

## 4. Ergebnisse

Statistische Leistungsvergleiche von *Zweit-* und *Drittklässlern* sowie *Zielgruppe* und *Patienten* wurden mithilfe nichtparametrischer statistischer Tests durchgeführt, deren - gegenüber den verteilungsgebundenen Verfahren - geringere Effizienz durch die Unabhängigkeit von der sonst erforderlichen Annahme der Normalverteilung der Merkmale aufgewogen wird.

### 4.1 Base-line, Datenlage

*Zielgruppe*: 82 Kinder mit normalem peripheren Hörvermögen, IQ > 90 jedoch ohne SES bzw. LRS.

*Patienten*: 37 *Patienten*/Kinder mit normalem peripheren Hörvermögen und IQ > 90.

Anamnestische Daten von Eltern (Fragebogen bzw. Gespräch mit untersuchendem Arzt).

24,6 % der *Zielgruppe* und 40,6 % der *Patienten* haben bzw. hatten häufig Mittelohrentzündungen, z.T. mit Adenotomie oder Parazentese, kein signifikanter Unterschied (Tab. 4.1.a. im Anhang)

84,2 % der *Zielgruppe* und 48,6 % der *Patienten* sind ohne Auffälligkeiten in der Sprachentwicklung (Tab. 4.1.1). Dieser Unterschied ist hochsignifikant ( $p = 0,01$ ).

Sprachentwicklung	<i>Zielgruppe</i>	Anteil in %	<i>Patienten</i>	Anteil in %
Keine Auffälligkeiten	69	84,2	18	48,6
SES	2	2,4	2	5,4
Sprachtherapie	9	11,0	14	37,8
Ergotherapie	1	1,2	3	8,1
Keine Angaben	1	1,2	0	0,0

Tab. 4.1.1 Sprachentwicklung ( $p = 0,01$ )

Zweisprachigkeit war kein Ausschlusskriterium. Der Unterschied im Anteil der zweisprachigen Kinder (*Zielgruppe* 6,1 % ; *Patienten* 8,1 %) ist nicht signifikant.

Im Artikulationsbefund sind beide Gruppen statistisch gleich (Verteilung Tab. 4.1.b im Anhang).

Anteil der Jungen: *Zielgruppe* 46,3 %, *Patienten* 67,6 %, signifikanter Unterschied, aber irrelevant für die weiteren Untersuchungen.

Der Parameter „kognitive Entwicklung“ wurde durch das Einschlusskriterium IQ > 90 für beide Gruppen kontrolliert.

Für die mit IQ > 90 gestutzte Verteilung liegt das Gesamtmittel beider Gruppen mit 108,5 über, die Standardabweichung mit 8,9 erwartungsgemäß unter den Referenzwerten des CPM:  $100 \pm 15$ .

Der Motorfree Visual Perception Test (MVPT-R) zeigt mit 108,5 (Std. 11,5) und 102,7 (Std. 13,0) einen signifikanten Unterschied ( $p = 0,05$ ) zwischen *Zielgruppe* und *Patienten*.

Der signifikante Unterschied im MVPT-R, lässt sich möglicherweise auf die etwas andere Aufgabenzusammensetzung zurückführen und dass auch die Leistungen der älteren Kinder altersnormiert bewertet werden.

Der Frostig Test der visuellen Wahrnehmung zeigt in keinem Untertest signifikante Unterschiede zwischen *Zielgruppe* und *Patienten*.

Die folgende Tabelle (Tab. 4.1.2) zeigt die Ergebnisse des Diagnostischen Rechtschreibtests (DRT) für *Zielgruppe* und *Patienten*. Die Auswertung dieses Tests umfasst neben dem quantitativ ausgewerteten Gesamtwert (Fehleranzahl) auch eine qualitative Bewertung nach Regel- und Wahrnehmungsfehlern. Diese unterschiedlichen Bewertungsmodi spiegeln sich in den Mittelwerten wider.

Diagnostischer Rechtschreibtest (DRT)				
	Mittelwert		Std.	
Prozentrang	<i>Zielgruppe</i>	<i>Patienten</i>	<i>Zielgruppe</i>	<i>Patienten</i>
Gesamtwert	54,5	7,8	26,1	8,4
Regelfehler	51,5	16,8	26,1	18,5
Wahrnehmungsfehler	60,0	9,6	27,0	9,9

Tab. 4.1.2 Leistungen im Diagnostischen Rechtschreibtest ( $p = 0,00$ )

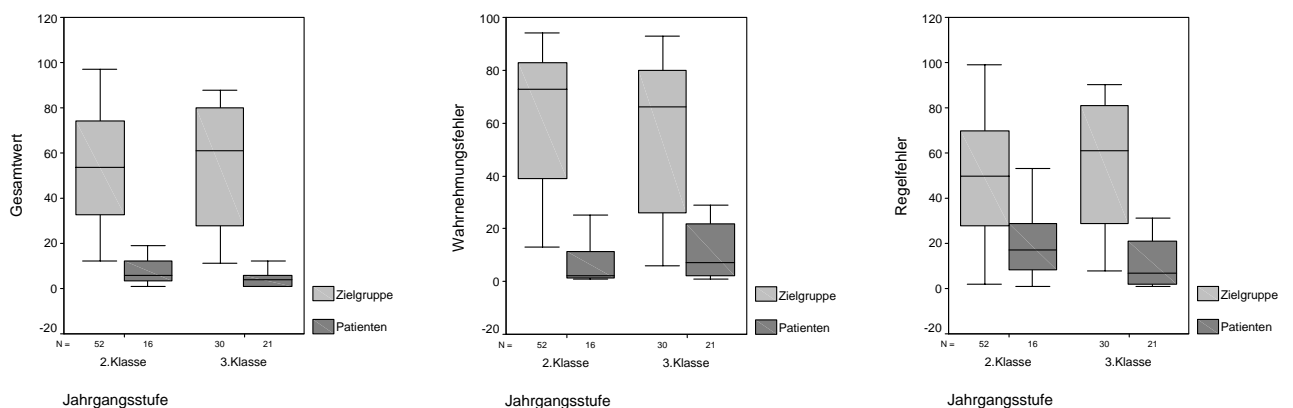


Abb. 4.1.1 Ergebnisse des DRT nach Klassenstufen

Auf der Ordinate sind jeweils die Prozentrangwerte für die quantitativ ausgezählte Fehleranzahl (Gesamtwert) oder die qualitativen Leistungen bezüglich Regelfehler (z.B. Groß-/Kleinschreibung) und Wahrnehmungsfehler (z.B. Lautfolgefehler), auf der Abzisse ist die Anzahl der Kinder, nach Klassenstufe unterteilt, dargestellt.

Die extrem hohe Standardabweichung bei den Ergebnissen der *Zielgruppe* reflektiert das starke Leistungsgefälle zwischen den Schülern der verschiedenen Schulklassen und Schulen. Der Unterschied zwischen *Zielgruppe* und *Patienten* ist hochsignifikant ( $p = 0,00$ ).

Auf den Zusammenhang zwischen auditiven und visuellen Leistungen und der Rechtschreibung der Kinder wird in Kapitel 4.5 eingegangen.

## **4.2 Audiometrische Untersuchungen von *Zielgruppe* und *Patienten***

Für die mit audiometrischen Verfahren untersuchten Leistungen liegen keine Normwerte vor. Aus diesem Grund wurde zuerst geprüft, ob bezüglich der Leistungen zwischen den Kindern der zweiten und dritten Klassenstufe Unterschiede bestehen oder ob es gerechtfertigt ist, die Gesamtgruppe zu betrachten. Die Leistungen der Kinder in den beiden Klassenstufen unterschieden sich nicht signifikant voneinander. Deshalb erfolgte für diese Tests die Auswertung jeweils für das gesamte Kollektiv (Tab.4.2.a im Anhang).

### **4.2.1 Freiburger Sprachverstehenstest (FSV)**

Im Freiburger Sprachverstehenstest (FSV) erreichten 93,9% der *Zielgruppe* und 94,6 % der *Patienten* ein 100 %iges Sprachverstehen. Der Unterschied zwischen beiden Gruppen war nicht signifikant (Tab. 4.2.1.a im Anhang).

### **4.2.2 Freiburger Sprachverstehenstest im Störschall (FSVS)**

Beim Sprachverstehen im sprachsimulierenden Rauschen von hinten (Nutzschall – Störschall – Verhältnis 65/60 dB) erreichten 90,3 % der *Zielgruppe* ein 90 %iges oder besseres Sprachverstehen. Die Leistungen der *Patienten* bei Sprachverstehen im sprachsimulierenden Rauschen waren deutlich eingeschränkter. Insgesamt erreichten nur 51,3 % ein 90 %iges oder besseres Sprachverstehen im Störschall. Die Verteilung der Leistungen für beide Gruppen finden sich in Tab. 4.2.2.a im Anhang.

Die Leistungen der beiden Gruppen unterschieden sich hochsignifikant ( $p = 0,00$ ) und sind in Abb. 4.2.2.1 dargestellt.

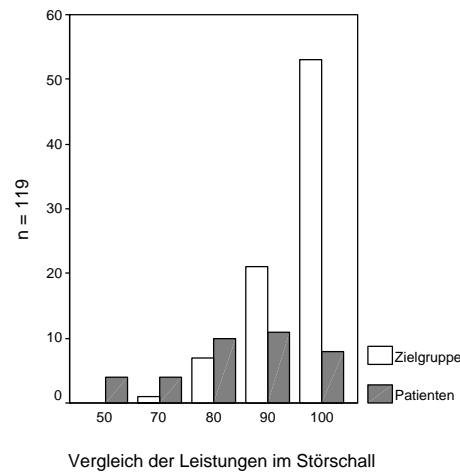


Abb. 4.2.2.1 Verteilung der Leistungen im Freiburger Sprachverstehenstest im Störschall von *Zielgruppe* und *Patienten*, gemessen in 10%-Schritten ( $p = 0,00$ )

### 4.2.3 Zeitkomprimierte Wörter

Für das Verstehen von zeitkomprimierten Wörtern liegen bislang ebenfalls keinerlei Richtwerte vor. Für die jeweils 16 zweisilbigen Wörter, die monaural rechts/links angeboten wurden, wurde die Fehleranzahl ermittelt. In der *Zielgruppe* erreichten die Zweitklässler für das rechte und linke Ohr Mediane von fünf, die Drittklässler von jeweils vier Fehlern. Der T-Test für gepaarte Stichproben ergibt für die gesamte Kindergruppe keinen signifikanten Unterschied zwischen den rechts- und linksohrigen Leistungen.

Zeitkomprimierte Wörter links					Zeitkomprimierte Wörter rechts			
<i>Zielgruppe</i> (n = 82)			<i>Patienten</i> (n = 37)		<i>Zielgruppe</i> (n = 82)		<i>Patienten</i> (n = 37)	
Fehler	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%
1	1	1,2			2	2,4		
2	8	9,8			4	4,9		
3	10	12,2	6	16,2	17	20,7	6	16,2
4	22	26,8	6	16,2	18	22,0	4	10,8
5	17	20,7	9	24,3	17	20,7	8	21,6
6	11	13,4	5	13,5	13	15,9	8	21,6
7	11	13,4	4	10,8	7	8,5	8	21,6
8	1	1,2	7	18,9	3	3,7	2	5,4
9	1	1,2			1	1,2		
10							1	2,7

Tab. 4.2.3.1 Unterschiede bei zeitkomprimierten Wörter links ( $p = 0,05$ ) und rechts ( $p = 0,01$ )

Insgesamt lagen die Mediane der *Patienten* bei den Zweitklässlern für das rechte und linke Ohr bei 6 Fehlern und bei den Drittklässlern für beide Ohren bei 5 Fehlern. Die Gesamtleistungen von *Zielgruppe*

und *Patienten* sind in Tab. 4.2.3.1 dargestellt. Es ergab sich ein signifikanter Unterschied zwischen beiden Gruppen für die linksohrigen Leistungen ( $p = 0,05$ ) und ein hochsignifikanter Unterschied für die rechtsohrigen Leistungen ( $p = 0,01$ ).

#### 4.2.4 Dichotisches Hörvermögen

Der in Anlehnung an Uttenweiler verwendete dichotische Hörtest wurde bei einer Lautstärke von 50 dB SPL angeboten. Die fünf simultan angebotenen Wortpaare (jeweils incl. Artikel) entsprachen bei korrekter Wiedergabe jedes Wortes einem Ergebnis von 100%. Jedes falsch oder nicht verstandene Wort entsprach einer um 10% reduzierten Leistung. Wurde von den zusammengesetzten Hauptwörtern ein Teil des Wortes korrekt wiedergegeben, erhielt das Kind hierfür noch 5%. Der Unterschied zwischen rechts- und linksohrigen Leistungen in der Gesamtgruppe ist erwartungsgemäß hochsignifikant ( $p = 0,01$ ) zugunsten besserer Leistungen mit dem rechten Ohr.

Die Ergebnisse waren bei beiden Gruppen für das rechte und linke Ohr unterschiedlich.

Rechts erreichten 89% der *Zielgruppe* und 67,6% der *Patienten* ein 100 %iges Verstehen. Es bestand zwischen beiden Gruppen ein hochsignifikanter Unterschied ( $p = 0,01$ ).

Uttenweiler rechts 50 dB				
<i>Zielgruppe (n=82)</i>			<i>Patienten (n = 37)</i>	
Verstehen	Anzahl	%	Anzahl	%
80 %	3	3,7	2	5,4
85 %	3	3,7		
90 %	2	2,4	5	13,5
95 %	1	1,2	5	13,5
100 %	73	89,0	25	67,6

Tab. 4.2.4.1 Dichotisches Hörvermögen, rechtsohrige Leistungen ( $p = 0,01$ )

Links lag das 100 %ige Verstehen der *Zielgruppe* bei 80,5 %. Die *Patienten* erreichen 64,9 %. Dieses Ergebnis zeigt keinen signifikanten Unterschied zwischen den beiden Gruppen. Die Verteilung der linksohrigen Leistungen findet sich in Tab. 4.2.4.a im Anhang.

#### 4.3 Psychometrische Untersuchungen der auditiven Leistungen: *Zielgruppe* und *Patienten*

Für die Untersuchung der auditiven Leistungen mit normierten Verfahren können die Ergebnisse von Ziel- und *Patientengruppe* ebenfalls für das jeweilige Gesamtkollektiv gegenübergestellt werden, da entweder altersbezogene Normwerte (T-Werte) vorlagen oder die in den Tests erreichten Rohwerte

miteinander verglichen wurden. Eine Ausnahme bildet hier der Heidelberger Lautdifferenzierungstest (H-LAD).

### 4.3.1 Lautdiskrimination

Für den H-LAD liegen nur orientierende Normierungswerte für die Klassenstufen 2 und 4 vor, so dass für die vorliegende Auswertung von den Rohwerten ausgegangen wurde. Die Leistungen von Drittklässlern werden in dieser Untersuchung erstmalig beschrieben. Ebenso gibt es bislang keine Gegenüberstellung von Normal- und *Patientendaten*.

#### Leistungen der *Zielgruppe*

Die Zweit- und Drittklässler der *Zielgruppe* unterschieden sich im Untertest H1 auditiv, der die auditive Diskriminationsfähigkeit für Wörter und für Silben prüft, signifikant voneinander ( $p = 0,05$ ). Die Zweitklässler erreichten einen MW = 22,7 (STD  $\pm$  2,1), die Drittklässler einen MW = 23,7 (STD  $\pm$  1,5). Eine Verteilung der Rohpunktwerte für beide Gruppen findet sich in Tab. 4.3.1.a im Anhang.

Im Untertest H1 kinästhetisch (H1 kin.), in dem von den Kindern die Wiedergabe der zu diskriminierenden Wortpaare verlangt wird, unterschieden sich Leistungen der Zweit- und Drittklässler hochsignifikant ( $p = 0,00$ ). Die Leistungen sind in Tab. 4.3.1.1 dargestellt. Die Zweitklässler erreichten einen MW = 18,9 (STD  $\pm$  3,0), die Drittklässler einen MW = 21,5 (STD  $\pm$  2,8).

H - LAD - Gesamtwert (H 1 kin)				
Rohwert	Anzahl Kl. 2	%	Anzahl Kl. 3	%
8	1	1,9		
12	1	1,9		
13	2	3,8		
14	1	1,9	1	3,3
15	1	1,9		
16	2	3,8		
17	4	7,7	2	6,7
18	7	13,5	2	6,7
19	9	17,3	2	6,7
20	5	9,6	4	13,3
21	7	13,5	2	6,7
22	11	21,2	4	13,3
23	1	1,9	4	13,3
24			5	16,7
25			4	13,3
Gesamt	52	100,0	30	100,0

Tab. 4.3.1.1 Vergleich der Leistungen der *Zielgruppe* (Kl. 2 vs. Kl. 3) im H-1 kin. ( $p = 0,05$ )

Da in diesen Untertest auch die auditive Merkfähigkeit einget, entspricht die deutlich bessere Leistung der älteren Drittklässler den Erwartungen.

Kein signifikanter Unterschied zwischen zweiter und dritter Klasse bestand im Untertest H2, der die Differenzierung und Analyse von Konsonantenhäufungen im Anlaut prüft. Die Zweitklässler erreichen einen MW = 10,3 (STD ± 1,9), die Drittklässler einen MW = 10,9 (STD ± 1,7). Die genaue Verteilung findet sich in Tab. 4.3.1.b im Anhang.

### Vergleich Zielgruppe vs. Patienten

Abb. 4.3.1.1 zeigt den Vergleich der Leistungen von *Zielgruppe* und *Patienten*. Alle drei Untertests des H-LAD ergaben für beide Klassenstufen hochsignifikante Unterschiede ( $p = 0,00$ ).

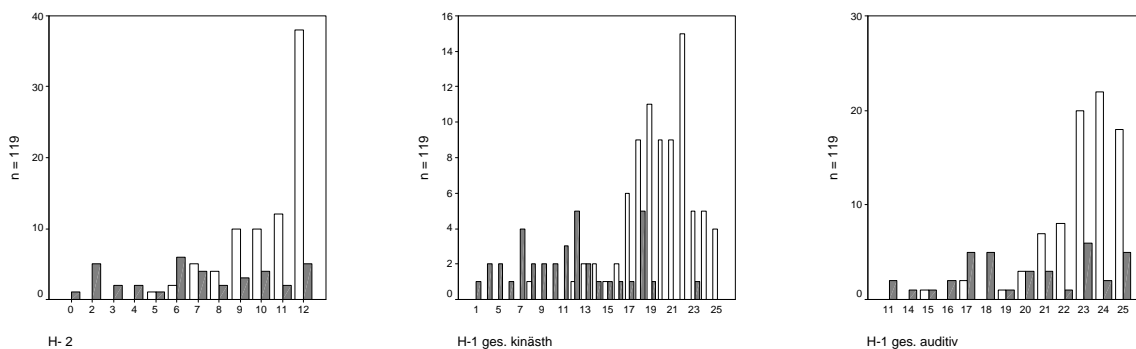


Abb. 4.3.1.1 Vergleich der erzielten Rohwerte der drei Untertests des H-LAD ( $p = 0,00$ ) für das gesamte Kollektiv (*Zielgruppe* weiß / *Patienten* grau).

Zur besseren Übersichtlichkeit wird in (Tab. 4.3.1.2) nur eine Übersicht über die Mittelwerte der beiden Gruppen für die drei Untertests des H-LAD dargestellt, die sich ebenfalls hochsignifikant unterscheiden ( $p = 0,00$ ).

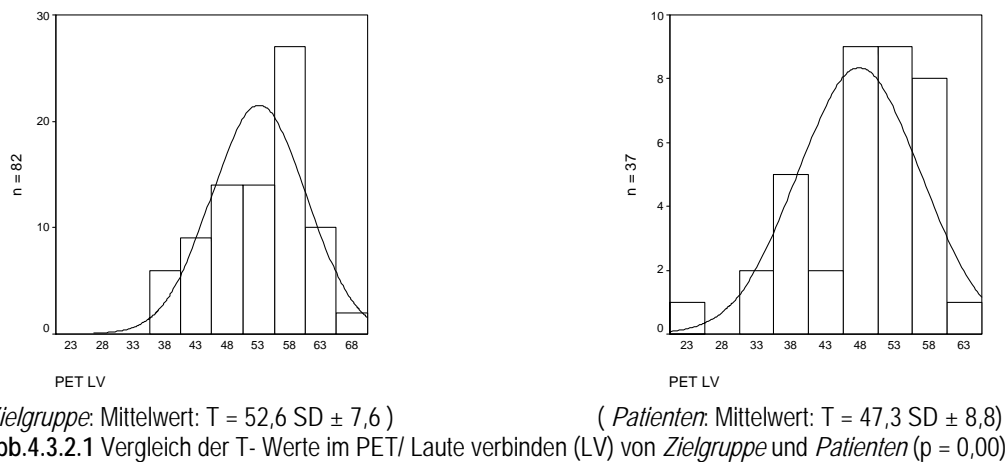
Gruppenvergleich	Mittelwert	STD
H-1 aud. <i>Zielgruppe</i>	23,0	2,0
<i>Patienten</i>	19,8	3,9
H-1 kin. <i>Zielgruppe</i>	19,9	3,2
<i>Patienten</i>	11,3	5,2
H-2 <i>Zielgruppe</i>	10,5	1,8
<i>Patienten</i>	7,0	3,5

Tab. 4.3.1.2 MW und STD von *Zielgruppe* und *Patienten* im H-LAD ( $p = 0,00$ )

### 4.3.2 Auditive Synthese und Ergänzung

#### *Psycholinguistischer Entwicklungstest – Laute verbinden -*

Die Einschätzung der Leistungen in der auditiven Synthese erfolgte mittels zweier Untertests aus dem Psycholinguistischen Entwicklungstest (PET). Die Normierung erfolgt in T-Werten mit einem Mittelwert von 50 und einer Standardabweichung von 10. Das bedeutet, dass der Bereich  $T = 40 - 60$  als durchschnittlicher oder normaler Bereich gilt. Diese T-Werte ermöglichen, dass die Leistungen der *Zielgruppe* ( $n = 82$ ) und der *Patienten* ( $n = 37$ ) direkt miteinander verglichen werden können.

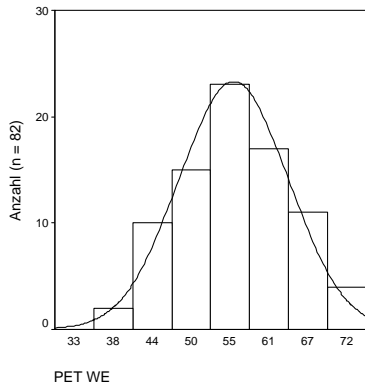


Erwartungsgemäß sollte ein Mittelwert von 50 erreicht werden. Dieser wurde im Laute verbinden (LV) bei der *Zielgruppe* mit einem MW = 52,6 leicht überschritten. Die *Patientengruppe* lag mit einem MW = 47,3 leicht darunter. Die beiden Gruppen unterschieden sich im Untertest LV hochsignifikant ( $p = 0,00$ ) voneinander (Abb.4.3.2.1). Die Verteilung der T-Werte des Untertests beider Gruppen findet sich in Tab. 4.3.2.a im Anhang.

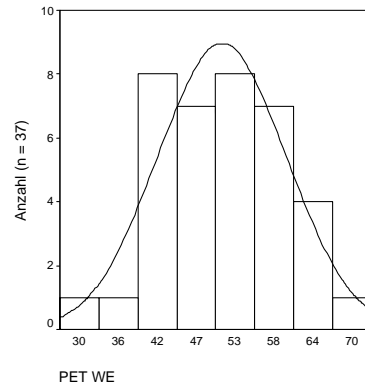
#### *Psycholinguistischer Entwicklungstest - Wörter ergänzen -*

Im Untertest Wörter ergänzen (WE) erreichten die Kinder der *Zielgruppe* einen MW = 55,8;  $SD \pm 7,9$ , die *Patienten* einen MW = 50,9;  $SD \pm 9,2$  s. Abb. 4.3.2.2. Dieser Unterschied war mit ( $p = 0,01$ ) ebenfalls hochsignifikant.





(Zielgruppe: Mittelwert: T = 55,8 SD ± 7,9)



(Patienten: Mittelwert: T = 50,9 SD ± 9,2)

Abb.4.3.2.2 Vergleich der T-Werte im PET Wörter ergänzen (WE) von Zielgruppe und Patienten\*\*

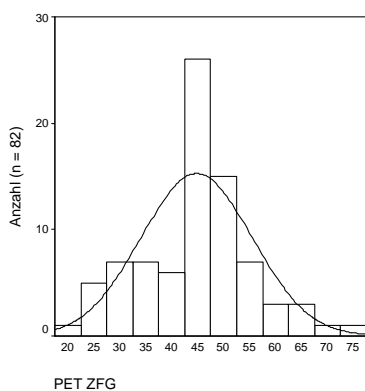
Die Verteilung der T-Werte des Untertests beider Gruppen findet sich in (Tab. 4.3.2.b) im Anhang.

### 4.3.3 Auditive Sequenzierung und Speicherung

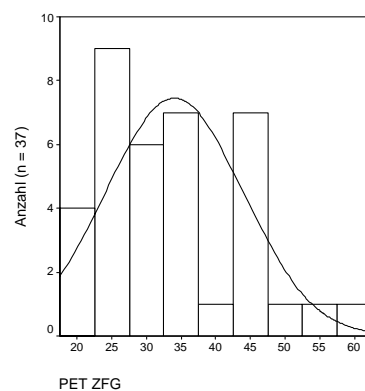
#### Zahlenfolgedächtnis (ZFG)

Die Überprüfung der auditiven Merkfähigkeit stellt einen weiteren Aspekt der Untersuchung dar. Das Zahlenfolgedächtnis (ZFG), ein Untertest des PET, prüft den mechanischen Aspekt dieser Merkfähigkeit. Auch bei diesem Untertest werden die alterstransformierten T-Werte der Zielgruppe mit denen der Patienten verglichen.

Entsprechend der Normierung des PET wird auch hier ein Mittelwert von T = 50, SD± 10 erwartet. In der Zielgruppe liegt der Mittelwert mit T = 44,7, SD± 10,7 etwas darunter. Die Patienten erreichten mit einem T = 34,0, SD± 9,9 einen hochsignifikant schlechteren Mittelwert ( $p = 0,00$ ), siehe Abb. 4.3.3.1.



(Zielgruppe: Mittelwert: T = 44,7 SD ± 10,7)



(Patienten: Mittelwert: T = 34,0 SD ± 9,9)

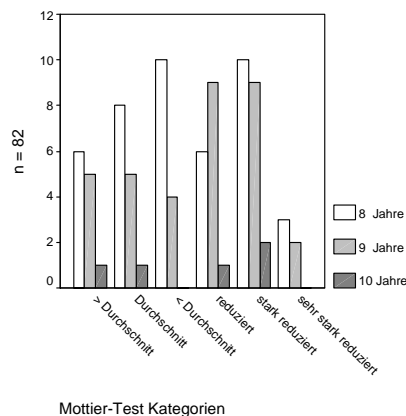
Abb. 4. 3.3.1 Vergleich der T-Werte im PET Zahlenfolgedächtnis (ZFG) von Zielgruppe und Patienten\*\*

Die T-Werte der Zielgruppe und Patienten werden in Tab.4.3.3.a im Anhang dargestellt.

### **Mottier – Test**

#### **Leistungen der Zielgruppe**

Weitere Sequenzierungsleistungen, allerdings in Verbindung mit einer phonematischen Differenzierung, wurden mit dem Mottier-Test erfasst. Für diesen Test liegen bislang nur sehr grobe Orientierungswerte vor. Bohny (1981) unterteilte die Leistungen der Kinder nach Lebensjahren in vier Kategorien („Durchschnitt“, „reduziert“, „stark reduziert“, „sehr stark reduziert“).



**Abb. 4.3.3.2** Leistungen der Zielgruppe im Mottier-Test

Diese Kategorien wurden in der vorliegenden Untersuchung aufgrund der erbrachten Leistungen um zwei erweitert, und zwar „> Durchschnitt“ für Rohwerte über dem Durchschnittswert. Eine Kategorie „< Durchschnitt“ wurde für die Rohwertpunkte eingeführt, die zwischen reduziert und Durchschnitt lagen. Die Verteilung der Leistung der Zielgruppe nach diesen Kategorien ist in Abb. 4.3.3.2 dargestellt. Eine genaue Aufteilung der Rohwertpunkte ist in Tab.4.3.3.b im Anhang aufgeführt. Der Unterschied in den Leistungen zwischen den Kindern der zweiten und dritten Klasse ist nicht signifikant.

#### **Vergleich von Zielgruppe und Patienten**

Die Leistungen der Zielgruppe und Patienten waren im Mottier-Test hochsignifikant ( $p = 0,00$ ) unterschiedlich, mit deutlich schlechteren Leistungen der Patienten. In Abb. 4.3.3.3 sind die Leistungen aus Gründen der Übersichtlichkeit ebenfalls entsprechend diesen Kategorien dargestellt. Der Unterschied in den Rohwerten war ebenfalls hochsignifikant ( $p = 0,00$ ) und in Tab. 4.3.3.b im Anhang dargestellt.

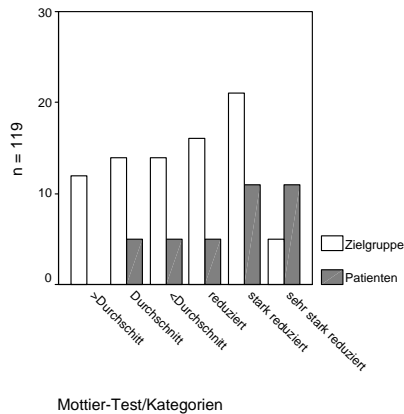


Abb. 4.3.3.3 Ergebnisse im Mottier-Test *Zielgruppe* vs. *Patienten* ( $p = 0,00$ )

### Heidelberger Sprachentwicklungstest ( HSET/IS)

Der Untertest Imitation von Satzstrukturen (IS) des Heidelberger Sprachentwicklungstests prüft ebenfalls das sprachauditive Kurzzeitgedächtnis, allerdings für semantisch relevantes Material. Das bedeutet, die Wiedergabe des vorgesprochenen Satzes durch das Kind erfordert neben Merkfähigkeit und phonematischer Differenzierung auch eine Fähigkeit, morphologisch-syntaktische Strukturen zu erfassen. Die erzielten Rohwerte werden in altersnormierte T-Werte transformiert. Ein Vergleich der Leistungen der Zweit- und Drittklässler der *Zielgruppe* ergab keinen signifikanten Unterschied.

### Vergleich *Zielgruppe* vs. *Patienten*

Eine Verteilung der T-Werte für *Zielgruppe* und *Patienten* findet sich in Tab.4.3.3.c im Anhang. Die Leistungen der *Zielgruppe* unterschieden sich hochsignifikant ( $p = 0,00$ ) von denen der *Patienten* (Abb. 4.3.3.4).

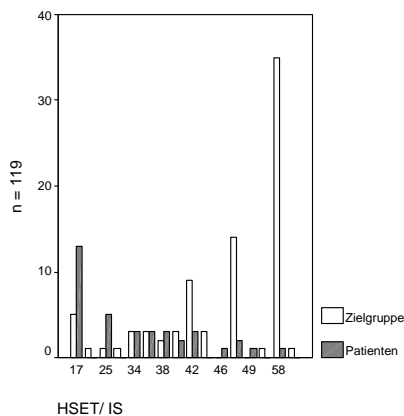


Abb. 4.3.3.4 Gegenüberstellung der T-Werte des HSET/ Imitation von Satzstrukturen (IS) *Zielgruppe* vs. *Patienten* ( $p = 0,00$ )

Die Kinder der *Zielgruppe* erreichten mit einem  $MW = 47,3$ ,  $SD \pm 12,1$  ein etwas unter dem erwarteten  $T = 50$  liegendes Ergebnis. Die *Patienten* blieben mit einem  $MW = 30,3$ ,  $SD \pm 12,1$  deutlich darunter.

### Fehler in der Wiedergabe von Dativsätzen

Bei der Auswertung dieses Untertests ergab sich ein interessanter Aspekt. Vier der insgesamt 12 Sätze enthalten eine Dativkonstruktion, z.B. „Das Fahrrad wird von *dem* Omnibus an die Wand geschoben.“ Eine Wiedergabe dieses Satzes unter Verwendung des Akkusativs anstelle des Dativs, also „Das Fahrrad wird von *den* Omnibus an die Wand geschoben“, ist laut Auswertungskriterien als Fehler zu werten.

In der Untersuchung zeigte sich, dass sehr viele Kinder der *Zielgruppe* durchgängig den Dativ durch den Akkusativ ersetzen. Diese Beobachtung erfolgte sowohl bei den jüngeren als auch bei den älteren Kindern. Da auch die Normierung dieses Entwicklungstests bei der Altersgruppe der 8,11-jährigen Kinder endet, bedeutet dies, dass 51,3 % des untersuchten Kollektivs sogar über der maximalen Altersnorm lagen, also sprachlich noch kompetenter sein müssten. Aufgrund der Leistungen der Kinder in der Phonemdiskrimination im H-LAD/H1 ist nicht davon auszugehen, dass der Unterschied zwischen „m“ und „n“ nicht wahrgenommen wurde. Vielmehr scheint dies ein Fehler im Erfassen der morphologisch-syntaktischen Struktur zu sein. Die Verwendung des Akkusativs anstelle des Dativs ist in der Berliner Umgangssprache häufig vorhanden, so dass bei der Auswertung möglicherweise eine Dialektkorrektur vorgenommen werden müsste. Aus diesem Grund erfolgte die Auswertung dieses Untertests einmal nach Testauswertungskriterien und einmal ohne Berücksichtigung des „Dativfehlers“. Die Unterschiede zwischen beiden Testauswertungen waren sowohl für die älteren als auch für die jüngeren Kinder der *Zielgruppe* hochsignifikant ( $p = 0,00$ ), s. Abb. 4.3.3.5.

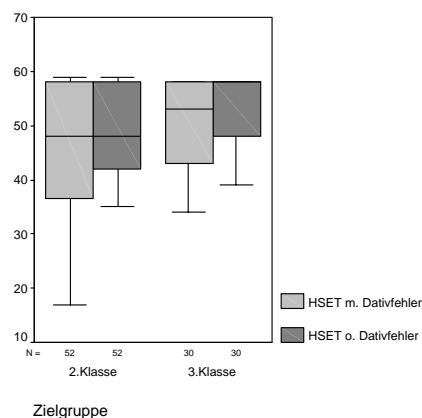
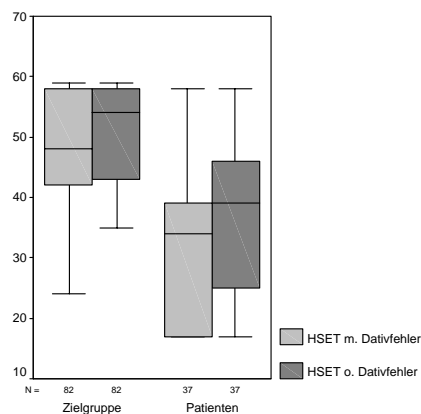


Abb.4.3.3.5 Die T-Werte der *Zielgruppe* im HSET/Imitation von Satzstrukturen unterschieden sich in beiden Klassenstufen hochsignifikant ( $p = 0,00$ ) bei einer Auswertung mit und ohne Dativfehler

Die gleichen Beobachtungen konnten auch an den *Patienten* gemacht werden. Hier trat der „Dativfehler“ noch etwas häufiger auf. Im Mittel ergab sich bei der *Zielgruppe* ein um 3,3 T-Wertpunkte besseres Ergebnis, wenn der Dativfehler unberücksichtigt blieb. Bei den *Patienten* waren es sogar 5,6 T-Wertpunkte. Das bedeutet, auch die *Patienten* zeigten im T-Test einen hochsignifikanten Unterschied ( $p = 0,00$ ) in den Leistungen „mit und ohne Dativfehler“.



**Abb.4.3.3.6** Die T- Werte von *Zielgruppe* vs. *Patienten* im HSET/IS unterschieden sich hochsignifikant ( $p = 0,00$ ) bei einer Auswertung mit und ohne Dativfehler

Insgesamt besteht ein hochsignifikanter Unterschied ( $p = 0,00$ ) zwischen den Leistungen in der Imitation von Satzstrukturen zwischen *Zielgruppe* und *Patienten* unabhängig vom Dativ. Die gesamten Leistungen sind in Abb. 4.3.3.6 dargestellt.

#### 4.4 Zusammenhang der auditiven Leistungen untereinander

Die Korrelationen der auditiven Testergebnisse (Tab. 4.4.1) zeigen, dass die Leistung im Freiburger Sprachverstehenstest die geringsten Korrelationen mit den anderen Testergebnissen aufweist. Die psychometrischen Verfahren korrelieren insgesamt sehr hoch miteinander. Die Begründung hierfür könnte sein, dass sich bestimmte auditive Aspekte, wie Diskrimination, Lautsynthese und auditive Merkspanne in mehreren Verfahren wiederfinden. Als Einzelergebnis ist hierbei hervorzuheben, dass der Mottier-Test mit dem Zahlenfolgegedächtnis (PET/ZFG) deutlich höher korreliert als mit den beiden Testverfahren (PET/LV und H-LAD H2), welche Lautsynthese, bzw. -analyse prüfen. D.h. die auditive Merkspanne fließt in diesen Test offenbar stärker ein als die phonematische Analyse. Die Ergebnisse zeigen auch, dass die drei Untertests des H-LAD hohe Korrelationen aufweisen ( $r = 0,5$  bis  $r = 0,69$ ).

	FSVS	Zeitk W rechts	Zeitk W links	Dich. Hör rechts	Dich. Hör links	Mott Test	PET ZFG	HSET IS	PET LV	PET WE	H1 Aud	H1 Kin	H2
FSV	0,01	- 0,16	- 0,14	0,07	<b>0,20</b>	- 0,02	0,04	0,09	0,17	- <b>0,24</b>	- 0,00	- 0,08	- 0,06
FSVS		- <b>0,22</b>	- <b>0,26</b>	0,11	0,13	<b>0,20</b>	<b>0,25</b>	<b>0,28</b>	0,17	0,13	<b>0,18</b>	<b>0,27</b>	<b>0,33</b>
Zeitk W rechts			<b>0,63</b>	- 0,04	- 0,07	- 0,09	- <b>0,19</b>	- <b>0,19</b>	- <b>0,20</b>	- 0,01	- 0,16	- <b>0,21</b>	- <b>0,25</b>
Zeitk W links				- 0,07	- <b>0,19</b>	- 0,16	- 0,18	- 0,15	- 0,18	- 0,07	- 0,13	- <b>0,23</b>	- <b>0,27</b>
Dich. Hör rechts					<b>0,60</b>	0,15	<b>0,21</b>	<b>0,18</b>	<b>0,21</b>	0,17	0,09	0,15	0,13
Dich. Hör links						<b>0,21</b>	<b>0,24</b>	<b>0,25</b>	<b>0,21</b>	<b>0,19</b>	0,08	0,18	0,06
Mott Test							<b>0,70</b>	<b>0,47</b>	<b>0,40</b>	0,15	<b>0,45</b>	<b>0,43</b>	<b>0,30</b>
PET ZFG								<b>0,50</b>	<b>0,40</b>	<b>0,26</b>	<b>0,38</b>	<b>0,47</b>	<b>0,37</b>
HSET IS									<b>0,41</b>	<b>0,33</b>	<b>0,59</b>	<b>0,56</b>	<b>0,51</b>
PET LV										<b>0,28</b>	<b>0,31</b>	<b>0,38</b>	<b>0,38</b>
PET WE											<b>0,19</b>	<b>0,34</b>	<b>0,37</b>
H1 Aud												<b>0,61</b>	<b>0,50</b>
H1 Kin													<b>0,69</b>

Tab. 4.4.1 Spearman Korrelationskoeffizienten der auditiven Testergebnisse – (signifikante Werte im Fettdruck)

#### 4.5 Auswirkungen aller Leistungen auf das Rechtschreibergebnis

Zunächst wurden die Korrelationen zwischen IQ, visueller Wahrnehmung und den 14 auditiven Testverfahren betrachtet (Tab. 4.5.1)

	IQ	MVPT	Frostig gesamt	DRT RF	DRT WF
IQ		<b>0,36</b>	<b>0,37</b>	0,15	0,17
MVPT			<b>0,33</b>	<b>0,28</b>	<b>0,32</b>
Frostig gesamt				0,11	0,15
DRT RF					<b>0,69</b>
FSV	0,43	0,14	0,18	0,01	0,03
FSVS	- 0,12	<b>0,19</b>	0,05	<b>0,36</b>	<b>0,40</b>
Zeitkomp. W. r.	0,01	0,01	- <b>0,10</b>	- 0,13	- <b>0,21</b>
Zeitkomp.. W.l.	0,04	0,04	- 0,17	- 0,17	- <b>0,21</b>
Dichot. Hör r.	<b>0,19</b>	0,14	- 0,01	0,16	0,15
Dichot. Hör l.	0,11	0,05	0,12	<b>0,19</b>	<b>0,21</b>
Mottier	0,05	<b>0,24</b>	0,11	<b>0,35</b>	<b>0,40</b>
PET/ZFG	<b>0,25</b>	<b>0,46</b>	<b>0,21</b>	<b>0,35</b>	<b>0,53</b>
HSET IS	0,11	<b>0,26</b>	0,13	<b>0,53</b>	<b>0,58</b>
PET/LV	0,11	<b>0,20</b>	0,17	<b>0,37</b>	<b>0,45</b>
PET/WE	0,04	0,09	0,04	<b>0,31</b>	<b>0,36</b>
H-LAD H1 aud.	<b>0,19</b>	0,09	<b>0,24</b>	<b>0,30</b>	<b>0,36</b>
H-LAD H1 kin	0,17	<b>0,29</b>	0,07	<b>0,44</b>	<b>0,58</b>
H-LAD H2	- 0,03	0,04	0,03	<b>0,32</b>	<b>0,54</b>

Tab. 4.5.1 Korrelationskoeffizienten von IQ, Ergebnissen der visuellen Wahrnehmungstests und der Leistungen in den auditiven Untersuchungsverfahren (signifikante Werte im Fettdruck)

Alle acht psychometrischen auditiven Untersuchungsverfahren korrelieren hochsignifikant mit den Rechtschreibleistungen in Bezug auf Wahrnehmungs- und Regelfehler, wobei die Koeffizienten bei den Wahrnehmungsfehlern bei einigen Tests deutlich höher sind als bei den Regelfehlern. Dieser Zusammenhang entspricht den Erwartungen, da die Leistungen in der Diskrimination und Lautanalyse sich stärker in Fehlern der Phonem-Graphem-Konversion ähnlich klingender Phoneme widerspiegeln (z.B. „o“ → „u“).

Um zu zeigen, dass die berechneten Korrelationen den jeweiligen Variablen selbst und nicht der Wirkung dritter möglicherweise intervenierender Variablen dieser Untersuchung zuzuschreiben sind, wurden partielle Korrelationen (Spearman rho) gerechnet. Hier zeigt sich in keinem Fall eine wesentliche Differenz zwischen dem gewöhnlichen und partiellen Korrelationskoeffizienten. D.h. die beobachteten Korrelationen sind allein dem Zusammenhang jeweils der beiden Variablen zuzurechnen. Tab.4.5.2 zeigt, dass sich die Korrelationskoeffizienten zwischen den auditiven Untersuchungsverfahren und den Wahrnehmungsfehlern (DRT/WF) nur minimal verändern, wenn als Kontrollbedingung die anderen, für das Rechtschreibergebnis relevanten Leistungen herangezogen werden.

Korrelation mit DRT/WF	Spearman rho	Kontrolle		
		IQ	MVPT-R	Frostig-Ges.
FSVS	0,40	0,42	0,38	0,39
Zeitkomp. W. rechts	-0,21	-0,20	-0,21	-0,18
Zeitkomp.. W.l.	-0,21	-0,20	-0,21	-0,17
Dichot. Hör r.	0,15	0,03	0,03	0,06
Dichot.Hör l.	0,21	0,19	0,21	0,20
Mottier	0,40	0,41	0,37	0,39
PET/ZFG	0,53	0,49	0,46	0,50
HSET IS	0,58	0,56	0,53	0,56
PET/LV	0,45	0,42	0,41	0,42
PET/WE	0,36	0,35	0,34	0,35
H-LAD H1 aud.	0,36	.389	0,37	0,38
H-LAD H1 kin	0,58	0,59	0,58	0,60
H-LAD H2	0,54	0,56	0,56	0,55

Tab. 4.5.2 Partielle Korrelationen für Wahrnehmungsfehler

In Tab. 4.5.3 ergibt sich ein vergleichbares Ergebnis für die Regelfehler.

Korrelation mit DRT/RF	Spearman rho	Kontrolle		
		IQ	MVPT-R	Frostig-Ges.
FSVS	0,36	0,35	0,30	0,32
Zeitkomp. W. rechts	-0,13	-0,15	-0,15	-0,13
Zeitkomp.. W.links	-0,17	-0,17	-0,18	-0,14
Dichot. Hör rechts	0,16	0,10	0,10	0,12
Dichot.Hör links	0,19	0,15	0,16	0,16
Mottier	0,35	0,34	0,31	0,33
PET/ZFG	0,35	0,31	0,27	0,32
HSET IS	0,53	0,50	0,47	0,50
PET/LV	0,37	0,35	0,33	0,35
PET/WE	0,31	0,30	0,29	0,30
H-LAD H1 aud.	0,30	0,28	0,26	0,28
H-LAD H1 kin	0,44	0,38	0,37	0,39
H-LAD H2	0,32	0,32	0,32	0,31

Tab. 4.5.3 Partielle Korrelationen für Regelfehler

#### 4.6 Auditive Tests vs. Rechtschreibleistungen für die 2. und 3. Klassenstufe

Um herauszufinden welche der auditiven Untersuchungsverfahren besonders starken Einfluss auf die Rechtschreibung hatten, wurde eine schrittweise lineare Regression (auditive Tests auf Wahrnehmungs- und Regelfehler) getrennt für 2. (68 Kinder) und 3. Klasse (51 Kinder) berechnet. Das Ergebnis zeigte, dass bezüglich der Regelfehler das „Sätze nachsprechen“ für beide Klassenstufen von Bedeutung ist, also das sprachauditive Kurzzeitgedächtnis eine Rolle spielt. Bei den älteren Schülern leistet das „Wörter ergänzen“ einen Beitrag zur Varianzaufklärung der Regelfehler. Kinder mit guten Rechtschreibfähigkeiten könnten bei der Analyse des akustischen Eindrucks eines „verstümmelten“ Wortes richtig abgespeicherte Wortbilder zur Analyse zu Hilfe nehmen, was eine gute Graphem-Phonem- Zuordnung repräsentiert (Tab. 4.6.1).

Verfahren	Regelfehler <sup>1</sup>			
	2. Klasse		3. Klasse	
	R <sup>2</sup>	p-value	R <sup>2</sup>	p-value
HSET / IS	0,19	0,000	0,36	0,000
PET / WE			0,46	0,006

Tab. 4.6.1 Regression: auditive Leistungen vs. Regelfehler

<sup>1</sup> die Tafel zeigen nur die Tests mit signifikantem Einfluss



Deutlich anders ist die Zusammensetzung der Untersuchungsverfahren, die mit den Wahrnehmungsfehlern in Zusammenhang stehen (Tab. 4.6.2).

Verfahren	Wahrnehmungsfehler <sup>1</sup>			
	2. Klasse		3. Klasse	
	R <sup>2</sup>	p-value	R <sup>2</sup>	p-value
H-LAD H1 kin	0,35	0,000	0,36	0,000
HSET / IS	0,42	0,000	0,46	0,006
PET / LV	/	/	0,63	0,007
SVSS	0,46	0,000		
PET / WE	0,51	0,000		

Tab. 4.6.2 Regression: auditive Leistungen vs. Wahrnehmungsfehler

Für beide Klassenstufen steht hier mit dem „H-LAD H1 kin“ ein Test im Vordergrund, der die korrekte Wiedergabe der zu diskriminierenden Wort- oder Silbenpaare verlangt. Dieses setzt neben Analysefähigkeit und Speicherung, den kinästhetisch-motorisch korrekten Bewegungsentwurf der Wörter voraus, eine Leistung die für die Unterscheidung ähnlich klingender Laute, z.B. „d/g“, „o/u“ notwendig ist. Dass Kinder am Beginn des Schriftspracherwerbs beim Schreiben häufig lautierend mitsprechen, unterstreicht die Bedeutung dieser Leistung. So besteht auch bei den Zweitklässlern noch eine hochsignifikante Korrelation Spearman Rho = - 0,32 (p = 0,01) zwischen den Rechtschreibleistungen und der Artikulation, bei den Drittklässler ist dieser Zusammenhang nicht mehr signifikant. Schriftspracherwerb ist ohne ein funktionierendes sprachauditives Kurzzeitgedächtnis kaum vorstellbar, weshalb auch in dieser Analyse die „Imitation von Sätzen“ bedeutsam ist. Im nächsten Schritt unterscheiden sich die Klassenstufen allerdings entscheidend. Während für die Drittklässler mit dem Untertest „Laute verbinden“ ein Leistungsaspekt der phonologischen Bewusstheit auf die Rechtschreibleistung einwirkt (McBride-Chang 1995), ist es für jüngere Kinder das Verstehen von Sprache im Störschall, allerdings ist der Varianzanteil hierfür nicht sehr hoch. Ähnlich verhält es sich mit der Auswirkung vom „Wörter ergänzen“ bei den jüngeren Kindern, die eine eher komplexere Anforderung darstellt.

#### 4.7 Erstellung eines auditiven Leistungsprofils der Zielgruppe

Ein Teilziel der vorliegenden Untersuchung ist die Erstellung eines Leistungsprofils für die auditive Verarbeitung und Wahrnehmung von Grundschulern der Klassenstufen zwei und drei. Die notwendigen

<sup>1</sup> die Tafel zeigt nur die Tests mit signifikantem Einfluss

Voraussetzungen einer altersentsprechenden kognitiven und visuellen Entwicklung sind bei der vorliegenden Kohorte (n = 82) gegeben. Die Einschätzung, ob die auditiven Leistungen als altersentsprechend zu bewerten sind, ist insbesondere für die audiologischen Untersuchungsverfahren aufgrund fehlender Normierungswerte nur subjektiv möglich. Auch die psychometrischen Untersuchungsverfahren wurden anhand einer aktuellen Stichprobe bezüglich der vorgegebenen Normwerte kritisch betrachtet, da z.B. im ZFG die *Zielgruppe* deutlich unter dem erwarteten MW von 50 blieb. Objektive Außenkriterien, anhand derer sich Kinder ohne bzw. mit Störungen in der AVW unterscheiden, sind nur orientierend vorhanden. So besteht im deutschsprachigen und anglo-amerikanischen Raum Übereinstimmung, dass Störungen in der AVW Schwierigkeiten im Laut- und Schriftspracherwerb zur Folge haben können (Günther et al. 1992, Schulte-Körne et al. 1998, Sloan 1998, Chermak 1998).

Um zu einer Leistungsbewertung und damit auch zu einer Definition zu gelangen, wann von einer AVWS bei einem Kind gesprochen werden kann, erfolgte eine subjektive Kategorisierung der auditiven Leistungen unserer ansonsten „gesunden“ Kinder der *Zielgruppe*. Bezüglich der Sensitivität war es notwendig, einen „cut-off“-Punkt zu definieren, der so gewählt ist, dass der Prozentsatz der gesunden Kinder, der zur Risikogruppe für AVWS wird, auf einer mit internationalen Untersuchungen vergleichbaren Evidenz basiert. Die vergleichbare Untersuchung von Dornitz und Schow (2000) legte als „cut-off“ aus ähnlichen Überlegungen je nach Test eine bzw. zwei Standardabweichungen zugrunde, da bei einer zu weichen Kriterienwahl ein großer Teil ansonsten unauffälliger Kinder der AVWS-Risikogruppe zugeordnet würde.

Auditive Untersuchungsverfahren	Leistung	Dimension
Freiburger Sprachverstehenstest (FSV)	< 100	%
Freiburger Sprachverstehenstest im Störschall (FSVS)	< 90	%
Zeitkomprimierte Wörter (rechts)	≥ 8	Fehler
Zeitkomprimierte Wörter (links)	≥ 7	Fehler
Dichotisches Hören (rechts)	≤ 90	%
Dichotisches Hören (links)	≤ 85	%
Mottier-Test	≤ 15	Rohwert
Zahlenfolgedächtnis (ZFG/PET)	≤ 29	T - Wert
Sätze nachsprechen (HSET/IS)	≤ 30	T - Wert
Laute verbinden (PET/LV)	≤ 43	T - Wert
Wörter ergänzen (PET/WE)	≤ 46	T - Wert
Heidelberger Lautdifferenzierungstest H-LAD/ H1 auditiv	≤ 20	Rohwert
H-LAD/ H1 kinästhetisch	≤ 16	Rohwert
H2	≤ 8	Rohwert

Tab. 4.7.1 Cut-off-Werte für die auditiven Testverfahren

Zur Vereinheitlichung wurde in dieser Untersuchung als „Grenzwert“ oder „cut-off“ jeweils derjenige genommen, der von 90 % der *Zielgruppe* noch erreicht wurde. Eine Unterschreitung dieses Grenzwertes entsprach einem Versagen in diesem Untersuchungsverfahren. Die Werte für die insgesamt 14 auditiven Verfahren sind tabellarisch aufgelistet (Tab. 4.7.1).

Werden auf der Grundlage dieser Werte die Leistungen der Kinder der *Zielgruppe* und der *Patienten* ermittelt, so ergibt sich bezüglich des Versagens in einzelnen Tests folgende Verteilung (Tab 4.7.2). Der Unterschied zwischen beiden Gruppen ist hochsignifikant ( $p = 0,00$ ).

Versagen der Kinder in den auditiven Untersuchungsverfahren				
Anzahl der Tests	<i>Zielgruppe</i>		<i>Patienten</i>	
	Anzahl	%	Anzahl	%
0	37	45,1	1	2,7
1	24	29,3	4	10,8
2	8	9,8	3	8,1
3	7	8,5	4	10,8
4	4	4,9	5	13,5
5	2	2,4	8	21,6
6			2	5,4
7			6	16,2
8			3	8,1
11			1	2,7
Gesamt	82	100	37	100

Tab. 4.7.2 *Zielgruppe* und *Patienten* unterschieden sich hochsignifikant ( $p = 0,00$ ) in der Anzahl der Unterschreitungen der Cut-off-Werte in den auditiven Untersuchungsverfahren

Dieses Ergebnis zeigt, dass nur knapp die Hälfte der *Zielgruppe* (45 %) in allen Tests die erforderlichen Leistungen erbrachte. In der *Patientengruppe* wurde diese Leistung allerdings nur von einem Kind erbracht. Es zeigte aber auch, dass knapp 16% der *Zielgruppe* in 3 – 5 Tests versagten. Bei den *Patienten* versagten dagegen 86% in 3 – 11 Tests.

Die Leistungen der *Zielgruppe* machen deutlich, dass es möglich ist, mit unzureichenden Leistungen in einzelnen auditiven Tests ohne Schwierigkeiten im Schriftspracherwerb zurecht zu kommen. Kinder mit einer Rechtschreibstörung (DRT-Prozentrang < 10) waren aus der *Zielgruppe* von vornherein ausgeschlossen worden. Darüber hinaus sind alle verwendeten subjektiven Testverfahren in hohem Maße aufmerksamkeitsabhängig. Das bedeutet, dass eine kurze Unaufmerksamkeit schon zu einem Fehlerquotienten in dem jeweiligen Test führen kann, der einem Versagen entspricht. Diese Überlegungen ließen es notwendig erscheinen, nicht nur die Kinder als auditiv unauffällig zu klassifizieren, die in keinem Test versagen. Da es zu diesem Zeitpunkt keine Re-Test- Analyse gibt, mit der belegt werden

könnte, dass die jeweiligen Kinder immer in den zwei oder drei gleichen Tests versagen, musste auch hierbei mit einem Toleranzbereich gearbeitet werden. Aus diesem Grund wurde keine Gewichtung vorgenommen, in welchen Tests ein Kind versagen kann oder nicht, sondern alle Untersuchungsverfahren wurden gleichwertig betrachtet. Die Verteilung der Daten legte aber nahe, dass ein Kind auch in zwei der 14 Tests noch versagen kann, ohne dass von einem Problem in der auditiven Verarbeitung und Wahrnehmung gesprochen werden sollte. Dies bedeutet dann, dass 84% der *Zielgruppe* und 22% der *Patienten* als unauffällig bezüglich ihrer auditiven Leistungen eingestuft werden können.

#### **4.7.1 Bewertung der Rechtschreibergebnisse auf dieser Grundlage**

Die lineare Regressionsanalyse zeigte, dass der IQ und die visuellen Leistungen im Frostig-Test weder auf die Gesamtleistung im DRT noch auf die Regel- oder Wahrnehmungsfehler signifikanten Einfluss hatten. Lediglich die visuellen Leistungen der Kinder im MVPT-R beeinflussten alle drei Rechtschreibergebnisse signifikant. Die kodierten Leistungen der Kinder in den 14 auditiven Untertests entsprechend der „cut-off“ Werte und der Häufigkeit des Versagens ( $\leq 2$  Tests und  $> 2$  Tests) wurden zu der Variablen „Versagen in 14 auditiven Tests“ zusammengefasst.

Diese Variable „Versagen in den 14 auditiven Test“ und die Leistungen im MVPT-R erklären 35,6 % der Gesamtvarianz bei der Zielgröße „DRT-Gesamtwert“. Bei den „Wahrnehmungsfehlern“ erklären die beiden Variablen 39,9 % und bei den „Regelfehlern“ 21,4 % der Gesamtvarianz. Da die Wahrnehmungsfehler am stärksten von auditiven Leistungen beeinflusst werden, ist davon auszugehen, dass der deutlich höhere Varianzanteil auf diese Leistungen zurückzuführen ist. Im Gegensatz dazu haben Regelfehler nur bedingt etwas mit Hörverarbeitung zu tun und sind von daher von diesen Leistungen entsprechend weniger beeinflusst.

### **4.8 Sensitivitätsanalyse der auditiven Testverfahren**

#### **4.8.1 Das Problem**

Die in der diagnostischen Praxis übliche Routine zur Identifikation auditiv gestörter Kinder besteht gegenwärtig in der kompletten Durchführung, der in dieser Untersuchung verwendeten 14 Testverfahren. Dieses Vorgehen ist notwendig, da bislang keine Aussagen getroffen werden konnten, welche Verfahren besonders effektiv und welche möglicherweise überflüssig sind. Die Behandlung dieses Problems erfolgt in folgenden Schritten:

1. Festlegung der cut-offs für die einzelnen Testverfahren zusammen mit
2. Bestimmung der Anzahl von cut-off Überschreitungen für auditive Störung, wie zuvor beschrieben.
3. ROC-Kurven der einzelnen Tests
4. Optimierung der cut-offs
5. Das Reihenfolgeproblem: Welche Einsparungen hinsichtlich des Diagnoseaufwands lassen sich durch intelligente Anordnung der einzelnen Tests (Reihenfolge der Durchführung) gewinnen?
6. Kritik der Vorgehensweise

Wenn man einmal von rein behavioristischen Argumentationen „*es ist, was der Test misst*“ absieht, erzwingt das Fehlen eines Außenkriteriums für *auditive Störung* eine iterative Lösung des mit den Fragen 1 bis 5 umschriebenen Problems. Diese ist in dem Diagramm in Abb. 4.8.2.1 dargestellt:

#### 4.8.2 Iterative Vorgehensweise, Bestimmung optimaler Cut-off-Werte

Die mit Punkt 4. formulierte Frage wurde auf iterative Weise gelöst. Die dabei verwendeten Zyklen sind in der folgenden Abbildung schematisch dargestellt.

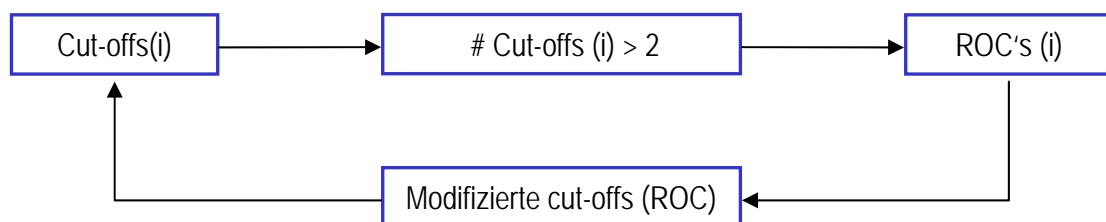


Abb. 4.8.2.1 Bestimmung der optimalen Cut-off-Werte

Jeder Zyklus ( $i = 1, 2, \dots$ ) besteht aus folgenden Operationen

- a. Anwendung der cut-offs ( $i$ )
- b. Diagnose "auditive Störung" bei mehr als 2 cut-off Überschreitungen in den 14 Tests (dispositiv)
- c. ROC-Kurven für jeden der 14 Tests mit der Indikatorvariablen nach Kriterium b.
- d. Bestimmung "optimaler" cut-offs aus den ROC-Kurven

Das Verfahren und insbesondere seine mögliche Konvergenz sind abhängig von den konkreten Daten. Dabei bedeutet Konvergenz die Entwicklung hin zur Konstanz der individuellen cut-offs bei weiteren Iterationen.

### 4.8.3 Das Reihenfolgenproblem

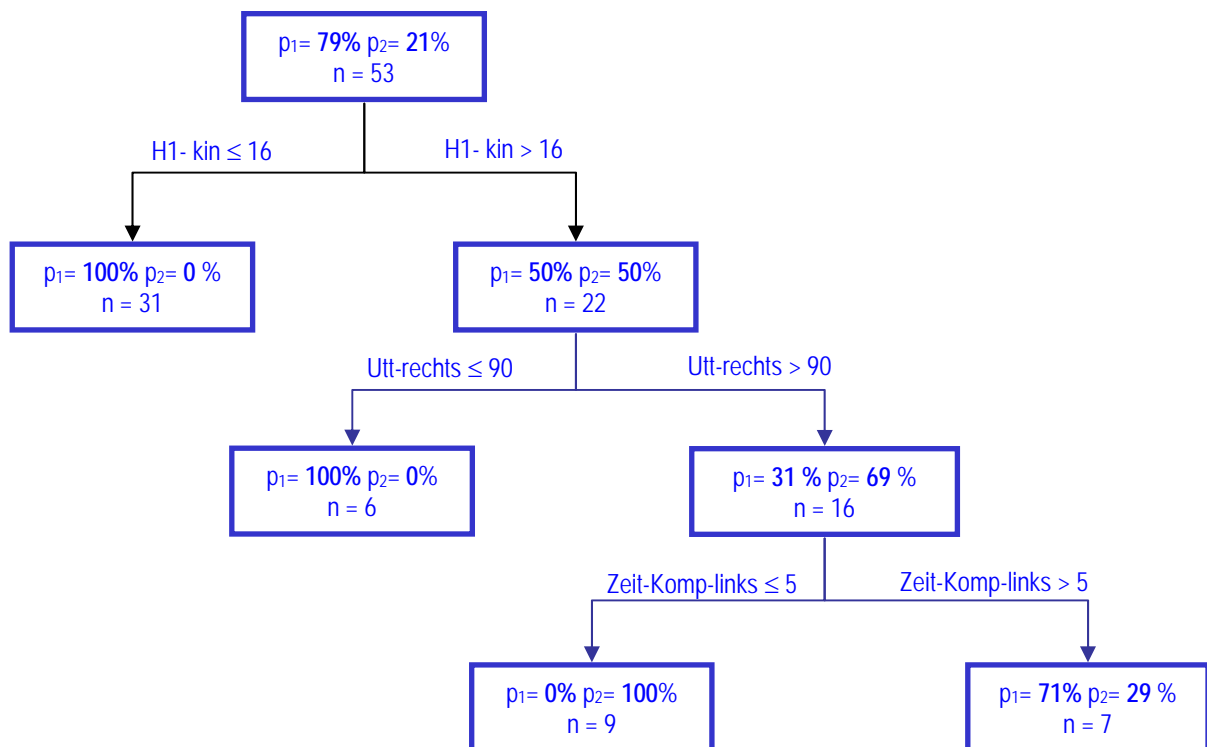
Nach unserem bisherigen Kenntnisstand stellen auditive Störungen kein eindimensionales Phänomen dar, sondern modellieren sich in einer Reihe unterschiedlicher Facetten, deren statistische Verteilung noch weitgehend unbekannt ist. Entsprechend ist von den heute verfügbaren Tests davon auszugehen, dass sie jeweils nur eine oder einige dieser Facetten bzw. Teildimensionen abzubilden in der Lage sind. Wenn man von einer möglichen Gleichverteilung aller Facetten im Störungsprofil absieht, folgt hieraus u.a. die Notwendigkeit, die Trennschärfe der gesamten Diagnose durch eine optimale Reihenfolge der Einzeltests zu optimieren. Die konkrete Forschungsaufgabe besteht also darin, jene Reihenfolge der Anwendung der Einzeltests zu bestimmen, mit der bei minimaler Anzahl der durchgeführten Tests die Trennschärfe des Gesamttests maximal ist. Es zeigt sich, dass die erforderliche Anzahl der Tests für Teilgruppen weit unter der des vollen Programms aller 14 Tests lag. Der logische Hintergrund hierfür ist in der Redundanz aller Tests zu sehen.

Die Bestimmung dieser optimalen Reihenfolge sowie die Verringerung der Zahl überhaupt notwendiger Tests erfolgt unter Verwendung einer Baumanalyse, deren Vorgehensweise hier kurz beschrieben werden soll. Knapp  $\frac{1}{4}$  der untersuchten *Patienten* zeigen keine Auffälligkeiten in der auditiven Wahrnehmung (Versagen  $\leq 2$  Tests). Auf dieser Basis wurde für die Zusammensetzung des Evaluierungskollektivs der Datensatz der 42 Kinder, die in mehr als 3 Tests versagt hatten, durch 11 randomisiert ausgewählte auditiv unauffällige Kinder ergänzt. Damit entspricht diese Kohorte im Verhältnis dem in der Klinik vorgestellten *Patienterkollektiv*.

## 4.9 Baumanalysen

### 4.91 Evaluierungskollektiv

Die in Grafik 4.9.1.1 dargestellten Baumanalyse zielen darauf ab, die Proviso-Testgesamtheit von auffälligen und auditiv unauffälligen Kindern auf der Basis der vorliegenden 14 Tests derart in homogene Teilgesamtheiten zu zerlegen, dass diese entweder nur auditiv gestörte oder nur nicht gestörte Kinder enthalten.



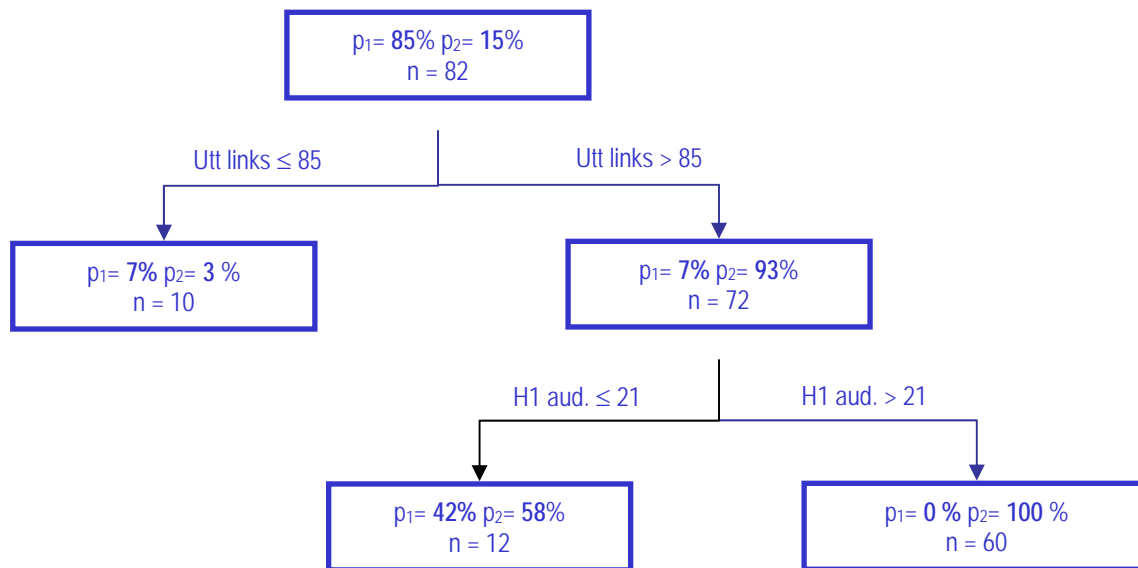
**Legende:**  $p_1$ : Anteil > 2 Tests auffällig  $p_2$ : Anteil ≤ 2 Tests auffällig

Abb. 4.9.1.1 Baumanalyse des Evaluierungskollektivs

In dieser Zusammensetzung zeigt der Untertest H1 kin die größte Trennschärfe. Aufgrund dieses Verfahrens können bereits 58 % der Kinder als eindeutig auditiv auffällig ausgewiesen werden. Dieses Verfahren beinhaltet phonematische Differenzierungsleistungen und Sprachauditives Kurzzeitgedächtnis. Durch Ergänzung mit dem dichotischen Hörtest (rechts) verbleiben nur noch 30% der Gesamtstichprobe bezüglich ihrer auditiven Auffälligkeiten unklar. Mit Hilfe des zeitkomprimierten Worttestes (links) können bis auf 7 Kinder, das entspricht 13% des Evaluierungskollektivs, eindeutig in auditiv auffällig oder unauffällig klassifiziert werden.

#### 4.9.2 Zielgruppe

Bezieht man die Baumanalyse ausschließlich auf die Kinder der *Zielgruppe*, die nur zu einem sehr geringen Prozentsatz (15 %) auditive Auffälligkeiten zeigten, so erhält eine andere Testkonstellation zur Trennung der auditiv auffälligen von den unauffälligen Gültigkeit (s. Grafik 4.9.2.1)



Legende:  $p_1$ : Anteil > 2 Tests auffällig  $p_2$ : Anteil  $\leq$  2 Tests auffällig

Abb. 4.9.2.1 Baumanalyse der *Zielgruppe*

Die hohe Trennschärfe, die das dichotische Hören links in der *Zielgruppe* zeigt, kann so interpretiert werden, dass bei Kindern dieser Altersgruppe die Fähigkeit der binauralen Separation, nämlich unterschiedliche Informationen auf beiden Ohren gleich gut wahrzunehmen, eine grundlegende Fähigkeit darstellt. Immerhin erreichten 80,5 % dieser Gruppe auf dem linken Ohr ein 100%iges Verstehen. Da bei diesem Test allerdings dreisilbige Wörter verwendet werden, ist davon auszugehen, dass Komponenten wie Wortschatz in diese Leistungen mit einfließen. Um dies zu klären, wäre es notwendig, einen dichotischen Test mit semantisch inhaltsarmem Material, wie Zahlen durchzuführen, den es gegenwärtig nur in englischer Sprache gibt. Im Gruppenvergleich [*Zielgruppe* vs. *Patienten*] zeigte der Test Dichotisches Hören linksohrig keinen signifikanten Unterschied zwischen gesunden Kindern und *Patienten*. Der zweite Test, der in der *Zielgruppe* signifikant trennte, prüft mit der auditiven Diskrimination von Wörtern oder Silben, die sich jeweils in einem Phonem unterscheiden, ebenfalls eine grundlegende Leis-



tung. Insgesamt unterschritten nur 17 % der *Zielgruppe* den in der Sensitivitätsanalyse kritischen Wert von [ H1 aud.  $\leq 21$ ]. Ob diese zwei Leistungen für normalentwickelte Kinder dieser Altersgruppe die grundlegenden auditive Funktionen darstellen, müsste an einer größeren Gruppe von gesunden Kindern überprüft werden.

#### **4.9.3 Extremgruppenvergleich**

Erwartungsgemäß hängt die Sensitivität der verwendeten Untersuchungsverfahren sehr stark von der Gruppenzusammensetzung ab, was auch durch eine ergänzende Analyse mittels eines Extremgruppenvergleiches dokumentiert wird.

Hierzu wurde die Gruppe der 38 Kinder, die in keinem der auditiven Testverfahren versagt hatten, den 31 Kindern gegenübergestellt, die in vier oder mehr Tests versagt hatten. Die Testkonstellation, die sich hier als signifikant trennscharf erwies, umfasste im ersten Schritt den H-LAD H1 aud. [H1 aud  $\leq 21$  vs. H1 aud.  $> 21$ ], im zweiten Schritt durch den Mottier-Test [Mottier RP  $\leq 21$  vs. Mottier RP  $> 21$ ], im dritten Schritt konnten mit dem H-LAD H2 [H2  $\leq 7$  vs. H2  $> 7$ ] fünf weitere Kinder zu 100 % als auditiv auffällig eingestuft werden. Auch bei dieser Analyse gelingt keine 100%-ige Trennung. Es verbleiben drei Kinder, die weiter untersucht werden müssten.

#### **4.10 Bewertung der Baumanalysen**

Die durchgeführten Baumanalysen zeigen, dass die qualitative und quantitative Aussagekraft der verwendeten 14 auditiven Untersuchungsverfahren von der Zusammensetzung des Kollektivs und von der Position, die das jeweilige Verfahren in der Reihenfolge hat, abhängen. Da die verschiedenen Testverfahren einander überlappende Leistungen messen, kann ein effektives Screening mit 3 – 4 Testverfahren im klinischen Alltag durch eine evidenzbasierte Reihenfolge optimiert werden. Dies setzt allerdings voraus, dass die zu untersuchende Kohorte eine ähnliche Zusammensetzung aufweist wie das vorliegende Evaluierungskollektiv. Die 3 Testverfahren, die sich für die Einteilung dieser Gruppe als besonders wirksam erwiesen haben, umfassen Diskrimination und Gedächtnisleistungen sowie den Aspekt des Sprachverstehens bei verminderter Redundanz.

#### **4.11 Sensitivität und Spezifität der einzelnen auditiven Untersuchungsverfahren**

Zur Bewertung eines Tests hinsichtlich seiner Trennschärfe bezüglich eines dichotomen Kriteriums werden traditionell die beiden Wahrscheinlichkeiten der Koinzidenz von Testergebnis und Realität in

Form der Sensitivität und Spezifität sowie die positive (PPV)<sup>1</sup> und negative Prädiktion (NPV)<sup>2</sup> verwendet, deren Schätzung durch die entsprechenden bedingten relativen Häufigkeiten erfolgt. Mit den absoluten Häufigkeiten a, b, c, und d der folgenden Tafel

		Kind ist auditiv		
		auffällig	unauffällig	
Test	positiv	a	b	a+b
	negativ	c	d	c+d
		a+c	b+d	

gilt:

$$\text{Sensitivität} := \frac{a}{a+c}$$

$$\text{Spezifität} := \frac{d}{b+d}$$

$$\text{PPV} := \frac{a}{a+b}$$

$$\text{NPV} := \frac{d}{c+d}$$

Die in der vorliegenden Arbeit untersuchten auditiven Tests liefern jeweils Messwerte auf einer fein gestuften Skala und benötigen somit hinsichtlich der dichotomen Entscheidung "ja" / "nein" bezüglich auditiver Störungen die Festlegung eines cut-points. Die Wahl dieses Punktes entscheidet bei gegebener Güte des Tests über dessen Sensitivität, Spezifität, etc. Eine einfache Methode zur Bestimmung des cut-points bietet die ROC<sup>1</sup>-Kurve. Es liegt auf der Hand, dass etwa Sensitivität und Spezifität nicht unabhängig voneinander maximiert werden können. Tendenziell sind beide Parameter gegenläufig.

Die systemische Variation der Parameter Sensitivität und Spezifität in dem Bereich von 0 bis 1 führt in einer Abbildung der Funktion "Sensitivität = f(1-Spezifität)" zu einer von (0;0) bis (1;1) verlaufenden Kurve. Die Güte des Tests kann gemessen werden durch die Area under the Curve (AUC), d.h. durch die Fläche unter dieser Kurve. Für jeden der Tests kann nun an seiner ROC-Kurve der "optimale" cut-point festgelegt werden.

---

<sup>1</sup> PPV = Positive Predictive Value

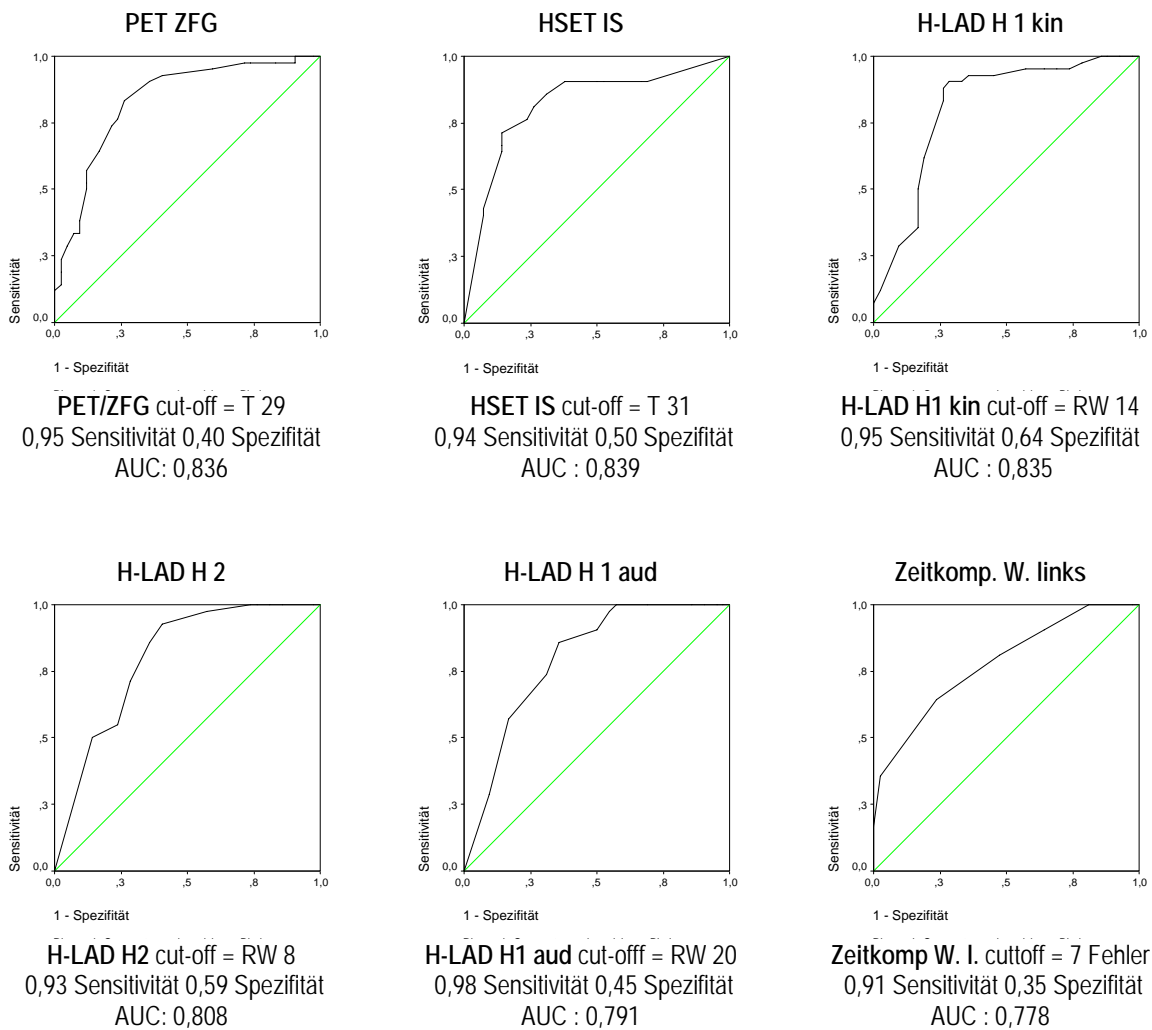
<sup>2</sup> NPV = Negative Predictive Value

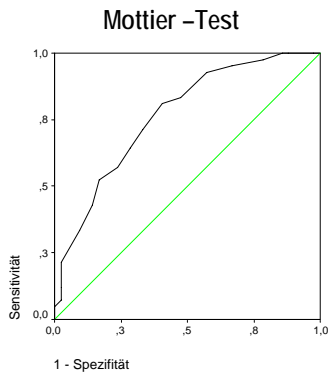
<sup>1</sup> Receiver Operating Characteristic

Eine optimale Anpassung sollte bewirken, dass einerseits nur eine möglichst geringe Anzahl tatsächlich auffälliger Kinder fälschlicherweise als unauffällig bezeichnet und andererseits unauffällige Kinder möglichst keiner weiteren Behandlung unterzogen werden.

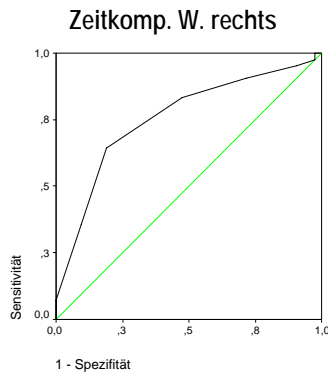
Die ROC-Kurven der hier untersuchten Tests finden sich in den folgenden Abbildungen 4.11.1 –14.

**Abb. 4.11.1 - 14 ROC- Kurven für die auditiven Testverfahren**

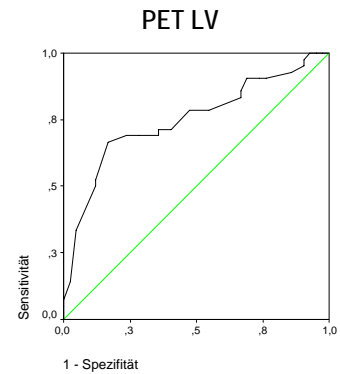




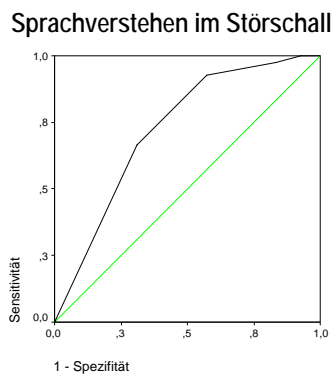
**Mottier cut-off = RW 15**  
0,93 Sensitivität 0,43 Spezifität  
AUC: 0,766



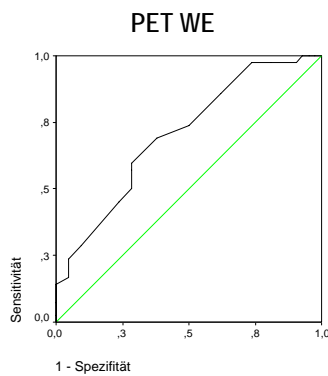
**Zeitkomp W. r. cut-off = 7 Fehler**  
0,91 Sensitivität 0,29 Spezifität  
AUC : 0,755



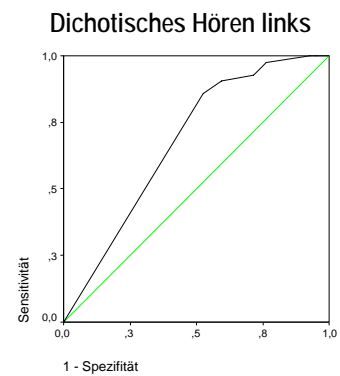
**PET/LV cut-off = T 42**  
0,91 Sensitivität 0,31 Spezifität  
AUC : 0,750



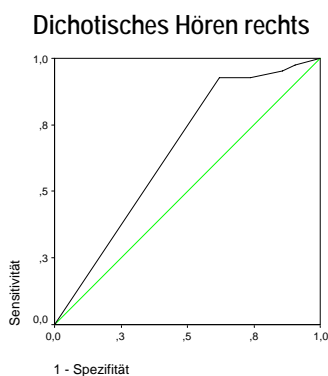
**SVS cut-off = 85 %**  
0,90 Sensitivität 0,43 Spezifität  
AUC: 0,727



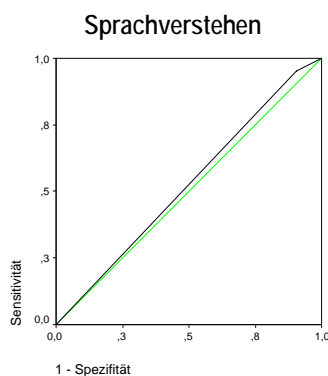
**PET/WE cut-off = T 45**  
0,95 Sensitivität 0,26 Spezifität  
AUC : 0,711



**Dichot. Li. cut-off = 88 %**  
0,93 Sensitivität 0,28 Spezifität  
AUC : 0,692



**Dichot. Re. cut-off = 98 %**  
0,93 Sensitivität 0,38 Spezifität  
AUC: 0,657



**Sprachverstehen cutoff = 95 %**  
0,95 Sensitivität 0,19 Spezifität  
AUC : 0,528

Für acht der verwendeten Untersuchungsverfahren war der in der ROC-Kurve als optimaler cut-off ermittelte Wert mit dem zuvor festgelegten 90%-Kriterium identisch. Bei den anderen Testverfahren war die Abweichung vom festgelegten cut-off-Wert so gering, dass die unter 4.8.2 beschriebene iterative Vorgehensweise zur Optimierung der cut-offs nicht notwendig war.

Vier der verwendeten Testverfahren erreichten AUC-Werte  $> 0,8$ , mit Ergebnissen bezüglich der Sensitivität  $> 90$ . Dadurch ist sichergestellt, dass auditiv auffällige Kinder mit einer hohen Wahrscheinlichkeit identifiziert werden können. Die Spezifität dieser vier Verfahren mit Werten zwischen 0,64 und 0,19 lässt keine zufriedenstellende Identifikation der auditiv unauffälligen Kinder zu.

#### 4.12 Explorative Faktorenanalyse

Die Testverfahren, die bei den verschiedenen Kollektiven in ihrer Spezifität unterschiedlich wirksam sind, sollten sich entsprechend den Konsensuspapieren (ASHA, 1996; Ptok et al., 2000) übergeordneten auditiven Funktionen zuordnen lassen. Diese Submodalitäten der auditiven Verarbeitung und Wahrnehmung umfassen u.a. auditive Aufmerksamkeit, auditive Diskrimination, Identifikation, Separation, Selektion, Ergänzung, Merkfähigkeit etc. (Schow et al, 2000; Jerger & Musiek, 2000; Nickisch, 2002). Um zu prüfen inwieweit sich die in dieser Untersuchung verwendete Testbatterie auf derartig grundlegende Faktoren reduzieren lässt, erfolgte für die 14 auditiven Untersuchungsverfahren anhand des Gesamtkollektivs ( $n = 119$ ) eine explorative Faktorenanalyse als Hauptkomponentenmodell mit Varimax-Rotation. Die Komponenten und ihre Ladungen sind in Tab.4.12.1 zusammengestellt.

Untersuchungsverfahren	Komponente			
	1	2	3	4
HSET Imitation von Satzstrukturen	,83			
H-LAD H1- kin	,82			
Mottier – Test	,79			
H-LAD H2	,77			
H-LAD H1 aud	,77			
PET/ ZFG	,74			
PET /LV	,71			
Zeitkomprimierte Wörter links		-,88		
Zeitkomprimierte Wörter rechts		-,86		
Sprachverstehen im Störschall		,50		
Dichotisches Hören Uttenweiler links			,84	
Dichotisches Hören Uttenweilerrechts			,83	
Sprachverstehen				,85
PET/WE				,60

Tab. 4.12.1 Rotierte Komponentenmatrix

Die vier extrahierten Komponenten in Tab. 4.12.2 erklären insgesamt 70 % der Gesamtvarianz. Die erste Komponente ist dabei mit 33,8 % der stärkste Faktor und besteht mit ähnlichen Ladungen aus nahezu allen psychometrischen Testverfahren (ausgenommen PET/WE).

Erklärte Gesamtvarianz			
Rotierte Summe der quadrierten Ladungen			
Komponente	Gesamt	% der Varianz	Kumulierte %
1	4,73	33,8	33,8
2	2,07	14,8	48,6
3	1,76	12,6	61,2
4	1,24	8,8	70,0

Tab. 4.12.2 Erklärte Gesamtvarianz der Hauptkomponentenanalyse

Diese Tests prüfen verschiedene Aspekte der auditiven Diskrimination und Analyse, sowie des auditiven Gedächtnisses. Hierin zeigt sich deutlich, dass sich diese beiden Submodalitäten (Diskrimination und Merkfähigkeit) mit den Testverfahren, die gegenwärtig in der Diagnostik Verwendung finden, nicht trennen lassen.

Die zweite Komponente besteht aus den Tests mit zeitkomprimierten Wörtern rechts und links, sowie aus dem Sprachverstehen im Störschall, das in angloamerikanischen Untersuchungen als Leistung in der auditiven Figur-Grundwahrnehmung bezeichnet wird (Keith 2000). Gemeinsam ist diesen Untersuchungsverfahren, dass bedingt durch eine Verzerrung des sprachlichen Signals dessen Redundanz und damit die Verständlichkeit erheblich eingeschränkt ist und vom Individuum ein Schließen oder Ergänzen des akustischen Signals gefordert ist, um zum Verstehen zu kommen. Die in angloamerikanischen Untersuchungen verwendeten zeitkomprimierten Sprachsignale werden in der Regel als Einsilber mit 70 % Zeitkomprimierung angeboten, allerdings bei Erwachsenen mit z.T. umschriebenen neurologischen Störungen (Beasley et al. 1972). Nach Bellis (1997) haben bezüglich der fehlenden Redundanz, die für die Untersuchung von Kindern verwendeten Wörter, bei denen die Frequenzen über 1000 Hz weggefiltert werden, einen ähnlichen Effekt, wie die Zeitkompression des Sprachsignals. Die Reduktion des sprachlichen Signals, die entweder durch Überlagerung durch ein Störgeräusch oder durch Verzerrung zustande kommt, stellt einen Faktor dar, der als Submodalität „Ergänzung auditiver Informationen“ (auditory closure) analog zu der Untersuchung von Schow & Chermak (1999) klar zu umreißen ist.

Die dritte Komponente, die das Verstehen dichotisch angebotener Wortpaare rechts und links beinhaltet, kann als binaurale Integration/Separation bezeichnet werden. Die Benennung dieser Leistung erfolgt in der angloamerikanischen Literatur uneinheitlich. So bezeichnen Schow & Chermak (1999) die Leistungen im dichotischen Worttest als „binaurale Separation“. Dornitz & Schow, (2000) benutzen für diese Leistung, allerdings mit dichotisch angebotene Zahlenpaaren, den Begriff „binaurale Integration“. Im Unterschied dazu wird das Verstehen von dichotischen Sätzen als binaurale Separation bezeichnet. Hierbei handelt es sich allerdings um eine etwas andere Aufgabenstellung. Im Unterschied zu den mit gleicher Lautstärke angebotenen Wort- oder Zahlenpaaren, wird der Testperson bei den dichotischen

Sätzen der zu identifizierende „Zielsatz“ mit 30 dB und der auszublendende konkurrierende Satz mit 50 dB Lautstärke angeboten. Schow et al. (2000) sprechen sich letztendlich dafür aus, diese Leistungen als einen binauralen Faktor zu verstehen.

Der vierte Faktor, der sich aus dem Sprachverstehen und dem Wörter ergänzen zusammensetzt, lässt sich zu keiner auditiven Submodalität zusammenfassen. Die Aufgabe „Wörter ergänzen“ aus dem PET erfordert von der Aufgabenstellung her, eine Leistung, die dem Schließen von auditiven Informationen entsprechen sollte. Allerdings zeigt die klinische Praxis, dass dieser Test besonders Kindern, die über einen eingeschränkten Wortschatz verfügen, Probleme bereitet. Da das Sprachverstehen ohne Verzerrungen allen untersuchten Kindern zu knapp 95 % korrekt gelingt, könnten diese Verfahren eher das Sprachverständnis auf Wortebene repräsentieren als eine auditive Leistung im engeren Sinne.

#### **4.13 Konfirmatorische Faktorenanalyse**

Im Gegensatz zur explorativen bietet die konfirmatorische Faktorenanalyse die Möglichkeit, theoretisch begründete Modelle einer Prüfung zu unterziehen, an deren Ende das verwendete Modell akzeptiert oder verworfen wird. Neben der explorativen Faktorenanalyse bieten die Kovarianzstrukturmodelle die Möglichkeit, pfadanalytische Methoden auf latente Variablen anzuwenden.

In der vorliegenden Analyse wurde ein Dreifaktormodell gewählt, mit dem – theoretisch – alle 12 der 14 verwendeten Tests durch 3 Dimensionen (Faktoren) „erklärt“ und die zugehörigen Ladungen geschätzt werden. Abbildung 4.13.1 zeigt das Ergebnis der Berechnung.

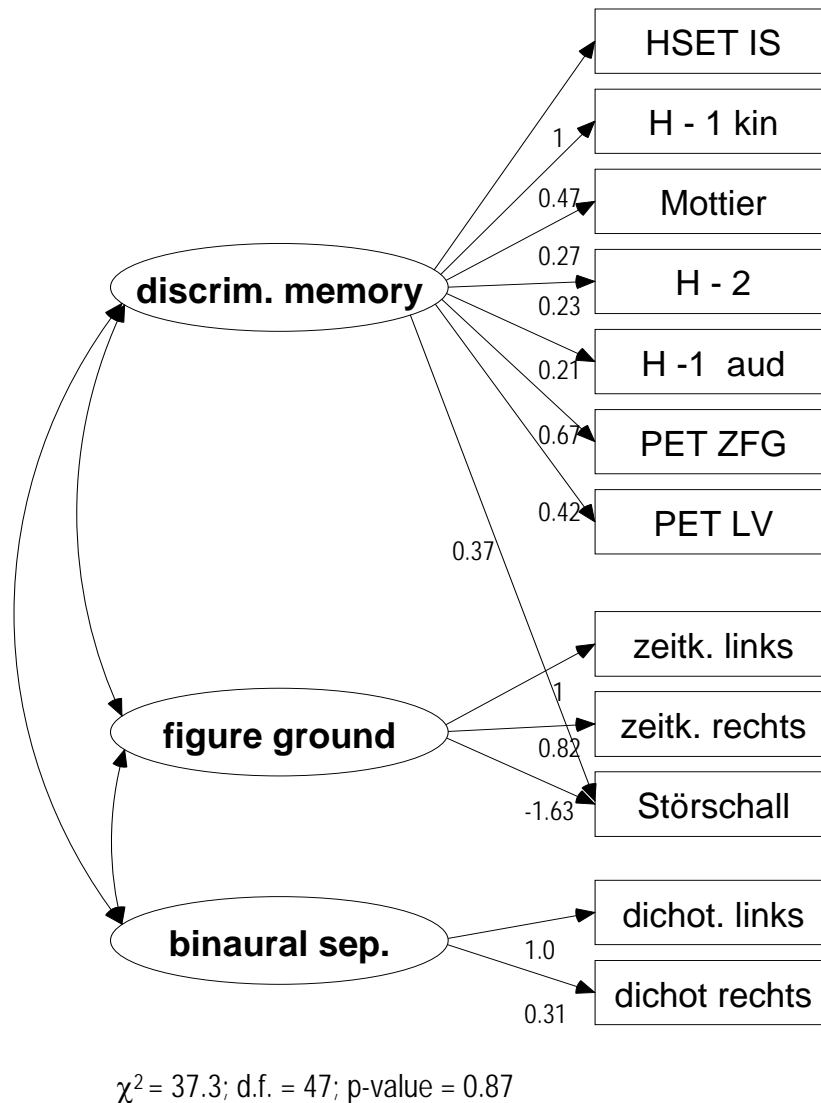


Abb. 4.13.1 Konfirmatorische Faktorenanalyse als Dreifaktormodell (Amos 4). Die drei gewählten Faktoren entsprechen den der explorativen Faktorenanalyse mit dem größten Varianzanteil: Discrimination/memory (Diskriminatioin/ Gedächtnis), Figure ground (Figur-Grundwahrnehmung), binaural separation (bianurale Integration bzw. Seperation)

Die vorangegangene explorative Faktorenanalyse legte nahe, dass im wesentlichen drei Faktoren die 14 Testvariablen - mit Ausnahme des Sprachverstehens - aufklären. Die in Abb. 4.13.1 dargestellte konfirmatorische Faktorenanalyse zeigt die statistische Kompatibilität der Dreifaktorstruktur. Im Gegensatz zur explorativen Faktorenanalyse zeigte sich hier jedoch, dass ein befriedigender „fit“ des Modells nur dann zu erreichen ist, wenn die Variable "Störschall" sowohl den Faktor „Discrimination and memory“ als auch „Figure ground“ lädt, eine Information, die sich mit der explorativen Form so nicht gewinnen



lässt. Daneben zeigte die Analyse auch, dass ein befriedigender „fit“ des Dreifaktormodells nur für nicht-orthogonale Faktoren zu erreichen ist. Alle drei Faktoren zeigen eine Kovarianz untereinander und erweisen sich somit als inhaltlich nicht disjunkt. Als einzige Variable zeigte das Dichotische Hören rechts keine Signifikanz für den Faktor der binauralen Separation