

# ***DEUTSCH PLUS***

**Bericht zum Probelauf 2004**

**Herausgeber:**

Senatsverwaltung für Bildung, Jugend und Sport  
Beuthstraße 6-8  
10117 Berlin

**Verantwortlich:**

Tom Stryck  
Referat I D: Schul- und Qualitätsentwicklung,  
Schulforschung, Fort- und Weiterbildung

**Autor:**

Wolfgang Wendt (I D 12)

Berlin, Januar 2005

## INHALT

### ZUSAMMENFASSUNG

### ANLAGE, DURCHFÜHRUNG UND DATENGRUNDLAGE DER UNTERSUCHUNG

- A Ziel und Durchführung des Probelaufs
- B Beschreibung der Stichprobe
- C Struktur des Erhebungsinstrumentariums und Aufbau des Datensatzes

### ERGEBNISSE ZUM ERSTEN TEIL VON *DEUTSCH PLUS*

- D *DEUTSCH PLUS*/Teil I-2: Fragen an die Eltern
- E *DEUTSCH PLUS*/Teil I-3: Fragen an das Kind
- F *DEUTSCH PLUS*/Teil I-4: Auswertung (Fazit aus dem ersten Teil)

### ERGEBNISSE ZUM ZWEITEN TEIL VON *DEUTSCH PLUS*

- G *DEUTSCH PLUS*/Teil II-1: Passiver Wortschatz
- H *DEUTSCH PLUS*/Teil II-2: Aufgabenverständnis
- I *DEUTSCH PLUS*/Teil II-3: Aktive Äußerungen
- J *DEUTSCH PLUS*/Teil II-4: Gesamtauswertung der Beobachtungssituationen (Fazit aus dem zweiten Teil)

### ERGEBNISSE ZUM PARALLELTEST *KISTE*

- K *KISTE*/Prolog: Rahmenbedingungen
- L *KISTE*: Ergebnisse zur Sprachkompetenz

### ZUSAMMENFÜHRUNG DER EINZELERGEBNISSE

- M Zusammenhänge zwischen den verschiedenen Ergebnissen zur Sprachkompetenz
- N Fazit

## ZUSAMMENFASSUNG

Im neuen Schulgesetz ist im Rahmen der Schuleingangsuntersuchung eine Sprachstandsfeststellung vorgeschrieben, deren Ergebnis dazu führen kann, dass Eltern verpflichtet werden, ihre Kinder in einen Sprachförderkurs zu schicken. Es soll sichergestellt werden, dass die Kinder am künftigen Unterricht erfolgreich mitarbeiten können.

**DEUTSCH PLUS**, das Instrument zur Sprachstandsfeststellung, basiert auf einem **Screeningverfahren** und orientiert sich an *Fit in Deutsch*, das das Niedersächsische Kultusministerium im Schuljahr 2002/2003 erstmals eingesetzt hat. **DEUTSCH PLUS** ist ein **Schwellentest**, d.h. es wird überprüft, ob die sprachlichen Fähigkeiten der Kinder ober- oder unterhalb einer Schwelle liegen, die mindestens erreicht werden muss, um dem Unterricht folgen zu können.

Die im Herbst 2004 eingesetzte Version von **DEUTSCH PLUS** ist die zweite, denn die ursprüngliche Fassung wurde in einem **Probelauf** getestet und anschließend optimiert. Im Probelauf sollten **Praxistauglichkeit** und **Validität (Gültigkeit)** überprüft werden. Zur Einschätzung der Validität wurde als Maßstab der seit langem eingeführte **Sprachtest KISTE** ausgewählt, dessen beide Subtests *Wortschatz (WO)* und *Satzbildung (SB)* **parallel zu DEUTSCH PLUS** einzusetzen waren.

**Die Auswertung des Probelaufes verlief zweigleisig:** Zum einen wurde das Erhebungsinstrument anhand empirisch-statistischer Methoden der Testkonstruktion überprüft, zum anderen wurden die Durchführenden (Testleiter/innen) gebeten, Kommentare zum Verfahren abzugeben, um Hinweise auf Praktikabilität, Aussagekraft und Verbesserungsmöglichkeiten zu erhalten. Hiervon wurde in erfreulicher Weise vielfach Gebrauch gemacht. Es gab zahlreiche generelle Aussagen oder Anmerkungen zu speziellen Aspekten der ersten Version von **DEUTSCH PLUS**. Mechthild Pieler (LISUM) hat sich der mühseligen Arbeit unterzogen, alle Äußerungen zusammenzustellen, zu sichten, zu gruppieren und hieraus Hinweise für Verbesserungen abzuleiten. Neben diesen schriftlichen Rückmeldungen gab es am 21. Juni 2004 ein Auswertungstreffen mit den beteiligten Grundschulen und Kindertagesstätten sowie den Sonderpädagogen/innen, die dankenswerter Weise als zusätzliche Testleiter/innen fungiert hatten. Auf diesem Treffen wurden erste Ergebnisse der statistischen Auswertung vorgestellt und die Erfahrungen aller Beteiligten diskutiert, was weitere sehr nützliche Informationen zum Erhebungsinstrument erbrachte.

Die Weiterentwicklung von **DEUTSCH PLUS** zur zweiten Version beruht also auf **zwei Säulen**: Der Auswertung der mündlichen und schriftlichen Rückmeldungen zum Verfahren und der empirisch-statistischen Überprüfung. Der vorliegende Bericht legt Rechenschaft über diese methodische Überprüfung ab. Er ist daher naturgemäß **ein technischer Report**, dennoch um Allgemeinverständlichkeit bemüht.

Das zentrale Ergebnis lautet:

**Hinsichtlich der beiden Punkte Praxistauglichkeit und Validität zeigte sich die Eignung von DEUTSCH PLUS.**

Außer der Validität als dem obersten der Testgütekriterien sind auch die Reliabilität und die Objektivität (von Durchführung und Auswertung) wichtig, um die Qualität eines Messinstrumentes beurteilen zu können. Bei der Schätzung der **Reliabilität** (mittels des Konsistenzkoeffizienten  $\alpha$ ) ergaben sich zufriedenstellende bis sehr gute Werte. Zwar war auch die **Ob-**

**ektivität** in ausreichendem Maße gegeben, konnte aber an den Stellen verbessert werden, an denen einige Durchführende (Testleiter/innen) bei der Formulierung des Ergebnisses von den vorgegebenen Auswertungskriterien abgewichen waren.

**Zu DEUTSCH PLUS brauchte demnach keine Alternative gesucht zu werden**, aber der Probelauf ergab eine Reihe von Hinweisen zur Optimierung, die zu einer veränderten Struktur des Erhebungsinstrumentes führten, die wir am Ende der Zusammenfassung skizzieren.

In den Abschnitten M und N am Ende des Berichtes werden die Ergebnisse des Probelaufes zusammengeführt und Schlussfolgerungen gezogen. Einige ausgewählte und wichtige Punkte hieraus:

### Zur Stichprobe

Am Probelauf nahmen **N=384 Kinder aus Vorklassen (55%) und Kindertagesstätten (45%)** teil. Die Stichprobe ist nicht repräsentativ, sondern Kinder aus sozial belasteten Gegenden sind überrepräsentiert, was im Einklang mit den Zielen des Probelaufs steht. Es gibt mehr West- als Ostschulen, bei den Kitas ist es umgekehrt. Aufgrund zu vieler fehlender Testwerte beruht die eigentliche über die soziodemografischen Angaben hinausgehende Analyse auf den Daten von **N=364** Kindern.

### Struktur des Erhebungsinstrumentariums in seiner ersten Version

Insgesamt bestanden die beim Probelauf eingesetzten Erhebungsinstrumente aus neun Modulen, nämlich drei Mal drei Unterteilen.

D+/I-1: Informationen zum Kind (Angaben aus den Akten)

D+/I-2: Fragen an die Eltern

D+/I-3: Fragen an das Kind

D+/II-1: Passiver Wortschatz

D+/II-2: Aufgabenverständnis

D+/II-3: Aktive Äußerungen

K/WO: Aktiver Wortschatz

K/SB: Satzbildung

K/ZU: Satzbildung/Zusatzauswertung grammatikalischer Fehler.

### Fragen an die Eltern

Anhand einer Reihe von Fragen an die Eltern sollten diese die Sprachkompetenz ihres Kindes einschätzen. Die höchste Einstufung erfolgte für die Deutschkompetenz der von ihrer Herkunft her deutschsprachigen Kinder. In erheblichem Abstand folgte dann die Sprachkompetenz der jeweiligen Erstsprache der Kinder nichtdeutscher Herkunftssprache und an letzter Stelle schließlich deren Deutschkompetenz:

D/D >> ndH/ndH > D/ndH

(Deutschkompetenz beurteilt durch ihre deutschsprachigen Eltern

>> Erstsprachkompetenz beurteilt durch ihre nichtdeutschsprachige Eltern

> Deutschkompetenz beurteilt durch ihre nichtdeutschsprachigen Eltern).

### **Auswertung von DEUTSCH PLUS/Teil I**

Die Testleiter/innen sollten zwei Urteile abgeben. Das eine, das sich auf die Auswertung der Elternangaben stützte, das andere, dem die Antworten der Kinder zugrundelagen. Zum ersten war kein präzises Auswertungskriterium vorgegeben. Das mag mit ein Grund dafür sein, dass es beim ersten Urteil zu zahlreichen Ausfällen kam (rund 20% der Fälle) und ferner (in über 10% der Fälle) zu Widersprüchen zwischen den beiden Einschätzungen. Größenordnungsmäßig allerdings deckten sich beide Urteile insoweit, als dass beide Male in rund 80% der Fälle keine Sprachförderung für notwendig gehalten wurde.

Wie groß die Gruppe der zu Fördernden im Herbst 2004 sein wird, ist aus derartigen Werten schwer zu schätzen, da die Stichprobe für Berlin nicht repräsentativ ist. Da die förderungsbedürftigen Kinder in der Stichprobe überproportional vertreten sind, liefert die Stichprobe höchstwahrscheinlich Obergrenzen.

### **DEUTSCH PLUS/Teil II-1: Passiver Wortschatz**

Auf zwei sog. Wimmelbildern waren Dinge oder Vorgänge zu identifizieren, wobei die Anzahl richtig gezeigter Nomen (mindestens 20) und Verben (mindestens 8) als Schwelle definiert wurde. Maximal möglich richtig zu nennen waren 24 Nomen und 10 Verben. 43% der Kinder blieben unterhalb der festgelegten Schwelle.

Die Reliabilität des Moduls ist sehr hoch. Ohne bedeutsame Einbußen für die Gütetestkriterien kann auf das erste Wimmelbild verzichtet werden.

### **DEUTSCH PLUS/Teil II-2: Aufgabenverständnis**

Anhand von Spielsachen waren einfache und komplexe Anweisungen umzusetzen. Als Schwelle für eine erfolgreiche Bearbeitung dieses Moduls galt die Umsetzung von mindestens fünf (von maximal sechs) einfachen und drei (von maximal vier) komplexen Anweisungen. 57% der Kinder blieben unterhalb der Schwelle.

Die Reliabilität des Moduls ist mittelhoch.

### **DEUTSCH PLUS/Teil II-3: Aktive Äußerungen**

Das Material dieses Moduls besteht aus den beiden Bildserien "Was ist hier los?" und "Glas auf dem Schrank". Während die zweite Serie eine kleine Geschichte wiedergibt (die Bilder sind dem Kind in der richtigen Reihenfolge vorzulegen), besteht die erste aus Einzelkarten. Das Kind kann eine der beiden Serien auswählen. Entweder äußert sich das Kind dann spontan zu den Bildern oder es wird ein Impuls gegeben, der nicht zuviel Wortschatz anbietet. Protokolliert werden die Äußerungen des Kindes und hinsichtlich dreier Aspekte ausgewertet. Die Schwelle gilt dann als erreicht, wenn alle drei Kriterien erfüllt sind. 40% der Kinder blieben bei diesem Modul, das sicherlich das schwierigste der dreien ist, unterhalb der Schwelle.

Einerseits ist zwar die Reliabilität des Moduls zufriedenstellend, andererseits sollte die Auswertungsobjektivität verbessert werden, denn die Auswertung ist in weit höherem Maße als bei den anderen Bestandteilen von den Sichtweisen der Testleiter/innen abhängig.

## **DEUTSCH PLUS/Teil II-4: Gesamtauswertung von Teil II**

Von den Kindern, die den Teil II von *DEUTSCH PLUS* durchliefen, sollten aufgrund der Testergebnisse rund 70% einer Sprachförderung zugeführt werden. Bezogen auf die Ausgangsmenge aller Kinder ergibt sich ein schätzungsweise Förderbedarf bei rund 15% der Kinder, ein Wert, der angesichts der nichtrepräsentativen Zusammensetzung der Stichprobe eine Obergrenze darstellen dürfte, gleichwohl realistisch zu sein scheint.

### **KISTE: Ergebnisse zur Sprachkompetenz**

Es gibt drei zentrale Ergebnisse:

- Die Sprachkompetenz der Stichprobenkinder ist unterdurchschnittlich.<sup>1</sup>
- Die Heterogenität der Leistungen ist sehr hoch.
- Die Leistungen im Subtest Satzbildung sind besser als im Subtest Wortschatz.

### **Validität (Korrelationen zwischen den Merkmalen aus *DEUTSCH PLUS* und aus *KISTE*):**

Generell gilt, dass die Validitätsüberprüfung sehr zufriedenstellend ausfällt. Die Korrelationen weisen nahezu durchweg eine Größenordnung auf, wie sie zu wünschen ist, nämlich eine mittelhohe. Wären die Koeffizienten zu klein, dann müsste die Validität der Module von *DEUTSCH PLUS* angezweifelt werden. Wären sie zu hoch, so wäre dies deswegen seltsam, weil beide Test unterschiedliche Ziele verfolgen (Schwellentest versus differenzierte Diagnostik).

## **Zum Aufbau der zweiten Version von *DEUTSCH PLUS***

Die Hinweise, die sich aus den empirisch-statistischen Ergebnissen und aus den Rückmeldungen der Beteiligten ergaben, führten zu einer Weiterentwicklung der Sprachstandsfeststellung. Die zweite Version von *DEUTSCH PLUS* besteht nunmehr aus vier Teilen, von denen die letzten drei Teile B, C und D Beobachtungssituationen mit dem Kind sind, während einleitend im Teil A zentrale Informationen zusammengestellt werden:

- Teil A: Basisinformationen zur Sprachbeherrschung des Kindes**
- Teil B: Beobachtungssituation *Passiver Wortschatz [Wimmelbild]***
- Teil C: Beobachtungssituation *Aufgabenverständnis [Kuscheltier]***
- Teil D: Beobachtungssituation *Aktive Äußerungen [Bildkarten]***

Begleitend wird bei allen Aufgabenstellungen das kommunikative Verhalten und die Aussprache der Kinder beobachtet.

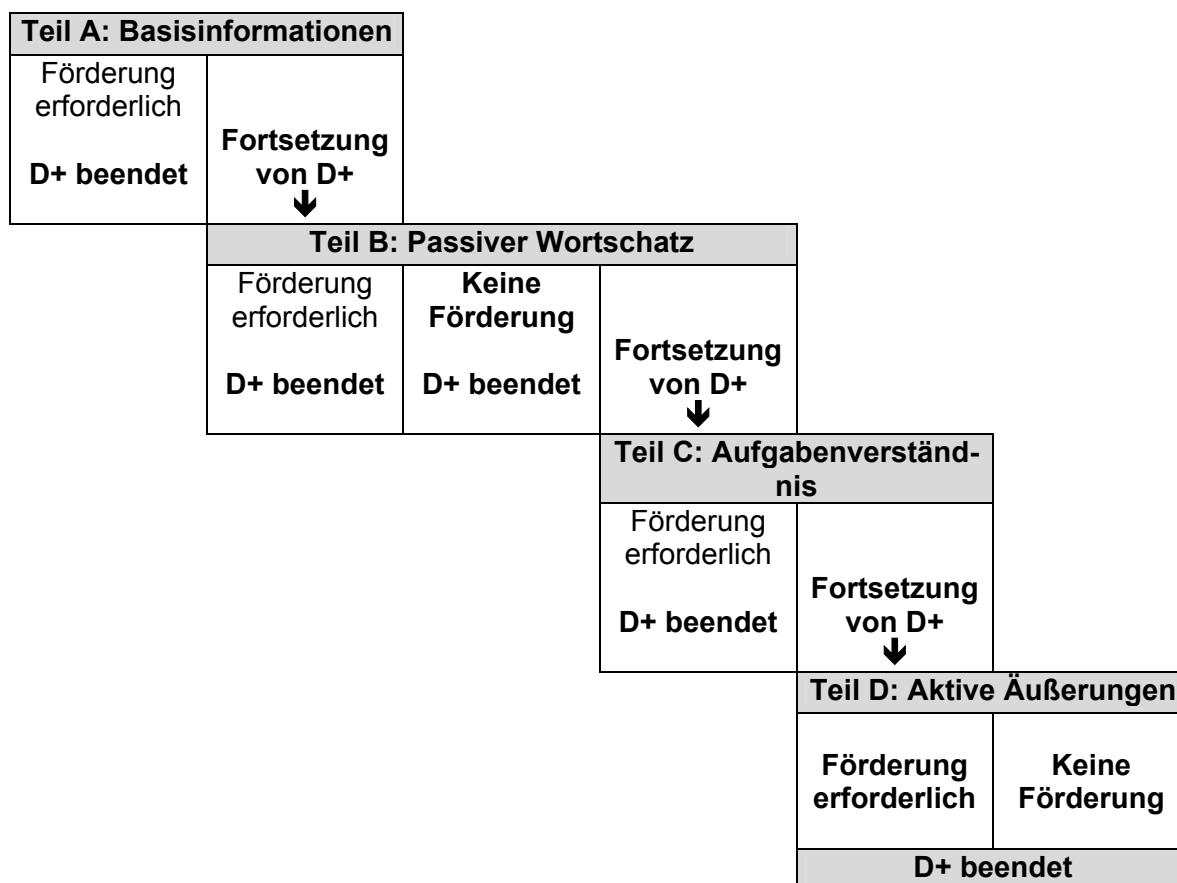
Da es offensichtlich weniger gravierend ist, wenn ein Kind, das es nicht nötig hätte, gesondert sprachlich gefördert zu werden, dennoch einem Förderkurs zugewiesen wird, als die umgekehrte Fehlentscheidung, nämlich einem förderbedürftigen Kind die Förderung vorzuenthalten, gilt bei Durchführung und Auswertung von *DEUTSCH PLUS*, **sich im Zweifelsfalle für die Förderung zu entscheiden.**

---

<sup>1</sup> Eine derartige Aussage ist möglich, da KISTE ein an einer Eichstichprobe normierter Test ist; vgl. auch Kapitel C.

Diese Perspektive prägt die Ablauf- und Entscheidungslogik der vier Teile von *DEUTSCH PLUS/Version 2*, die im nachstehenden Ablaufschema dargestellt wird:

Nach jedem Teil ist es möglich, eine Entscheidung **für** die Förderung zu fällen, aber nur nach den Teilen B und D - also nur nach persönlicher Begegnung mit dem Kind - kann geschlussfolgert werden, das Kind bedürfe keiner Förderung.



## ANLAGE, DURCHFÜHRUNG UND DATENGRUNDLAGE DER UNTERSUCHUNG

### A ZIEL UND DURCHFÜHRUNG DES PROBELAUFES

Im neuen Schulgesetz ist im Rahmen der Schuleingangsuntersuchung eine Sprachstandsfeststellung vorgeschrieben, deren Ergebnis dazu führen kann, dass Eltern verpflichtet werden, ihre Kinder in einen Sprachförderkurs zu schicken. Es soll sichergestellt werden, dass die Kinder am künftigen Unterricht erfolgreich mitarbeiten können.

*DEUTSCH PLUS*, das Instrument für die Sprachstandsfeststellung, basiert auf einem Screeningverfahren. Es orientiert sich an dem Instrument *Fit in Deutsch*, das das Niedersächsische Kultusministerium im Schuljahr 2002/2003 erstmals eingesetzt hat. Beim Screening geht es darum, mit möglichst geringem Aufwand, aber zugleich zuverlässig die Kinder heraus zu finden, für die eine sofortige Förderung noch vor dem Schulbeginn notwendig ist. Ziel einer derartigen Sprachstandsmessung ist also zu überprüfen, ob die sprachlichen Fähigkeiten der Kinder ober- oder unterhalb einer Schwelle liegen, die mindestens erreicht werden muss, um dem Unterricht folgen zu können.

Das Messinstrument zum Sprachstand kann hingegen keine hinreichende Auskunft darüber geben, in welchen einzelnen der untersuchten Bereichen welche spezifische Förderung erfolgen muss und welche Lernangebote die Lernentwicklung des einzelnen Kindes am besten unterstützen würden. Eine solche genaue Erfassung des individuellen Sprachstandes erfordert eine differenziertere Diagnose als es *DEUTSCH PLUS* leisten soll und zu leisten imstande ist.

*DEUTSCH PLUS* besteht in der ersten Version, die Gegenstand des Probelaufes ist, aus zwei Teilen (Einzelheiten weiter unten im Kapitel C):

#### **Teil 1: Gespräche mit den Eltern und mit dem Kind**

#### **Teil 2: Beobachtungssituationen zu einzelnen Sprachaspekten.**

Diese Zweiteilung beruht auf dem Wunsch, den Gesamtaufwand dadurch zu verringern, dass dem eigentlichen Test (der relativ umfangreiche Teil 2) mit dem Teil 1 ein Filter vorgeschaltet wird. Der idealtypische Ablauf sollte folgendermaßen aussehen:

In Gesprächen mit den Eltern und dem Kind, die sich auf zentrale sprachrelevante Eckdaten konzentrierten und im Teil 1 dokumentiert würden, lasse sich feststellen, welche Kinder die deutsche Sprache so gut beherrschten, dass sie keine vorschulischen Fördermaßnahmen benötigten. Ebenso werde in den Gesprächen deutlich, welche Kinder einsprachig in einer anderen Sprache als Deutsch seien und bei denen auf jeden Fall eine Sprachförderung vor dem Schulbeginn erforderlich sei. In beiden Fällen wäre eine Durchführung des zweiten Teils nicht mehr notwendig, weil eine eindeutige Entscheidung getroffen werden könne. Bei allen Kindern, bei denen diese eindeutige Entscheidung nach den Gesprächen nicht möglich sei (Klärungsgruppe), würde auch der Teil 2 der Sprachstandsfeststellung (Beobachtungssituationen zu einzelnen Sprachaspekten) durchgeführt, um abzuklären, ob eine Fördermaßnahme erforderlich sei.

Vor dem flächendeckenden Einsatz im Rahmen der kommenden Schuleingangsuntersuchung im Herbst 2004 war das Instrument zu erproben. Zwei Fragen standen im Vordergrund:

1. Ist **DEUTSCH PLUS** praxistauglich?
2. Ist **DEUTSCH PLUS** ausreichend valide (gültig)?

Um Schätzwerte für die Validität zu erhalten, musste ein Paralleltest, ein zweites und bereits erprobtes Instrument, eingesetzt werden, das zielähnlich zu **DEUTSCH PLUS** ist und dessen Gütekennwerte bekannt sind. Nach der Prüfung mehrerer Sprachtests fiel die Wahl auf **KISTE** aus dem Beltz-Verlag. Zum Einsatz kommen sollten allein die beiden Subtests *Wortschatz* und *Satzbildung*, die die größte inhaltliche Affinität zu den in **DEUTSCH PLUS** thematisierten Aspekten aufweisen. Mit dieser Beschränkung ließ sich zugleich die Belastung für die beteiligten Kinder und Testleiter/innen reduzieren.

Der Probelauf erfolgte in zwei Schüben: Vor den Osterferien 2004 in elf Grundschulen, nach den Osterferien in elf Kindertagesstätten. Die Erprobung sollte folgendermaßen ablaufen:

1. Der erste Teil von **DEUTSCH PLUS** wird in den Schulen und in den Kitas von den Vorklassenleitern/innen und Erziehern/innen alleine bestritten.
2. Der zweite Teil von **DEUTSCH PLUS** wird mit Unterstützung einer weiteren Person aus der Sonderpädagogik/Förderschwerpunkt Sprache durchgeführt, um das Protokollieren und Beobachten zu erleichtern und zu vertiefen. Der/die Sonderpädagoge/in soll zudem den Paralleltest (zwei Untertests von **KISTE**) durchführen.

Abweichend von der geplanten späteren Praxis sollte im Probelauf unabhängig vom Ergebnis des ersten Teils auf jeden Fall auch der zweite Teil eingesetzt werden, um überprüfen zu können, wie weit auf den ersten Teil Verlass ist: Würde es - so hier die Leitfrage - Konstellationen geben, in denen die Ergebnisse aus Teil 1 und Teil 2 sich widersprechen? Könnte es sein, dass nach dem ersten Teil ein Kind keine Förderung benötige, gleichwohl aber nach dem zweiten? Und wenn ja, wie häufig würden derartige Widersprüche auftreten? Daher lautete die Planung, in jeder der beteiligten Vorklassen und Kita-Gruppen bei jeweils genau zehn Kindern die beiden Instrumentarien vollständig einzusetzen, d.h. beide Teile von **DEUTSCH PLUS** und die beiden ausgewählten Subtests von **KISTE**. Die Testrealität sah zuweilen anders aus; vielleicht konnte nicht immer deutlich gemacht werden, dass nicht die Kinder, sondern der Test getestet wurde. Kapitel C beschreibt den Datensatz, der der Auswertung zugrunde liegt.

#### **Hinweis zu den Tabellen:**

*Kaum ein Frage-, kaum ein Erhebungsbogen wird vollständig ausgefüllt. Die in den nachstehenden Kapiteln wiedergegebenen Tabellen beziehen sich daher ggf. auf stark schwankende Fallzahlen. So fehlen z.B. zum Merkmal Alter bei 45 Kindern die Angaben, zum Merkmal Geschlecht bei 41 Kindern. Würden beide Merkmale in einer Tabelle gleichzeitig berücksichtigt (etwa um die Frage zu beantworten, ob die Altersverteilung bei den Mädchen und bei den Jungen identisch ist), dann erhielten wir eine Tabelle, in der mindestens 45 Kinder nicht enthalten sind, i.d.R. sogar noch mehr, da in die Tabelle nur die Kinder eingehen können, für die wir die Angaben zu beiden Merkmalen kennen. Mit anderen Worten: Die von Tabelle zu Tabelle unterschiedlichen Bezugsgrößen sind empirische Normalität.*

## B BESCHREIBUNG DER STICHPROBE

Aus den Grundschulen und den Kindertagesstätten kamen Unterlagen zu insgesamt 384 Kindern zurück. Für den Probelauf wurde keine Repräsentativität angestrebt, sondern darauf geachtet, dass das gesamte Spektrum abgedeckt würde und vor allem jene Kinder überproportional vertreten wären, bei denen am ehesten ein Förderbedarf zu erwarten war. Dementsprechend wurde an ausgewählte Grundschulen und Kitas herangetreten, deren Teilnahme freiwillig war. Im Folgenden sollen einige Charakteristika der Stichprobe dargestellt werden, die aus diesem Vorgehen resultierte.

### Geografisches: Test- und Wohnort

#### B.1 Tabelle: Wo wurde die Testung durchgeführt?

Testort	Häufigkeiten	
Schule	213	55%
Kita	171	45%
Insgesamt	384	100%

Etwas über die Hälfte der Stichprobenkinder wurde in der Grundschule getestet, etwas weniger als die Hälfte in den Kitas. Dies entspricht zwar nicht der zu erwartenden Aufteilung im Herbst 2004, ist aber für die Fragen, auf die der Probelauf antworten soll, ohne Belang, denn die Tauglichkeit des Instrumentariums muss sich unabhängig vom Einsatzort erweisen.

Ein Vergleich der Bezirke, in denen die Kinder wohnen und in denen der Testort liegt, zeigt das kaum überraschende und hier nicht dokumentierte Ergebnis, dass i.d.R. die Kinder in ihrem Wohnbezirk auch die Vorklasse oder die Kita besuchen. Aber wie verteilen sich die Testorte auf die Wohnbezirke?

#### B.2 Tabelle: Testort und Wohnbezirk des Kindes.

Kinder aus ↓	Insgesamt		davon wurden getestet			
			in der Schule		in der Kita	
01 Mitte	65	18%	18	9%	47	30%
02 Friedrichshain-Kreuzberg	123	33%	76	36%	47	30%
03 Pankow	1				1	1%
04 Charlottenburg-Wilmersdorf	2	1%	2	1%		
05 Spandau	23	6%	10	5%	13	8%
06 Steglitz-Zehlendorf	33	9%	33	16%		
07 Tempelhof-Schönberg	44	12%	42	20%	2	1%
08 Neukölln	31	8%	31	15%		
10 Marzahn-Hellersdorf	33	9%			33	21%
12 Reinickendorf	13	4%			13	8%
<b>Insgesamt</b>	<b>368</b>	<b>100%</b>	<b>212</b>	<b>100%</b>	<b>156</b>	<b>100%</b>

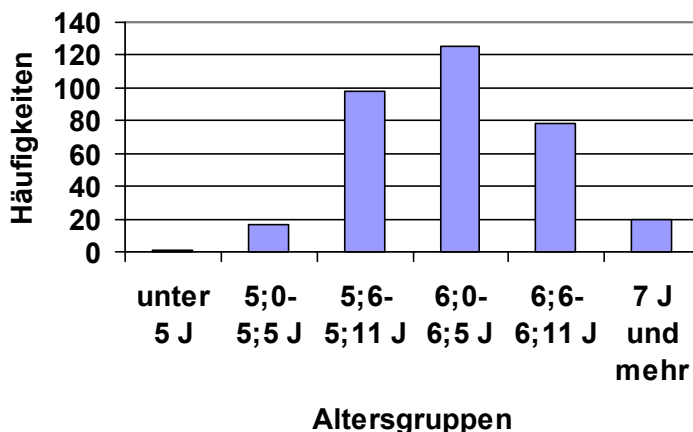
Die Tabelle B.2 ist folgendermaßen zu lesen: Von insgesamt 368 Kindern liegen die Angaben zum Bezirk, in den sie wohnen, und zum Testort vor. 18% von ihnen wohnen in Mitte; von diesen 65 Kindern wurden 18 in der Schule und 47 in der Kita getestet. Die 18 Kinder entsprechen einem Anteil von 9% der 212 Kinder, die insgesamt in der Schule getestet wurden. (Ein Vergleich mit der Tabelle B.1 zeigt, dass wir von einem "Schultestkind" nicht den Wohnort kennen; daher taucht dieses Kind in der Tabelle B.2 nicht auf.) Entsprechend beträgt mit 47 Kindern aus Mitte der Anteil an den "Kitatestkindern" 30%.

Die Tabelle B.2 zeigt, dass es für den Testort Kita drei Schwerpunkte gibt: Mitte (30%), Friedrichshain-Kreuzberg (30%) und Marzahn-Hellersdorf (21%). Die Kinder, die in den Schulen getestet wurden, stammen vor allem ebenfalls aus Friedrichshain-Kreuzberg (zu 36%), aus Steglitz-Zehlendorf (16%), Tempelhof-Schöneberg (20%) und Neukölln (15%). Es gibt also eine "Westlastigkeit" bei den Schulen als Testort, während die Kitakinder eher aus der östlichen Stadthälfte stammen. Diese unterschiedliche Akzentsetzung ist als Merkposten zu behalten, falls die Analysen Unterschiede zwischen den Schul- und den Kitakindern zeigen sollten. Derartige mögliche Unterschiede könnten dann u.U. weniger auf Unterschiede in den Institutionen als auf Unterschiede in den Wohngebieten beruhen.

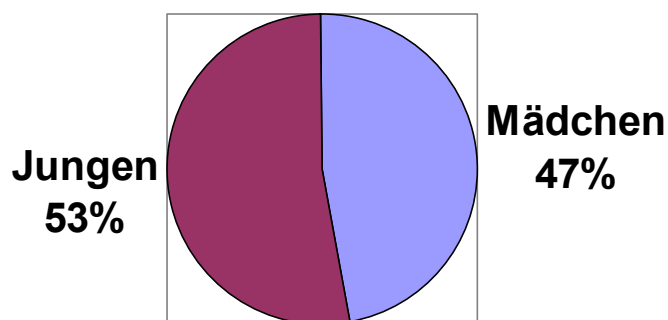
### Die Kinder: Soziodemografisches

Wie sich die Stichprobe hinsichtlich der beiden relevanten Merkmale Alter und Geschlecht zusammensetzt, zeigen die Abbildungen B.3 und B.4.

#### **B.3 Abbildung: Verteilung der Kinder aus dem Probelauf nach Alter (gruppiert).**



Wie zu erwarten finden wir nahezu alle Kinder in der Altersspanne von fünfeneinhalb bis sechseinhalb Jahren, knapp 80%. Jungens bilden mit 53% in der Stichprobe die Mehrheit. Berücksichtigen wir beide Merkmale zugleich, so stellen wir fest, dass die Jungens im Schnitt etwas älter sind als die Mädchen, ohne dass allerdings die Mittelwerte sich statistisch signifikant unterscheiden (hierzu keine Tabelle).

**B.4 Tabelle: Verteilung der Kinder nach Geschlecht.**

Interessanterweise differieren die Altersmittelwerte der Kita- und der Schulkinder ebenfalls nicht, wohingegen die Zusammensetzung nach Geschlecht sich stark unterscheidet: 52% der Kitakinder sind Mädchen, von den Schulkindern sind 57% Jungen. Dies wird ggf. im weiter oben bereits erwähnten Sinne bei späteren Analysen zu berücksichtigen sein.

Zu 287 Kindern liegt die Angabe zur Staatsangehörigkeit vor (60% von ihnen deutsch, 21% türkisch), zu 335 Kindern das Geburtsland (94% Deutschland, 0,3% Türkei). Von den 275 Kindern, von denen wir beides kennen, sind die Angaben in der Tabelle B.5 aufgeführt.

**B.5 Tabelle: Verteilung der Kinder nach Staatsangehörigkeit und Geburtsland.**

Geburtsland	Deutschland	Türkei	Sonstiges	Insgesamt
<b>Staatsangehörigkeit</b>				
Deutschland	164 64%		7 36%	171 62%
Türkei	53 21%	1 100%		54 20%
Sonstiges	38 15%		12 63%	50 18%
<b>Insgesamt</b>	<b>255 100%</b>	<b>1 100%</b>	<b>19 100%</b>	<b>275 100%</b>

In den Zahlen der Tabelle B.5 schlägt sich das bekannte Phänomen nieder, dass die Ausländer zu einem großen Teil bereits in der dritten Generation hier leben. Sie behalten ihre ausländische Staatsangehörigkeit, aber Deutschland ist ihr Lebensmittelpunkt. Wie sich das auf Sprachgewohnheiten auswirkt, ist Gegenstand der nachstehenden Tabellen.

## Familie und Sprache

Tabelle B.6 dokumentiert die Angaben der Eltern zur Erstsprache des Kindes und der Sprache, die vornehmlich zu Hause gesprochen wird. Es zeigt sich, dass die Verteilungen weitgehend identisch sind, und eine entsprechende (hier nicht wiedergegebene) Kreuztabellierung zeigt: Erst- und Familiensprache sind weitestgehend identisch. Sind also Erst- und damit auch die häusliche Verkehrssprache nicht deutsch, so bedeutet dies, dass die Sprachförderung nicht durch Einflüsse zu Hause verstärkt wird.

### **B.6 Tabelle: Erstsprache der Kinder und überwiegende Verkehrssprache zu Hause.**

Erstsprache	Häufigkeiten	Verkehrssprache	Häufigkeiten
deutsch	159 47%	deutsch	145 51%
türkisch	95 30%	türkisch	82 29%
arabisch	40 12%	arabisch	31 11%
Osteuropa <sup>#</sup>	25 8%	Osteuropa <sup>#</sup>	20 7%
andere	10 3%	andere	9 3%
<b>Insgesamt</b>	<b>322 100%</b>	<b>Insgesamt</b>	<b>287 100%</b>

Bei der Zweisprachigkeit unterscheidet *DEUTSCH PLUS* zwischen "Zwei Sprachen verstehen" und als Steigerung "Zwei Sprachen sprechen". Tabelle B.7 zeigt, wie viele Kinder nach den Angaben der Eltern in die daraus resultierenden Kategorien fallen.

### **B.7 Tabelle: Zweisprachigkeit der Kinder: Verstehen und Sprechen.**

	Verstehen	Nein	Ja	Insgesamt
Sprechen				
Nein		120 100%	14 7%	134 41%
Ja			191 93%	191 59%
Insgesamt		120 100%	205 100%	325 100%

Interessanterweise, so die Schlussfolgerung aus der Tabelle B.7, gibt es kaum Unterschiede zwischen den beiden Stufen Verstehen und Sprechen. Von den 205 Kindern, die eine zweite Sprache verstehen, sind 93% auch in der Lage sie zu sprechen. Das bedeutet, dass in einer überarbeiteten Fassung von *DEUTSCH PLUS* auf diese Unterscheidung verzichtet werden kann. Im Übrigen zeigen die Zahlen, wie stark die Stichprobe von den Gesamtberliner Verhältnissen abweicht: Dass es mehr zweisprachige als einsprachige Kinder gibt, dürfte wohl kaum generell gelten.

<sup>#</sup> incl. Ex-UdSSR

**C Struktur des Erhebungsinstrumentariums und Aufbau des Datensatzes**

Die Sprachstandsfeststellung von "DEUTSCH PLUS" besteht - wie im Kapitel A skizziert - in der hier behandelten ersten Version aus zwei Teilen, einem Filterteil und einer weiterreichenden Prüfung sprachlicher Basiskompetenzen. Der erste, der Filterteil soll die Kinder in drei Gruppen einteilen: Diejenigen, die eindeutig keiner Sprachförderung bedürfen, solche, bei denen es klar ist, dass sie gefördert müssen, und schließlich die Kinder, bei denen eine Entscheidung über die Förderbedürftigkeit aufgrund der bislang vorliegenden Informationen nicht möglich ist. Diese dritte Gruppe soll dann den Teil 2 von DEUTSCH PLUS durchlaufen.

Jeder der beiden Teile von DEUTSCH PLUS/Version 1 besteht jeweils wiederum aus drei Unterteilen:

**D+/I-1: Informationen zum Kind.**

Dies sind Daten, die sich gewissermaßen aus der Aktenlage ergeben. Zu ihnen gehören die Merkmale Alter, Geschlecht und Staatsangehörigkeit, die im Kapitel B vorgestellt wurden.

**D+/I-2: Fragen an die Eltern.**

Zum Einen handelt es sich um wichtige Ergänzungen zu den einleitenden Merkmalen (Erst- und Familiensprache sowie die Zweisprachigkeit gehören dazu, vgl. Kapitel B.), zum Anderen um direkte Fragen nach sprachlichen Fähigkeiten. Der Test selber macht keine präzise Vorgabe zur Schwelle.

**D+/I-3: Fragen an das Kind.**

Anhand der Antworten des Kindes auf 14 Fragen soll festgestellt werden, ob das Kind sachlich angemessene Wörter benutzt, sich flüssig und grammatisch weitgehend richtig auszudrücken vermag. Die Schwelle liegt bei mindestens 10 angemessenen Formulierungen.

**D+/II-1: Passiver Wortschatz.**

Es werden zwei sog. Wimmelbilder betrachtet um festzustellen, ob Nomen und Verben, die in Fragen an das Kind enthalten sind, richtig aufgefasst wurden. Die Schwelle liegt bei 20 richtig erkannten Nomen und 8 Verben.

**D+/II-2: Aufgabenverständnis.**

Verbunden mit Spielsachen werden einfache und komplexe Aufforderungen an das Kind gerichtet. Kriterium ist, wie viele dieser Aufforderungen das Kind richtig versteht: Es müssen mindestens 5 einfache und 3 komplexe Anweisungen richtig umgesetzt werden.

**D+/II-3: Aktive Äußerungen.**

Dem Kind werden alternativ zwei Bilderserien vorgegeben, zwischen denen es wählen kann. Zu der gewählten Bildserie soll es sich äußern. Die Aufgabe wird als erfolgreich bewältigt gewertet, wenn mindestens zwei Drittel der Äußerungen des Kindes länger als drei Wörter sind und wenn es zugleich vorwiegend sachlich angemessene Nomen und Verben benutzt.

*KISTE* wurde als Paralleltest zur Validierung eingesetzt. *KISTE* besteht aus einer Reihe von Subtests, von denen zwei wg. ihrer Zielähnlichkeit zu *DEUTSCH PLUS* ausgewählt wurden, nämlich *Aktiver Wortschatz* und *Satzbildung*. Aufgrund der Möglichkeit, beim Subtest *Satzbildung* eine Zusatzauswertung vorzunehmen, wovon eine Reihe der Testleiter/innen Gebrauch machte, ergeben sich auch für den Paralleltest drei Unterteile:

**K/WO: Aktiver Wortschatz.**

Den Kindern werden Fragen gestellt, die Antworten gezählt und in Punkte umgerechnet. Dabei wird je nach Komplexität der Frage die Anzahl der kindlichen Äußerungen mit dem Faktor 1 oder dem Faktor 2 multipliziert; Beispiel: "Was kann man alles essen?" (Faktor 1) "Was kann man alles lernen?" (Faktor 2) Die Summe der Punkte drückt die Leistung des Kindes aus. Diese Punktsomme lässt sich in sog. T-Werte und Prozenträge<sup>2</sup> umrechnen, die getrennt für die 5- und die 6-Jährigen dem Testmanual entnommen werden können. Damit existiert für diesen Subtest ein altersunabhängiger Ergebniswert.

**K/SB: Satzbildung.**

Aus vorgegebenen Wortpaaren muss das Kind Sätze bilden; Beispiele: "spielen - Kind", "arbeiten - verdienen", "schlafen - müde". Bewertet werden Inhalt und grammatikalische Korrektheit. Auch hier werden je nach Anforderungsgrad der Vorgaben unterschiedlich viele Punkte vergeben. Das Ergebnis des Subtests wird ausgedrückt in Semantikpunkten, Grammatikpunkten und in der Gesamtsumme beider Punktarten. Alle drei Resultate lassen sich ebenfalls in altersspezifische T-Werte und Prozenträge umrechnen.

**K/ZU: Satzbildung/Zusatzauswertung grammatikalischer Fehler.**

Der Test *KISTE* bietet die Möglichkeit, über die eben skizzierte Auswertung hinaus genauer die Äußerungen des Kindes zu untersuchen, z.B. auf fehlende oder falsche Artikel, auf falsche Flexionen von Verben etc. Eine derartige Zusatzauswertung ist bedeutsam, wenn es um gezielte Diagnostik und das Bestimmen von Fördermaßnahmen geht, reicht aber über das Ziel von *DEUTSCH PLUS* als Schwellentest weit hinaus. Dennoch hat eine Reihe von Testleiter/innen diese Zusatzauswertung vorgenommen, was für die vorliegend dokumentierte Auswertung zusätzliche Möglichkeiten eröffnet.

Insgesamt können wir somit beim Erhebungsinstrumentarium des Probelaufes neun Module, nämlich drei Mal drei Unterteile, unterscheiden. Der dazugehörige Datensatz enthält viele Lücken, von denen die Tabelle C.1 Zeugnis ablegt.

---

<sup>2</sup> T-Werte und Prozenträge werden bei der Entwicklung eines Tests anhand von Eichstichproben berechnet. Die Eichstichprobe soll möglichst repräsentativ für die Zielgruppe des Tests sein. Die Rohwerte, die unmittelbaren Ergebnisse des Tests, werden umgerechnet, um die Ergebnisse informationhaltiger zu machen:

T-Werte sind so normiert, dass ihr Mittelwert 50 und ihre Standardabweichung 10 betragen. Dem T-Wert ist also unmittelbar zu entnehmen, wie weit das Ergebnis unter oder über dem Mittelwert liegt. Prozenträge geben die relative Position an. Ein Prozenrang von z.B. 41 bedeutet, dass 40% der Kinder schlechtere und 59% bessere Leistungen erzielen.

Derartige Werte werden häufig für spezifische Teilgruppen angegeben. In unserem Falle ist davon auszugehen, dass die sprachlichen Fähigkeiten stark vom Alter abhängen. Daher wurden für *KISTE* altersspezifische T-Werte und Prozenträge berechnet.

**C.1 Tabelle: Überblick über die Vollständigkeit des Datensatzes differenziert nach den neun Modulen.**

	Alles fehlt		Teile vorhanden		Alles vorhanden	
<i>DEUTSCH PLUS</i> , Teil I						
D+ I-1: Informationen zum Kind	0	0%	286	74%	98	26%
D+I-2: Fragen an die Eltern	33	9%	200	52%	151	39%
D+I-3: Fragen an das Kind	28	7%	0	0%	356	93%
<i>DEUTSCH PLUS</i> , Teil II						
D+II-1: Passiver Wortschatz	138	36%	2	0,5%	244	63,5%
D+II-2: Aufgabenverständnis	138	36%	1	0,3%	245	63,7
D+II-3: Aktive Äußerungen	139	36%	192	22%	53	14%
<i>KISTE</i>						
K/WO: Aktiver Wortschatz	131	34%	0	0%	253	66%
K/SB: Satzbildung	131	34%	0	0%	253	66%
K/ZU: SB/Zusatzauswertung	261	68%	0	0%	123	32%

Zu bedenken ist, dass von Testung zu Testung die Lücken an anderen Stellen auftreten können, d.h. es gibt die unterschiedlichsten Konstellationen innerhalb des Datensatzes. Da für die weiteren Analysen und die Zielsetzung des Probelaufes das erste und das letzte Modul von nachgeordneter Bedeutung sind (beim ersten lassen sich die Informationen im Ernstfall auch nachträglich einholen, beim letzten geht es um detaillierte Diagnostik), wurde für alle Fälle untersucht, welche der sieben anderen Module vollständig fehlten, teilweise oder ganz vorhanden waren. Nur von 22 Kindern lag ein vollständiger Datensatz vor. Da bei zwanzig Fällen die Lücken so erheblich waren, dass eine Auswertung bei ihnen keinen Sinn machte (bei siebzehn von ihnen lagen keinerlei Ergebnisse zu *DEUTSCH PLUS* vor), wurden diese von der weiteren Analyse ausgeschlossen, so dass der endgültige Datensatz **N = 364** Fälle umfasst.

Bei der Aufbereitung des Datensatzes ergab sich ein Problem mit den Urteilen der Testleiter/innen am Ende der Module. Das Urteil richtete sich nicht immer nach den vorgegebenen Kriterien, wobei es zwei Typen von Abweichungen gab.

Typ 1: Die Schwelle wurde nicht erreicht und dennoch wurde attestiert, das Kriterium sei erfüllt.

Typ 2: Die Schwelle wurde erreicht und dennoch wurde für eine Fortsetzung der Testung votiert bzw. sich für eine Förderung ausgesprochen.

Abweichungen des ersten Typs sind gravierender, denn eine unnötige Förderung ist weniger schlimm als eine notwendige nicht durchzuführen. Da im Probelauf überprüft werden sollte, wie gut *DEUTSCH PLUS*, wie gut also die ihm eigenen Kriterien im Hinblick auf das Bestimmen von Förderbedarf funktionieren, werden den Auswertungen korrigierte Ergebnisvariable zugrundegelegt, also solche, die sich strikt nach den im Erhebungsinstrument beschriebenen Auswertungskriterien richten. Bei jedem Modul geben wir die Abweichungen zwischen dem ursprünglichen Testleiter/innen- (*TL*) und dem kriteriumsorientierten Urteil (*PC-korrigiert*) an.

Wenn im Folgenden Ergebnisse aus dem Probelauf dargestellt werden, so ist zu beachten, dass diese zwar nicht repräsentativ für Berlin sind, aber für Berlin mögliche Ergebnisse darstellen. Und von diesen einen Eindruck zu gewinnen, war neben der Prüfung auf Praktikabilität und Validität Ziel des Probelaufes.

## ERGEBNISSE ZUM ERSTEN TEIL VON DEUTSCH PLUS

### D DEUTSCH PLUS/Teil I-2: FRAGEN AN DIE ELTERN

Vom zweiten Modul des ersten Teils haben wir bereits einige Variablen im Kapitel B vorgestellt: Erst- und familiäre Verkehrssprache sowie die Zweisprachigkeit. Neben weiteren Aspekten wie Geburtsland der Eltern sind es drei Fragen, die unmittelbar etwas mit einer eventuellen sprachlichen Förderbedürftigkeit zu tun haben: Die beiden parallel gestalteten Fragen "Wie gut beherrscht das Kind die deutsche Sprache?" und "Wie gut beherrscht das Kind die nicht deutsche Erstsprache?" sowie die Frage nach Besonderheiten oder Problemen in der Sprachentwicklung. Die beiden ersten Fragen sind in elf Punkte gegliedert, zu denen die Eltern angeben müssen, ob sie auf ihr Kind zutreffen oder nicht, wie z.B. das Kind verstehe Aufforderungen, kenne Lieder etc. Tabelle D.1 unterscheidet zwischen den Kindern, deren Erstsprache Deutsch oder eine andere Sprache ist, und ob die Eltern Aussage zur Kompetenz ihres Kindes in Deutsch bzw. in ihrer nichtdeutschen Erstsprache machen.

#### **D.1 Tabelle: Wie schätzen die Eltern die Sprachkompetenz ihres Kindes ein?**

	N	MW	s
Insgesamt	343	9,2	2,4
Erstsprache Deutsch Kompetenz in Deutsch	151	10,4	0,9
Erstsprache nicht-Deutsch Kompetenz in Deutsch	168	8,0	2,7
Erstsprache nicht-Deutsch Kompetenz in Erstsprache	139	8,9	1,8

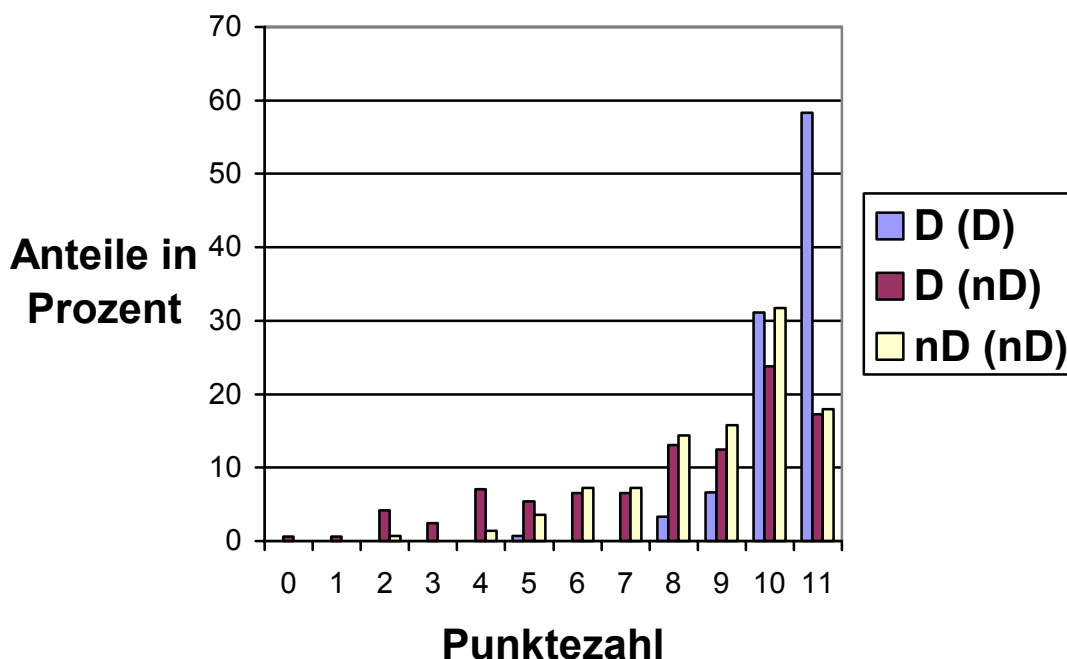
Die Mittelwerte zeigen, dass die deutschen Eltern die sprachliche Kompetenz ihres Kindes am höchsten einstufen (MW=10,4), höher auch als die Kompetenz der nichtdeutschen Kinder in deren Erstsprache (MW=8,9). Erstaunlich ist, wie gering die Differenz in den Bewertungen ist, wenn die nichtdeutschen Eltern die Sprachkompetenz ihrer Kinder in Deutsch (MW= 8,0) und in ihrer Erstsprache (MW=8,9) beurteilen sollen. In beiden Kombinationen werden selten die maximal möglichen Punktwerte vergeben. Wie die Punkteverteilungen der drei Kombinationen aussehen, wie also die Mittelwerte sich zusammensetzen, zeigt Abbildung D.2.

Nahezu 60% der Eltern vergeben für die Sprachkompetenz ihrer erstsprachlich deutschen Kinder den Höchstwert von 11 Punkten; in den beiden anderen Kombinationen sind es nicht einmal 20%.

Auch wenn das Ergebnis auf der subjektiven Einschätzung der Eltern beruht, ist es doch durchaus plausibel. So ist die soziale Zusammensetzung der dH- und der ndH-Schülerschaft zu Ungunsten letzterer verschieden, und ferner ist es schwieriger, bei ndH-Kindern in einer ansonsten deutschsprachigen Umgebung die eigene andere Erstsprache zu entwickeln als beim von vornherein deutschsprachigen Nachwuchs. Diese Konvergenz inhaltlicher Überlegungen und elterlicher Einschätzungen führt zur Frage, wie verlässlich die Elternurteile sind. Wir kommen auf diese Frage zurück.

**D.2 Abbildung: Wie schätzen die Eltern die Sprachkompetenz ihres Kindes ein?  
Verteilung der Punktwerte.**

D (D) : Eltern zur Deutschkompetenz, Erstsprache der Kinder: Deutsch  
 D (nD) : Eltern zur Deutschkompetenz, Erstsprache nicht-Deutsch  
 nD(nD) : Eltern zur Kompetenz in der nichtdeutschen Erstsprache



Interessanterweise korrelieren die Einschätzungen der Eltern von nD-Kindern, wenn sie die sprachliche Kompetenzen in Deutsch und in ihrer (jeweiligen) Erstsprache beurteilen, nur zu .25 (SPEARMANs  $\rho$ ). Der Wert ist zwar aufgrund der Fallzahlen hochsignifikant von Null verschieden, aber sehr niedrig. Aus der Sicht der Eltern können also die Entwicklungen der Erstsprache und von Deutsch bei ihren Kindern durchaus unterschiedlich verlaufen. Auch dies ein durchaus plausibles Ergebnis.

Die Frage nach sprachlichen Problemen oder Besonderheiten bejahen knapp 20% der 351 Eltern, die eine Antwort gaben. Dabei wird diese Frage häufiger von nichtdeutschen als von deutschen Eltern bejaht.

**E DEUTSCH PLUS/Teil I-3: FRAGEN AN DAS KIND**

Im dritten Modul des ersten Teils werden vierzehn Fragen an das Kind gerichtet und die Antworten daraufhin beurteilt, ob sie den Anforderungen nach Angemessenheit, Korrektheit und Flüssigkeit genügen. Wie Rückmeldungen der Testleiter/innen zeigten, weisen einige Fragen den Nachteil auf, durch kurze Antworten - wie z.B. Ja oder Nein - im Prinzip angemessen bearbeitbar zu sein, so dass sich kaum ein verlässliches Bild der sprachlichen Fähigkeiten gewinnen lässt, wenn das Kind nicht schon vorher bekannt ist. Statistisch kommt dies dadurch zum Ausdruck, dass alle Fragen zu mindestens 80% angemessen beantwortet wurden. Und es gibt einige Fragen, die durch sehr hohe Mittelwerte und sehr niedrige Streuungen gekennzeichnet sind, d.h. fast alle Kinder haben sie richtig beantwortet, die Items differenzieren nicht zwischen den Kindern. Hierzu zählen die Fragen

- 01: "Wie heißt du?"  
MW= 0,98 s= 0,15
- 04: "Hast du eine Schwester? Hast du einen Bruder? Wie heißt sie?/Wie heißt er?"  
MW= 0,97 s= 0,18
- 09: "Was trinkst du gern?"  
MW= 0,96 s= 0,19

Die Items mit (relativ) niedrigen Mittelwerten, die also stärker zwischen den Kindern differenzieren, sind:

- 03: "Wo wohnst du?"  
MW= 0,83 s= 0,38
- 06: "Was macht ihr [das Kind samt Freund/in] zusammen? Wo trefft ihr euch?"  
MW= 0,82 s= 0,38
- 07: "Was hast du gestern gemacht?"  
MW= 0,83 s= 0,38
- 10: "Was spielst du gerne? Womit spielst du gern?"  
MW= 0,83 s= 0,38
- 13: "Was möchtest du in der Schule lernen?"  
MW= 0,81 s= 0,40

Für jede angemessene Antwort bekommt das Kind einen Punkt zugewiesen, so dass ein Kind maximal 14 Punkte erreichen konnte. Der Mittelwert über alle 354 Kinder, von denen wir Angaben vorliegen haben, beträgt 12,5 - ist also im Einklang mit den skizzierten Teilergebnissen sehr hoch. Rund die Hälfte der Kinder erreicht die Maximalzahl von 14.

(Der Erhebungsbogen sieht im Anschluss an die vierzehn Fragen Platz für weitere Äußerungen des Kindes vor. In 7% der Bögen finden sich diverse Eintragungen.)

Die Testleiter/innen sind gehalten, die Anzahl angemessener Antworten zusammenzuzählen. Vergleicht man nun die Anzahl der Kreuze, die bei den einzelnen Fragen die Angemessenheit der Antworten markieren, mit der von den Testleitern/innen gebildeten Summe, so gibt es in immerhin 18 von 354 Fällen (rund 5%) Abweichungen, von denen zwei zu Ungunsten der Kindeleistungen sind und in allerdings nur einem Fall bei der Auswertung des ersten Teiles zu einem anderen Ergebnis führen würde. Dennoch ist zu überlegen, ob der Erhebungsbogen so gestaltet werden kann, dass die Häufigkeiten von Rechenfehlern verringert wird. Für die weitere Auswertung wird die korrigierte Variable zugrundegelegt.

**F DEUTSCH PLUS/Teil I-4: AUSWERTUNG**

Die Testleiter/innen sind gehalten, am Ende des ersten Teiles zwei Urteile abzugeben: Zum Einen ihre Schlussfolgerung aus den Fragen an die Eltern, wofür kein konkretes Kriterium vorgegeben wurde; zum Anderen die Auswertung der Fragen an das Kind mit dem explizit genannten Kriterium, das Kind müsse zehn der vierzehn Fragen angemessen beantwortet haben. Die erste Bewertung fehlt in 75 (21%) der Erhebungsbögen, die zweite in 10 (3%). Diese unterschiedlich hohen Ausfallquoten können sowohl auf das Fehlen eines expliziten Kriteriums im ersten Fall zurückzuführen sein als auch darauf, dass bei einigen Kindern (13) ein Elterngespräch (aus unbekanntem Gründen) nicht stattgefunden hat.<sup>3</sup>

Das Ergebnis aufgrund der Elternbefragung lautet: In 20% der Fälle ist eine Förderung erforderlich (59 von 289 bei 75 fehlenden Angaben). 80% der Kinder bedürfen keiner Sprachförderung.

**Tabelle F.1** zeigt, dass die Entscheidung der Testleiter/innen, ob ein Kind die Förderung benötigt oder nicht, weitgehend verträglich mit anderen Kriterien ist, aber nicht völlig widerspruchsfrei. Gravierend erscheint, wenn aufgrund des Elterngesprächs die Deutschkompetenz eines Kindes zwar nur hinsichtlich einer kleinen Anzahl von Aspekten als vorhanden eingeschätzt wird (0 bis 9 Punkte), dennoch aber eine Förderung für nicht erforderlich gehalten wird. Es gibt Extremfälle, da ein Kind nur ein oder zwei Punkte bekommt, dennoch aber keine Förderung vorgeschlagen wird.

Das Ergebnis aufgrund der Kinderbefragung lautet (TL-Urteil): In 18% der Fälle ist eine Fortsetzung der Sprachstandsfeststellung erforderlich (62 von 354 bei 10 fehlenden Angaben). 82% der Kinder bedürfen keiner Sprachförderung.

**Tabelle F.2** dokumentiert, welche Abweichungen es zwischen dem ursprünglichen TL-Urteil und einer strikten Anwendung des Kriteriums "10 von 14" gibt.<sup>4</sup>

Es zeigt, dass auch bei Erfüllung des Kriteriums in 23 Fällen entschieden wird, die Sprachstandsfeststellung fortzusetzen. Es gibt Kinder, bei denen wahrscheinlich aufgrund von Erfahrung oder von Auffälligkeiten während des Gespräches, die nicht im Erhebungsbogen dokumentiert sind, die Testleiter/innen für eine Fortsetzung plädieren, weil die Informationsbasis nicht solide genug ist. Derartige Entscheidungen sollten in den Hinweisen, wie *DEUTSCH PLUS* anzuwenden ist, durchaus positiv herausgestellt werden, denn eine hieraus resultierende Fehlklassifikation ist nicht so gravierend, wie ein Kind fälschlicherweise von der Förderung auszuschließen.

<sup>3</sup> Allerdings ist auch noch folgendes etwas seltsame Phänomen zu berücksichtigen: In 42 Erhebungsbögen wurden alle Fragen an das Kind bearbeitet, d.h. mit einem entsprechenden Kreuz versehen und deren Summe gebildet, dennoch aber kein entsprechendes Urteil am Ende von Teil I vermerkt. In diesen Fällen wurde die Bewertung, da eindeutig ableitbar, nachgetragen.

<sup>4</sup> Die Unterschiede, die zwischen der Addition der Testleiter/innen und der Nachzählung bestehen, sind geringfügig. Wichtig ist im vorliegenden Fall allein das eine Kind, das bei der Nachzählung das Kriterium verfehlt. Da es ein Einzelfall in der Stichprobe ist, lässt sich nicht vorhersagen, wie hoch der Anteil an Fehlklassifikationen sein wird, wird das Erhebungsinstrument flächendeckend eingesetzt.

**F.1 Tabelle:** Schlussfolgerung der Testleiter/innen (TL) aus dem Modul "Fragen an die Eltern" (D+/I-2) und ihr Zusammenhang mit weiteren Angaben.

TL-Schlussfolgerung Kriteriumsmerkmale	Förderung nicht erforder- lich	Förderung er- forderlich	Insgesamt
<b>Erstsprache</b>			
deutsch	129 100%	0 0%	129 100
nichtdeutsch	84 60%	57 40%	141 100%
<b>Familiensprache</b>			
deutsch	127 99%	1 1%	128 100%
nichtdeutsch	63 56%	50 44%	113 100%
<b>Zweisprachigkeit/Verstehen<sup>#</sup></b>			
Nein	0 0%	5 100%	5 100%
Ja	84 62%	52 38%	136 100%
<b>Zweisprachigkeit/Sprechen<sup>#</sup></b>			
Nein	3 27%	8 73%	11 100%
Ja	81 62%	49 38%	130 100%
<b>Eltern zur D-Kompetenz</b>			
0 - 9 Punkte	41 45%	51 55%	92 100%
10 Punkte	72 91%	7 9%	79 100%
11 Punkte*	114 100%	0 0%	114 100%
<b>Eltern zur nD-Erstsprachenkomp.<sup>#</sup></b>			
0 - 9 Punkte	36 53%	32 47%	68 100%
10 Punkte	24 60%	16 40%	40 100%
11 Punkte*	18 78%	5 22%	23 100%

**F.2 Tabelle:** Schlussfolgerung der Testleiter/innen (TL) aus dem Modul "Fragen an das Kind" (D+/I-3) und ihr Zusammenhang mit dem vorgegebenen Kriterium, mindestens 10 von 14 Fragen adäquat beantwortet zu haben.

TL-Schlussfolgerung Kriteriumsorientiertes Urteil	Förderung nicht erforderlich	Fortsetzung des Tests	Insgesamt
<b>Förderung nicht erforderlich</b> (10 von 14 erreicht)	292 82%	23 7%	315 89%
<b>Fortsetzung des Tests</b> (10 von 14 nicht erreicht)	0 0%	39 11%	39 11%
<b>Insgesamt</b>	292 82%	62 18%	354 100%

Dieses Kapitel abschließend wenden wir uns der Frage zu, wie weit die beiden Teilurteile (Schlussfolgerungen aufgrund der Elternaussagen und aufgrund der Kinderantworten) miteinander verträglich sind; Tabelle F.3.

<sup>#</sup> Aussagen der Eltern von ndH-Kindern

\* Vgl. Kapitel D. Maximal möglich sind 11 Punkte.

**F.3 Tabelle: Vergleich der beiden TL-Teilurteile am Ende des ersten Teils von DEUTSCH PLUS.** (In Klammern die Werte, falls das Kriterium "10 von 14" strikt angewendet wird.)

Eltern-Ergebnis Kind-Ergebnis	Förderung nicht erforder- lich	Förderung er- forderlich	Insgesamt
<b>Keine Fortsetzung, keine Förderung er- forderlich</b>	219 76% (222 77%)	23 8% (37 13%)	242 84% (259 90%)
<b>Fortsetzung der Te- stung</b>	11 4% ( 8 3%)	36 12% (22 7%)	47 16% (30 10%)
<b>Insgesamt</b>	230 80%	59 20%	289 100%

In 289 Fällen liegen alle erforderlichen Angaben vor, um die Zuordnung der beiden Teilurteile vornehmen zu können.

Legen wir die ursprünglichen TL-Urteile zugrunde, so sind die 219 plus 36 Fälle, in denen entweder beide Male für eine Förderung entschieden wird, oder in denen eine Fortsetzung der Testung bzw. eine Förderung angezeigt erscheint, miteinander kompatibel. In 88% der Fälle erhalten wir in der Stichprobe eine stimmige Entscheidung am Ende des ersten Teils. Widersprüche zwischen den beiden Teilurteilen treten in 12% der Fälle auf. Ob sie gravierende Konsequenzen haben oder nicht, hängt davon ab, wie der Konflikt gelöst wird. Im Zweifelsfalle, so muss die grundlegende Richtlinie lauten, ist für eine Teilnahme an der Förderung zu entscheiden.

Das Bild verändert sich etwas, beachten wir strikt das vorgegebene Kriterium. Wir wissen bereits aus Tabelle F.2, dass dann für 89% der Kinder die Entscheidung gegen eine Förderung fällt im Vergleich zu 82% nach dem TL-Urteil, die hier vorsichtiger bewerten. Tabelle F.3 zeigt, dass die Unstimmigkeit mit der Schlussfolgerung aus dem Elterngespräch zunimmt. Zu überlegen ist demnach, ob nicht das Kriterium erhöht werden sollte, dass z.B. 11 von 14 Antworten angemessen sein müssen.

## ERGEBNISSE ZUM ZWEITEN TEIL VON DEUTSCH PLUS

### G DEUTSCH PLUS/Teil II-1: PASSIVER WORTSCHATZ

Der Teil D+/II-1 wurde - abweichend von der Planung - 120 Kindern nicht vorgegeben, bei denen also ein Vergleich der Ergebnisse aus dem ersten und dem zweiten Teil nicht möglich ist. Die folgenden Werte beziehen sich somit auf N=244 Fälle.

Auf zwei sog. Wimmelbilder waren Dinge oder Vorgänge zu identifizieren, wobei die Anzahl richtig gezeigter Nomen (mindestens 20) und Verben (mindestens 8) als Schwelle definiert wurde. Maximal möglich richtig zu nennen waren 24 Nomen und 10 Verben. Tabelle G.1 gibt die wichtigsten Ergebnisse wieder.

#### G.1 Tabelle: Anzahl richtig gezeigter Nomen und Verben sowie das (strikt kriteriumsorientierte) Gesamtergebnis aus dem Modul D+/II-1.

Aspekt	Häufigkeit	Anteil
<b>Richtig gezeigte Nomen</b>		
0 - 12	15	6%
13 - 19	82	34%
20 - 23	116	48%
24	31	13%
	<b>MW= 19,7</b>	<b>s= 3,8</b>
<b>Richtig gezeigte Verben</b>		
0 - 5	14	6%
6 - 7	38	16%
8 - 9	109	45%
10	83	34%
	<b>MW= 8,5</b>	<b>s= 1,6</b>
<b>Modul insgesamt bearbeitet</b>		
nicht erfolgreich	106	43%
erfolgreich <sup>5</sup>	138	57%

Das Teilkriterium Nomen erreichen 40% der getesteten Kinder nicht, das Teilkriterium Verben 22%. Da beide Teilkriterien erfüllt sein mussten, damit der Teil D+/II-1 als erfolgreich bearbeitet gewertet wurde, bleiben insgesamt sogar 43% der Kinder unterhalb der Schwelle.

Von besonderem Interesse ist es, ob sich das Modul ohne Schaden für die Zuverlässigkeit des Ergebnisses kürzen lässt, ob es also reicht, statt zweier nur ein Bild vorzugeben. Hierzu die Tabelle G.2:

Wir können die jeweils 12 (Nomen) plus 5 (Verben) Fragen zu den Wimmelbildern als Items von zwei Subskalen auffassen, die wir D2-BILD1 und D2-BILD2 nennen wollen. Im Probelauf sind diese beiden Teilskalen zusammengenommen worden; die bisherige Gesamtskala sei mit D2-GANZ bezeichnet. Dann zeigt Tabelle G.2 zweierlei:

<sup>5</sup> Verwendet wurde die korrigierte Ergebnisvariable, d.h. die Variable, die sich bei strikter Anwendung des vorgegebenen Kriteriums ergibt.

- Aufgrund der geringeren Länge jeder der Subskalen geht deren innere Konsistenz (gemessen an CRONBACHs  $\alpha$ ) erwartungsgemäß zurück, allerdings vor allem bei der zweiten Subskala nicht in erheblichem Maße.
- Von den beiden Subskalen korreliert die zweite mit ihren beiden Bestandteilen (Nomen und Verben) höher mit der Gesamtskala als die erste.

**G.2 Tabelle:** Reliabilität (Interne Konsistenz) und Korrelationen der dem Modul D+/II-1 zugrundeliegenden (Teil)Skalen.

Skala	Reliabilität $\alpha$	$\alpha$ (standardisiert)
D2-GANZ	.86	.86
D2-BILD1	.73	.77
D2-BILD2	.79	.78
<b>Korrelationen</b> (SPEARMANs $\rho$ )		
D2-GANZ/NOMEN	D2-NOMEN1 .84	D2-NOMEN2 .93
D2-GANZ/VERBEN	D2-VERB1 .69	D2-VERB2 .89

Aus statistischer Sicht könnte demnach auf die Vorgabe des ersten Bildes verzichtet werden. Wo aber sollte die Schwelle gesetzt werden, wird allein das Bild 2 eingesetzt? Maximal können 12 Nomen und 5 Verben richtig gezeigt werden. Die Mittelwerte betragen 9,6 richtig gezeigte Nomen ( $s=2,3$ ) und 2,3 richtig gezeigter Verben ( $s=1,1$ ). Tabelle G.3 dokumentiert - behalten wir die bisherige Entscheidung, die aufgrund beider Bilder getroffen wurde, bei -, wie sich Punktezahlen beim zweiten Wimmelbild und Erfolg bei diesem Modul verteilen.

**G.3 Tabelle:** Anzahl richtig gezeigter Nomen und Verben beim zweiten Bild sowie Ergebnis aus dem Modul D+/II-1.

Wimmelbild 2 Aspekt	D+/II-1 nicht erfolgreich	D+/II-1 erfolgreich
<b>Richtig gezeigte Nomen</b>		
0 - 7	49 100%	0 0%
8 - 9	40 89%	5 11%
10	14 38%	23 62%
11	3 6%	48 94%
12	0 0%	62 100%
	106	138
<b>Richtig gezeigte Verben</b>		
0 - 2	32 100%	0 0%
3	26 76%	8 24%
4	23 33%	46 67%
5	25 23%	84 77%
	106	138

Eine Betrachtung der Tabelle G.3 lehrt, dass als Mindestkriterium 4 Verben richtig gezeigt werden müssen. (Darunter ist die Quote der Nichterfolgreichen größer als die der Erfolgreichen.) Aber wo sollte die Grenze für die Nomen gezogen werden?

**G.4 Tabelle:** Die Erfolgsquoten beim Modul D+/II-1 nach altem und nach den beiden neuen Kriterium. (N=244 Fälle (100%), die dieses Modul bearbeitet haben.)

Nach altem Kriterium	D+/II-1 nicht erfolgreich		D+/II-1 erfolgreich			
<b>Nur bei Wimmelbild 2</b>						
Kriterium 1: Richtig gezeigt ≥ 10 Nomen und 4 Verben						
Nicht erfolgreich	96	39%	13	5%	109	45%
erfolgreich	10	4%	125	51%	135	55%
Kriterium 2: Richtig gezeigt ≥ 11 Nomen und 4 Verben						
Nicht erfolgreich	104	43%	33	13%	137	56%
erfolgreich	2	1%	105	43%	107	44%
	106	43%	138	57%		

Das Ergebnis, das uns Tabelle G.4 mit einem Vergleich der Erfolgsquoten bei altem und neuen Kriterien zeigt, lautet: Insgesamt liegt das leichtere Kriterium (mindestens 10 Nomen) dichter am alten Resultat (erfolgreich sind 55% bzw. 57% der Kinder). Und die Gesamtzahl an "Fehlklassifikationen" - nehmen wir die alte Zuordnung als Bezugspunkt - ist bei diesem Kriterium mit 9% (5% + 4%) niedriger als bei dem etwas schwereren (14% = 13% + 1%).

Abschließend vergleichen wir das ursprüngliche TL-Urteil mit dem strikt kriteriumsorientierten; vgl. Tabelle G.5. In 3% der Fälle weichen die beiden Bewertungsmodi voneinander ab. Im Gegensatz zur Schlussfolgerung am Ende des ersten Teils ist diesmal das kriteriumsorientierte konservativer als das TL-Urteil in dem Sinne, dass für mehr Kinder eine Förderung befürwortet wird (106 Kinder gegenüber 102).

**G.5 Tabelle:** Vergleich der beiden Urteile am Ende des ersten Moduls von *Deutsch Plus*, Teil II.

kriteriumsorientiert TL-Urteil	nicht erfolgreich		erfolgreich		Insgesamt	
nicht erfolgreich	100	41%	2	1%	102	42%
erfolgreich	6	2%	136	56%	142	58%
Insgesamt	106	43%	138	57%	244	100%

**H DEUTSCH PLUS/Teil II-2: AUFGABENVERSTÄNDNIS**

Der Teil D+/II-2 wurde 119 Kindern nicht vorgegeben. Die folgenden Werte beziehen sich somit auf N=245 Fälle. Anhand von Spielsachen waren einfache und komplexe Anweisungen umzusetzen. Als Schwelle für eine erfolgreiche Bearbeitung dieses Moduls galt die Umsetzung von mindestens fünf (von maximal sechs) einfachen und drei (von maximal vier) komplexen Anweisungen. Tabelle H.1 gibt die wichtigsten Ergebnisse wieder. Zugrundeliegt wiederum die kriteriumsorientierte Sichtweise.

**H.1 Tabelle: Anzahl richtig umgesetzter einfacher und komplexer Anweisungen sowie das (kriteriumsorientierte) Gesamtergebnis vom Modul D+/II-2.**

	Häufigkeit	Anteil
<b>Richtig verstanden:</b>		
<b>Einfache Anweisungen</b>		
0 - 4	136	56%
5	42	17%
6	67	27%
	MW= 4,2	s= 1,5
<b>Komplexe Anweisungen</b>		
0 - 2	11	4%
3	22	9%
4	212	87%
	MW= 3,8	s= 0,6
<b>Modul insgesamt bearbeitet</b>		
nicht erfolgreich	140	57%
erfolgreich	105	43%

Waren beim ersten Modul des zweiten Teiles von *DEUTSCH PLUS* mit 57% noch die Mehrheit der Kinder erfolgreich, so dreht sich dieses Verhältnis exakt um: Nun erfüllen 57% der Kinder nicht das Kriterium. Darin kommt der gesteigerte Schwierigkeitsgrad der Aufgabenstellung zum Ausdruck.

Die interne Konsistenz des Moduls D+/II-2 ist mit  $\alpha=.62$  nur mäßig hoch.

Wie weit TL- und kriteriumsorientiertes Urteil voneinander abweichen, zeigt Tabelle H.2. Abweichungen treten in 2% der Fälle auf und zwar ausschließlich in der Richtung, dass die Kriteriumsorientierung konservativer ist.

**H.2 Tabelle: Vergleich der beiden Urteile am Ende des zweiten Moduls von *Deutsch Plus*, Teil II.**

<b>kriteriumsorientiert TL-Urteil</b>	<b>nicht erfolgreich</b>	<b>erfolgreich</b>	<b>Insgesamt</b>
<b>nicht erfolgreich</b>	136 56%	0 0%	136 56%
<b>erfolgreich</b>	4 2%	105 43%	109 44%
<b>Insgesamt</b>	140 57%	105 43%	245 100%

## I DEUTSCH PLUS/Teil II-3: AKTIVE ÄUßERUNGEN

Der Teil D+/II-3 wurde 119 Kindern nicht vorgegeben. Die folgenden Werte beziehen sich somit auf N=245 Fälle.

Das Material dieses Moduls besteht aus den beiden Bildserien "Was ist hier los?" und "Glas auf dem Schrank". Während die zweite Serie eine kleine Geschichte wiedergibt (die Bilder sind in der richtigen Reihenfolge dem Kind vorzulegen), besteht die erste aus Einzelkarten. Das Kind kann eine der beiden Serien auswählen. Entweder äußert sich das Kind dann spontan zu den Bildern oder es wird ein Impuls gegeben, der nicht zuviel Wortschatz anbietet. Protokolliert werden die Äußerungen des Kindes und hinsichtlich dreier Aspekte ausgewertet:

- Struktur. Kriterium: Mindestens zwei Drittel der kindlichen Äußerungen müssen länger als drei Wörter sein.
- Wortschatz/Nomen. Kriterium: Verwendung von überwiegend sachlich angemessenen Nomen.
- Wortschatz/Verben. Kriterium: Verwendung von überwiegend sachlich angemessenen Verben.

Die Schwelle gilt dann als erreicht, wenn alle drei Kriterien erfüllt sind. Tabelle I.1 gibt die wichtigsten Ergebnisse wieder. Grundlage ist wiederum die Kriteriumsorientierung.

### I.1 Tabelle: Erfüllen der Teilkriterien im Modul D+/II-3 sowie (kriteriumsorientiertes) Gesamtergebnis. Siehe Text.

	Häufigkeit	Anteil
<b>Teilkriterium Struktur</b>		
nicht erfüllt	65	27%
erfüllt	180	73%
<b>Teilkriterium Nomen<sup>#</sup></b>		
nicht erfüllt	48	20%
erfüllt	196	80%
<b>Teilkriterium Verben<sup>#</sup></b>		
nicht erfüllt	54	22%
erfüllt	190	78%
<b>Modul insgesamt bearbeitet</b>		
nicht erfolgreich	99	40%
erfolgreich	146	60%

<sup>#</sup> Von einem Kind kennen wir zwar nicht die Teilergebnisse zu den Kriterien *Nomen* und *Verben* (daher hier N=244), aber der Erhebungsbogen enthält die Angaben zu den Mehrwortäußerungen und das Testleiterurteil am Ende des Moduls (daher hier N=245).

Während für jedes Teilkriterium die Erfolgsquoten recht hoch sind, so ergibt deren Zusammenschau einen größenordnungsmäßig ähnlichen Anteil von Kindern, die die Schwelle überschreiten, wie beim ersten Modul von D+/II.

Die Häufigkeiten, mit denen die beiden Bildserien gewählt wurden, sind stark unterschiedlich. Tabelle I.2 liefert zusätzlich zu der Präferenz die jeweils damit erzielten Erfolgsquoten.

**I.2 Tabelle: Erfüllen der Teilkriterien im Modul D+/II-3 sowie Gesamtergebnis.**  
Siehe Text.

	Bildserie 1		Bildserie 2		Beide Bildserien	
<b>Teilkriterium Struktur</b>						
nicht erfüllt	52	36%	4	8%	9	17%
erfüllt	91	64%	45	92%	44	83%
	143	100%	49	100%	53	100%
<b>Teilkriterium Nomen</b>						
nicht erfüllt	23	16%	13	27%	12	23%
erfüllt	119	84%	36	73%	41	77%
	142	100%	49	100%	53	100%
<b>Teilkriterium Verben</b>						
nicht erfüllt	32	23%	8	16%	14	26%
erfüllt	110	77%	41	84%	39	74%
	142	100%	49	100%	53	100%
<b>Modul insgesamt bearbeitet</b>						
nicht erfolgreich	64	45%	16	33%	19	36%
erfolgreich	79	55%	33	67%	34	64%
	143	100%	49	100%	53	100%

Die Kinder, die die zweite Bildserie gewählt haben oder beide bearbeiteten, erzielen ein weit aus besseres Gesamtergebnis in diesem Modul als die anderen Kinder. Diese Aussage lässt sich aber nicht gleichmäßig auf die drei Teilkriterien übertragen. Diese Kinder sind zwar deutlich besser bei der Anzahl der Mehrwortäußerungen, erfüllen aber in geringerem Maße das Teilkriterium *Nomen*. Beim Teilkriterium *Verben* liegen die drei Kindergruppen nicht weit auseinander. Hierfür mag es zwei Gründe geben: Zum Einen unterscheiden sich vielleicht die Kinder, die sich durch eine Bildgeschichte (Serie 2) angezogen fühlen, von jenen, die die Einzelbilder kommentieren. Zum Anderen liegt es am Material selber, denn eine Bildgeschichte provoziert von sich aus längere Äußerungen als Einzelbilder. Treffen diese Überlegungen zu, so könnte dies bedeuten, dass man den Kindern nicht gerecht wird, wird bei beiden Bildserien derselbe Maßstab angelegt. Dies wäre dann problematisch, würde zwischen Bildmaterial und Kindertyp ein systematischer Zusammenhang bestehen.

Das dritte Modul hängt in weit stärkerem Maße von den Einschätzungen der Testleiter/innen ab als die beiden voranstehenden. Die Äußerungen der Kinder sind unter verschiedenen Aspekten zu beurteilen. Dies drückt sich bei der teststatistischen Analyse darin aus, dass wir weitgehend nur über aggregierte Daten verfügen (die Angaben zu Struktur und Wortschatz), wenn nicht die Mühsal einer detaillierten inhaltlichen Analyse der Wortprotokolle der kindlichen Äußerungen auf sich genommen wird, wobei hier vorausgesetzt werden müsste, dass in der Testsituation tatsächlich wörtlich und genau protokolliert wird. Betrachten wir die drei Teilkriterien als eine Form aggregierter Items der Skala *Aktive Äußerungen*, so ergibt die

Reliabilitätsberechnung einen Konsistenzkoeffizienten von  $\alpha=.67$ , also dennoch zufriedenstellend hoch.

Die stärkere Abhängigkeit der Werte von der Auswertung durch die jeweiligen Testleiter/innen mag mit ein Faktor sein, warum im dritten Modul weit stärker als bei den beiden vorausgegangenen die TL- und die kriteriumsorientierten Urteile auseinanderklaffen, wie Tabelle I.3 zeigt.

**I.3 Tabelle: Vergleich der beiden Urteile am Ende des dritten Moduls von *Deutsch Plus*, Teil II.**

<b>kriteriumsorientiert TL-Urteil</b>	nicht erfolgreich	erfolgreich	Insgesamt
nicht erfolgreich	78 32%	14 6%	92 38%
erfolgreich	17 7%	132 55%	149 62%
Insgesamt	95 39%	146 61%	241 100%

In 13% der Fälle (23 von 241) stimmen die beiden Resultate nicht überein, wobei das kriteriumsorientierte Urteil wieder etwas konservativer ist als das der Testleiter/innen (17 vs. 14 widersprüchliche Zuordnungen).<sup>6</sup> Das Ausmaß der Unterschiede ist erstaunlich, denn das kriteriumsorientierte Urteil nutzt nichts anderes als die Angaben der Testleiter/innen, die diese aber anders auswerten oder bis zu einem gewissen Grade nicht berücksichtigen, wobei wir die Gründe nicht kennen. Das dritte Modul ist jenes, das sowohl an die Kinder wie an die Auswertenden die größten Anforderungen aller drei Module stellt. Es ist durchaus möglich, dass durch die Testsituation das Kind ein Mehr an Falschem und Richtigem produziert, als durch die auswertenden Abfragen von *DEUTSCH PLUS* erfasst wird. Und dieses Mehr bringt die Testleiter/innen zu ihrer abweichenden Einschätzung.

<sup>6</sup> An Tabelle I.3 zeigt sich, welchen Effekt fehlende Angaben haben können. Von vier Kindern, die nach den vorgegebenen Kriterien nicht erfolgreich waren, kennen wir nicht die TL-Einschätzung. Daher drücken sie in der Tabelle I.1 die Erfolgsquote auf 60%, während sie in Tabelle I.3 aufgrund der fehlenden Angaben auf 61% steigt und damit den Unterschied zwischen den beiden Urteilen verringert.

## J DEUTSCH PLUS/Teil II-4: GESAMTAUSWERTUNG DER BEOBACHTUNGSSITUATIONEN (FAZIT AUS DEM ZWEITEN TEIL)

Betrachten wir die Gesamtauswertung des zweiten Teils von *DEUTSCH PLUS*, so stehen wir vor der Schwierigkeit, dass es mehrere Bezugspunkte gibt: Zum Einen die ursprünglichen Testleiter/innen-Urteile, zum Anderen die strikt kriteriumsorientierten. Der erste Bezugspunkt verkompliziert sich zusätzlich dadurch, dass die Entscheidungen am Ende eines jeden Moduls auf das Endblatt übertragen werden mussten und bei diesen Übertragungen Ungenauigkeiten vorkamen (knapp 2% der Fälle pro Modul und insgesamt - da teilweise dieselben Fälle betroffen waren - bei 4%). Bei der Anwendung der übertragenen Teilergebnisse auf das Gesamturteil wurde sich aber an das vorgegebene Kriterium (insgesamt erfolgreich nur dann, wenn alle drei Teilergebnisse positiv sind) ohne Ausnahme gehalten. Würden hingegen die Teilergebnisse, die am Ende der jeweiligen Module stehen, herangezogen, müsste das abschließende Urteil in drei Fällen anders lauten. Da die unmittelbar am Ende eines Moduls stehenden Teilergebnisse von keinem Übertragungsfehler betroffen sind, ziehen wir bei den nachstehenden Auswertungen diese heran, sofern Teilergebnisse eine Rolle spielen. Doch zunächst verschaffen wir uns einen Überblick über die abschließenden Bewertungen.

### J.1 Tabelle: Ist laut *DEUTSCH PLUS*/Teil II Teilnahme an Sprachfördermaßnahmen erforderlich?

Teilnahme erforderlich?	Häufigkeiten	
TL-Urteil		
Nein	78	32%
Ja	163	68%
	241	100%
kriteriumsorientiert		
Nein	71	29%
Ja	173	71%
	244	100%

Auch insgesamt gilt, dass wir ein konservativeres Ergebnis (im Sinne von mehr Empfehlungen zur Förderung) bei der strikt kriteriumsorientierten Entscheidung erhalten als beim TL-Urteil. Die in *DEUTSCH PLUS* enthaltenen Kriterien sind also vorsichtiger als diejenigen, die im Probelauf die Auswertenden vorgenommen haben.

Da die Stichprobe bei weitem nicht repräsentativ ist, ist es sehr schwierig, abzuschätzen, wie viele Entscheidungen für eine Förderung bei einem flächendeckenden Einsatz fallen. Gehen wir aufgrund der Tabelle F.3, die die Widersprüche in den bis dahin erzielten Ergebnissen aufzeigt, von 20% Kindern aus, die in den Teil II kommen, und von diesen wiederum rund 70% einer Förderung bedürfen, dann erhielten wir als Schätzwert einen Anteil von knapp 15% förderbedürftiger Kinder, ein Wert, der trotz aller Unwägbarkeiten eine realistische obere Grenze darstellen dürfte.

Allerdings gibt es ein erhebliches Problem, das auch für die Schätzung Relevanz besitzt, aber darüber hinaus reicht und in der nachstehenden Tabelle J.2 dokumentiert wird: Die Abweichungen zwischen den Einschätzungen aus dem ersten und dem zweiten Teil von *DEUTSCH PLUS*.

**J.2 Tabelle:** Zusammenhang zwischen den Urteilen am Ende der beiden Teile von DEUTSCH PLUS.

a) Urteil am Ende von D+/I-3 (Fragen an das Kind) und das Gesamturteil.

D+/II: Gesamtbewertung	TL-Urteil				
	Keine Förderung erforderlich		Förderung erforderlich		Insgesamt
D+/I-3: Fragen an das Kind					
TL-Urteil					
Keine Förderung notwendig	72	31%	109	46%	181 77%
Fortsetzung der Testung	4	2%	50	21%	54 23%
Insgesamt	76	32%	159	68%	235 100%
	Kriteriumsorientiertes Urteil:				
	Keine Förderung erforderlich		Förderung erforderlich		Insgesamt
Kriteriumsorientiertes Urteil					
Keine Förderung notwendig	70	30%	133	57%	203 86%
Fortsetzung der Testung	1	0%	31	13%	32 14%
Insgesamt	71	30%	164	70%	235 100%

Tabelle J.2 macht das Ausmaß des Problems deutlich. Bezogen auf die Menge von N=235 Kindern, von denen wir Daten aus dem zweiten Teil haben, kommen Teil I und Teil II in rund der Hälfte der Fälle zu unterschiedlichen Ergebnissen. Dabei treten weniger Abweichungen auf, wenn wir für beide Urteile die Einschätzung der Testleiter/innen heranziehen, als wenn wir die strikte Ausrichtung an den vorgegebenen Kriterien zugrundelegen. Das dürfte auf einer Tendenz beruhen, die eigene Sichtweise möglichst konsistent zu halten.

Wiederholen wir den Vergleich, indem wir nun das Ergebnis aus der Elternbefragung betrachten, so steht allein das TL-Urteil zur Verfügung; vgl. Tabelle J.2-b.

**J.2 Tabelle:** b) Urteil am Ende von D+/I-2 (Fragen an die Eltern) und das Gesamturteil.

D+/II: Gesamtbewertung	TL-Urteil (Kriteriumsorientiertes Urteil)				
	Keine Förderung erforderlich		Förderung erforderlich		Insgesamt
D+/I-2: Fragen an die Eltern					
TL-Urteil					
Keine Förderung notwendig	59	32%	71	39%	130 71%
	(55)	(30%)	(75)	(41%)	(130 71%)
Förderung erforderlich	2	1%	50	28%	52 29%
	(3)	(2%)	(49)	(27%)	(52 29%)
Insgesamt	61	33%	121	67%	182 100%
	(58)	(32%)	(124)	(68%)	(182 100%)

Das Bild, das sich aus J.2-b ergibt, ist etwas günstiger als zuvor. Das dürfte daran liegen, dass die rund einhundert Kinder, die nicht in den zweiten Teil der Sprachstandsfeststellung gekommen sind, solche waren, die aufgrund des Elterngesprächs als nicht förderungsbedürftig eingeschätzt wurden. Wären sie tatsächlich den Modulen des zweiten Teils ausgesetzt worden, wie es die Planung des Probelaufs vorsah, dann hätten wir verlässlichere und genauere Aussagen zu dem Phänomen widersprüchlicher Ergebnisse, die auf die Grenzen der Validität verweisen, treffen können.

In jedem Falle jedoch hat das Problem ein Ausmaß, das Konsequenzen für die Gestaltung des Erhebungsinstrumentes verlangt. Wir kommen im letzten Kapitel darauf zurück.

## ERGEBNISSE ZUM PARALLELTEST *KISTE*

### **K** *KISTE*/Prolog: RAHMENBEDINGUNGEN

Zu Beginn von *KISTE*, der als Paralleltest zu *DEUTSCH PLUS* vorgegeben wurde, werden die Durchführenden gebeten, die Sprechfreudigkeit des Kindes ebenso einzuschätzen wie ihren Eindruck des kindlichen Verständnisses vom Geforderten. Die Einschätzung der Erzieher/innen bzw. Vorklassenleiter/innen einerseits und der Sonderpädagogen/innen, die für die Durchführung von *KISTE* verantwortlich zeichneten, stimmen weitgehend überein.<sup>7</sup>

Leider liegen nur von einhundert Kindern die Daten zu diesen Rahmenbedingungen des Testens vor. Bei einer dreistufigen Einschätzung (gering, mittel, hoch) gilt ungefähr, dass die Hälfte der Kinder in der Mitte verortet wurden und der Rest gleichermaßen in der oberen wie in der unteren Kategorie. Dies gilt allerdings nur für das Merkmal *Sprechfreudigkeit*. Hinsichtlich des Verständnisses wurde die Hälfte der Kinder in der unteren Kategorie eingestuft und ein Drittel in der mittleren. Hier besteht sicherlich ein Zusammenhang mit dem Ergebnis, das im vorstehenden Kapitel wiedergegeben wurde, dass nämlich am Ende von D+/Teil II bei 70% der Kinder ein Förderbedarf festgestellt wurde, dies wiederum Ausdruck der speziellen Stichprobe, die dem Probelauf zugrundelag.

Korrelieren wir Sprechfreudigkeit und Verständnis mit den Ergebnissen aus dem dazugehörigen Test *KISTE*, so ergeben sich mittelhohe Koeffizienten, die zwischen .33 und .67 schwanken, also durchaus etwas zur Erhellung der gezeigten Leistungen beitragen können. Daher ist zu prüfen, wie weit auch für *DEUTSCH PLUS* etwas von der Testatmosphäre festgehalten werden sollte.

---

<sup>7</sup> SPEARMANs  $\rho$  beträgt .77 und ist hochsignifikant von Null verschieden.

**L KISTE: ERGEBNISSE ZUR SPRACHKOMPETENZ**

Aus *KISTE* wurden zwei Subtests ausgewählt, die den Kindern außer *DEUTSCH PLUS* vorgegeben waren: Aktiver Wortschatz (WO) und Satzbildung (SB). Hierzu mehr im Kapitel C, das eine kurze Erläuterung zu den T-Werten und Prozenträngen enthält, die in die folgenden Darstellungen der Ergebnisse mit eingehen.

**L.1 Tabelle: Ergebnisse aus den Subtests Aktiver Wortschatz (WO) und Satzbildung (SB) von *KISTE*.**

<b>WO</b>	<b>Punkte</b> N=236	<b>T-Werte<sup>#</sup></b> N=192	<b>Prozentränge<sup>#</sup></b> N=192
Minimum	3	5	0
Maximum	128	79	100
Mittelwert	44	39	24
Streuung	20	12	26
<b>SB: Semantik</b>			
Minimum	0	30	2
Maximum	21	5	100
Mittelwert	9	48	43
Streuung	7	13	36
<b>SB: Grammatik</b>			
Minimum	0	31	3
Maximum	13	72	99
Mittelwert	4	44	32
Streuung	4	11	31
<b>SB: Gesamtsumme</b>			
Minimum	0	27	0
Maximum	34	73	99
Mittelwert	12	45	37
Streuung	10	13	33

<sup>#</sup> Die Anzahl der Fälle ist für die T-Werte und Prozentränge deshalb niedriger als für die Rohwerte, weil nicht für alle Altersstufen Umrechnungstabellen zur Verfügung standen.

Eine Betrachtung der Tabelle L.1 zeigt dreierlei:

- Die Leistungen der Kinder sind im Mittel unterdurchschnittlich, wie ein Blick auf die T-Werte lehrt, die so genormt sind, dass deren Mittel bei 50 (und deren Streuung bei 10) liegt. Alle in der Stichprobe erzielten Mittelwerte liegen mit 39, 48, 44 und 45 darunter.
- Zugleich ist die Heterogenität unserer Stichprobe sehr groß. Die Streuungen der Punkte (Rohwerte) bewegen sich beim Subtest SB in allen seinen Facetten in derselben Größenordnung wie die Mittelwerte und erreichen bei WO immerhin einen halben Mittelwert.
- Die Kinder sind im Subtest SB deutlich besser als im Subtest WO.

Über die Standardauswertung des Subtests SB hinaus bietet KISTE die Möglichkeit, die Äußerungen der Kinder anhand weiterer Aspekte zu analysieren, was differenzierte diagnostische Möglichkeiten eröffnet und Hinweise auf Ansatzpunkte für eine Förderung liefert. Dies ist im vorliegenden Zusammenhang, in dem der Schwellentest *DEUTSCH PLUS* untersucht wird, kaum von Belang, kann aber das Verständnis unserer Ergebnisse erhöhen. Tabelle L.2 geht auf die drei zentralen Merkmale ein, zu denen die Zusatzauswertung führt:

- **Anz SAS:** Anzahl der semantisch akzeptablen Sätze
- **Quo FSAS:** Fehlerquotient aus der Summe aller grammatikalischen Fehler bezogen auf die Anzahl der semantisch akzeptablen Sätze
- **Quo WSAS:** Wörterquotient aus der Summe aller Wörter in den semantisch akzeptablen Sätzen bezogen auf die Anzahl der semantisch akzeptablen Sätze.

Tabelle L.2 fasst die zentralen Parameter der drei Ergebnismerkmale zusammen.

**L.2 Tabelle: Ergebnisse aus der Zusatzauswertung des Subtests *Satzbildung* (SB) von *KISTE*. (Abkürzungen siehe Text.)**

<b>SB/ Zusatzauswertung</b>	<b>Anz SAS N=107</b>	<b>Quo FSAS N=102</b>	<b>Quo WSAS N=91</b>
Minimum	0	0,1	1
Maximum	15	24	13
Mittelwert	7,0	2,0	5,8
Streuung	4,1	3,0	2,1

Ein Wert von 0,1 des Merkmals *Quo FSAS* heißt, dass nur 0,1 Fehler pro Satz gemacht wurden. Dementsprechend bedeutet *Quo WSAS* = 13, dass das betreffende Kind im Schnitt Sätze aus dreizehn Wörtern gebildet hat.

Auch bei den Ergebnissen der Zusatzauswertung können wir die große Heterogenität der Werte in unserer Stichprobe beobachten.

## ZUSAMMENFÜHRUNG DER EINZELERGEBNISSE

### M ZUSAMMENHÄNGE ZWISCHEN DEN VERSCHIEDENEN ERGEBNISSEN ZUR SPRACHKOMPETENZ

Eine zentrale Rolle im nachfolgend beschriebenen Auswertungsschritt spielen Korrelationskoeffizienten. Von ihnen genügt zum Verständnis Folgendes zu beachten: Korrelationskoeffizienten nehmen Werte zwischen -1 und +1 an. Die Größe des Koeffizienten ist ein Gradmesser der Stärke des Zusammenhanges, das Vorzeichen beschreibt dessen Richtung (gleichsinnig proportional (je größer das eine, desto größer das andere) oder umgekehrt proportional). Die Koeffizienten dürfen nicht als Prozentwerte missinterpretiert werden. Faustregel für deren Größe in den Tabellen dieses Kapitels ist: Sind sie dem Betrag nach größer als .20, so sind sie statistisch signifikant von Null verschieden, was nicht heißt, dass auch ein inhaltlich bedeutsamer Zusammenhang vorliegt. Bei der Vielzahl an hier berechneter Korrelationskoeffizienten ist damit zu rechnen, dass eine Vielzahl durch Zufall verschieden von Null sind, ohne dass dies eine reale Grundlage hat. Wir werden in jeder Tabelle versuchen müssen, Muster herauszufinden, Konstellationen, die inhaltlich plausibel sind.

#### Korrelation der Indices aus DEUTSCH PLUS

##### M.1 Tabelle: Korrelationen der Kriteriumsvariable aus DEUTSCH PLUS.

D+	Eltern	Kinder	II-1: N	II-1: V	II-2: 1	II-2: 2	II-3:2/3	II-3: N	II-3: V
Eltern	---	.50	.54	.42	.46	.22	.43	.29	.32
Kinder		---	.28	.23	.37	.15	.27	.16	.22
II-1: N			---	.66	.62	.27	.38	.24	.28
II-1: V				---	.54	.26	.32	.15	.25
II-2: 1					---	.20	.32	.18	.20
II-2: 2						---	.37	.21	.29
II-3:2/3							---	.24	.30
II-3: N								---	.68
II-3: V									---

Erläuterung der Abkürzungen:

- Eltern** Einschätzung der Deutschkompetenz des Kindes durch die Eltern  
**Kinder** Auswertung der Fragen an das Kind ("Kriterium 10 von 14")  
**II-1: N** Passiver Wortschatz: Anzahl richtig gezeigter Nomen  
**II-1: V** Passiver Wortschatz: Anzahl richtig gezeigter Verben  
**II-2: 1** Aufgabenverständnis: Anzahl richtig verstandener einfacher Anweisungen  
**II-2: 2** Aufgabenverständnis: Anzahl richtig verstandener komplexer Anweisungen  
**II-3: 2/3** Aktive Äußerungen: Anzahl der Mehrwortäußerungen  
**II-3: N** Aktive Äußerungen: Überwiegend sachlich angemessene Nomen  
**II-3: V** Aktive Äußerungen: Überwiegend sachlich angemessene Verben  
 Die Merkmale sind dergestalt codiert, dass größere Werte eine höhere Sprachkompetenz bedeuten.

Wir wollen zunächst jeden der einzelnen Teile von DEUTSCH PLUS für sich betrachten.

Der Zusammenhang zwischen den Ergebnissen aus der Eltern- und der Kindbefragung (D+/Teil I) ist mit .50 mittelhoch, während er bei D+/II-1 mit .66 deutlich höher liegt. Diese Korrelation zwischen den Unterteilen Nomen und Verben drückte sich in dem hohen Konsistenzkoeffizienten von  $\alpha=.86$  aus; vgl. Kapitel G. In den Merkmalen *Anzahl richtig gezeigter Nomen* und *Anzahl richtig gezeigter Verben* kommt also cum grano salis dieselbe sprachliche Kompetenz zum Ausdruck. Dies gilt nicht für die beiden Unterteile von D+/II-2, die mit nur .20 untereinander korrelieren. Offensichtlich geht die Anforderung, komplexe Anweisungen richtig zu verstehen, entschieden über die Fähigkeit, einfache Anweisungen richtig zu verstehen, hinaus. Bei einer Überarbeitung von *DEUTSCH PLUS* sollten demnach zunächst beide Aspekte weiterhin Berücksichtigung finden; vgl. aber Kommentar zur Tabelle M.4. Einer späteren Untersuchung, die auf einer breiteren empirischen Grundlage stattfindet, ist es dann vorbehalten zu entscheiden, ob beide Aspekte erhalten bleiben und ob sie dann ggf. getrennt zu werten sind oder ob die neu zu berechnende interne Konsistenz ausreichend ist, um beide weiterhin zu einem in sich hinreichend homogenen (Sub-)Test zusammenzufassen.

Im Einklang mit den bisherigen Werten stehen die Korrelationskoeffizienten von D+/II-3: Ein hoher Zusammenhang zwischen den Bestandteilen *Nomen* und *Verben* (.68), ein niedriger dieser beiden mit dem Merkmal *Anzahl der Mehrwortäußerungen* (.24 bzw. .30). Das Merkmal *Anzahl der Mehrwortäußerungen* erfasst einen komplexeren Zusammenhang als die Aspekte überwiegender Verwendung sachlich angemessener Nomen und Verben. Auch hier gilt das eben Gesagte: Alle Unterteile belassen und nach dem ersten großen Einsatz überprüfen, ob eine getrennte Auswertung, also ein Aufbrechen in Subtests, angezeigt ist.

Vergleichen wir in der Tabelle M.1 die Korrelationen aus der Eltern- und der Kinderbefragung mit den Subtests aus dem Teil II von *DEUTSCH PLUS*, so fallen die durchweg höheren Korrelationen der Elternaussagen auf, die also eher als die Kinderitems das zu erfassen scheinen, was der Teil II misst.

Auf eine weitergehende Analyse der Subtests von Teil II wird verzichtet; wir wenden uns stattdessen der Tabelle M.2 zu, die die Korrelationen zwischen den Subtests und den Urteilen am Ende jeder der Teile angibt, die also die Zusammenhänge zwischen den einzelnen Bestandteilen einerseits und den Testleiter/innen- bzw. kriteriumsorientierten Urteilen andererseits darstellt.<sup>8</sup>

## **M.2 Tabelle: Korrelationen der Kriteriumsvariablen aus *DEUTSCH PLUS* mit deren Ergebnisaussagen.**

D+	1A-X	1B-X	1B-Y	21-X	21-Y	22-X	22-Y	23-X	23-Y	II-X	II-Y
<b>Eltern</b>	-.62	-.46	-.39	.52	.53	.42	.40	.48	.47	-.48	-.42
<b>Kinder</b>	-.44	-.55	-.58	.25	.27	.31	.31	.20	.28	-.31	-.29
<b>II-1: N</b>	-.49	-.20	-.20	.83	.85	.53	.53	.48	.46	-.64	-.63
<b>II-1: V</b>	-.44	-.14	-.23	.60	.60	.46	.45	.40	.38	-.47	-.49
<b>II-2: 1</b>	-.35	-.28	-.30	.50	.52	.86	.85	.35	.34	-.67	-.67
<b>II-2: 2</b>	-.16	-.11	-.13	.23	.24	.24	.25	.31	.34	-.27	-.25
<b>II-3:2/3</b>	-.34	-.33	-.28	.33	.32	.32	.32	.60	.73	-.41	-.38
<b>II-3: N</b>	-.21	-.26	-.15	.19	.24	.18	.18	.36	.60	-.26	-.32
<b>II-3: V</b>	-.27	-.25	-.07	.30	.28	.18	.18	.40	.65	-.29	-.34

<sup>8</sup> Hinweis: Die negativen Korrelationskoeffizienten kommen dadurch zustande, dass bei den Subtests größere Werte eine größere sprachliche Kompetenz bedeuten, bei den Urteilen jedoch umgekehrt herum der größere Werte den Förderbedarf anzeigt; der Zusammenhang ist demnach ein umgekehrt proportionaler.

*Erläuterung der Abkürzungen* (vgl. auch Tabelle M.1):

- 1A-X** Urteil der Testleiter/innen (TL) abgeleitet aus den Fragen an die Eltern
- 1B-X** Urteil der Testleiter/innen (TL) abgeleitet aus den Fragen an das Kind
- 1B-Y** Kriteriumsorientiertes Urteil abgeleitet aus den Fragen an das Kind  
Bei den zusammenfassenden Urteilen am Ende von Teil I bedeutet 0 "Keine Förderung erforderlich" und 1 die negative Einschätzung der Sprachkompetenz.
- 21-X** Passiver Wortschatz: TL-Urteil am Ende des Moduls
- 21-Y** Passiver Wortschatz: Kriteriumsorientiertes Urteil am Ende des Moduls
- 22-X** Aufgabenverständnis: TL-Urteil am Ende des Moduls
- 22-Y** Aufgabenverständnis: Kriteriumsorientiertes Urteil am Ende des Moduls
- 23-X** Aktive Äußerungen: TL-Urteil am Ende des Moduls
- 23-Y** Aktive Äußerungen: Kriteriumsorientiertes Urteil am Ende des Moduls  
Die Urteile am Ende jeden Moduls sind mit 0 "Kriterium verfehlt" und 1 "Kriterium erreicht" codiert.
- II-X** Gesamtbewertung der Testleiter/innen am Ende des zweiten Teils
- II-Y** Kriteriumsorientierte Gesamtbewertung am Ende des zweiten Teils  
Bei den zusammenfassenden Urteilen am Ende von Teil II bedeutet 0 "Keine Förderung erforderlich" und 1 die negative Einschätzung der Sprachkompetenz.

Der Probelauf hatte als eines der ersten auffälligen Ergebnisse erbracht, dass die Durchführenden sich in zuweilen erheblichem Umfang nicht an die vorgegebenen Auswertungskriterien hielten. Es wurde daher zwischen den Testleiter/innen(TL)- und den kriteriumsorientierten Urteilen unterschieden. Erstere sind in der Tabelle M.2 durch ein X, letztere durch ein Y gekennzeichnet.

Die beiden Typen von Urteilen korrelieren in etwa gleich hoch mit den dazugehörigen Subtests - mit einer Ausnahme: Für den Teil D+/II-3 liegen die Korrelationen der kriteriumsorientierten Urteile deutlich über jenen der Testleiter/innen, d.h. in diesem Teil mit den höchsten Anforderungen, dessen Auswertung zugleich wesentlich stärker von den Durchführenden abhängt als die Auswertung der anderen Teile, weichen die TL-Urteile weit stärker von den Ergebnissen der Subtests ab als die kriteriumsorientierten. Mit .36 und .40 bei den Aspekten *Nomen* und *Verben* liegen die Korrelationen sogar ausgesprochen niedrig, d.h. die Testleiter/innen machen ihr Urteil nur noch in geringem Umfang von den fixierten Leistungen der Kinder abhängig. Für ihre Einschätzung spielen demnach Aspekte eine Rolle, die wir dem vorliegenden Material nicht entnehmen können.

Betrachten wir die Gesamturteile zum zweiten Teil von *DEUTSCH PLUS*, so ist es wiederum das dritte Modul (D+/II-3), zu dem der geringste Zusammenhang besteht - auf TL-Seite -.41, -.26 und -.29, bei der Kriteriumsorientierung -.38, -.32 und -.34 -, Korrelationskoeffizienten, die erstaunlicherweise teilweise von jenen übertroffen werden, die für die Eltern- und Kinderaussagen aus dem ersten Teil gelten (-.48 und -.31 bzw. -.42 und -.29). Gestaltung und Rolle des Subtests ist folglich zu überprüfen. Jedenfalls sind Durchführung, Auswertung und Einbindung in das Gesamturteil zu präzisieren.

Auffällig ist ferner, dass der Subtest D+/II-2 (*Anzahl richtig verstandener komplexer Äußerungen*) in nur geringem Zusammenhang mit dem Gesamturteil des zweiten Teils von *DEUTSCH PLUS* steht. Ist er nicht bedeutsam für eine Grobeinteilung, wie sie ein Schwellentest vornimmt? Anhand der Ergebnisse aus dem Paralleltest *KISTE* wird seine Tauglichkeit zu prüfen sein. Auch hier muss aus Platzgründen auf eine vollständige Analyse der Tabelle M.2 verzichtet werden. Wir wenden uns zunächst dem Paralleltest zu, bevor wir den Zusammenhang zwischen *DEUTSCH PLUS* und *KISTE* betrachten.

### Korrelation der Indices aus *KISTE*

Die Tabelle M.3 enthält die Koeffizienten zu den Korrelationen der verschiedenen Ergebnisvariablen aus dem Paralleltest *KISTE*.

#### M.3 Tabelle: Korrelationen der Kriteriumsvariable aus *KISTE*.

<i>KISTE</i>	<b>WO</b>	<b>SB-GR</b>	<b>SB-SEM</b>	<b>SB-SUM</b>	<b>Anz SAS</b>	<b>Q-FSAS</b>	<b>Q-WSAS</b>
<b>WO</b>	---	.58	.57	.60	.64	-.58	.21
<b>SB-GR</b>		---	.84	.92	.75	-.80	.13
<b>SB-SEM</b>			---	.98	.93	-.69	.25
<b>SB-SUM</b>				---	.92	-.79	.22
<b>Anz SAS</b>					---	-.78	.24
<b>Q-FSAS</b>						---	-.06
<b>Q-WSAS</b>							---

Erläuterung der Abkürzungen:

- WO:** Aktiver Wortschatz/Anzahl der Punkte  
**SB-GR:** Satzbildung/Anzahl der Grammatikpunkte  
**SB-SEM:** Satzbildung/Anzahl der Semantikpunkte  
**SB-SUM:** Satzbildung/Gesamtzahl der SB-Punkte  
**Anz SAS:** Anzahl der semantisch akzeptablen Sätze  
**Q FSAS:** Fehlerquotient aus der Summe aller grammatikalischen Fehler bezogen auf die Anzahl der semantisch akzeptablen Sätze  
**Q WSAS:** Wörterquotient aus der Summe aller Wörter in den semantisch akzeptablen Sätzen bezogen auf die Anzahl der semantisch akzeptablen Sätze.

Vergleichen wir die Korrelationskoeffizienten von WO mit den SB-Bestandteilen, so sind diese mit .58, .57 und .60 deutlich niedriger als die SB-Koeffizienten mit .84, .92 und .98 untereinander. Die beiden Subtests WO und SB von *KISTE* erfassen demnach nachweislich unterschiedliche Aspekte sprachlicher Kompetenz.

Es zeigt sich, dass von der Zusatzauswertung das letzte Merkmal, also die Quantität des sprachlich Produzierten bezogen auf die Anzahl semantisch akzeptabler Sätze, einen nur äußerst geringen Zusammenhang zu den ansonsten erfassten Aspekten aufweist. Dies bedeutet: Der Aufwand der entsprechenden Auswertung lohnt die Mühe nicht oder es wird etwas qualitativ Unterschiedliches an sprachlicher Kompetenz erfasst. Anders sieht es hingegen mit dem Nenner *Anz SAS* und dem ebenfalls hierauf bezogenen Fehlerquotienten *Q-FSAS* aus, die beide substantiell mit den anderen Merkmalen korrelieren.

### Korrelation der Indices aus *DEUTSCH PLUS* und *KISTE*

Die Tabelle M.4 gibt Auskunft darüber, wie stark die Zusammenhänge zwischen den Subtests von *DEUTSCH PLUS* und des Paralleltests *KISTE* sind. Die Tabelle liefert somit Hinweise, welche der einzelnen Bestandteile von *DEUTSCH PLUS* besonders gut sprachliche Kompetenz im Sinne der Subtests von *KISTE* messen, also wie valide sie sind.<sup>9</sup>

<sup>9</sup> Nicht näher betrachtet wird die letzte Spalte *Q-WSAS* mit den Ergebnissen aus der Zusatzauswertung, die - wie sich bereits in der Tabelle M.3 zeigte - keinen Zusammenhang mit den übrigen Aspekten aufweisen.

#### **M.4 Tabelle: Korrelationen der Kriteriumsvariable aus DEUTSCH PLUS und aus KISTE.**

(Zu den Abkürzungen siehe Tabellen zuvor.)

<b>D+ ↓</b>	<b>WO</b>	<b>SB-GR</b>	<b>SB-SEM</b>	<b>SB-SUM</b>	<b>Anz SAS</b>	<b>Q-FSAS</b>	<b>Q-WSAS</b>
<b>Eltern</b>	.51	.63	.56	.60	.63	-.55	.20
<b>Kinder</b>	.37	.40	.38	.39	.40	-.26	.17
<b>II-1: N</b>	.56	.57	.52	.55	.42	-.43	.06
<b>II-1: V</b>	.44	.47	.42	.45	.27	-.32	.00
<b>II-2: 1</b>	.51	.48	.46	.49	.46	-.40	-.01
<b>II-2: 2</b>	.22	.30	.29	.30	.13	-.15	-.02
<b>II-3:2/3</b>	.39	.43	.43	.45	.36	-.33	.03
<b>II-3: N</b>	.28	.29	.30	.31	.38	-.39	.02
<b>II-3: V</b>	.28	.35	.32	.35	.29	-.36	.04

Generell gilt, dass die Validitätsüberprüfung sehr zufriedenstellend ausfällt. Die Korrelationen weisen nahezu durchweg eine Größenordnung auf, wie sie zu wünschen ist, nämlich eine mittelhohe. Wären die Koeffizienten zu klein, dann müsste die Validität der Module von *DEUTSCH PLUS* angezweifelt werden. Wären sie zu hoch, so wäre dies deswegen seltsam, weil beide Test unterschiedliche Ziele verfolgen (Schwellentest versus differenzierte Diagnostik).

Der ansonsten vielleicht bemerkenswerteste Befund, den die Tabelle M.4 liefert, sind die Korrelationen zwischen den Ergebnissen der Elternbefragung und den Subtests aus *KISTE*. Es sind die jeweils höchsten Koeffizienten (mit der einen Ausnahme von .56 bei der Kombination D+/II-1 und WO). Die Angaben der Eltern scheinen demnach valider zu sein als häufig behauptet wird. Allerdings wissen wir nicht, in welchem Umfang aufgrund vorgängiger Kenntnisse die Durchführenden die Fragen selber beantwortet haben anstatt die Eltern zu befragen. Aber auch bei teilweiser Substitution der Elternangaben durch Testleiter/innen-Einschätzungen bedeutet dies natürlich, dass persönliche Urteile in jedem Falle eine bedeutsame Informationsquelle darstellen, die es zu nutzen gilt.

Mittelhohe Korrelationskoeffizienten finden wir ferner bei den Modulen D+/II-1 (bei beiden Auswertungsaspekten *Nomen* und *Verben*) und D+/II-2. Bei letzterem aber nur beim Merkmal *Anzahl richtig verstandener einfacher Anweisungen*, was sich im Einklang mit der Vermutung befindet, die wir bei der Diskussion der Tabelle M.2 äußerten, dass nämlich das zweite Merkmal, *Anzahl richtig verstandener komplexer Anweisungen*, keinen sehr hohen Beitrag zur Wirksamkeit des Schwellentestes leistet.

Die Korrelationen zwischen *KISTE* und dem dritten und komplexesten Modul des zweiten Teils von *DEUTSCH PLUS*, *Aktive Äußerungen*, fallen niedriger aus als bei den eben besprochenen, am höchsten noch beim Merkmal *Anzahl der Mehrwortäußerungen*, also einem auch aus inhaltlicher Sicht bedeutsamen Aspekt.

Nachdem wir die einzelnen Bestandteile von *DEUTSCH PLUS* untersucht, also eine Feinanalyse betrieben haben, wenden wir uns nun der Tabelle M.5 zu, der die Fragestellung zugrundeliegt, wie valide die aus diesen Bestandteilen gewonnenen globaleren Einschätzungen der sprachlichen Kompetenz am Ende eines jeden Moduls und zusammenfassend am Ende des Teils II sind.<sup>10</sup> bevor wir auf das Endurteil eingehen, betrachten wir der Reihe nach

<sup>10</sup> Wiederum wird *Q-WSAS* nicht näher betrachtet (die letzte Zeile mit den Teilergebnissen aus der Zusatzauswertung).

die fünf Module: I-A (Elternbefragung), I-B (Kinderbefragung), II-1 (Passiver Wortschatz), II-2 (Aufgabenverständnis), II-3 (Aktive Äußerungen). Da für die Auswertung der Elternbefragung keine eng umrissenen Vorgaben gemacht wurden, liegt für das erste Modul als Ergebnis nur die Testleiter/innen-Bewertung vor, während für die anderen vier Module jeweils zwei Urteile zu berücksichtigen sind, nämlich diejenigen der Testleiter/innen und diejenigen, die sich strikt an die vorgegebenen Auswertungskriterien halten. Von besonderem Interesse ist, ob diese strikt kriteriumsorientierten Urteile valider (im Sinne der Korrelation mit KISTE) sind als die durch subjektive Faktoren modifizierten Einschätzungen der Testleiter/innen.

#### **M.5 Tabelle: Korrelationen der Kriteriumsvariablen aus KISTE mit den Ergebnisaussagen aus DEUTSCH PLUS.**

(Abkürzungen siehe Tabellen zuvor.)

D+ ↓	1A-X	1B-X	1B-Y	21-X	21-Y	22-X	22-Y	23-X	23-Y	II-X	II-Y
WO	-.41	-.26	-.29	.48	.52	.45	.43	.43	.43	-.50	-.46
SB-GR	-.52	-.41	-.31	.52	.51	.47	.46	.51	.52	-.55	-.53
SB-SEM	-.43	-.37	-.31	.42	.43	.44	.43	.48	.52	-.50	-.49
SB-SUM	-.48	-.40	-.32	.47	.47	.47	.46	.51	.53	-.54	-.52
AnzSAS	-.37	-.27	-.20	.31	.34	.42	.41	.45	.50	-.40	-.35
Q-FSAS	.36	.18	.11	-.39	-.42	-.35	-.36	-.49	-.55	.47	.47
Q-WSAS	-.11	-.12	-.10	.10	-.06	-.09	-.08	-.01	-.05	.09	.17

Zunächst jedoch sei festgehalten, dass sich wiederum die Bewertung, die sich auf die Elternaussagen stützt, valider zu sein scheint als die Bewertung der Kinderantworten. Durchweg finden wir in der Spalte 1A-X höhere Korrelationen als in den Spalten 1B-X und 1B-Y. Ferner ist die Auswertung des zweiten Moduls (Kinderbefragung) valider (im Sinne der Korrelation mit KISTE), wenn das vorgegebene Kriterium durch die subjektive Einschätzung der Testleiter/innen abgemildert wird (Korrelationen in 1B-X  $\geq$  1B-Y). Interessanterweise wiederholt sich diese Konstellation bei den drei Modulen des zweiten Teils von DEUTSCH PLUS nicht. Gemessen an den Ergebnissen von KISTE sind die TL- und die kriteriumsorientierten Urteile in etwa gleich gut.

Sehr bedeutsam für die Güte von DEUTSCH PLUS sind die Unterschiede in den Korrelationen der KISTE-Subtests mit den Modulen von D+/II. So korreliert WO (Wortschatz) höher mit dem Modul 21 (Passiver Wortschatz) als mit den beiden anderen Modulen von D+/II. Das Modul 23 (Aktive Äußerungen) korreliert i.d.R. höher als die beiden anderen Module mit dem Subtest SB (Satzbildung). Bemerkenswerterweise gilt dies nicht für den Teilaspekt SB-GR (Satzbildung/Grammatik), der mit dem Modul D+/21 ebenso hoch korreliert wie mit dem Modul D+/23, während der Teilaspekt SB-SEM (und die Zusatzauswertungen) wiederum niedrigere Korrelationen als D+/23 aufweist. Alles in allem jedenfalls sind die eben skizzierten Muster ein deutlicher Hinweis auf die Validität von DEUTSCH PLUS.

Diese Aussage wird zusätzlich gestützt, betrachten wir die Korrelationen des den Teil II abschließenden Gesamturteils. Die Korrelationen sind - wie erwünscht (s.o.) - mittelhoch. Ein Vergleich der Varianten X und Y scheint auf eine tendenziell leicht höhere Validität der TL- als die der kriteriumsorientierten Urteile hinzuweisen, wobei die Differenzen wesentlich geringer als beim Modul D+/IB ausfallen und inhaltlich kaum relevant sind.

Eingangs von KISTE wird die Sprechfreudigkeit des Kindes und seine Verständnisfähigkeit eingeschätzt. Korrelieren wir (in den einhundert bis einhundertundfünfzig Fällen, für die entsprechende Daten vorliegen) die Einschätzungen mit den Urteilen am Ende der D+ - Module, so erhalten wir zwar i.d.R. geringere Koeffizienten als bei der Korrelation mit den KISTE-Subtests (vgl. Kapitel K), aber nahezu immer wird die Größenordnung von .20 bis .40 erreicht. Somit erfassen diese Einschätzungen durchaus einen wichtigen Aspekt der Testsituation und sollten bei einer Überarbeitung von DEUTSCH PLUS berücksichtigt werden.

**N FAZIT****Zusammenfassung<sup>11</sup>****(a) Ziel des Probelaufs**

Überprüft werden sollten die Praxistauglichkeit  
die Validität (Gültigkeit).

Hierzu wurde als Maßstab der seit langem eingeführte Sprachtest *KISTE* ausgewählt, dessen beiden Subtests *Wortschatz (WO)* und *Satzbildung (SB)* parallel zu *DEUTSCH PLUS* einzusetzen waren.

**(b) Beschreibung der Stichprobe**

Am Probelauf nahmen N=384 Kinder aus Vorklassen (55%) und Kindertagesstätten (45%) teil. Die Stichprobe ist nicht repräsentativ, sondern Kinder aus sozial belasteten Gegenden sind überrepräsentiert, was im Einklang mit den Zielen des Probelaufs steht. Es gibt mehr West- als Ostschulen, bei den Kitas ist es umgekehrt.

80% der Stichprobenkinder sind zwischen fünfeinhalb und sechseinhalb Jahren alt. Jungens bilden mit 53% die Mehrheit. Rund zwei Drittel der Kinder besitzen die deutsche Staatsangehörigkeit und ein Fünftel die türkische; aber nahezu alle Kinder sind in Deutschland geboren. Bei etwa der Hälfte der Kinder aus der Stichprobe ist Deutsch die Erstsprache und die Verkehrssprache zu Hause. (Die Differenz zu den zwei Dritteln mit der deutschen Staatsangehörigkeit dürfte durch die Aussiedlerkinder zustandekommen.)

**(c) Struktur des Erhebungsinstrumentariums und Aufbau des Datensatzes**

Insgesamt bestehen die beim Probelauf eingesetzten Erhebungsinstrumente aus neun Modulen, nämlich drei Mal drei Unterteilen.

D+/I-1: Informationen zum Kind  
D+/I-2: Fragen an die Eltern  
D+/I-3: Fragen an das Kind

D+/II-1: Passiver Wortschatz  
D+/II-2: Aufgabenverständnis  
D+/II-3: Aktive Äußerungen

K/WO: Aktiver Wortschatz  
K/SB: Satzbildung  
K/ZU: Satzbildung/Zusatzauswertung grammatikalischer Fehler.

Keine weitere Berücksichtigung aufgrund zu vieler fehlender Werte fanden zwanzig Fälle. Die weitere Auswertung stützt sich somit auf die Daten von N=364 Kindern.

<sup>11</sup> Die Zusammenfassung folgt in ihrer Durchbuchstabierung den Kapiteln A bis M. In den korrespondierenden Kapiteln finden sich die Einzelheiten zu den hier zwangsläufig vergrößernden Aussagen.

Am Ende jedes Moduls sollten die Testleiter/innen (TL) die zuvor eingesetzten Testitems auswerten und zu einem modulspezifischen Ergebnis zusammenfassen. Es zeigte sich, dass in einer ganzen Reihe von Fällen bei der Formulierung des Ergebnisses von den vorgegebenen Auswertungskriterien abgewichen wurde. Daher wurde für die Auswertung des Probelaufes jeweils zwischen dem TL- und dem kriteriumsorientierten Urteil unterschieden.

**(d) D+/I-2: Fragen an die Eltern**

Anhand einer Reihe von Fragen an die Eltern sollten diese die Sprachkompetenz ihres Kindes einschätzen. Die höchste Einstufung erfolgte für die Deutschkompetenz der von ihrer Herkunft her deutschsprachigen Kinder. In erheblichem Abstand folgte dann die Sprachkompetenz der jeweiligen Erstsprache der Kinder nichtdeutscher Herkunftssprache und an letzter Stelle schließlich deren Deutschkompetenz:

D/D >> ndH/ndH > D/ndH.

(Deutschkompetenz beurteilt durch ihre deutschsprachigen Eltern

- >> Erstsprachkompetenz beurteilt durch ihre nichtdeutschsprachige Eltern
- > Deutschkompetenz beurteilt durch ihre nichtdeutschsprachigen Eltern).

**(e) D+/I-3: Fragen an das Kind**

Der weitaus größte Teil der Kinder (80%) meisterte die hier vorgegebene Hürde, zehn von vierzehn Fragen angemessen zu beantworten. Es kam (bei rund 5% der Fälle) zu Fehlern in der Summenbildung.

**(f) D+/I-4: Auswertung von DEUTSCH PLUS/Teil I**

Die Testleiter/innen sollten zwei Urteile abgeben. Das eine, das sich auf die Auswertung der Elternangaben stützte, das andere, dem die Antworten der Kinder zugrundelagen. Zum ersten war kein präzises Auswertungskriterium vorgegeben. Das mag mit ein Grund dafür sein, dass es beim ersten Urteil zu zahlreichen Ausfällen kam (rund 20% der Fälle) und ferner (in über 10% der Fälle) zu Widersprüchen zwischen den beiden Einschätzungen.

Größenordnungsmäßig allerdings deckten sich beide Urteile insoweit, als dass beide Male in rund 80% der Fälle keine Sprachförderung für notwendig gehalten wurde. Aus den *Elternangaben* folgerten die Auswertenden aber ferner, dass in 20% auf jeden Fall eine Förderung erforderlich sei, während für die restlichen 20% aus dem Erhebungsteil *Kinderantworten* eine Fortsetzung der Testung empfohlen wurde.

Wie groß die Gruppe der zu Fördernden im kommenden Herbst sein wird, ist aus derartigen Werten schwer zu schätzen, da die Stichprobe für Berlin nicht repräsentativ ist. Da die förderungsbedürftigen Kinder in der Stichprobe überproportional vertreten sind, liefert die Stichprobe höchstwahrscheinlich Obergrenzen.

**(g) D+/II-1: Passiver Wortschatz**

Auf zwei sog. Wimmelbildern waren Dinge oder Vorgänge zu identifizieren, wobei die Anzahl richtig gezeigter Nomen (mindestens 20) und Verben (mindestens 8) als Schwelle definiert wurde. Maximal möglich richtig zu nennen waren 24 Nomen und 10 Verben. 43% der Kinder blieben unterhalb der festgelegten Schwelle.

Die Reliabilität des Moduls ist sehr hoch. Ohne bedeutsame Einbußen für die Gütetestkriterien kann auf das erste Wimmelbild verzichtet werden.

#### **(h) D+/II-2: Aufgabenverständnis**

Anhand von Spielsachen waren einfache und komplexe Anweisungen umzusetzen. Als Schwelle für eine erfolgreiche Bearbeitung dieses Moduls galt die Umsetzung von mindestens fünf (von maximal sechs) einfachen und drei (von maximal vier) komplexen Anweisungen. 57% der Kinder blieben unterhalb der Schwelle.

Die Reliabilität des Moduls ist mittelhoch.

#### **(i) D+/II-3: Aktive Äußerungen**

Das Material dieses Moduls besteht aus den beiden Bildserien "Was ist hier los?" und "Glas auf dem Schrank". Während die zweite Serie eine kleine Geschichte wiedergibt (die Bilder sind dem Kind in der richtigen Reihenfolge vorzulegen), besteht die erste aus Einzelkarten. Das Kind kann eine der beiden Serien auswählen. Entweder äußert sich das Kind dann spontan zu den Bildern oder es wird ein Impuls gegeben, der nicht zuviel Wortschatz anbietet. Protokolliert werden die Äußerungen des Kindes und hinsichtlich dreier Aspekte ausgewertet.<sup>12</sup> Die Schwelle gilt dann als erreicht, wenn alle drei Kriterien erfüllt sind. 40% der Kinder blieben bei diesem Modul, das sicherlich das schwierigste der dreien ist, unterhalb der Schwelle.

Einerseits ist zwar die Reliabilität des Moduls zufriedenstellend, andererseits sollte die Auswertungsobjektivität verbessert werden, denn die Auswertung ist in weit höherem Maße als bei den anderen Bestandteilen von den Sichtweisen der Testleiter/innen abhängig. Das führte z.B. dazu, dass hier weitaus häufiger als zuvor TL- und kriteriumsorientierte Urteile (s.o. Punkt (c)) auseinanderklaffen (in 13% der Fälle).

#### **(j) D+/II-4: Gesamtauswertung von Teil II**

Von den Kindern, die den Teil II von *DEUTSCH PLUS* durchliefen, sollten aufgrund der Testergebnisse rund 70% einer Sprachförderung zugeführt werden. Bezogen auf die Ausgangsmenge aller Kinder ergibt sich ein schätzungsweise Förderbedarf bei rund 15% der Kinder, ein Wert, der angesichts der nichtrepräsentativen Zusammensetzung der Stichprobe eine Obergrenze darstellen dürfte, gleichwohl realistisch zu sein scheint.

Eine besondere Schwierigkeit ergibt sich aus dem schwerwiegenden Umstand, dass bei etwa der Hälfte der Kinder die Ergebnisse aus Teil I und Teil II sich widersprechen! Abgemildert wird diese Aussage dadurch, dass ein derartiger Vergleich nur bei den N=235 (anstelle von 364) Kindern möglich war, zu denen wir die Ergebnisse beider Teile kennen.

---

<sup>12</sup> - Struktur. Kriterium: Mindestens zwei Drittel der kindlichen Äußerungen müssen länger als drei Wörter sein.  
- Wortschatz/Nomen. Kriterium: Verwendung von überwiegend sachlich angemessenen Nomen.  
- Wortschatz/Verben. Kriterium: Verwendung von überwiegend sachlich angemessenen Verben.

**(k) KISTE/Prolog: Rahmenbedingungen**

Im Paralleltest *KISTE* werden die Durchführenden gebeten, die Sprechfreudigkeit des Kindes ebenso einzuschätzen wie ihren Eindruck des kindlichen Verständnisses vom Geforderten.

Leider liegen nur von einhundert Kindern die Daten zu diesen Rahmenbedingungen des Testens vor. Bei einer dreistufigen Einschätzung (gering, mittel, hoch) gilt ungefähr, dass die Hälfte der Kinder in der Mitte verortet wurden und der Rest gleichermaßen in der oberen wie in der unteren Kategorie. Dies gilt allerdings nur für das Merkmal *Sprechfreudigkeit*. Hinsichtlich des Verständnisses wurde die Hälfte der Kinder in der unteren Kategorie eingestuft und ein Drittel in der mittleren, Ausdruck der speziellen Stichprobe, die dem Probelauf zugrundelag.

**(l) KISTE: Ergebnisse zur Sprachkompetenz**

Es gibt drei zentrale Ergebnisse:

- Die Sprachkompetenz der Stichprobenkinder ist unterdurchschnittlich.<sup>13</sup>
- Die Heterogenität der Leistungen ist sehr hoch.
- Die Leistungen im Subtest Satzbildung sind besser als im Subtest Wortschatz.

**(m) Zusammenhänge zwischen den verschiedenen Ergebnissen zur Sprachkompetenz****# Korrelationen innerhalb der Module von DEUTSCH PLUS:**

Die einzelnen Module von *DEUTSCH PLUS* bestehen aus mehreren Kriteriumsvariablen, anhand derer das jeweilige Modulergebnis bestimmt wird (Beispiel Modul D+/II-1: Anzahl richtig aufgefasster Nomen und Verben; vgl. Kapitel C). Die Korrelationen zwischen den Kriteriumsvariablen zeigen, dass die Module in sich hinreichend homogen sind, also tatsächlich als Subtests gewertet werden können, ein Ergebnis, das durch die Reliabilitätsberechnungen gestützt wird. Eine Einschränkung gibt es beim Modul D+/II-2; siehe nachstehend und weiter unten.

**# Korrelationen der Kriteriumsvariablen mit den Modulergebnissen:**

- Die Korrelationen zu den beiden Varianten der Modulergebnissen (Testleiter/innen- und kriteriumsorientierte Urteile) sind in etwa gleich hoch. Ausnahme Modul D+/II-3, bei dem die Korrelationen der kriteriumsorientierten Urteile deutlich über jenen der Testleiter/innen liegen, d.h. in diesem Teil mit den höchsten Anforderungen, dessen Auswertung zugleich wesentlich stärker von den Durchführenden abhängt als die Auswertung der anderen Teile, weichen die TL-Urteile weit stärker von den Ergebnissen der Subtests ab als die kriteriumsorientierten. Bei den Aspekten *Nomen* und *Verben* liegen die Korrelationen sogar ausgesprochen niedrig, d.h. die Testleiter/innen machen ihr Urteil nur noch in geringem Umfang von den fixierten Leistungen der Kinder abhängig. Für ihre Einschätzung spielen demnach Aspekte eine Rolle, die wir dem vorliegenden Material nicht entnehmen können.

---

<sup>13</sup> Eine derartige Aussage ist möglich, da *KISTE* ein an einer Eichstichprobe normierter Test ist; vgl. auch Kapitel C.

Das dritte Modul D+/II-3 besitzt auch den geringsten Zusammenhang mit dem Gesamturteil des Teils II.

- Auffällig ist ferner, dass beim Modul D+/II-2 das Merkmal *Anzahl richtig verstandener komplexer Äußerungen* in nur geringem Zusammenhang mit dem Gesamturteil des zweiten Teils von *DEUTSCH PLUS* steht. Ist er nicht bedeutsam für eine Grobeinteilung, wie sie ein Schwellentest vornimmt?

#### # **Validität (Korrelationen zwischen den Merkmalen aus *DEUTSCH PLUS* und aus *KISTE*):**

Generell gilt, dass die Validitätsüberprüfung sehr zufriedenstellend ausfällt. Die Korrelationen weisen nahezu durchweg eine Größenordnung auf, wie sie zu wünschen ist, nämlich eine mittelhohe. Wären die Koeffizienten zu klein, dann müsste die Validität der Module von *DEUTSCH PLUS* angezweifelt werden. Wären sie zu hoch, so wäre dies deswegen seltsam, weil beide Test unterschiedliche Ziele verfolgen (Schwellentest versus differenzierte Diagnostik).

Bemerkenswerte Einzelheiten:

- Die Korrelationen zwischen den Ergebnissen der Elternbefragung und den Subtests aus *KISTE* sind die jeweils höchsten. Die Angaben der Eltern scheinen demnach valider zu sein als häufig behauptet wird. Allerdings wissen wir nicht, in welchem Umfang aufgrund vorgängiger Kenntnisse die Durchführenden die Fragen selber beantwortet haben anstatt die Eltern zu befragen. Aber auch bei teilweiser Substitution der Elternangaben durch Testleiter/innen-Einschätzungen bedeutet dies natürlich, dass persönliche Urteile in jedem Falle eine bedeutsame Informationsquelle darstellen, die es zu nutzen gilt.
- Die Korrelationen des Merkmals *Anzahl richtig verstandener komplexer Anweisungen* aus dem Modul D+/II-2 sind recht niedrig, was sich im Einklang mit der bereits geäußerten Vermutung befindet, dass dieser Bestandteil keinen sehr hohen Beitrag zur Wirksamkeit des Schwellentestes leistet.
- Die Korrelationen zwischen *KISTE* und dem dritten und komplexesten Modul des zweiten Teils von *DEUTSCH PLUS*, *Aktive Äußerungen*, fallen ebenfalls ziemlich niedrig aus, am höchsten noch beim Merkmal *Anzahl der Mehrwortäußerungen*, also einem auch aus inhaltlicher Sicht bedeutsamen Aspekt.
- Gemessen an den Ergebnissen von *KISTE* sind die Testleiter/innen- und die kriteriumsorientierten Urteile in etwa gleich gut.
- Eingangs von *KISTE* wird die Sprechfreudigkeit des Kindes und seine Verständnissfähigkeit eingeschätzt. Korrelieren wir (in den einhundert bis einhundertundfünfzig Fällen, für die entsprechende Daten vorliegen) die Einschätzungen mit den Urteilen am Ende der D+ - Module, so erhalten wir zwar i.d.R. geringere Koeffizienten als bei der Korrelation mit den *KISTE*-Subtests (vgl. Kapitel K), aber nahezu immer wird die Größenordnung von .20 bis .40 erreicht. Somit erfassen diese Einschätzungen durchaus einen wichtigen Aspekt der Testsituation und sollten bei einer Überarbeitung von *DEUTSCH PLUS* berücksichtigt werden.

## Schlussfolgerungen

### 1. DEUTSCH PLUS ALS GRUNDSÄTZLICH GEEIGNETES SPRACHSTANDSINSTRUMENT

(a) Ziel des Probelaufes war, die Praktikabilität und Validität von *DEUTSCH PLUS* zu überprüfen. Hinsichtlich beider Punkte zeigte sich die Tauglichkeit der Sprachstandsfeststellung.

Außer der Validität als dem obersten der Testgütekriterien sind auch die Reliabilität und die Objektivität (von Durchführung und Auswertung) wichtig, um die Qualität eines Messinstrumentes beurteilen zu können. Bei der Schätzung der Reliabilität (mittels des Konsistenzkoeffizienten  $\alpha$ ) ergaben sich zufriedenstellende bis sehr gute Werte. Zwar ist bereits jetzt die Objektivität in ausreichendem Maße gegeben, sollte aber an einigen Stellen verbessert werden. Hierzu mehr weiter unten.

Zu *DEUTSCH PLUS* braucht demnach keine Alternative gesucht zu werden, aber der Probelauf ergab eine Reihe von Hinweisen zur Optimierung.

(b) Die nachstehenden Vorschläge für eine Überarbeitung orientieren sich an drei Aspekten:

- Praktikabilität: D+ ist möglichst kurz zu halten (da Individual-, und kein Gruppentest), um die Belastung für alle Beteiligten zu minimieren und zugleich die Akzeptanz zu maximieren.
- Funktion: D+ ist ein Schwellentest, kein Diagnostikum um Förderhinweise zu bekommen.
- Empirie: Was ist aus statistischer Sicht entbehrlich, ohne die Qualität von D+ zu senken?

### 2. EINZELTEILE ÜBERGREIFENDE VERBESSERUNGSVORSCHLÄGE

(a) **Eindeutigkeit der Durchführung:** Die Hinweise zur Anwendung des Erhebungsinstrumentes sind bei jedem der Module zu präzisieren. Eine besondere Herausforderung stellt das Modul *Aktive Äußerungen* (D+/II-3) dar, da dort der Spielraum der Testleiter/innen besonders groß ist (sowohl bei der Durchführung wie bei der Auswertung).

(b) **Eindeutigkeit der Auswertung:** Die Auswertungskriterien sind eindeutig(er) zu formulieren, um Unsicherheiten der Durchführenden zu minimieren und die Auswertungsobjektivität zu erhöhen.

In einer von Modul zu Modul unterschiedlichen, aber teilweise erheblichen Anzahl von Fällen kam es zu Differenzen zwischen dem Urteil der Testleiter/innen und dem Urteil, das sich bei strikter Anwendung der vorgegebenen Auswertungskriterien ergibt. Gemessen an *KISTE* kommt beiden Urteilstypen in etwa gleiche Validität zu, tendenziell den TL-Urteilen jedoch eine leicht höhere. Zugleich aber fallen i.d.R. die kriteriumsorientierten Urteile konservativer

in dem Sinne aus, dass sie eher für eine Förderung plädieren. Da im Zweifelsfalle die Förderung Priorität besitzt, sollten die Kriterien nicht aufgeweicht, aber eben präzisiert werden.

Falls die *Fragen an die Eltern* (D+/I-2) erhalten bleiben (vgl. hierzu Punkt 4/b weiter unten), muss an dieser Stelle das bislang fehlende Auswertungskriterium explizit formuliert werden. Eine besondere Bedeutung kommt dem bisherigen Item 8 zu, das verschiedene Aspekte der Sprachkompetenz in Deutsch thematisiert. Wir kommen darauf zurück.

**(c) Äußere Gestaltung:** Das LayOut ist an allen Stellen daraufhin zu überprüfen, ob es durch Übersichtlichkeit zur Eindeutigkeit und einfachen Handhabung beiträgt. Es ist nach Möglichkeiten zu suchen, die in nicht vernachlässigbarem Umfang vorkommenden Rechen- und Übertragungsfehler weitestgehend zu vermeiden.

Geschicktes LayOut sollte zudem zu einer Verringerung des Umfangs beitragen.

**(d) Testdauer verkürzen:** Im Teil II von *DEUTSCH PLUS* sollte die Sprachstandsmessung abgebrochen werden, sobald das Kriterium eines Moduls nicht erreicht wird, denn Erfolg wird in D+/II definiert als das Überschreiten aller drei Teilschwellen.

### 3. VORSCHLÄGE ZU EINZELNEN ELEMENTEN VON *DEUTSCH PLUS*

**(a) Kürzen auf der ersten Seite des Teils I:** Nur die unbedingt erforderlichen Daten sind zu erfassen. Daher können die doppelte Abfrage des Alters und das Geschwisteritem gestrichen werden.

**(b) Kürzen im Modul D+/I-2 (Fragen an die Eltern):** *DEUTSCH PLUS* ist zu ndH-lastig. So Manches bislang Erfragte spielt für die Auswertung keine Rolle und steht in keinem nachweisbaren (statistischen) Zusammenhang mit anderen Teilen oder Ergebnissen des Erhebungsinstrumentes.

Aus empirisch-statistischer Sicht können die Items 4 (In welcher Sprache spricht das Kind mit ...), 6 (In welchem Land sind die Eltern geboren?), 7 (Sprechen die Eltern zwei Sprachen?), 9 (Wie gut beherrscht das Kind die nicht deutsche Erstsprache?) ersatzlos entfallen.

Das Item 1 (In welchem Land ist das Kind geboren?) kann dem Eingangsblatt zugeordnet werden. Item 3 (Zweisprachigkeit des Kindes) kann ebenfalls entfallen. Zumindest die Ausdifferenzierung nach Verstehen und Sprechen besitzt keine empirisch nachweisbare Relevanz.

**(c) Modul D+/I-3 (Fragen an das Kind):** Zwar ist es für einen Schwellentest durchaus zulässig, ja inhärent, eine schiefe Verteilung aufzuweisen. Allerdings sollte dieses Modul - wenn es denn erhalten bleibt; vgl. weiter unten Punkt 4/b - mit zwei Zielen überarbeitet werden: Die Fragen sollten stärkere Anreize zum Sprechen über Ja/Nein-Antworten hinaus enthalten und sie sollten stärker zwischen den Kindern zu differenzieren imstande sein.

Darüber hinaus muss durch geeignetes LayOut einer falschen Aufaddierung der Punkte entgegen gewirkt werden.

**(d) Urteile am Ende von D+/Teil I neu strukturieren:** Falls der erste Teil erhalten bleibt, muss aus dem zweigeteilten Urteil am Ende vom Teil I (Eltern - Kinder) unbedingt ein einheitliches werden. Bei der bisherigen Lösung ist weder die Auswertungsobjektivität gegeben (mangelnde Bewertungskriterien) noch kommen die Testleiter/innen zu widerspruchsfreien Ergebnissen, wie der Probelauf zeigte.

**(e) Kürzen im Modul D+/II-1 (Passiver Wortschatz):** Ohne Einbußen für die Aussagekraft des Erhebungsinstrumentes kann in diesem Modul auf das erste Wimmelbild verzichtet werden.

**(f) Mögliches Kürzen im Modul D+/II-2 (Aufgabenverständnis):** Die Daten des Probelaufes weisen darauf hin, dass möglicherweise auf den Teilaspekt *Anzahl richtig verstandener komplexer Anweisungen* verzichtet werden kann. Während die Kriteriumsvariable *Anzahl richtig verstandener einfacher Anweisungen* in substanziellem Ausmaß sowohl mit dem Ergebnis aus dem Modul als auch mit dem Gesamturteil von Teil II und mit den Subtests von *KISTE* korreliert, liegen die entsprechenden Korrelationen des inhaltlich anspruchsvolleren Merkmals deutlich niedriger. Wahrscheinlich ist es für einen Schwellentest nicht so relevant wie die einfachere Variable. Die entsprechenden Items sollten unter inhaltlichen Gesichtspunkten überarbeitet und ihre Anzahl erhöht werden. Ihre Tauglichkeit ist dann in einem neuen Probelauf zu überprüfen; vgl. Punkt 5/a.

**(g) Mögliches Kürzen und erforderliches Präzisieren im Modul D+/II-3 (Aktive Äußerungen):**

Problematisch ist die mangelnde Vergleichbarkeit der beiden Bildserien, die die Testgrundlage bilden. Die Kinder können aussuchen. Die inhaltlichen Anforderungen und der Aufforderungscharakter der Bildserien sind sehr unterschiedlich. Leider war der Anteil an Kindern, der beide Bildserien bearbeitet hat, so niedrig, dass sich keine zuverlässigen Angaben zu diesem Punkt machen lassen. Die Wahlfreiheit ist jedenfalls sehr problematisch, da u.U. unterschiedliche Aspekte der Sprachkompetenz erfasst werden. Die Problematik vergrößert sich ggf. noch dadurch, dass die Bildserien unterschiedliche Kindertypen ansprechen, so dass sich die jeweils erzielten Ergebnisse überhaupt nicht mehr vergleichen lassen. Berücksichtigen wir zudem noch, dass die Testung möglichst kurz sein sollte, so wird dafür plädiert, auf die zweite Bildserie zu verzichten.

Das dritte Modul ist jenes, das sowohl an die Kinder wie an die Auswertenden die größten Anforderungen aller drei Module stellt. In keinem Modul wichen die Testleiter/innen- und die kriteriumsorientierten Urteile (siehe oben in der Zusammenfassung Punkt (c)) so stark ab wie hier. Es ist durchaus möglich, dass durch die Testsituation das Kind ein Mehr an Falschem und Richtigem produziert, als durch die auswertenden Abfragen von *DEUTSCH PLUS* erfasst wird. Und dieses Mehr bringt die Testleiter/innen zu ihrer abweichenden Einschätzung. Zugleich ist - gemessen an *KISTE* - die Validität beider Urteilstypen in etwa gleich hoch. Gleichwohl muss versucht werden, die Abhängigkeit der Auswertung von den Durchführenden zu verringern und durch präzisierte Auswertungshinweise diese zu entlasten.

Zugleich sollten die Aspekte, die zwar für eine gezielte Förderung von Bedeutung sein mögen (vgl. die Kategorien *Umschreibungen* etc.), aber für einen Schwellentest irrelevant sind, entfallen.

**(h) Rahmenbedingungen abfragen:** Am Ende des überarbeiteten Erhebungsinstrumentes sollten - in Anlehnung an KISTE - Aspekte der Testatmosphäre erfasst werden, die möglicherweise in Zusammenhang mit den gezeigten Leistungen stehen wie etwa Sprechfreudigkeit und generell die Fähigkeit des Kindes, die Anweisungen des Testes zu verstehen.

#### 4. MÖGLICHE GRUNDLEGENDE VERÄNDERUNGEN

**(a) Zentrale Informationen in einem Blatt zusammenstellen:**

Hinweise zur gesetzlichen Grundlage, grundsätzliche Bemerkungen zur Durchführung der Sprachstandsmessung, die nicht spezifisch für die einzelnen Module sind, sollten aus der bisherigen Textfassung herausgezogen und in einem kurzen **Informationsblatt** gesammelt werden, das dann in einer geringeren Auflage als das Erhebungsinstrument, das es zudem entlastet, gedruckt werden kann.

Ein derartiges Informationsblatt muss unübersehbar die Regel enthalten, dass im Zweifelsfall die Entscheidung für eine Förderung zu fallen habe.

**(b) D+/Teil 1 in modifizierter Form aus DEUTSCH PLUS ausgliedern:**

Der bisherige Teil I könnte aufgeteilt werden: Ein kurzes **Basisblatt** enthält die zentralen Informationen (das modifizierte Deckblatt (vgl. Punkt 3/a) sowie die Items aus dem Modul D+/I-2 (Fragen an die Eltern) wie unter Punkt 3/b beschrieben. Dabei wird es den Ausfüllenden überlassen, ob sie die Informationen aufgrund ihrer bisherigen Erfahrungen mit dem Kind zusammenstellen oder aus einem aktuellen Gespräch mit den Eltern schöpfen. Dies ist bereits im Probelauf praktiziert worden, wie sich aus einem Auswertungstreffen mit Beteiligten im Ende Juni 2004 ergab. (Ein entsprechend neu gefasstes "Basisblatt" muss die Informationsquelle (Erfahrung mit dem Kind und/oder Gespräch mit den Eltern) benennen.)

Auch das bisherige Item 8 (Wie gut beherrscht das Kind die deutsche Sprache?) sollte in einem derartigen Basisblatt enthalten sein. Seine Funktion wäre dann, bei den Ausfüllenden, die die Sprachkompetenz der Kinder einschätzen sollen, ein Minimum an identischen Gesichtspunkten und somit einigermaßen vergleichbare Beurteilungsmaßstäbe zu sichern. Wie insbesondere die Validitätsüberprüfung zeigte (vgl. Kapitel M und dort die Tabelle M.4), liegt mit den im Item 8 thematisierten Aspekten eine wichtige, weil aussagekräftige Informationsquelle vor, von der wir leider im Probelauf nicht feststellen konnten, wie weit die Elternangaben oder die Erfahrung der Testleiter/innen mit dem Kind die empirische Grundlage für die Einschätzungen bildeten.

Das bisherige Modul D+/I-3 (Fragen an das Kind) könnte in einer derartigen Überarbeitung die Rolle zufallen, als Einstieg in den Teil II zu dienen. (Auch bei einer dergestalt geänderten Funktion wären die Fragen im o.a. Sinne - vgl. Punkt 3/c - zu überarbeiten.)