12. Zeichne soviele Kreuze als möglich in das Gitternetz, so daß
- sie dieselbe Größe haben wie das gegebene;
  - sie sich nicht überlappen (sie dürfen sich berühren).



13. Jeder Sack wiegt das Gleiche, und die Waage ist im Gleichgewicht. Wieviel wiegt ein einzelner Sack?



14. Die fünf Buchstaben A, B, C, D, E sollen in ein 5x5 Gitter eingetragen werden. Jede Zeile und jede Spalte muß jeden Buchstaben genau einmal enthalten. Die ersten drei Zeilen sind gegeben. Vervollständige das Gitter mit zwei weiteren Zeilen.

A	B	C	D	E
C	A	D	E	B
B	C	E	A	D

15. In einem Teich befinden sich 1200 Fische. Ihre Zahl vergrößert sich jedes Jahr um 15 Fische für je 100 Fische. Wie viele Fische befinden sich nach einem Jahr im Teich?

16. Welches Quadrat paßt nicht dazu?



1



2



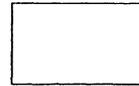
3



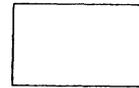
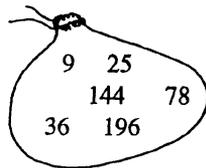
4



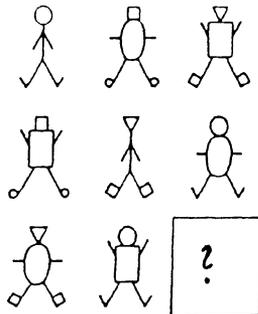
5



17. Welche Zahl gehört nicht in den Sack?



18. Gib die Nummer der Figur an, die das Muster vervollständigt; wähle aus den abgebildeten Figuren.



1



2



3



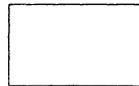
4



5



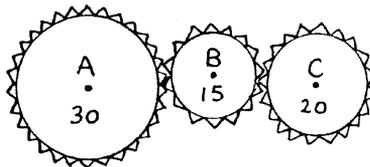
6



19. Wie lauten die nächsten beiden Zahlen in der Folge?

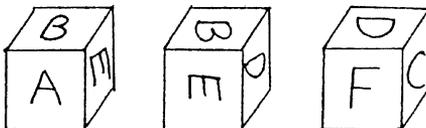
6 9 18 21 42 45 ... ..

20. Das Diagramm zeigt drei miteinander verbundene Zahnräder. Die Anzahl der Zähne steht innerhalb jedes Zahnrads. Wenn Zahnrad A 6 Umdrehungen macht, wie viele macht dann Zahnrad C?

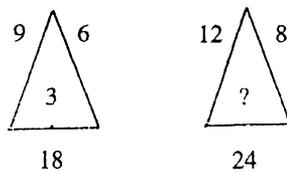


21. Hier sind drei Bilder eines Würfels. Auf den Seiten des Würfels befinden sich die Buchstaben A, B, C, D, E und F.

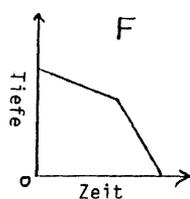
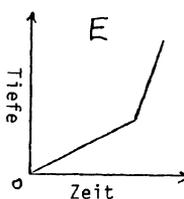
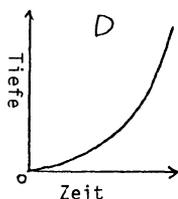
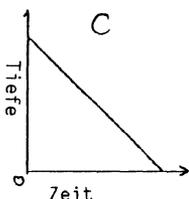
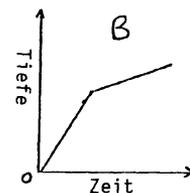
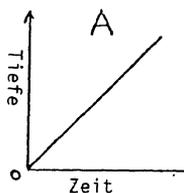
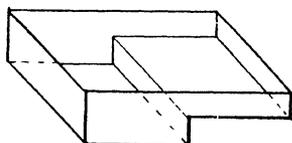
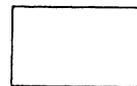
Welcher Buchstabe befindet sich auf der Seite, die dem Buchstaben E gegenüberliegt?



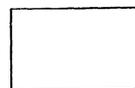
22. Wie lautet die fehlende Zahl?



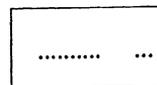
23. Das abgebildete Schwimmbecken wird mit gleichmäßig zufließendem Wasser gefüllt.  
Welcher der folgenden Graphen zeigt am besten die Zunahme der Wassertiefe in Abhängigkeit von der Zeit?



24. Thomas, Robert und Hans besitzen 575 DM, die zwischen ihnen aufgeteilt werden sollen.  
Sie beschließen folgende Aufteilung: Thomas erhält 19 DM mehr als Robert und Robert 17 DM mehr als Hans.  
Wieviel Geld bekommt Thomas?



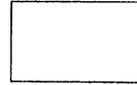
25. Drei Kinder mit den Vornamen Anne, Bernhard und Christine wohnen in der selben Straße.  
Ihr Alter beträgt 7, 9 und 10 Jahre und ihre Nachnamen lauten Schwarz, Grün und Weiß.  
Das Mädchen mit dem Nachnamen Schwarz ist drei Jahre älter als Christine, und das Kind mit dem Nachnamen Weiß ist 9 Jahre alt.  
Wie lautet der Nachname und das Alter von Bernhard?



26. Ein Unterseeboot hat Nahrungsmittel an Bord, die für eine Mannschaft von 25 Männern für 6 Monate reichen.  
Wie lange würden die Nahrungsmittel für 60 Männer reichen?



27. In einem Bonbonglas befinden sich 5 verschiedene Sorten von Bonbons. Wie viele Bonbons muß ein Kind mindestens aus dem Glas nehmen (mit geschlossenen Augen) um sicherzustellen, daß es zumindest drei von einer Sorte erhält?



28. Jedes Symbol in der untenstehenden Multiplikationsaufgabe steht für eine unterschiedliche ganze Zahl. Setze für jedes Symbol eine passende Zahl ein.

$$\square \cdot \begin{array}{|c|} \hline \blacksquare \\ \hline \end{array} = \begin{array}{|c|} \hline \bullet \\ \hline \end{array} \square \begin{array}{|c|} \hline \blacksquare \\ \hline \end{array}$$

