

# Zur prognostischen Validität des Heidelberger Auditiven Screenings in der Einschulungsdiagnostik HASE

## Abschlussbericht des Projektes EVER\*

Jeanette Roos und Hermann Schöler  
(Projektleitung)

Anke Treutlein  
(Wissenschaftliche Mitarbeiterin)

Februar 2007

\*gefördert durch die



sowie die

**Dürr-Stiftung Hamburg**  
und die  
**Günter-Reimann-Dubbers-Stiftung Heidelberg**

*E*valuation eines *V*orschulscreenings  
zur *E*rfassung von *R*isikokindern  
für Sprach- und Schriftspracherwerbsprobleme  
(EVER)

**Projekt EVER**

Pädagogische Hochschule Heidelberg

Keplerstr. 87, D-69120 Heidelberg

Tel. (06221) 477426

E-Mail: [Hermann.Schoeler@urz.uni-heidelberg.de](mailto:Hermann.Schoeler@urz.uni-heidelberg.de)

[Roos@ph-heidelberg.de](mailto:Roos@ph-heidelberg.de)

# Inhaltsverzeichnis

1	<b>Einleitung</b> .....	4
2	<b>HASE – das Heidelberger Auditive Screening in der Einschulungsdiagnostik</b> .....	5
2.1	Die vier Aufgaben .....	5
2.2	Zur Durchführung und Auswertung .....	6
3	<b>Fragestellung</b> .....	6
4	<b>Methode</b> .....	7
4.1	Untersuchungsgruppe .....	7
4.2	Durchführung der Untersuchungen .....	8
4.2.1	Durchführung von <i>HASE</i> .....	8
4.2.2	Schulleistungen und Deutschkenntnisse – Einschätzungen durch die Lehrpersonen .....	10
4.2.3	Prüfung der kognitiven Leistungsfähigkeit .....	10
4.2.4	Prüfung des Lesens und Rechtschreibens durch Schulleistungstests .....	10
4.3	Zur Bestimmung der Validität von HASE .....	11
4.3.1	Kennwerte für die prognostische Validität .....	11
4.3.2	Dichotomisierung des Prädiktors: Zur Bestimmung von Risikokindern .....	12
4.3.3	Dichotomisierung des Kriteriums: Problemkinder mit auffälligen Lese- Rechtschreibleistungen in der Schule .....	14
5	<b>Ergebnisse</b> .....	15
5.1	Prognostische Validität .....	15
5.1.1	Bestimmung des Risikos aufgrund der vorläufigen Grenzwerte .....	15
5.1.2	Bestimmung des Risikos aufgrund der Altersnorm .....	15
5.1.3	Bestimmung des Risikos aufgrund empirisch definierter Grenzwerte und Leistungsprofile .....	16
5.1.4	Bestimmung des Risikos aufgrund der kritischen Werte .....	18
5.2	Befunde zur Konstruktvalidität von HASE .....	20
5.2.1	<i>HASE</i> -Profil und Rechtschreibleistung .....	20
5.2.2	Leseleistungen .....	20
5.2.3	Klassifikation nach <i>HASE</i> und Testleistungen unter Kontrolle der Intelligenz .....	21
5.2.4	Klassifikation nach <i>HASE</i> und Deutschkenntnisse .....	21
5.2.5	Schulische und außerschulische Förderung .....	23
5.2.6	Zusammenhänge zwischen den Schriftsprachleistungen und den <i>HASE</i> -Aufgaben sowie der Klassifikation nach <i>HASE</i> .....	23
6	<b>Wertende Zusammenfassung: Prognostische Validität, kritische Werte und Leistungsprofile</b> .....	25
6.1	<i>HASE</i> ist ausreichend prognostisch valide .....	25
6.2	Die kritischen Werte für die <i>HASE</i> -Aufgaben .....	26
6.3	Zur Notwendigkeit einer instruktionsgemäßen Durchführung .....	27
7	<b>Ausblick: Zum Einsatz von <i>HASE</i> bei den Einschulungsuntersuchungen in Baden- Württemberg seit 2003</b> .....	27
8	<b>Literatur</b> .....	29
9	<b>Anhang</b> .....	30
9.1	Rohwerte-Verteilungen der <i>HASE</i> -Aufgaben .....	30
9.2	Fragebogen für Lehrerinnen und Lehrer .....	33

## 1 Einleitung

Die Forschungen der letzten Jahrzehnte haben gezeigt, dass Schwierigkeiten beim Schriftspracherwerb in der Schule bereits im Vorschulalter wirkungsvoll vorgebeugt werden kann. So können beispielsweise die in Dänemark entwickelten und evaluierten Programme (s. Lundberg, Frost & Petersen, 1988) zum Training der phonologischen Bewusstheit als einer Vorläuferfertigkeit für den Schriftspracherwerb das Risiko für nachfolgende Lese- und Rechtschreibschwierigkeiten bei Risikokindern bedeutsam mindern. In Deutschland konnten diese positiven Effekte durch die Trainingsstudien der Arbeitsgruppe um Wolfgang Schneider bestätigt werden (zsf. Weber, P. Marx & Schneider, 2007). Ein frühzeitiges Erkennen von Schwächen in bestimmten Vorläuferfertigkeiten oder Fähigkeiten kann also dazu führen, die Ausbildung von Lese- und Rechtschreibschwierigkeiten durch entsprechende Fördermaßnahmen zu verhindern oder zumindest abzumildern. Mit dem *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC)*, Jansen, Mannhaupt, H. Marx & Skowronek, 1999) steht seit Ende der 1990er Jahre auch ein Verfahren zur Verfügung, mit dem solche Risikokinder aufgefunden werden können, wobei die von den Autoren berichtete gute prognostische Validität in einer neuen Studie (P. Marx & Weber, 2006) nicht bestätigt werden konnte.

Die Erkenntnis, dass Schriftspracherwerbsschwierigkeiten bereits im Vorschulalter bei frühzeitiger Intervention vorgebeugt werden kann, führte 1997 eine Arbeitsgruppe im Gesundheitsamt des Rhein-Neckar-Kreises dazu, sich zum Ziel zu setzen, diagnostische Möglichkeiten für eine Früherkennung von Risikokindern zu entwickeln. Entsprechende diagnostische Verfahren sollten dann in die obligatorische Einschulungsuntersuchung (ESU) integriert werden. Diese in Baden-Württemberg mit allen Kindern eines Jahrganges durchgeführte ESU bietet die Möglichkeit, auch Kinder mit einem Schriftspracherwerbsrisiko auffinden zu können. Bedingung für ein solches Verfahren war eine zeitökonomische und objektive Durchführung (10 bis maximal 15 Minuten Dauer), gleichzeitig sollte es jedoch auch den Gütekriterien der Reliabilität und der Validität genügen, d. h. insbesondere ausreichende prognostische Validität aufweisen. Zur Bestimmung der prognostischen Validität sind Längsschnittstudien erforderlich.

In einer ersten Längsschnittstudie (Pilotversuch; s. Schöler, 2001) wurden in Heidelberg 208 Kinder bis zum Ende des zweiten Schuljahrs in ihrem Schriftspracherwerb beobachtet, die in der Einschulungsuntersuchung 1998 mit einer ersten Version eines Screenings zur Früherkennung von Risikokindern untersucht worden waren. Die Ergebnisse waren ermutigend, die Korrelationen zwischen den Prädiktoren (Leistungen in diesem Screening) und den Kriterien (den Schulleistungen) waren bedeutsam und variierten zwischen  $r = .29$  und  $r = .33$ . Das Gesundheitsamt der Stadt Mannheim führte dieses Screening im Jahr 2002 flächendeckend durch (Schäfer, Schöler, Roos, Grün-Nolz & Engler-Thümmel, 2003; Schöler, Roos, Schäfer, Dreßler, Grün-Nolz & Engler-Thümmel, 2002), obwohl aufgrund der Analysen der Pilotstudie zwischenzeitlich eine neue, inhaltlich und methodisch veränderte Version des Screenings, nämlich das *Heidelberger Auditive Screening in der Einschulungsuntersuchung (HASE)*, Brunner & Schöler, 2001/02) entwickelt worden war.

Die Kinder des Mannheimer Einschulungsjahrgangs 2003 (Schöler, Dutzi, Roos, Schäfer, Grün-Nolz & Engler-Thümmel, 2004) wurden dann erstmals mit *HASE* untersucht, das seither auch in einer Reihe anderer Gesundheitsamtsbezirke in Baden-Württemberg und über die Landesgrenzen hinaus Anwendung findet. Dieser erste, vollständig mit *HASE* untersuchte Einschulungsjahrgang 2003 dient der Bestimmung der prognostischen Validität des Screenings *HASE*.

## 2 HASE – das Heidelberger Auditive Screening in der Einschulungsdiagnostik

### 2.1 Die vier Aufgaben

Vor dem Hintergrund ihrer differenziellen Validität (Schöler, 1999) wurden die drei Aufgaben *Nachsprechen von Sätzen (NS)*, *Wiedergeben von Zahlen-Folgen (WZ)* und *Nachsprechen von Kunstwörtern (NK)* in das *Heidelberger Auditive Screening in der Einschulungsuntersuchung (HASE)*; Brunner & Schöler, 2001/02) einbezogen. Die vierte Aufgabe, das *Erkennen von Wortfamilien (EW)* ergänzt die Aufgaben zur Erfassung von auditiven Informationsverarbeitungsleistungen (deren Störung als zugrunde liegend für eine Reihe von Sprach- und Schriftspracherwerbsproblemen gelten) um die Erfassung der Fähigkeit, „vom Klang eines Wortes zu abstrahieren und trotz ähnlicher Klanggestalt nach semantischer Zusammengehörigkeit zu kategorisieren“ (Brunner & Schöler, 2001/02, S. 5). *EW* wurde aus dem *Heidelberger Vorschulscreening (HVS)*; Brunner et al., 2002) übernommen.

#### (1) *NS – Nachsprechen von Sätzen*

*NS* besteht aus zehn Sätzen zunehmender grammatischer Komplexität, von denen jeweils zwei Sätze Varianten eines Strukturtyps sind (z. B. Aufgabe 1a „Tina singt“; Aufgabe 1b „Peter rennt“; Aufgabe 5a „Der rote Stift liegt auf dem Sessel neben dem Heft“; Aufgabe 5b „Die große Lampe hängt über dem Tisch im Wohnzimmer“). Kann das Kind Variante a eines Strukturtyps korrekt reproduzieren, dann wird Variante b übersprungen, es erfolgt sofort die Vorgabe des nächst schwierigeren Items a.

#### (2) *WZ – Wiedergabe von Zahlen-Folgen*

Das Kind muss eine Zahlenfolge in der korrekten Reihenfolge reproduzieren. Die Anzahl der Zahlen pro Folge steigt von zwei bis sechs an, wobei nur einsilbige Zahlen zwischen „1“ und „10“ verwendet werden. Insgesamt können 10 Items vorgegeben werden, da pro Zahlenfolge zwei Items vorliegen. Auch hier wird die b-Version nur dann vorgegeben, wenn die a-Version nicht korrekt reproduziert wurde.

#### (3) *EW – Erkennen von Wortfamilien*

Das Kind muss erkennen, welche von drei ähnlich klingenden Wörtern (z. B. „Haus – Hans – Häuser“ oder „kommen – Koffer – kommt“) entsprechend dem gemeinsamen Wortstamm zusammengehören bzw. welches der drei Wörter nicht zur Wortfamilie gehört. Die Einführung erfolgt mit visuellen Beispielen, da somit die Aufforderung „Welches gehört nicht dazu“ besser verdeutlicht werden kann.

#### (4) *NK – Nachsprechen von Kunstwörtern sowie eines Zauberwortes*

Dem Kind werden neun mehrsilbige Wörter vorgesprochen, die es unmittelbar nachsprechen soll. Die Silbenzahl pro Wort nimmt hierbei zu. Vorgegeben werden zwei-, drei- und viersilbige Wörter (z. B. „mali“, „lakoma“, „pukawore“). Abschließend wird ein zehntes, ein sog. Zauberwort („abrakadabra“) vorgegeben, mit dem sich das Kind zum Testende eine kleine Belohnung hervorzaubert.

**Zur internen Validität.** Die Interkorrelationen zwischen den vier Aufgaben sind substantiell und alle statistisch bedeutsam. Die Koeffizienten variieren zwischen  $r = .26$  und  $r = .46$  und erreichen somit nur geringe bis mittlere Höhe (s. Tab. 1). Daher ist davon auszugehen, dass die vier Aufgaben ein unterschiedliches Leistungsspektrum erfassen.

*Tabelle 1* Interkorrelationen zwischen den *HASE*-Aufgaben (Tab. 11 in Schöler & Schäfer, 2004, S. 17)

		WZ	EW	NK
NS	r	<b>.37</b>	<b>.46</b>	<b>.38</b>
	N	6 593	3 321	6 596
WZ	r		<b>.26</b>	<b>.43</b>
	N		3 315	6 484
EW	r			<b>.30</b>
	N			3 303

## 2.2 Zur Durchführung und Auswertung

Die Vorgabe der einzelnen Items der vier Aufgaben erfolgt entweder computergestützt (dann werden sowohl die Instruktionen per PC angeboten, als auch die einzelnen Items über Lautsprecher vorgegeben, die Auswertung erfolgt automatisch und die Ergebnisse können sofort auf dem Bildschirm angezeigt werden und in einem Profilbogen ausgedruckt werden) oder die einzelnen Items werden über ein CD-Abspielgerät vorgegeben. Eine standardisierte Vorgabe der Items der beiden Aufgaben *WZ* und *NK* per CD ist unbedingt erforderlich, wie der Vergleich einer mündlichen mit einer Vorgabe per CD zeigt (s. Tab. 2).

*Tabelle 2* Vergleich der Leistungen bei den vier *HASE*-Aufgaben in Abhängigkeit von der Vorgabe der Items mit oder ohne CD (Daten aus Einschulungsuntersuchung 2003 des Landkreises Biberach; s. Tab. 10 in Schöler & Schäfer, 2004, S. 16)

Aufgabe (Itemzahl)	CD- Vorgabe	N	M	SD	t	p
NS (10)	nein	602	6.87	2.24	1.13	.26
	ja	863	7.00	2.18		
WZ (10)	nein	599	5.03	1.71	6.86	< .001
	ja	861	4.44	1.56		
EW (8)	nein	594	4.21	2.20	1.02	.31
	ja	834	4.32	2.08		
NK (10)	nein	601	7.63	1.82	4.21	< .001
	ja	860	7.20	1.92		

Für die einzelnen Aufgaben werden Risikowerte in Abhängigkeit vom Lebensalter (fünf oder sechs Jahre) angegeben. Sie sind zunächst als Schätzwerte zu verstehen, die sich aus Vergleichsuntersuchungen mit spracherwerbsauffälligen und -unauffälligen Kindern ergeben haben (Schöler, 1999). In der Handanweisung werden sie auch als vorläufig bewertet, da die Ergebnisse der hier berichteten Längsschnittstudie zur Bestimmung der prognostischen Validität des Screenings noch nicht vorlagen, die erst eine genaue Bestimmung der Risikowerte erlauben (s. u.). Zwischenzeitlich wurden Normen in Halbjahresschritten erstellt (Schöler & Schäfer, 2004).

## 3 Fragestellung

*HASE* ist als Screening konzipiert, die zuverlässige Vorhersage eines Risikokindes ist eine zentrale Anforderung. Korrelationen als Maßstab für die Güte der prognostischen Validität sind daher unzureichend. Der Fehler, ein Risikokind zu übersehen, muss möglichst vermieden werden. Der Fehler, dass zu viele Kinder „herausgesiebt“ werden, auch solche, die einer Förderung eigentlich nicht bedürfen, ist demgegenüber eher vernachlässigbar.

Zur Bestimmung der prognostischen Validität von *HASE* muss der Schriftspracherwerb längsschnittlich untersucht werden. Nur so kann geprüft werden, ob aufgrund der Leistung in

*HASE* eine Vorhersage in Hinblick auf Schriftspracherwerbsstörungen in der Grundschule auf individueller Ebene möglich ist. Die bisher vorläufigen Risikowerte sollen aufgrund der empirischen Befunde adjustiert werden. Neben einer Bestimmung der klassifikatorischen Vorhersagegüte sollen auch Aspekte der Konstruktvalidität geprüft werden.

## 4 Methode

### 4.1 Untersuchungsgruppe

An der Einschulungsuntersuchung 2003 (ESU 2003) haben knapp 3 000 Kinder teilgenommen. *HASE*-Daten liegen aber nur von 1 562 Kindern vor. Die Anzahl der Kinder verminderte sich aus verschiedenen Gründen: (1) Nicht alle Kinder, die an der Einschulungsuntersuchung teilnahmen, wurden auch mit *HASE* untersucht. (2) Bei zahlreichen Fällen war eine eindeutige Zuordnung der Einschulungsdaten des Gesundheitsamtes zu den späteren Schulleistungsbeurteilungen nicht möglich (z. B. weil Namen fehlten oder Kennnummern doppelt vergeben worden waren). (3) Die Einwilligung der Eltern zur Validierungsuntersuchung konnte erst eingeholt werden, als die Kinder bereits die 1. Klasse besuchten. Nicht alle Eltern haben diese Einverständniserklärungen auch zurückgegeben. Ein weiterer Versuch, Kinder zur Teilnahme zu gewinnen, deren Eltern auf die Bitte um Einverständnis im ersten Schuljahr nicht reagiert hatten, brachte keine nennenswerte Erweiterung der Stichprobe. (4) Vier Grundschulen lehnten eine Beteiligung ab. Kinder dieser Schulen konnten daher nicht in die Untersuchung einbezogen werden. (5) Von den identifizierbaren Kindern konnten nicht alle bis ans Ende der 3. Klasse begleitet werden (z. B. wegen Umzug, Klassenwiederholung oder Wechsel auf Förderschulen).

In der 1. Klasse konnten insgesamt 1 264 Kinder in die Studie einbezogen werden. Am Ende der 3. Klasse wurden noch 1 079 Kinder aus 90 Klassen in 30 Mannheimer Grundschulen untersucht (vgl. Tab. 3). Der Drop-out liegt demnach bei etwa 30 %.

*Tabelle 3* Anzahl der Kinder zu den einzelnen Erhebungszeitpunkten

	ESU 2003	Klasse 1	Klasse 2	Klasse 3
N	1 562	1 264	1 185	1 079
%	100	80.9	75.9	69.1

#### ***Zur Repräsentativität der Untersuchungsgruppe***

Mehr als die Hälfte (57.5 %), nämlich 1 172 der 2 039 Eltern der deutschen Kinder, die an der ESU 2003 teilnahmen, gaben Rückmeldungen: Davon erteilten 86.8 % ihr Einverständnis, 13.2 % verweigerten es. Von den Eltern der 875 ausländischen Kinder liegt von weniger als der Hälfte (46.7 %) eine Rückmeldung vor: 68.2 % davon erteilten, 31.8 % verweigerten die Teilnahme ihres Kindes an der Studie. Der Anteil der ausländischen Eltern, die ihr Einverständnis nicht erteilten, liegt damit fast dreimal so hoch wie bei den deutschen Eltern. Die ausländischen Eltern gaben zwar weniger häufig als deutsche Eltern Rückmeldung, sie waren jedoch vor allen Dingen seltener bereit, ihre Kinder an der Studie teilnehmen zu lassen. Dies verdeutlichen auch die folgenden Zahlen: Bei der ESU 2003 liegt der Anteil ausländischer Kinder bei 30.2 % und damit um zehn Prozentpunkte höher als in der Untersuchungsgruppe (20.8 %). Da die Informationen an die Eltern in Türkisch, Russisch oder Italienisch gegeben wurden, ist diese bedeutsam höhere Verweigerungsquote bei den ausländischen Eltern nicht auf mangelnde Deutschkenntnisse zurückzuführen.

## 4.2 Durchführung der Untersuchungen

### 4.2.1 Durchführung von *HASE*

Die Einschulungsuntersuchungen wurden zwischen Juli 2002 und Juni 2003 vom Bereich Gesundheit der Stadt Mannheim durchgeführt.<sup>1</sup> Diese enorme Altersspanne, die durch die fast einjährigen Untersuchungen gegeben ist, könnte eine Minderung der Validität bedingen, da sich gerade in der mittleren Kindheit deutliche Entwicklungsveränderungen innerhalb eines Jahres beobachten lassen. Um einen solchen Entwicklungseffekt bezüglich möglicher Leistungsveränderungen bei den vier *HASE*-Aufgaben über den Untersuchungszeitraum zu überprüfen, wurde der Untersuchungszeitraum grob nach Kalenderjahren aufgeteilt. 230 Kinder wurden zwischen Juli und Dezember 2002 untersucht, 1 322 zwischen Januar und Juni 2003 (bei zehn Kindern fehlen die Angaben zum Untersuchungsjahr). Beim Vergleich der Leistungen zeigt sich ein Entwicklungs- bzw. Alterseffekt nur bei *EW*: Die Leistungen der in 2003 untersuchten Kinder sind bedeutsam besser als diejenigen der in 2002 untersuchten Kindern (vgl. Tab. 4). Bei den Aufgaben *NS* und *NK* sind die Leistungen in beiden Untersuchungszeiträumen vergleichbar.

Tabelle 4 Leistungen bei den *HASE*-Aufgaben in den Jahren 2002 und 2003

	2002		2003		t	p
	M	SD	M	SD		
NS	6.97	2.42	7.04	2.25	.39	.70
WZ	5.94	1.82	4.88	1.49	9.62	< .001
EW	5.06	2.32	5.71	1.65	3.65	< .001
NK	8.35	1.67	8.29	1.46	.57	.57

Auffällig ist allerdings der Leistungsabfall bei *WZ*: Jüngere (im Jahr 2002 untersuchte) Kinder verfügen über eine bedeutsam höhere Zahlenspanne als ältere (im Jahr 2003 untersuchte) Kinder. Dies ist aber sehr unwahrscheinlich und lässt nur einen Schluss zu: Die Vorgabe der Zahlenfolgen erfolgte im Jahre 2002 und vermutlich auch in den ersten Monaten des Jahres 2003 nicht per CD, wie dies die Instruktion zwingend vorschreibt, sondern überwiegend durch die Untersucher/-innen (vgl. auch die deutlich geringeren Mittelwerte in Tab. 2). Eine verbale Vorgabe durch die untersuchenden Personen führt zu Erleichterungen für das Behalten der Zahlenfolgen z. B. durch unvermeidbare Rhythmisierungen und Prosodie.

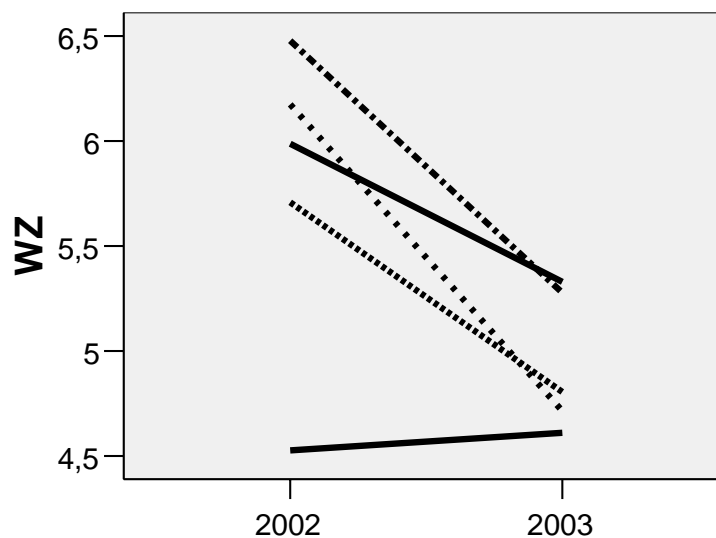
Die Leistungen variieren aber nicht nur in Abhängigkeit vom Untersuchungszeitpunkt sondern auch in Abhängigkeit vom Untersuchungsteam, sodass weitere Zweifel an einer einheitlichen instruktionsgemäßen Vorgabe der Aufgaben angebracht sind (s. Tab. 5). Diese Interpretation wird auch durch die signifikante Wechselwirkung zwischen Untersuchungsteam und Untersuchungsjahr nahe gelegt ( $F = 8.55$ ;  $p < .01$ ; s. Abb. 1). Nur bei einem der sechs Untersuchungsteams wurde offenbar zu jedem Zeitpunkt *WZ* instruktionsgemäß per CD durchgeführt. Bei den anderen Untersuchungsteams scheint erst nach und nach eine instruktionsgemäße Durchführung von *HASE* erfolgt zu sein.

<sup>1</sup> Die Daten wurden dem Projekt EVER zur Verfügung gestellt. Für die Zusammenarbeit möchten wir uns insbesondere bei Herrn Dr. Peter Schäfer bedanken. Unser Dank gilt auch der Leiterin des Bereichs Gesundheit Frau Dr. Holle Engler-Thümmel und Frau Dr. Pia Grün-Nolz, die auch bei auftretenden Problemen die Untersuchungen unterstützt haben.



*Tabelle 5* Leistungen bei den einzelnen HASE-Aufgaben in Abhängigkeit vom Untersuchungsteam

		Untersuchungsteam						F	p
		1	2	3	4	5	6		
NS	M	6.50	7.04	6.43	7.16	7.26	7.56	10.76	< .001
	(SD)	(2.35)	(2.29)	(2.46)	(2.29)	(1.96)	(2.10)		
WZ	M	5.04	5.63	4.88	4.62	5.46	4.94	14.82	< .001
	(SD)	(1.50)	(1.59)	(1.73)	(1.24)	(1.65)	(1.51)		
EW	M	4.76	5.32	5.25	5.88	5.98	5.72	10.53	< .001
	(SD)	(2.24)	(1.74)	(1.81)	(1.68)	(1.54)	(1.78)		
NK	M	7.52	8.29	8.02	8.48	8.50	8.55	11.91	< .001
	(SD)	(1.59)	(1.53)	(1.55)	(1.22)	(1.48)	(1.47)		



*Abbildung 1* Leistungen bei WZ in Abhängigkeit vom Untersuchungs-jahr und Untersuchungsteam

Der mögliche Einwand, dass die Teams in unterschiedlichen Stadtteilen eingesetzt worden sind und die Leistungsunterschiede somit durch verschiedene Sozialmilieus bedingt wären, lässt sich entkräften, da sich systematische Abweichungen in den Einzugsbereichen der Untersuchungsteams nicht bestätigen lassen. Die oben berichteten Befunde einer Leistungsabnahme mit zunehmendem Alter wären auch damit nicht zu erklären.

Nach allem ist davon auszugehen, dass zumindest die Mehrzahl der Untersuchungsteams HASE instruktionswidrig durchgeführt hat – die Verlaufsanalysen der WZ-Leistungen bei den einzelnen Untersuchungsteams lassen vermuten, dass erst ab Mai 2003 von einer instruktionsgemäßen Vorgabe mit CD ausgegangen kann. Bei der Beurteilung der Aufgaben muss also berücksichtigt werden, dass sie vermutlich nicht bei allen Kindern in gleicher Form durchgeführt wurden. Dieses instruktionswidrige Verhalten der Untersuchungsteams mindert die Prüfung der prognostischen Validität in hohem Maße, zumal die Aufgabenanzahl und damit die Varianz bei WZ gering sind.

#### 4.2.2 Schulleistungen und Deutschkenntnisse – Einschätzungen durch die Lehrpersonen<sup>2</sup>

Jeweils am Ende eines Schuljahres wurden die Schulleistungen der an der Längsschnittstudie teilnehmenden Kinder sowie weitere Informationen über die Lehrkräfte erhoben (s. Fragebogen für die Lehrkräfte im Anhang). Da in der 1. Klasse noch keine Schulnoten vergeben werden, sondern die Leistungsbeurteilung in Form eines Berichts erfolgt, wurden die Lehrkräfte in der 1. und 2. Klasse gebeten, die Lese- und Rechtschreibleistung anhand von fünf vorgegebenen Kategorien einzuschätzen (s. Kasten).

<i>Leseleistung</i>	<i>Rechtschreibleistung</i>
<ul style="list-style-type: none"> <li>▪ kann flüssig, sinnbetont lesen</li> <li>▪ kann langsam, selbständig erlesen</li> <li>▪ liest stockend, kann Sinn nicht entnehmen</li> <li>▪ kann Buchstaben zusammenziehen</li> <li>▪ hat unvollständige Buchstabenkenntnisse</li> </ul>	<ul style="list-style-type: none"> <li>▪ schreibt geübte Texte fehlerfrei</li> <li>▪ schreibt geübte Texte mit einigen Fehlern</li> <li>▪ kann lautgetreu schreiben</li> <li>▪ kann nur häufig geübte Wörter schreiben</li> <li>▪ schreibt „Buchstabensalat“</li> </ul>

Am Ende der 3. Klasse wurden die Deutschkenntnisse der Kinder auf folgender dreistufigen Skala eingeschätzt: (1) versteht alles, spricht gut; (2) versteht alles, spricht gebrochen; (3) versteht vieles nicht, spricht schlecht. Ausdrucksfähigkeit, Sprachverständnis, Wortschatz und Grammatik wurden nach Schulnoten bewertet. Am Ende der 2. und 3. Klasse wurde zusätzlich nach schulischen und/oder außerschulischen Fördermaßnahmen gefragt.

#### 4.2.3 Prüfung der kognitiven Leistungsfähigkeit

Am Ende der 3. Klasse wurde die kognitive Leistungsfähigkeit der Kinder durch das Prüfungssystem zur Schul- und Bildungsberatung (*PSB-R 4-6*; Lukesch, Kormann & Mayrhofer, 2002) erfasst. In zehn Untertests werden verschiedene Aspekte der kognitiven Leistungsfähigkeit untersucht, die den Thurstoneschen Primärfaktoren zuordenbar sind: Allgemeinwissen, Reasoning (Zahlenreihen, Buchstabenreihen und Figurale Reihen), Wortflüssigkeit, Gliederungsfähigkeit, Raumvorstellung, Gemeinsamkeiten finden, Zahlenaddition und Wahrnehmungsgeschwindigkeit (Zahlenvergleich). Der Test wurde als Gruppentest im Klassenverband durchgeführt.

#### 4.2.4 Prüfung des Lesens und Rechtschreibens durch Schulleistungstests

Am Ende der 3. Klasse wurden die Lese- und Rechtschreibleistungen ebenfalls durch Gruppentests im Klassenverband erhoben. Die Tests wurden an drei verschiedenen Schultagen zwei bis acht Wochen vor Beginn der Sommerferien von geschulten Hilfskräften durchgeführt.

##### ***Rechtschreiben***

Die Rechtschreibleistung wurde mit dem Diagnostischen Rechtschreibtest für dritte Klassen (*DRT 3*; Müller, 2004) geprüft, bei dem das Kind ein 44 Wörter umfassendes Lückendiktat schreiben muss. Anhand der Kategorisierung des Salzburger Lese- und Rechtschreibtest *SLRT* (Landerl, Wimmer & Moser, 1997) wurden auch die Rechtschreibfehler ausgewertet. Danach werden lautgetreue (Fehl-)Schreibungen, nicht-lautgetreue (Fehl-)Schreibungen sowie Fehler in der Groß-Klein-Schreibung unterschieden. Um die unterschiedlichen Klassenleistungsni-

<sup>2</sup> Für die Verteilung und Rücknahme der Fragebögen danken wir dem Staatlichen Schulamt Mannheim, insbesondere dem Leitenden Schulamtsdirektor *Botho Stern*, der die Untersuchungen immer konstruktiv unterstützt und gefördert hat. Unser Dank gilt auch den vielen Lehrerinnen und Lehrern, die immer sehr kooperativ die entsprechenden Beurteilungen vorgenommen haben.

veaus zu berücksichtigen, wurden diese Fehlerhäufigkeiten zur Gesamtfehlerzahl ins Verhältnis gesetzt.

### **Lesen**

Die Leseleistung wurde mit zwei Tests geprüft: (1) der Würzburger Leise Lese Probe (*WLLP*; Küspert & Schneider, 1998) und (2) Knuspels Leseaufgaben (*Knuspel-L*; H. Marx, 1998).

Bei der *WLLP* sollen die Kinder innerhalb von fünf Minuten möglichst viele der 140 vorgegebenen Wörter lesen, erfasst wird hierbei insbesondere die Lesegeschwindigkeit. Da in der *WLLP* getrennte Normen für Mädchen und Jungen angegeben sind – keine unseres Erachtens sinnvolle Differenzierung – erfolgen die Berechnungen mit den Rohwerten. In der 3. Klasse sollten durchschnittlich ca. 95 Wörter innerhalb der fünf Minuten korrekt gelesen werden.

Der *Knuspel-L* enthält vier Untertests (Hörverstehen, Rekodieren, Dekodieren, Leseverständnis), mit denen Aspekte des Leseverstehens überprüft und aus denen zwei Indikatoren gebildet werden: Score 1 umfasst die Leistungen beim Hörverstehen, Rekodieren und Dekodieren und gilt als ein Indikator für die Vorläuferfertigkeiten von verstehendem Lesen, Score 2 als Indikator für verstehendes Lesen; dazu werden die Leistungen beim Rekodieren, Dekodieren und Leseverständnis zusammengefasst.

## **4.3 Zur Bestimmung der Validität von HASE**

### **4.3.1 Kennwerte für die prognostische Validität**

Ein Screening soll das zuverlässige Auffinden entwicklungsgefährdeter Kinder ermöglichen. Für die Bestimmung der prognostischen Validität eines Screenings sind klassifikatorische Gütekriterien zu ermitteln. Zusammenhangsmaße zwischen Prädiktorwerten und Kriteriumsleistungen sind für die individuelle Risikobestimmung unzureichend. Bei einer klassifikatorischen Vorhersage werden sowohl der Prädiktor als auch das Kriterium dichotomisiert, sodass sich ein Vierfelderschema ergibt, das vier mögliche Ergebnisse einer Vorhersage enthält (s. Tab. 5; s. dazu auch Jansen, 2007).

*Tabelle 6* Vierfelderschema der möglichen Vorhersageergebnisse als Grundlage zur Berechnung der Güteindizes eines Screenings

		Leistung in der Schule (Kriterium)		
		Problem	unauffällig	
Screening (Prädiktor)	Risiko	<b>a</b> richtig positiv	<b>b</b> falsch positiv	Risikokinder
	unauffällig	<b>c</b> falsch negativ	<b>d</b> richtig negativ	Kinder ohne Risiko
		Problemkinder	unauffällige Kinder	

Voraussetzung zur Bestimmung der Güteindizes ist zum einen die Bestimmung der Kriterien für die *Prädiktoren*, d. h. hier die Grenzwerte für die *HASE*-Aufgaben und für das *HASE*-Profil, zum anderen muss das *Kriterium*, d. h. hier die Schriftsprachleistungen, definiert werden. Erst dann können Indizes für die Güte der prognostischen Validität bestimmt werden. Zur Quantifizierung der Möglichkeit des Screenings zwischen Risikokindern und altersgemäß entwickelten Kindern ohne Risiko zu unterscheiden, können verschiedene Güteindizes herangezogen werden: (1) Sensitivität, (2) Spezifität, (3) positive Korrektheit, (4) negative Korrektheit und (5) RAZ-Index.

- *Sensitivität*: Die Sensitivität entspricht dem prozentualen Anteil der Problemkinder in der Schule, die durch das Screening auch als Risikokinder vorhergesagt waren (richtig positiv) [Berechnung:  $(a/(a+c) \times 100)$ ].
- *Spezifität*: Mit Spezifität wird der prozentuale Anteil der Kinder bezeichnet, die mit *HASE* als unauffällig (kein Risiko) vorhergesagt wurden und auch keine Lese-Rechtschreibprobleme in der Schule zeigen (richtig negativ) [Berechnung:  $(d/(b+d) \times 100)$ ].
- *Positive Korrektheit*: Positive Korrektheit bezeichnet den prozentualen Anteil an „Richtig Positiven“ (a) an allen Risikokindern (a+b). Ein Wert von 50 bedeutet, dass 50 % der in *HASE* auffälligen Kinder in der Schule tatsächlich Schwierigkeiten im Lesen und Rechtschreiben entwickeln. Da ein Screening eher zu viele Kinder als Risikokinder einstufen sollte, als auch nur eines zu übersehen, ist dieser Index bei der Bewertung der Güte eines Screenings nur als zweitrangig zu betrachten [Berechnung:  $(a/(a+b) \times 100)$ ].
- *Negative Korrektheit*: Die Negative Korrektheit gibt den prozentualen Anteil der „Richtig Negativen“ (d) an allen Kindern ohne Risiko (c+d) wieder. Ein Wert von 50 bedeutet, dass 50 % der in *HASE* unauffälligen Kinder in der Schule keine Schwierigkeiten mit dem Lesen und Rechtschreiben entwickeln [Berechnung:  $(d/(c+d) \times 100)$ ].
- *RATZ-Index*: Der *relative Steigerungsbetrag gegenüber dem Zufall (RATZ-Index)* setzt die Zufallstrefferquote in Beziehung zur Gesamtrefferquote [Berechnung:  $((a+d)/N \times 100)$ ] und der Maximaltrefferquote [Berechnung: 1 minus absoluter Differenz zwischen Selektionsrate ( $((a-b)/N) \times 100$ ) und Grundrate ( $((a+c)/N) \times 100$ )]. Ein Wert zwischen 34 und 66 deutet auf eine gute, aber unspezifische Klassifikation, ein Wert über 66 auf eine sehr gute und spezifische Klassifikation hin (s. dazu auch H. Marx, 1992).

Neben der prognostischen Validität interessiert auch die Konstruktvalidität. Hierzu gehört die Prüfung, ob die Risikokinder bedeutsam schlechtere Lese- und Rechtschreibleistungen in der Schule erbringen als die Kinder ohne Risiko, und sich diese Leistungsdifferenzen z. B. nicht durch Intelligenzunterschiede erklären lassen.

#### 4.3.2 Dichotomisierung des Prädiktors: Zur Bestimmung von Risikokindern

**Vorüberlegungen.** Bei der Aufgabenauswahl für *HASE* wurde angenommen, dass eine auditive Informationsverarbeitungsstörung grundlegend für die Spracherwerbsdefizite ist (s. dazu Schöler, Fromm & Kany, 1998; Schöler, Fromm & Roos, 2003). Aufgrund der in *HASE* enthaltenen Aufgaben *NS*, *WZ* und *NK* ließen sich spracherwerbsgestörte und sprachunauffällige Kinder gut diskriminieren (Schöler, 1999). Da beim größten Teil der spracherwerbsgestörten Kinder ebenfalls Probleme beim Schriftspracherwerb auftreten, liegt die Annahme nahe, dass der Sprach- und der Schriftspracherwerbsstörung gleiche oder zumindest ähnliche Defizite in der auditiven Informationsverarbeitung zugrunde liegen. Ebenfalls ist anzunehmen, dass eine solche Störung nicht sprachenspezifisch sein kann, sondern dass der Spracherwerb unabhängig von der Muttersprache gestört ist. Ein Risiko für eine Schriftspracherwerbsstörung sollte daher ebenfalls bei Kindern anderer Muttersprachen als Deutsch mit diesen Aufgaben feststellbar sein. Die bisherigen Untersuchungen mit *HASE* (s. Schöler et al., 2002, 2004) zeigen zudem, dass Kinder mit unzureichenden Deutschkenntnissen keine Probleme haben, die beiden Aufgaben *WZ* und *NK* ähnlich gut wie deutschsprachige Kinder zu bearbeiten. Ein Risiko für Schriftspracherwerbsprobleme sollte sich daher durch geringe Leistungen insbesondere bei den Aufgaben *NS*, *WZ* und *NK* zeigen – und zwar unabhängig von den Deutschkenntnissen.

Kinder mit einem Sprachförderbedarf aufgrund unzureichender Deutschkenntnisse werden also die beiden Aufgaben *WZ* und *NK* ähnlich gut bearbeiten können wie spracherwerb-sunauffällige deutsche Kinder, sofern keine Spracherwerbsstörung vorliegt. Bei den sprach-abhängigen Aufgaben *NS* und *EW* sollten sie aber geringe Leistungen zeigen.

**Vorläufige kritische Werte aufgrund von Leistungsvergleichen zwischen spracherwerbsge-störten und sprachunauffälligen Kindern.** In einem ersten Schritt soll die Dichotomisierung des Prädiktors aufgrund der vorläufigen kritischen Werte bei den *HASE*-Aufgaben (s. Tab. 7) erfolgen, die in der Handanweisung angeführt sind.

*Tabelle 7* Vorläufige kritische Punktwerte bei den *HASE*-Aufgaben

	vorläufige kritische Punktwerte	
	5-jährige	6-jährige
NS	0-3	0-4
WZ	0-3	0-4
EW	0-4	0-5
NK	0-4	0-5

Als Risikokind ist definiert, wer bei allen vier *HASE*-Aufgaben im kritischen Punktwertebereich liegt oder in den beiden Aufgaben *NS* und *EW*.

**Kritische Werte aufgrund von Altersnormen.** Bei vielen Screenings werden bei der Bestimmung eines Risikos Altersnormen herangezogen. Damit wird die *Grundrate* automatisch auf den entsprechenden Grenzwert festgelegt – oft auf den Prozentrang 25 (unteres Quartil) oder auf den Prozentrang 16 (eine Standardabweichung unterhalb des Mittelwertes). Eine solche Festlegung des kritischen Wertes zur Dichotomisierung des Prädiktors ist nicht unproblematisch, denn sie impliziert, dass ein wesentlich höherer Anteil an Kindern mit einem Risiko identifiziert werden *muss*, als aufgrund der Prävalenzrate zu erwarten ist. Geht man von einem Anteil von etwa 5-7 % spracherwerbsgestörter Kinder und einer ähnlichen Rate von 5-10 % von lese-rechtschreibgestörten Kindern aus, dann werden bei einem PR 25 deutlich mehr Kinder als Risikokinder identifiziert. Da in einigen Regionen bei der Einschulungsuntersuchung ein Risiko nach dem Prozentrang 25 bestimmt wird, soll trotz der beschriebenen Problematik in einem zweiten Schritt auch die Dichotomisierung des Prädiktors aufgrund der Altersnormwerte (s. Tab. 8) erfolgen.

*Tabelle 8* Kritische Punktwerte aufgrund der Altersnorm (*PR* = 25; s. Schöler & Schäfer, 2004)

	Kritische Punktwerte			
	5;0 – 5; 5 Jahre	5;6 – 5;11 Jahre	6;0 – 6;5 Jahre	6;6 – 6;11 Jahre
NS	0-4	0-5	0-4	0-4
WZ	0-2	0-3	0-3	0-4
EW	0-1	0-3	0-3	0-6
NK	0-5	0-6	0-6	0-3

Als Risikokind ist definiert, wenn mehr als zwei *HASE*-Aufgaben im kritischen Punktwertebereich liegen. Nach Auskunft des Bereichs Gesundheit der Stadt Mannheim liegt dieses Leistungsprofil der Risikobestimmung bei der ESU zugrunde.

***Kritische Werte aufgrund eines empirischen Suchprozesses.*** In einem dritten Schritt soll das Problem der Dichotomisierung des Prädiktors empirisch durch einen Adjustierungsprozess gelöst werden: Durch einen Vergleich der Problemkinder mit unauffälligen Kindern werden kritische Punktwerte und HASE-Leistungsprofile gesucht, mit denen die Klassifikation in Kinder mit und ohne Risiko so erfolgen kann, dass Problemkinder gut vorhergesagt werden können.

#### **4.3.3 Dichotomisierung des Kriteriums: Problemkinder mit auffälligen Lese-Rechtschreibleistungen in der Schule**

Die Differenzierung zwischen Risikokindern und unauffälligen Kindern ist dann sinnvoll, wenn auf individueller Ebene eine Vorhersage der späteren Lese- und Rechtschreibleistungen möglich ist. Dabei soll nicht der genaue Leistungsstand eines Kindes vorhergesagt werden, sondern lediglich, ob die Lese- und Rechtschreibleistungen in einem kritischen, also auffälligen Bereich liegen oder ob sie als angemessen zu werten sind. Für die Bestimmung der Problemkinder in der Schule mit auffälligen Lese-Rechtschreibleistungen werden sowohl die Einschätzungen der Lehrerinnen und Lehrer als auch die Leistungen in den durchgeführten Tests herangezogen.

Als problematisch gelten Kinder,

- die in eine Sonder- bzw. Förderschule eingeschult wurden oder dorthin im ersten Schuljahr wechselten,
- die innerhalb der ersten zwei Schuljahre eine Klasse wiederholt haben,
- deren Lese- oder Schreibleistung in der 1. Klasse mit „hat unvollständige Buchstabenkenntnisse“ bzw. „schreibt Buchstabensalat“ bewertet wurden,

und Kinder mit folgenden Schriftspracheleistungen in der 3. Klasse:

- Schulnote 4 oder schlechter,
- unterdurchschnittliche Lese- und Rechtschreibtestleistungen ( $T < 40$ ).

Schwierigkeiten beim Lesen und Schreiben sollten sich in der 3. Klasse deutlich bemerkbar machen. Häufig werden lese-rechtschreibschwache Kinder erst im Laufe des dritten Schuljahres erkannt, da sie in Abhängigkeit von der Methodik des Anfangsunterrichts zum Teil gut kompensieren können, beispielsweise über Auswendiglernen der zu lernenden Wörter und Texte. Während im Anfangsunterricht je nach Methode möglicherweise nur auf lautgetreues Schreiben Wert gelegt wird, zählt in der 3. Klasse auch die orthographische Korrektheit. Erst in der 3. Klasse können demnach Schwierigkeiten beim Lesen und Schreiben, auch durch die Notengebung, valider bestimmt werden als in den ersten beiden Klassenstufen. Die Noten der 3. Klasse sind zwar abhängig vom Spektrum der jeweiligen Klassenleistungen – und diese unterscheiden sich teilweise dramatisch (s. Treutlein, Zöllner, Roos & Schöler, 2007; Zöllner, Roos & Schöler, 2006), andererseits werden die Lehrkräfte aber vermutlich Kinder mit auffälligem Schriftspracherwerb eher erkennen, als dies standardisierte Schulleistungstests vermögen. Denn bei der Vorgabe von Gruppentests kann eine Reihe situativer Bedingungen das Ergebnis eher verfälschen, als die längerfristige Beobachtung und Bewertung durch die Lehrperson. Bei einer Bewertung mit der Note 4 oder schlechter können Schwierigkeiten beim Lesen und/oder Schreiben angenommen werden. Der Vergleich zwischen solchermaßen auffälligen und unauffälligen Kindern dient dem Auffinden von Leistungsprofilen bei HASE, die zwischen Problemkindern und unauffälligen Kindern diskriminieren können.

## 5 Ergebnisse

### 5.1 Prognostische Validität

#### 5.1.1 Bestimmung des Risikos aufgrund der vorläufigen Grenzwerte

Werden die angegebenen kritischen Punktwerte zur Bestimmung des Prädiktors (Risiko oder kein Risiko) herangezogen, dann fallen die Güteindizes Sensitivität, Positive Korrektheit sowie der RATZ-Index nicht hoch aus (vgl. Tab. 9). Sehr gut sind die Werte für die Spezifität und die Negative Korrektheit, die nahezu durchgehend die maximale Ausprägung erreichen.

*Tabelle 9* Güteindizes der Vorhersage durch die vorläufigen kritischen Punktwerte (*HASE* Handanweisung) in den ersten drei Schuljahren

Klasse		Sensitivität	Spezifität	Positive Korrektheit	Negative Korrektheit	RATZ-Index
1	Lesenote	8	99	53	86	5.8
	Rechtschreibnote	9	99	53	88	7.2
	Einschätzung Lesen	0	98	0	98	-1.8
	Einschätzung Rechtschreiben	5	98	10	97	3.5
2	Deutschnote	6	99	42	89	4.6
	Lesenote	1	90	42	10	-1.0
	Rechtschreibnote	5	99	37	87	3.1
	Einschätzung Lesen	0	98	0	100	-1.8
	Einschätzung Rechtschreiben	14	98	11	99	12.7
3	Deutschnote	8	99	63	87	6.0
	Lesenote	8	99	50	91	6.9
	Rechtschreibnote	5	100	75	77	3.7
	Rechtschreibleistung ( <i>DRT 3</i> )	6	99	53	85	4.5
	Lesegeschwindigkeit ( <i>WLLP</i> )	5	99	53	83	3.6
	Vorläuferfertigkeiten ( <i>Knuspel-L Score 1</i> )	7	99	53	86	5.0
	Leseverständnis ( <i>Knuspel-L Score 2</i> )	5	99	53	88	3.4

#### 5.1.2 Bestimmung des Risikos aufgrund der Altersnorm

Erwartungsgemäß erreichen die Güteindizes auch bei der Bestimmung des Risikos aufgrund der Altersnorm, wenn als kritischer Wert der Prozentrang 25 festgelegt wird, keine zufriedenstellenden Größen bei der Sensitivität, der Positiven Korrektheit und dem RATZ-Index (vgl. Tab. 10). Die Werte für die Spezifität und die Negative Korrektheit sind wiederum überzeugend hoch.

**Tabelle 10** Güteindizes der Vorhersage durch die Altersnorm ( $PR = 25$ ) in den ersten drei Schuljahren

Klasse		Sensitivität	Spezifität	Positive Korrektheit	Negative Korrektheit	RATZ-Index
1	Lesenote	13	98	50	89	10.5
	Rechtschreibnote	15	98	53	90	12.2
	Einschätzung Lesen	8	98	5	99	5.5
	Einschätzung Rechtschreiben	5	98	5	98	2.6
2	Deutschnote	11	98	36	93	9.2
	Lesenote	2	86	59	6	-0.9
	Rechtschreibnote	8	98	32	90	5.2
	Einschätzung Lesen	0	98	0	100	-2.5
	Einschätzung Rechtschreiben	25	98	9	99	23.1
3	Deutschnote	13	99	63	91	10.8
	Lesenote	12	99	44	94	10.3
	Rechtschreibnote	8	99	75	82	5.7
	Rechtschreibleistung ( <i>DRT 3</i> )	8	99	47	89	6.3
	Lesegeschwindigkeit ( <i>WLLP</i> )	6	99	41	87	4.3
	Vorläuferfertigkeiten ( <i>Knuspel-L</i> Score 1)	12	99	78	87	10.4
	Leseverständnis ( <i>Knuspel-L</i> Score 2)	10	99	47	91	8.1

### 5.1.3 Bestimmung des Risikos aufgrund empirisch definierter Grenzwerte und Leistungsprofile

Im Folgenden werden zunächst die *HASE*-Leistungen einzelner Problemgruppen (s. Abschn. 4.3.3) mit den unauffälligen Kindern verglichen, um kritische Werte und Leistungsprofile aufzufinden. Anschließend werden dann die Güteindizes der empirisch erfolgten Bestimmung der Klassifikation dargestellt.

**Förderschulkinder.** Von den Kindern, die im Herbst 2003 eingeschult wurden, wurden 2 % ( $N = 37$ ) in eine Sonder- bzw. Förderschule eingeschult oder wechselten im Laufe der 1. Klasse dorthin. Nur zwei dieser Förderschulkinder wurden im Laufe der ersten zwei Schuljahre auf eine Grundschule umgeschult. Am Ende der 2. Klasse besuchten 2.6 % der Kinder eine Förderschule ( $N = 48$ ). Die Leistungen der Grundschulkinder liegen bei allen *HASE*-Aufgaben signifikant höher als bei den Förderschulkindern (vgl. Tab. 11; s. auch die Rohwertverteilungen in Tab. A1 im Anhang).

**Tabelle 11** Durchschnittliche Leistung der Grund- und der Förderschulkinder bei den *HASE*-Aufgaben

	Grundschule	Förderschule	t	df	p
	M (SD)	M (SD)			
NS	7.07 (2.25)	5.16 (2.51)	5.49	1 559	< .001
WZ	5.05 (1.58)	4.40 (1.58)	2.62	1 554	.01
EW	5.64 (1.75)	3.96 (2.54)	3.35	25.6	< .001
NK	8.35 (1.42)	6.24 (2.41)	5.59	40.8	< .001

**Klassenwiederholung in den ersten beiden Schuljahren.** In den ersten zwei Schuljahren wiederholen 4.1 % der Grundschulkinder eine Klasse. Kinder, die eine Klasse wiederholt ha-



ben, erreichten bei der Einschulungsuntersuchung bedeutsam geringere Leistungen bei den *HASE*-Aufgaben als die anderen Kinder (s. Tab. 12; s. auch die Rohwertverteilungen in Tab. A2 im Anhang).

*Table 12* Leistungen der Kinder, die eine Klasse wiederholt haben, und der Kinder, die nicht wiederholt haben, in den vier *HASE*-Aufgaben

	Grundschule	Förderschule	t	df	p
	M (SD)	M (SD)			
NS	7.07 (2.26)	5.88 (2.24)	4.14	1 559	< .001
WZ	5.07 (1.59)	4.17 (1.35)	4.48	1 554	< .001
EW	5.63 (1.79)	4.39 (1.36)	3.85	1 142	< .001
NK	8.32 (1.48)	7.66 (1.55)	3.52	1 546	< .001

***Auffällige Lese- oder Schreibleistung in der 1. Klasse.*** Grundschulkindern, die am Ende der 1. Klasse von den Lehrkräften als problematisch hinsichtlich ihrer Lese-Rechtschreibleistungen beurteilt werden, erzielten bei der ESU bedeutsam geringere Leistungen bei den *HASE*-Aufgaben als Kinder, die nach dem Urteil der Lehrpersonen als unauffällig bewertet werden (vgl. Tab. 13; s. auch die Rohwertverteilungen in Tab. A3 im Anhang).

*Table 13* Leistungen der unauffälligen und der Kinder mit Lese-Rechtschreibschwierigkeiten bei den *HASE*-Aufgaben

	Grundschule	Förderschule	t	df	p
	M (SD)	M (SD)			
NS	7.53 (2.04)	5.80 (2.40)	8.90	208.1	< .001
WZ	5.18 (1.59)	4.48 (1.49)	5.42	1 234	< .001
EW	5.83 (1.68)	4.69 (1.92)	6.39	946	< .001
NK	8.43 (1.41)	7.75 (1.63)	5.70	1 225	< .001

***Kritische Werte und Leistungsprofile.*** Der Rohwert, bei dem sich Problemkinder nach dem Urteil der Lehrperson eindeutig von unauffälligen Kindern unterscheiden lassen, gilt als kritischer Wert. Bei Überschreiten dieses Grenzwerts ist die Wahrscheinlichkeit für einen unauffälligen Schriftspracherwerb größer als die Wahrscheinlichkeit, ein Problemkind zu werden. Berücksichtigt wird dabei vor allem die Leistungsverteilung bei den Grundschulkindern, bei Uneindeutigkeiten werden die Leistungsverteilungen der Förderschulkinder zur Entscheidung hinzugezogen.

Bei *EW*, *NK* und *WZ* sind die kritischen Werte eindeutig festlegbar, bei *NS* muss die Leistungsverteilung der Förderschulkinder mitbetrachtet werden. Danach gelten die folgenden kritischen Werte:

	Kritischer Wert
NS	0-7
WZ	0-5
EW	0-5
NK	0-8

Diese kritischen Werte liegen höher als die vorläufigen Grenzwerte (*HASE-Manual*) und auch als Prozentrang 25 der Altersnorm (Schöler & Schäfer, 2004). Es gilt hier, Leistungsprofile aufzufinden, mit denen sich Problemkinder in der 3. Klasse vorhersagen lassen. Diese Suche orientiert sich an den oben beschriebenen theoretisch angenommenen Profilen (vgl. Abschn. 4.3.2). Bei den Problemkindern in der 3. Klasse finden sich die folgenden Leistungsprofile gehäuft:

- auffällige Leistungen in allen vier Aufgaben,
- auffällige Leistungen in drei von vier Aufgaben,
- auffällige Leistungen in *NS* und *EW*, nicht jedoch in *NK* und *WZ*,
- auffällige Leistungen in *NS* und *WZ*, nicht jedoch in *EW* und *NK*.

Diese empirisch gefundenen Leistungsprofile bilden im Folgenden die Basis für die Dichotomisierung des Prädiktors.

#### 5.1.4 Bestimmung des Risikos aufgrund der kritischen Werte

Bei Zugrundelegung der empirisch bestimmten Grenzwerte sind die Güteindizes für die prognostische Validität deutlich erhöht (vgl. Tab. 14). Sie bleiben zudem über die Klassenstufen relativ stabil. Die Werte für die negative Korrektheit liegen fast alle dicht am Maximum mit Werten zwischen 87 % und 99 %. Die Werte für die Spezifität zeigen, dass etwa Zweidrittel der Kinder durch den Prädiktor korrekt vorhergesagt werden. Die *RATZ*-Indizes variieren stärker, wobei nahezu alle Werte als eine gute, einige sogar als eine sehr gute Vorhersage gelten. Die Klassifikation erlaubt demnach eine gute Diskriminierung zwischen Problemkindern und unauffälligen Kindern.

*Tabelle 14* Güteindizes aufgrund der empirischen Verteilungen

		Sensitivität	Spezifität	Positive Korrektheit	Negative Korrektheit	RATZ-Index
1. Klasse	Lesenote	74.8	66.0	28.0	93.7	57.9
	Rechtschreibnote	69.1	64.4	22.6	93.3	48.6
	Einschätzung Lesen	75.0	60.0	4.0	99.1	57.8
	Einschätzung Rechtschreiben	75.7	60.0	6.4	98.6	58.6
2. Klasse	Deutschnote	78.3	65.3	22.2	96.0	64.1
	Lesenote	76.5	64.6	18.9	96.2	61.2
	Rechtschreibnote	67.7	64.8	21.8	93.2	46.7
	Einschätzung Lesen	50.0	60.8	0.2	99.8	17.7
	Einschätzung Rechtschreiben	80.0	61.3	2.9	99.5	67.0
3. Klasse	Deutschnote	74.8	67.3	25.2	94.8	59.3
	Lesenote	76.1	65.7	18.5	96.4	61.4
	Rechtschreibnote	64.5	69.5	38.1	87.0	42.5
	Rechtschreibleistung ( <i>DRT 3</i> )	64.9	66.2	24.6	91.7	43.2
	Lesegeschwindigkeit ( <i>WLLP</i> )	58.4	65.8	25.5	88.8	32.7
	Vorläuferfertigkeiten ( <i>Knuspel-L Score 1</i> )	73.0	70.1	36.8	91.6	56.4
	Leseverständnis ( <i>Knuspel-L Score 2</i> )	77.5	68.4	28.7	94.9	63.7

Aufgrund des Ergebnis-Verlaufs der einzelnen Untersuchungsteams können einigermaßen genau die Monate bestimmt werden, ab denen vermutlich die HASE-Vorgabe instruktionsgemäß, d. h. per CD vorgegeben wurde. Bei einem Team konnte alle Daten übernommen werden, bei zwei Teams ab März 2003 und bei den drei anderen ab Mai 2003. Tabelle 15 zeigt beispielhaft die Vierfeldertafeln bei Dichotomisierung des Prädiktors und des Kriteriums Deutschnote (3. Klasse) bzw. Leseverständnis (*Knuspel-L* Score 2).

*Tabelle 15* Verteilung der Kinder nach HASE-Leistung (Risiko vs. kein Risiko) und Kriteriumsleistungen: (a) Deutschnote in der 3. Klasse; (b) Leseverständnis (*Knuspel-L* Score 2)

(a)		Deutschnote 3. Klasse		
		Problem	unauffällig	
HASE	Risiko	<b>73</b> richtig positiv	<b>222</b> falsch positiv	<b>295</b> Risikokinder
	unauffällig	<b>15</b> falsch negativ	<b>367</b> richtig negativ	<b>382</b> Kinder ohne Risiko
		<b>88</b> Problemkinder	<b>589</b> unauffällige Kinder	<b>677</b>
(b)		Leseverständnis ( <i>Knuspel-L</i> Score 2)		
		Problem	unauffällig	
HASE	Risiko	<b>86</b> richtig positiv	<b>194</b> falsch positiv	280 Risikokinder
	unauffällig	<b>16</b> falsch negativ	<b>351</b> richtig negativ	367 Kinder ohne Risiko
		<b>102</b> Problemkinder	<b>545</b> unauffällige Kinder	<b>647</b>

Bei Zugrundelegung dieser „bereinigten“ Daten (s. Tabelle 16) fallen die resultierenden Güteindizes für die Vorhersage der Leistungen in der 3. Klasse deutlich besser aus und sind alleamt als zufriedenstellend bis gut zu bewerten.

*Tabelle 16* Güteindizes bei Zugrundelegung eines reduzierten, „bereinigten“ Datensatzes für die Vorhersage der Leistungen in der 3. Klasse

Leistungen in der 3. Klasse	Sensitivität	Spezifität	Positive Korrektheit	Negative Korrektheit	RATZ-Index
Deutschnote	83	62	25	96	70
Lesenote	82	61	19	97	68
Rechtschreibnote	69	64	36	88	46
Rechtschreibleistung ( <i>DRT 3</i> )	67	60	22	92	42
Lesegeschwindigkeit ( <i>WLLP</i> )	66	61	27	89	39
Vorläuferfertigkeiten ( <i>Knuspel-L</i> Score 1)	82	66	38	93	68
Leseverständnis ( <i>Knuspel-L</i> Score 2)	84	64	31	96	72

## 5.2 Befunde zur Konstruktvalidität von HASE

**Vorbemerkungen.** Im Folgenden wird varianzanalytisch überprüft, ob sich die Testleistungen in Abhängigkeit der Vorhersage nach *HASE* unterscheiden. Leider wurde *HASE* bei der ESU nicht bei allen Kindern vollständig vorgegeben, sondern nur bei Kindern, bei denen das Untersuchungsteam vermutete, dass sie auch schwierigere Aufgaben lösen können. Insbesondere bei Kindern mit unzureichenden Deutschkenntnissen wurde *EW* des Öfteren nicht durchgeführt. Aufgrund ausgelassener Aufgaben kann das entsprechende Leistungsprofil nicht bei allen Kindern erstellt werden. Kinder, bei denen Aufgaben fehlen, können daher auch nicht klassifiziert werden. Sollte die Annahme der Untersuchungsteams zutreffen, dass die Kinder in den nicht durchgeführten Aufgaben auffällige Leistungen erbracht hätten, werden sich weitere Risikokinder unter den aufgrund fehlender Werte nicht klassifizierbaren Kindern befinden. Dies sollte sich dann auch in den Leistungsvergleichen zeigen.

### 5.2.1 HASE-Profil und Rechtschreibleistung

Kinder, die in *HASE* unauffällig sind, erzielen bedeutsam höhere Leistungen im *DRT 3* als Risikokinder und Kinder, die nicht eindeutig klassifizierbar sind (vgl. Tab. 17). Risikokinder und Kinder ohne eindeutige Klassifikation produzieren häufiger nicht-lautgetreue Fehlschreibungen und auch mehr Fehler bei der Groß-Klein-Schreibung als die anderen Kinder. Lautgetreue Schreibweisen werden von unauffälligen und Risikokindern gleich häufig gezeigt (vgl. Tab. 17).

*Tabelle 17* Unterschiede in der Rechtschreibleistung (*DRT 3*) in Abhängigkeit von der Klassifikation nach *HASE*

		N	M (SD)	F	p	$\epsilon^2$
<i>DRT 3</i> ( <i>T</i> -Werte)	unauffällige Kinder	569	53.2 (8.7)	62.91	< .001	.11
	Risikokinder	353	47.2 (8.7)			
	nicht klassifizierbare Kinder	97	46.3 (8.7)			
lautgetreue Fehler (relativer Anteil)	unauffällige Kinder	562	.67 (.17)	12.02	< .001	.02
	Risikokinder	352	.63 (.17)			
	nicht klassifizierbare Kinder	97	.58 (.18)			
nicht-lautgetreue Fehler (relativer Anteil)	unauffällige Kinder	562	.12 (.12)	12.81	< .001	.03
	Risikokinder	352	.16 (.13)			
	nicht klassifizierbare Kinder	97	.16 (.13)			
Groß-Klein- Schreibungs-Fehler (relativer Anteil)	unauffällige Kinder	566	.28 (.17)	17.11	< .001	.03
	Risikokinder	352	.32 (.17)			
	nicht klassifizierbare Kinder	97	.38 (.17)			

### 5.2.2 Leseleistungen

Auch bei den Lesetestleistungen (*WLLP*, *Knuspel-L*) ergeben sich signifikante Leistungsunterschiede in Abhängigkeit von der Klassifikation nach *HASE*. Risikokinder lösen deutlich weniger Aufgaben als die Kinder ohne Risiko. Die geringsten Leistungen erzielen die nicht-klassifizierbaren Kinder (vgl. Tab. 18).

**Tabelle 18** Lesetestleistungen (*WLLP*, *Knuspel-L*) in Abhängigkeit von der Klassifikation nach *HASE*

		N	M (SD)	F	p	$\epsilon^2$
<i>WLLP</i> (Rohwert)	unauffällige Kinder	569	98.0 (20.4)	43.80	< .001	.08
	Risikokinder	353	87.0 (20.4)			
	nicht klassifizierbare Kinder	97	83.0 (20.4)			
Vorläuferfertigkeiten ( <i>Knuspel-L</i> Score 1)	unauffällige Kinder	572	51.5 (8.8)	128.51	< .001	.20
	Risikokinder	353	43.2 (8.8)			
	nicht klassifizierbare Kinder	94	40.6 (8.8)			
Lesefähigkeit ( <i>Knuspel-L</i> Score 2)	unauffällige Kinder	566	52.3 (8.3)	126.13	< .001	.20
	Risikokinder	348	44.4 (8.3)			
	nicht klassifizierbare Kinder	93	42.2 (8.3)			

### 5.2.3 Klassifikation nach *HASE* und Testleistungen unter Kontrolle der Intelligenz

Um zu prüfen, ob die beobachteten Leistungsunterschiede zwischen den nach *HASE* klassifizierten Gruppen lediglich als Intelligenzunterschiede zu interpretieren sind, wurde die Intelligenztestleistung kontrolliert. Auch bei Kontrolle der Intelligenz bleiben die Leistungsunterschiede beim *DRT 3*, der *WLLP* und dem *Knuspel-L* zwischen den drei Gruppen bestehen: Die Risikokinder und die nicht-klassifizierbaren Kinder erreichen bedeutsam geringere Leistungen als die Kinder ohne Risiko (vgl. Tab. 19). Die Leistungen sind sogar annähernd vergleichbar, ob mit oder ohne Kontrolle der Intelligenz (vgl. Tab. 17 und 18).

**Tabelle 19** Testleistungen (*DRT-3*, *WLLP*, *Knuspel-L*) in Abhängigkeit von der Klassifikation nach *HASE* unter Kontrolle der Intelligenz

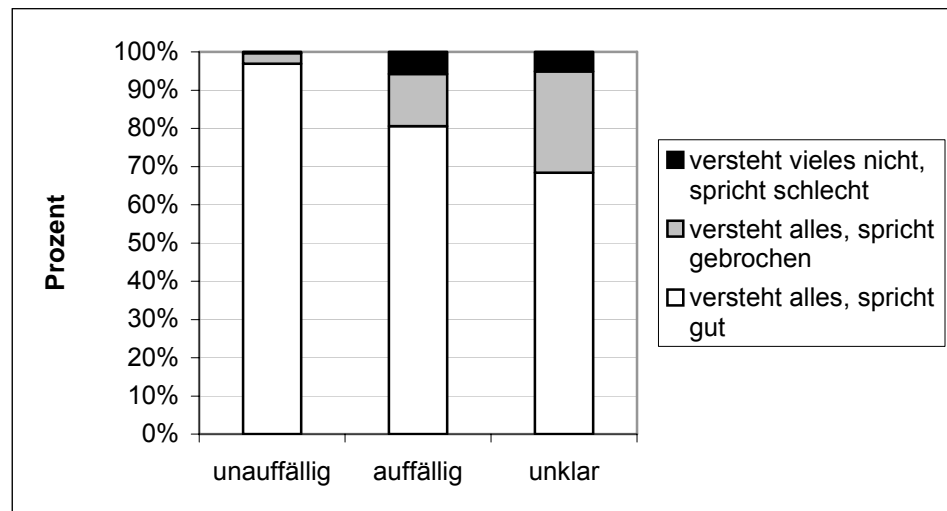
		N	M (SD)	F	p	$\epsilon^2$
<i>DRT 3</i>	unauffällige Kinder	545	52.1 (7.9)	23.17	< .001	.05
	Risikokinder	330	48.9 (7.9)			
	nicht klassifizierbare Kinder	93	47.8 (7.8)			
<i>WLLP</i>	unauffällige Kinder	545	98.2 (18.3)	12.77	< .001	.03
	Risikokinder	330	87.5 (18.4)			
	nicht klassifizierbare Kinder	93	83.9 (18.0)			
Vorläuferfertigkeiten ( <i>Knuspel-L</i> Score 1)	unauffällige Kinder	550	51.4 (8.0)	74.68	< .001	.13
	Risikokinder	328	43.4 (8.0)			
	nicht klassifizierbare Kinder	91	40.8 (7.9)			
Lesefähigkeit ( <i>Knuspel-L</i> Score 2)	unauffällige Kinder	544	51.3 (7.6)	70.89	< .001	.13
	Risikokinder	324	45.8 (7.6)			
	nicht klassifizierbare Kinder	90	43.7 (7.5)			

### 5.2.4 Klassifikation nach *HASE* und Deutschkenntnissen

Die nach *HASE* differenzierten Gruppen haben nach Einschätzung der Lehrpersonen in der 3. Klasse auch unterschiedliche Deutschkenntnisse ( $\chi^2(4) = 108.60$ ;  $p < .001$ ): Kinder, die in *HASE* keine Auffälligkeiten zeigten, werden signifikant häufiger mit „versteht alles, spricht gut“ beurteilt, während bei den Risikokindern und den nicht-klassifizierbaren Kindern häufiger die Kategorien „versteht alles, spricht gebrochen“ oder „versteht vieles nicht, spricht schlecht“ vorkommen (vgl. Abb. 1). Wie die Güteindizes zeigen, ist sogar auf individueller Ebene eine Vorhersage der Deutschkenntnisse in der 3. Klasse anhand der Leistungen in *HASE* recht gut möglich (vgl. Tab. 20).

*Tabelle 20* Güteindizes der Vorhersage nach *HASE* in Bezug auf die Deutschkenntnisse

Deutschkenntnisse (drei Kategorien)	Sensitivität	79.8
	Spezifität	65.9
	Positive Korrektheit	19.5
	Negative Korrektheit	96.9
	Ratz-Index	67.2



*Abbildung 1* Deutschkenntnisse in der 3. Klasse und Klassifikation nach *HASE*

Auch die Beurteilung von Ausdrucksfähigkeit, Sprachverständnis, Wortschatz und Grammatik anhand der Notenskala variiert bedeutsam in Abhängigkeit von der Vorhersage nach *HASE*. Kinder, die in *HASE* unauffällig waren, werden häufiger mit der Note 1 oder 2 bewertet und seltener mit einer schlechteren Note. Demgegenüber werden Risikokinder und Kinder, mit denen nicht alle Aufgaben durchgeführt wurden, häufiger mit 3 oder schlechter bewertet und sehr selten mit 1 oder 2 (vgl. Abb. 2).

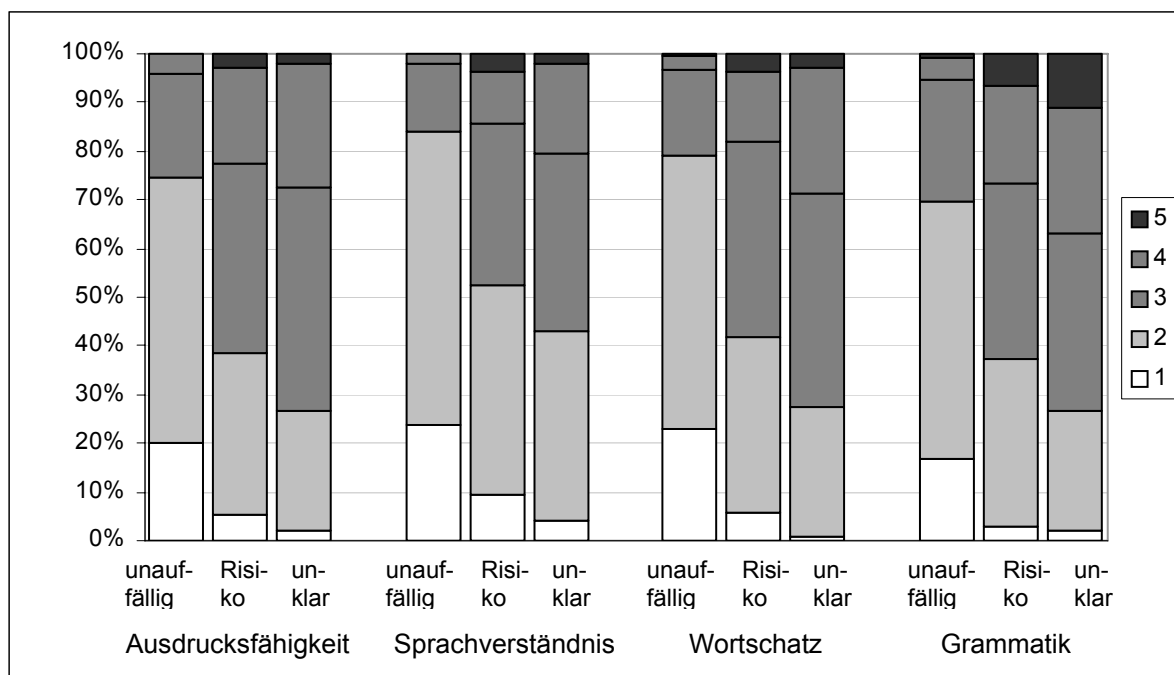


Abbildung 2 Beurteilung der Sprachkenntnisse in der 3. Klasse anhand von Noten in Abhängigkeit von der Vorhersage nach HASE

**5.2.5 Schulische und außerschulische Förderung**

Kinder, die in HASE unauffällige Leistungen zeigten, erhalten in der 2. und 3. Klasse seltener schulische und/oder außerschulische Förderung als Risikokinder und nicht-klassifizierbare Kinder. Abbildung 3 illustriert die Verteilung der schulischen Maßnahmen. Die gleiche Verteilung ergibt sich bei außerschulischen Maßnahmen.

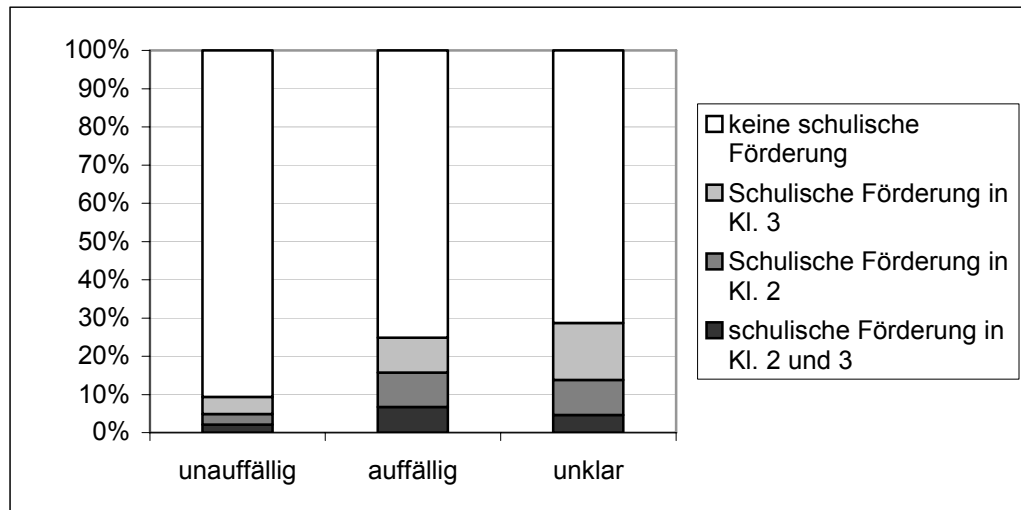


Abbildung 3 Schulische Lese-Rechtschreibförderung in Abhängigkeit von der Vorhersage durch HASE

**5.2.6 Zusammenhänge zwischen den Schriftsprachleistungen und den HASE-Aufgaben sowie der Klassifikation nach HASE**

Die vier HASE-Aufgaben korrelieren mit den diversen Leistungen und Leistungsbewertungen in der 3. Klasse in unterschiedlichem, aber statistisch bedeutsamem Maße (s. Tab. 21). Erwartungsgemäß stehen lediglich die Beurteilungen der Ablenkbarkeit und der Konzentration durch die Lehrkräfte in keiner signifikanten Beziehung zu den HASE-Aufgaben.

Die Aufgabe NS korreliert durchgehend am höchsten mit den verschiedenen Leistungswerten in der 3. Klasse. Ähnliche Korrelationen zu den Schulleistungen sind in der Literatur ansonsten nur bei der Intelligenz bekannt. Bei Auspartialisierung der Intelligenz reduziert sich zwar meist die Höhe der Koeffizienten, die Korrelationen zwischen NS und den schulischen bzw. Testleistungen bleiben aber substanziell und statistisch hoch signifikant (s. Tab. 21).

Abbildung 21 Interkorrelationen zwischen den HASE-Aufgaben und verschiedenen Testleistungen, Leistungs- und Verhaltensbewertungen in der 3. Klasse (links) sowie Partialkorrelationen bei Kontrolle der Intelligenz (fett: statistisch signifikant)

		Korrelationen				Partialkorrelationen			
		NS	WZ	EW	NK	NS	WZ	EW	NK
Deutschnote	r	<b>-.48</b>	<b>-.26</b>	<b>-.32</b>	<b>-.22</b>	<b>-.39</b>	<b>-.14</b>	<b>-.21</b>	<b>-.13</b>
	p	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
	N	1064	1063	847	1054	1003	1002	796	993
Lesenote	r	<b>-.41</b>	<b>-.26</b>	<b>-.29</b>	<b>-.22</b>	<b>-.30</b>	<b>-.14</b>	<b>-.18</b>	<b>-.12</b>
	p	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
	N	1058	1057	841	1048	1003	1002	796	993

- Forts. n. Seite -

- Fortsetzung Tab. 21 -

		Korrelationen				Partialkorrelationen			
		NS	WZ	EW	NK	NS	WZ	EW	NK
Rechtschreibnote	r	<b>-.38</b>	<b>-.22</b>	<b>-.24</b>	<b>-.18</b>	<b>-.27</b>	<b>-.10</b>	<b>-.12</b>	<b>-.08</b>
	p	< .001	< .001	< .001	< .001	< .001	.002	.001	.012
	N	1056	1055	839	1046	1003	1002	796	993
Mathematiknote	r	<b>-.39</b>	<b>-.24</b>	<b>-.28</b>	<b>-.20</b>	<b>-.28</b>	<b>-.11</b>	<b>-.16</b>	<b>-.10</b>
	p	< .001	< .001	< .001	< .001	< .001	< .001	< .001	.002
	N	1065	1064	847	1055	1003	1002	796	993
Ablenkbareit	r	-.03	-.02	.03	-.06	-.01	.00	.06	-.04
	p	.287	.483	.331	.057	.724	.953	.111	.163
	N	1056	1055	842	1046	1003	1002	796	993
Konzentration	r	.01	<b>.06</b>	.05	.04	.00	.05	.03	.03
	p	.699	.049	.187	.228	.958	.108	.328	.369
	N	1057	1056	842	1047	1003	1002	796	993
Sprachfertigkeiten	r	<b>-.45</b>	<b>-.14</b>	<b>-.26</b>	<b>-.13</b>	<b>-.44</b>	<b>-.11</b>	<b>-.24</b>	<b>-.10</b>
	p	< .001	< .001	< .001	< .001	< .001	< .001	< .001	.001
	N	1059	1058	841	1049	1003	1002	796	993
Ausdrucksfähigkeit	r	<b>-.52</b>	<b>-.27</b>	<b>-.33</b>	<b>-.23</b>	<b>-.44</b>	<b>-.17</b>	<b>-.23</b>	<b>-.15</b>
	p	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
	N	1060	1059	842	1050	1003	1002	796	993
Sprachverständnis	r	<b>-.49</b>	<b>-.25</b>	<b>-.32</b>	<b>-.18</b>	<b>-.41</b>	<b>-.15</b>	<b>-.23</b>	<b>-.10</b>
	p	< .001	< .001	< .001	< .001	< .001	< .001	< .001	.002
	N	1060	1059	842	1050	1003	1002	796	993
Wortschatz	r	<b>-.54</b>	<b>-.23</b>	<b>-.37</b>	<b>-.20</b>	<b>-.47</b>	<b>-.13</b>	<b>-.28</b>	<b>-.12</b>
	p	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
	N	1060	1059	842	1050	1003	1002	796	993
Grammatik	r	<b>-.51</b>	<b>-.20</b>	<b>-.35</b>	<b>-.21</b>	<b>-.43</b>	<b>-.08</b>	<b>-.26</b>	<b>-.13</b>
	p	< .001	< .001	< .001	< .001	< .001	.016	< .001	< .001
	N	1060	1059	842	1050	1003	1002	796	993
DRT 3	r	<b>.39</b>	<b>.23</b>	<b>.26</b>	<b>.22</b>	<b>.28</b>	<b>.11</b>	<b>.14</b>	<b>.12</b>
	p	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
	N	1018	1017	807	1008	1001	1001	796	993
Knuspel-L Score 1	r	<b>.55</b>	<b>.29</b>	<b>.36</b>	<b>.26</b>	<b>.47</b>	<b>.18</b>	<b>.26</b>	<b>.17</b>
	p	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
	N	1018	1017	809	1008	1001	1001	796	993
Knuspel-L Score 2	r	<b>.52</b>	<b>.30</b>	<b>.34</b>	<b>.24</b>	<b>.44</b>	<b>.19</b>	<b>.24</b>	<b>.15</b>
	p	< .001	< .001	< .001	< .001	< .001	< .001	< .001	< .001
	N	1006	1005	800	996	990	990	796	990
WLLP	r	<b>.36</b>	<b>.16</b>	<b>.22</b>	<b>.18</b>	<b>.25</b>	.02	<b>.09</b>	<b>.08</b>
	p	< .001	< .001	< .001	< .001	< .001	.554	.013	.017
	N	1018	1017	806	1008	1001	1001	796	993



Zwischen der Klassifikation nach *HASE* und verschiedenen Maßen der Schriftsprachleistungen bestehen ebenfalls mittlere, statistisch hochsignifikante Zusammenhänge. Bei Kontrolle der Intelligenz erhöhen sich die Koeffizienten sogar geringfügig (vgl. Tab. 22).

*Tabelle 22* Biseriale Korrelationen zwischen der Vorhersage nach *HASE* und verschiedenen Außenkriterien (für alle Koeffizienten gilt  $p < .001$ )

Außenkriterium	Klasse 1		Klasse 2		Klasse 3	
	Intelligenz		Intelligenz		Intelligenz	
	nicht kontrolliert	kontrolliert	nicht kontrolliert	kontrolliert	nicht kontrolliert	kontrolliert
Deutschnote	-	-	-.45	-.47	-.45	-.48
Lesenote	-.45	-.47	-.42	-.44	-.42	-.44
Rechtschreibnote	-.40	-.42	-.38	-.41	-.40	-.43
Einschätzung Lesen	-.37	-.39	-.32	-.34	-	-
Einschätzung Rechtschreiben	-.30	-.32	-.29	-.31	-	-
<i>DRT 3</i>	-	-	-	-	.39	.40
<i>WLLP</i>	-	-	-	-	.31	.32
<i>Knuspel-L Score 1</i>	-	-	-	-	.51	.53
<i>Knuspel-L Score 2</i>	-	-	-	-	.51	.53

## 6 Wertende Zusammenfassung: Prognostische Validität, kritische Werte und Leistungsprofile

Mit dem *Heidelberger Auditiven Screenings in der Einschulungsdiagnostik (HASE)* sollen Kinder mit einem Risiko für Schriftspracherwerbsprobleme bereits im Vorschulalter aufgefunden werden. Ziel des Projektes EVER war die Bestimmung der prognostischen Validität dieses Screenings, denn nur bei ausreichender Vorhersagegüte lässt sich der Einsatz eines solchen Verfahrens rechtfertigen. Um dieses Ziel erreichen zu können, ist die längsschnittliche Beobachtung von Kindern erforderlich, die mit dem Screening untersucht worden sind. Da *HASE* in der Einschulungsuntersuchung 2003 in Mannheim flächendeckend eingesetzt wurde, konnte die Leistungsentwicklung der Mannheimer Kinder, insbesondere beim Lesen und Rechtschreiben, in den ersten drei Grundschuljahren beobachtet werden – eine notwendige Voraussetzung, um die prognostische Validität bestimmen zu können.

### 6.1 *HASE* ist ausreichend prognostisch valide

Obwohl die Itemzahl bei den Aufgaben gering ist, und damit ausreichende Reliabilität schwieriger erreichbar ist, und obwohl massive Durchführungsfehler durch die Untersuchungsteams in Mannheim vorgekommen sind, ist *HASE* als Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten geeignet. Seine prognostische Validität ist als ausreichend zu bewerten. Die bisher in *HASE* publizierten vorläufigen kritischen Werte sind, dies zeigen die erhobenen Daten, zu ersetzen. Die Altersnormen erlauben keine zuverlässige Vorhersage über die Entwicklung von Lese-Rechtschreibschwierigkeiten – dies war mit den Altersnormen allerdings auch nicht beabsichtigt.

Bei einem Screening muss der Anteil der als Nicht-Risiko klassifizierten Kinder, die aber später zu Problemfällen werden, möglichst gering gehalten werden. Dieser Klassifikationsfehler ist bei *HASE* ausreichend klein. Weniger relevant ist bei einem Screening die Zahl der Kinder, die als Risikokinder klassifiziert werden, die aber tatsächlich später keine Probleme aufweisen und unauffällig bleiben. Diese Positive Korrektheit ist bei *HASE* gering aus-

geprägt. Dies verdeutlicht umso mehr die Notwendigkeit, dass bei einem auffälligen Befund in *HASE* eine weitere ausführliche Diagnostik unumgänglich ist. Bei der Früherkennung von Risikokindern kommt es in erster Linie darauf an, keine gefährdeten Kinder zu übersehen, sprich falsch-negative Kinder zu vermeiden, was mit *HASE* gelingt.

Alle anderen Güteindizes fallen zufriedenstellend aus. Die Vorhersage eines Risikos nach *HASE* wird sowohl bei den Bewertungen der Lese-Rechtschreibfertigkeiten durch die Lehrerinnen und Lehrer als auch bei den Untersuchungen mit den Schulleistungstests bestätigt. Die Vorhersage aufgrund der Klassifikation nach *HASE* ist ähnlich gut wie die Vorhersage mit dem zehn Monate vor der Einschulung durchgeführten *BISC* (Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten, Jansen, Mannhaupt, H. Marx & Skowronek, 1999), für das am Ende der 2. Klasse RATZ-Indizes zwischen 33 und 65 berichtet werden (H. Marx, Jansen & Skowronek, 2000). Für *HASE* liegen die Werte am Ende der 3. Klasse, also für einen Vorhersagezeitraum von mehr als drei Jahren, zwischen 39 und 72 leicht besser in einem ähnlichen Bereich.

Die anhand der Klassifikation bestimmten Kinder mit und ohne ein Risiko für Lese-Rechtschreibprobleme unterscheiden sich in ihren Lese- und Rechtschreibleistungen bedeutsam. Die Kinder, die wegen fehlender Aufgaben nicht klassifizierbar waren, weisen meist die schwächsten Leistungen auf; dies spricht dafür, dass die Annahme der Untersuchungsteams zutreffend war, dass die Kinder die nicht durchgeführten Aufgaben vermutlich nicht hätten bearbeiten können bzw. bei diesen Aufgaben nur geringe Leistungen gezeigt hätten.

Die Korrelationen zwischen der Vorhersage nach *HASE* und schulischen Leistungen fallen – wie in ähnlichen Verfahren – moderat, aber statistisch bedeutsam aus (beim *BISC* werden Korrelationen zu Leistungen am Ende der 2. Klasse von  $r = .53$  bis  $r = .60$  berichtet). Bei Betrachtung der Beziehungen zwischen den einzelnen *HASE*-Aufgaben und den Sprach- und Schriftspracherwerbsleistungen fallen insbesondere die mittelhohen Beziehungen der Aufgabe *Nachsprechen von Sätzen (NS)* auf. Dies spricht noch einmal deutlich für die diagnostische Eignung des Nachsprechens von Sätzen im Rahmen eines Screenings, bei dem Fertigkeiten und Fähigkeiten, wie Leistungen des Kurz- bzw. Arbeitsgedächtnisses, semantisches und grammatisches Wissen, konfundiert sind.

Zusammenfassend kann *HASE* als ein prognostisch und inhaltlich valides Screening bewertet werden, das zur Frühdiagnostik von Schwierigkeiten beim Lesen und Rechtschreiben tauglich ist. Eine differenzierte und differenzierende Diagnostik kann aber durch ein Screening nicht ersetzt werden.

## 6.2 Die kritischen Werte für die *HASE*-Aufgaben

Die folgende Tabelle 23 enthält die kritischen Werte bei den vier *HASE*-Aufgaben, Tabelle 24 zeigt die Leistungsprofile von Risikokindern.

Tabelle 23 Kritische Werte bei den *HASE*-Aufgaben

Rohwert	NS	WZ	EW	NK	
10	unauffällige Leistung	unauffällige Leistung		unauffällige Leistung	
9					
8	auffällige Leistung	unauffällige Leistung	unauffällige Leistung	auffällige Leistung	
7					
6		auffällige Leistung	auffällige Leistung		auffällige Leistung
5					
4					
3					
2					
1	auffällige Leistung	auffällige Leistung	auffällige Leistung		
0					

Bei den sieben Leistungsprofilen (s. Tab. 24) ist die Aufgabe *Nachsprechen von Sätzen* (NS) sechsmal auffällig (ausgenommen Profil 2). Dies stützt die Annahme, dass mit dieser Aufgabe sehr gut zwischen spracherwerbsgestörten und sprachunauffälligen Kindern diskriminiert werden kann.

**Tabelle 24** Kritische Leistungsprofile zur Bestimmung der Risikokinder (gerasterte Felder: auffällige Leistung, weiße Felder: unauffällige Leistung)

		NS	WZ	EW	NK
Aufgabenprofile von Risikokindern	1				
	2				
	3				
	4				
	5				
	6				
	7				

Nur bei den Profilen 6 und 7 sind lediglich zwei Aufgaben auffällig, ansonsten kann davon ausgegangen werden, dass bei drei auffälligen Aufgaben ein Risiko anzunehmen ist. Aufgabenprofil 6 legt theoretisch einen Deutschförderbedarf nahe, da die beiden sprachabhängigen Aufgaben *NS* und *EW* auffällig, die Aufgaben *WZ* und *NK* unauffällig sind. Letztere sind Indikatoren für die Funktionalität der phonologischen Schleife des Arbeitsgedächtnisses und können bei einer Sprach- oder Schriftspracherwerbsstörung gemindert sein.

### 6.3 Zur Notwendigkeit einer instruktionsgemäßen Durchführung

Die enorme Abhängigkeit der Leistungen von den Untersuchungsteams hat noch einmal verdeutlicht, dass eine standardisierte Vorgabe der Aufgaben, insbesondere der Aufgabe *WZ*, unverzichtbar ist, um die Objektivität und Reliabilität, und damit auch die prognostische Validität des Screenings, zu gewährleisten. Die Leistungsdifferenzen, die durch eine Vorgabe mit oder ohne CD auftreten, sind ausreichend belegt. Gerade bei *WZ* nivelliert diese Leistungsdifferenz den Unterschied zwischen den Leistungen eines Kindes mit einem Risiko und einem unauffälligen Kind. Eine Untersucherin oder ein Untersucher sind außerstande, die Vorgabe verbal so zu gestalten, dass sich keine Erleichterungen durch Rhythmisierungen und Clusterungen ergeben.

## 7 Ausblick: Zum Einsatz von *HASE* bei den Einschulungsuntersuchungen in Baden-Württemberg seit 2003

*HASE* wird seit 2003 in acht Gesundheitsamtsbezirken in Baden-Württemberg im Rahmen der Einschulungsuntersuchung bei ca. 25 000 Kindern durchgeführt (Dr. Schäfer, Fachbereich Gesundheit der Stadt Mannheim, persönl. Mitteilung am 15.02.2007), wobei nur die drei Aufgaben *NS*, *WZ* und *NK* vorgegeben werden. *EW* hat sich bei vielen Kindern als problematisch und als zu schwierig erwiesen, insbesondere bei Kindern mit unzureichenden Deutschkenntnissen ergaben sich Instruktionsschwierigkeiten.

Als kritische Werte bei den einzelnen *HASE*-Aufgaben sind die 10. Perzentile der deutschsprachigen Kinder definiert: Bei *NS* und *NK* entspricht der Prozentrang 10 jeweils dem Rohwert 6, bei *WZ* dem Rohwert 5. Liegt die jeweilige Leistung unter diesen Grenzwerten, gilt sie als auffällig. Erreicht oder übertrifft ein Kind bei allen drei Aufgaben nicht diese Grenzwerte, dann gilt es als Risikokind mit einem Therapiebedarf. Dies gilt auch, wenn nur bei *NS* und *NK* die Grenzwerte nicht erreicht werden. Nach Schäfer wird bei 4 bis 7 % aller Kinder dieses Antwortmuster gefunden, bei dem ein Verdacht auf eine Sprach- bzw. Schriftspracherwerbsstörung besteht. Ist nur die Leistung bei *NS* auffällig, wird ein pädagogischer Förderbedarf angenommen, denn meist wären dann nur die Deutschkenntnisse unzureichend.

In Baden-Württemberg wird – wie ebenfalls in den meisten Bundesländern – die Einschulungsuntersuchung künftig um etwa ein Jahr vorgezogen. Konzeptionen wie „Schulanfang auf neuen Wegen“ und „Schulreifes Kind“ erfordern diese Vorverlegung. Risikokinder sollen frühzeitiger entdeckt und entsprechenden Interventionen zugewiesen werden. Nach Schäfer hat sich in den bisher durchgeführten vorgezogenen Pilotuntersuchungen gezeigt, dass *HASE* (bzw. die drei Aufgaben *NS*, *WZ* und *NK*) auch mit vier- bis fünfjährigen Kindern problemlos durchgeführt werden kann. In einer in Vorbereitung befindlichen Kabinettsvorlage wird daher der Einsatz der drei Aufgaben *NS*, *WZ* und *NK* aus *HASE* auch bei der Neugestaltung der „Einschulungsuntersuchung“ empfohlen. Es bleibt zu hoffen, dass dann die Durchführung instruktionsgemäß erfolgt, damit die im vorliegenden Bericht dargestellte Prognosemöglichkeit des Screenings auch erhalten bleibt.

## 8 Literatur

- Brunner, M. & Schöler, H. (2001/2002). *HASE – Heidelberger Auditives Screening in der Einschulungsuntersuchung*. Wertingen: Westra.
- Jansen, H. (2007). Screenings zur Früherkennung von Lese-Rechtschreibschwierigkeiten. In H. Schöler & A. Welling (Hrsg.), *Handbuch der Sonderpädagogik, Band 1 Sonderpädagogik der Sprache* (S. 618-632). Göttingen: Hogrefe.
- Jansen, H., Mannhaupt, G., Marx, H. & Skowronek, H. (1999). *Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC)*. Göttingen: Hogrefe.
- Küspert, P. & Schneider, W. (1998). *Würzburger Leise Leseprobe (WLLP): Ein Gruppenlesetest für die Grundschule*. Göttingen: Hogrefe.
- Landerl, K., Wimmer, H. & Moser, E. (1997). *Salzburger Lese- und Rechtschreibtest (SLRT)*. Göttingen: Hogrefe.
- Lukesch, H., Kormann, A. & Mayrhofer, S. (2002). *PSB-R 4-6 – Prüfsystem für Schul- und Bildungsberatung für 4. bis 6. Klassen – revidierte Fassung*. Göttingen: Hogrefe.
- Lundberg, I., Frost, J. & Petersen, O. P. (1988). Effects of an extensive program for stimulating phonological awareness in preschool children. *Reading Research Quarterly*, 23, 263-284.
- Marx, H. (1992). Methodische und inhaltliche Argumente für und wider eine frühe Identifikation und Prädiktion von Lese-Rechtschreibschwierigkeiten. *Diagnostica*, 38, 249-268.
- Marx, H. (1998). *Knuspels Leseaufgaben (KNUSPEL-L). Gruppenlesetest für Kinder Ende des ersten bis vierten Schuljahres*. Göttingen: Hogrefe.
- Marx, P. & Weber, J. (2006). Vorschulische Vorhersage von Lese- und Rechtschreibschwierigkeiten. Neue Befunde zur prognostischen Validität des Bielefelder Screenings (BISC). *Zeitschrift für Pädagogische Psychologie*, 20, 251-259.
- Müller, R. (2004). *DRT 3 – Diagnostischer Rechtschreibtest für 3. Klassen*. Göttingen: Hogrefe.
- Schäfer, P., Schöler, H., Roos, J., Grün-Nolz, P. & Engler-Thümmel, H. (2003). Einschulungsuntersuchung 2002 in Mannheim – Sprachentwicklungsstand bei Schulbeginn. *Gesundheitswesen*, 65, 676-682.
- Schöler, H. (2001). *Zur Früherkennung von Schriftspracherwerbsproblemen im Rahmen der Einschulungsuntersuchungen* (Arbeitsberichte aus dem Forschungsprojekt Differentialdiagnostik Nr. 10). Heidelberg: Pädagogische Hochschule, Erziehungs- und Sozialwissenschaftliche Fakultät.
- Schöler, H. & Schäfer, P. (2004). *HASE Heidelberger Auditives Screening in der Einschulungsuntersuchung – Itemanalysen und Normen* (Arbeitsbericht aus dem Forschungsprojekt Differentialdiagnostik, Nr. 17). Heidelberg: Pädagogische Hochschule, Institut für Sonderpädagogik, Abt. Psychologie in sonderpädagogischen Handlungsfeldern.
- Schöler, H., Dutzi, I., Roos, J., Schäfer, P., Grün-Nolz, P. & Engler-Thümmel, H. (2004). *Einschulungsuntersuchungen 2003 in Mannheim* (Arbeitsberichte aus dem Forschungsprojekt Differentialdiagnostik Nr. 16). Heidelberg: Pädagogische Hochschule, Institut für Sonderpädagogik, Abt. Psychologie in sonderpädagogischen Handlungsfeldern.
- Schöler, H., Roos, J., Schäfer, P., Dreßler, A., Grün-Nolz, P. & Engler-Thümmel, H. (2002). *Einschulungsuntersuchungen 2002 in Mannheim* (Arbeitsberichte aus dem Forschungsprojekt Differentialdiagnostik Nr. 13). Heidelberg: Pädagogische Hochschule, Institut für Sonderpädagogik, Abt. Psychologie in sonderpädagogischen Handlungsfeldern.
- Treutlein, A., Zöllner, I., Roos, J. & Schöler, H. (2007). Die Zusammensetzung einer Grundschulklasse als Risikofaktor für den Schriftspracherwerb. *Empirische Pädagogik*, 21, (i. Dr.).
- Weber, J., Marx, P. & Schneider, W. (2007). Die vorschulische Förderung der phonologischen Bewusstheit. In H. Schöler & A. Welling (Hrsg.), *Handbuch der Sonderpädagogik, Band 1 Sonderpädagogik der Sprache* (S. 746-761). Göttingen: Hogrefe.
- Zöllner, I., Roos, J. & Schöler, H. (2006). Einfluss soziokultureller Faktoren auf den Schriftspracherwerb im Grundschulalter. In A. Schröder-Lenzen (Hrsg.), *Risikofaktoren kindlicher Entwicklung. Migration, Leistungsangst und Schulübergang* (S. 45-65). Wiesbaden: Verlag für Sozialwissenschaften.

## 9 Anhang

### 9.1 Rohwerte-Verteilungen der HASE-Aufgaben

Abbildung A1 Rohwerte-Verteilungen bei den HASE-Aufgaben von Grund- und Förderschulkindern

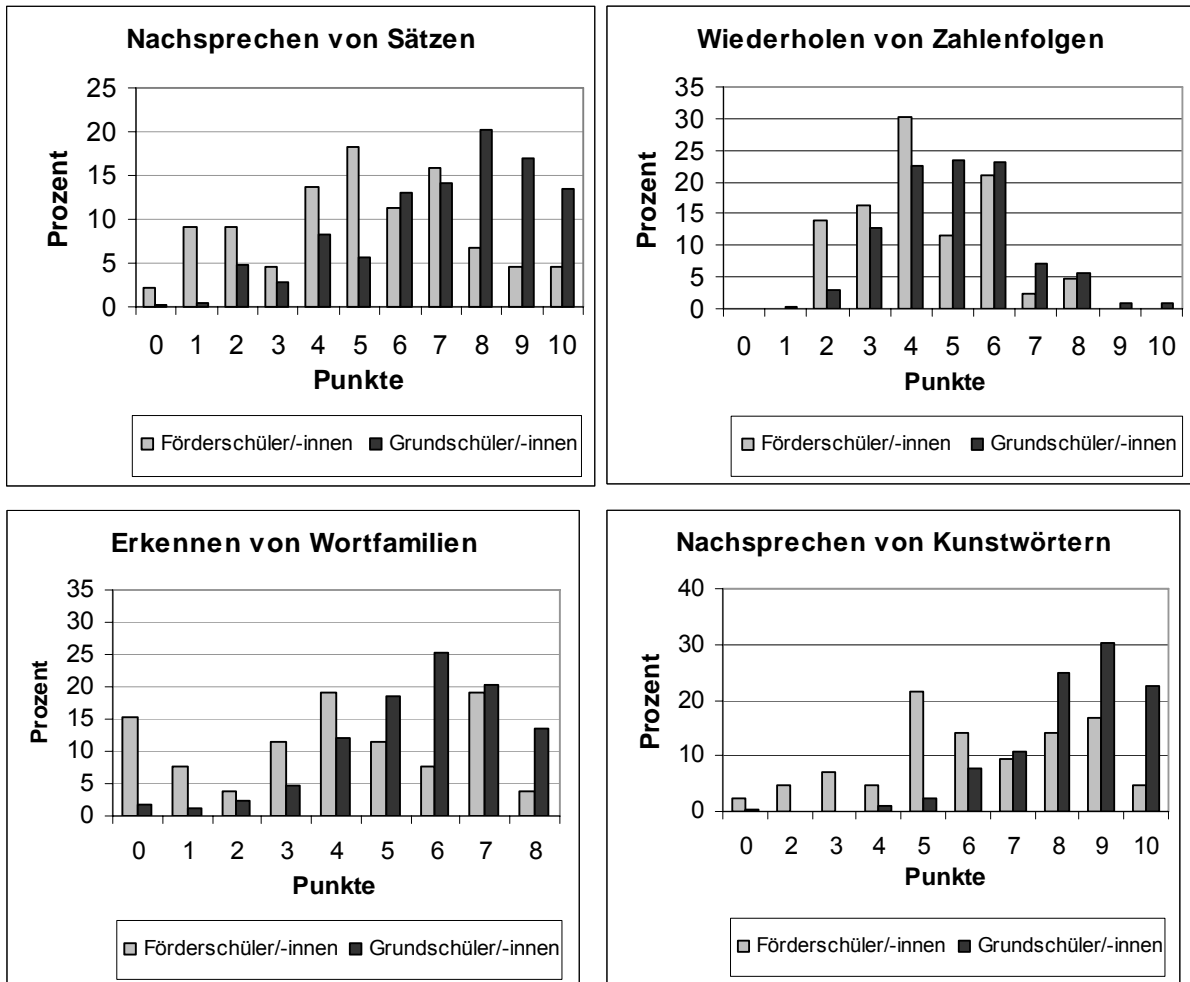


Abbildung A2 Rohwerte-Verteilungen bei den HASE-Aufgaben von Kindern, die eine Klasse wiederholt haben oder nicht

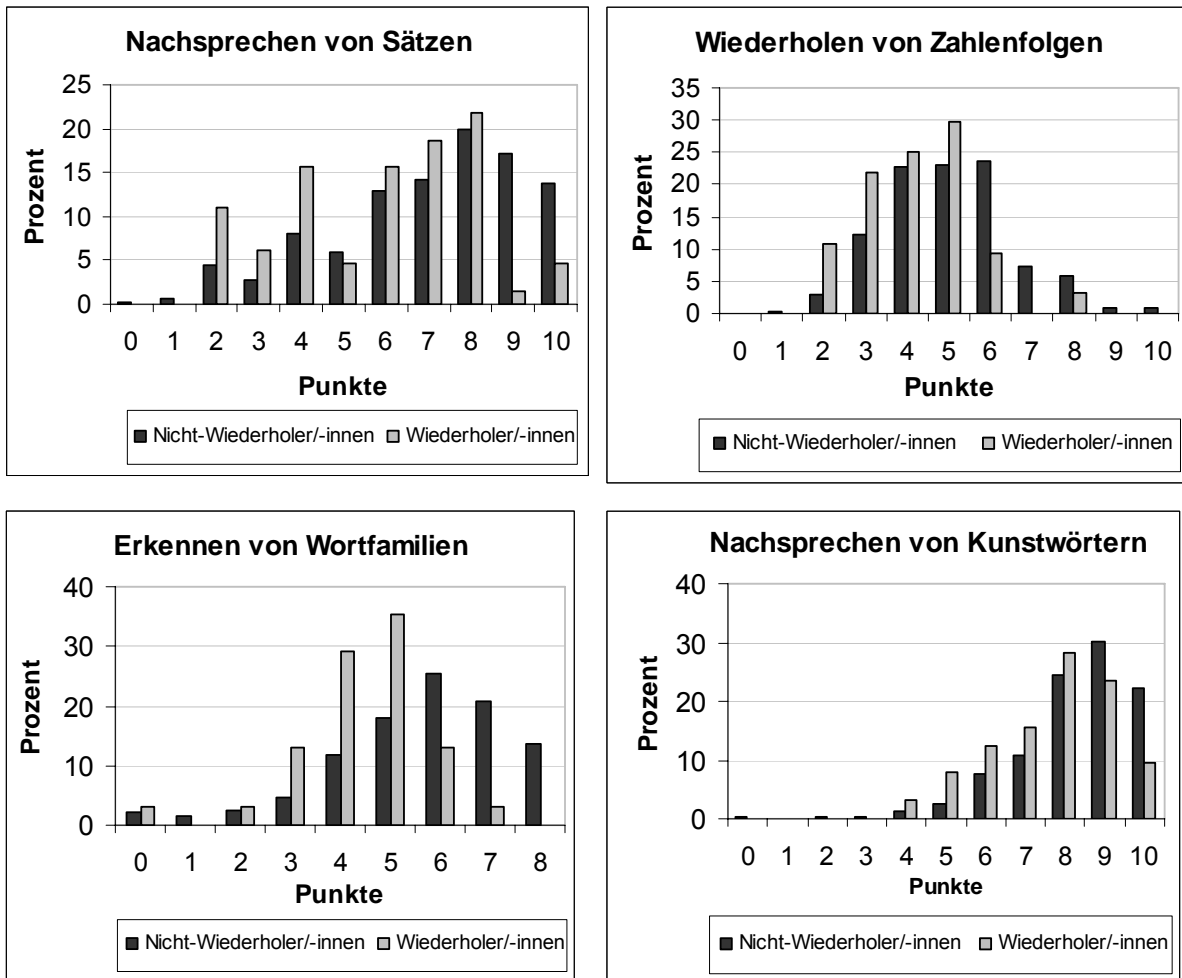
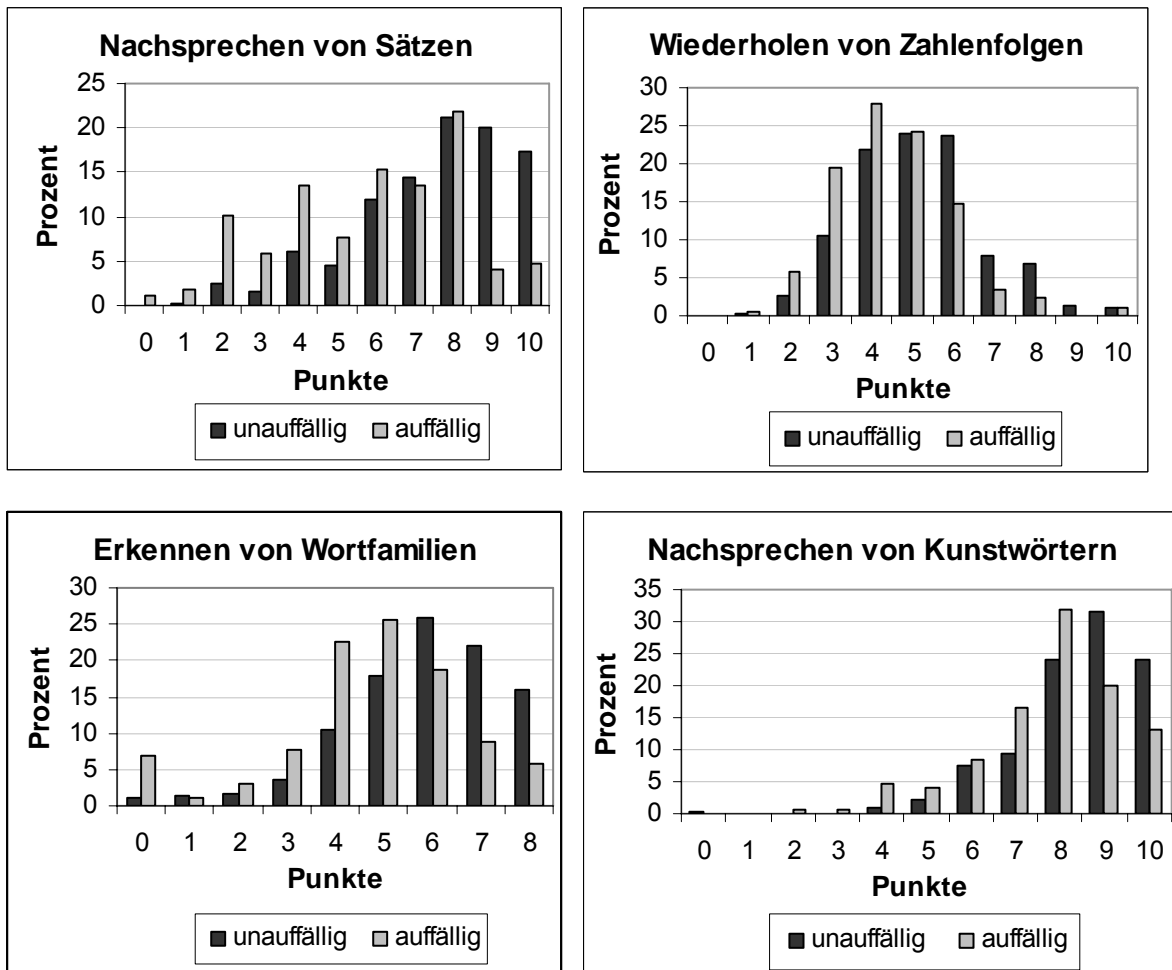


Abbildung A3 Rohwerte-Verteilungen bei den HASE-Aufgaben der unauffälligen und der lese-rechtschreibschwachen Kinder am Ende der 1. Klasse (Bewertung durch die Lehrperson)





**9.2 Fragebogen für Lehrerinnen und Lehrer**



Einschulungsjahrgang **2003**  
**3. Klasse**

Schulleistungsbeurteilung 2006

<b>Name, Vorname</b>	<b>Schule</b>	<b>Klasse</b>
Geburtstag		

**Beurteilung der Leistungen im laufenden Schuljahr**

**Leistungsbeurteilung in der 3. Klasse** (Bitte nur ganze Noten angeben)

Zeugnisnote in Deutsch.....	<input type="checkbox"/>	Lesenote (bitte einschätzen)	<input type="checkbox"/>
Rechtschreibnote (bitte einschätzen)	<input type="checkbox"/>		

Zeugnisnote in Mathematik.....	<input type="checkbox"/>
--------------------------------	--------------------------

**Versetzung am Ende der 3. Klasse**

Das Kind wird in die 4. Klasse  versetzt  nicht versetzt

**Ablenkbarkeit und Konzentration im vergangenen Schuljahr**

Ablenkbarkeit  selten  manchmal  oft  
 Konzentrationsfähigkeit  gering  mittel  hoch

**Förderungsmaßnahmen des Lesens/Rechtschreibens im vergangenen Schuljahr**

durch spezielle schulische Maßnahmen  ja  nein  
 durch außerschulische Maßnahmen  ja  nein  nicht bekannt

**Sprachfertigkeiten im Deutschen**

Bitte nur ein Kästchen ankreuzen!  Versteht alles, spricht gut <input type="checkbox"/> Versteht alles, spricht gebrochen <input type="checkbox"/> Versteht vieles nicht, spricht schlecht <input type="checkbox"/>	Beurteilen Sie bitte anhand von Noten!  Ausdrucksfähigkeit ..... <input type="checkbox"/> Sprachverständnis..... <input type="checkbox"/> Wortschatz..... <input type="checkbox"/> Grammatik..... <input type="checkbox"/>
---	---

**Vielen Dank!**