

**Kreuzvalidierung des Heidelberger
Auditiven Screenings in der
Einschulungsuntersuchung
(HASE)**

**Anke Treutlein
Jeanette Roos
Hermann Schöler**

Abt. Psychologie in sonderpädagogischen Handlungsfeldern
Institut für Sonderpädagogik
Pädagogische Hochschule Heidelberg

Kreuzvalidierung des Heidelberger Auditiven Screenings in der Einschulungsuntersuchung (HASE)

Anke Treutlein¹, Jeanette Roos²
und Hermann Schöler³

Bericht Nr. 26

April 2011



Projekt EVER* (Leitung: Prof. Dr. Jeanette Roos und Prof. Dr. Hermann Schöler)
Pädagogische Hochschule Heidelberg

¹Universität Stuttgart

²Fakultät I - Fach Psychologie

³Fakultät I - Institut für Sonderpädagogik

Abteilung Psychologie in sonderpädagogischen Handlungsfeldern

Keplerstr. 87, D - 69120 Heidelberg - ☎ (06221) 477-426

Email: k40@ix.urz.uni-heidelberg.de

Website: http://www.ph-heidelberg.de/wp/schoeler/Aktuelle_Forschungen.html

*Die längsschnittliche Untersuchung (Projekt EVER) wurde gefördert von
der Dürr-Stiftung Hamburg, der Günter-Reimann-Dubbers-Stiftung Heidelberg
sowie der Pädagogischen Hochschule Heidelberg.

Inhalt

	Zusammenfassung	III
	Abstract	III
1	Einleitung	1
1.1	Das Heidelberger Auditive Screening in der Einschulungsuntersuchung	1
1.2	Fragestellung	3
2	Methode	3
2.1	Untersuchungsgruppe und Durchführung von HASE	3
2.2	Zur Bestimmung der Validität von HASE	4
3	Ergebnisse	6
4	Diskussion	7
5	Literatur	9

Zusammenfassung

Das *Heidelberger Auditive Screening in der Einschulungsuntersuchung* (HASE) gilt als prognostisch und inhaltlich valide für die Früherkennung von Sprach- und Schriftspracherwerbsproblemen. Eine Kreuzvalidierung stand bislang aus. An einer Stichprobe von 453 Kindern wurde die Vorhersagekraft für Schwierigkeiten im Lesen und Rechtschreiben am Ende der 2. Klasse überprüft. Auch bei dieser zweiten Längsschnittstudie erweist sich HASE als zufrieden stellend prognostisch valide.

Abstract*

The *Heidelberger Auditives Screening in der Einschulungsuntersuchung* (HASE, Heidelberg Auditive Screening for the examination before primary school enrolment) is a screening with prognostic and content validity for an early detection of difficulties in language and literacy acquisition. Up to now cross-validity has not been tested. On the basis of a sample of 453 children it was tested, if HASE could predict reading and spelling difficulties at the end of the second grade. In this second longitudinal study, too, HASE has shown satisfactory prognostic validity.

1 Einleitung

Bei etwa jedem 10. Kind sind deutliche Schwierigkeiten beim Lesen- und Rechtschreibenlernen in der Grundschule zu beobachten (Weber & Marx, 2008). Da wesentliche Vorläuferfertigkeiten bereits im Vorschulalter erworben werden und auch wirksame Interventionsmöglichkeiten bereits im letzten Vorschuljahr vorliegen (z. B. Trainingsprogramme zur Förderung der phonologischen Bewusstheit, s. Lundberg, Frost & Petersen, 1988 zum dänischen Trainingsprogramm oder zsf. Weber, P. Marx & Schneider, 2007 zum deutschen Trainingsprogramm) ist ein frühzeitiges Erkennen von Schwächen in bestimmten Vorläuferfertigkeiten oder bestimmten Fähigkeiten sinnvoll, um Lese- und Rechtschreibschwierigkeiten durch entsprechende Fördermaßnahmen vorzubeugen oder sie – bei Vorliegen von legasthenen Störungsformen – zumindest zu mindern.

Eine Möglichkeit der Intervention könnte nun beispielsweise darin bestehen, flächendeckend Förderungen solcher Vorläuferfertigkeiten durchzuführen. In allen Kindertageseinrichtungen würden dann diese Förderprogramme präventiv eingesetzt, eine individuelle Diagnostik würde sich erübrigen. Eine andere Möglichkeit besteht darin, zielgenauer diejenigen Kinder herauszufinden, die einer besonderen Intervention, sei es eine Förderung oder eine Therapie, bedürfen. Die erste Variante hat den Nachteil, dass bei einer flächendeckenden Intervention oft diejenigen, die einen erhöhten Förder- oder Therapiebedarf aufweisen, davon weniger gut profitieren als diejenigen, bei denen keine Intervention erforderlich ist. Daher wird bspw. auch bei der Durchführung der Förderprogramme zur phonologischen Bewusstheit empfohlen (s. Olofsson & Lundberg, 1985; Schneider, Roth, Küspert & Ennemoser, 1998; s. auch Roos, Treutlein, Zöller & Schöler, 2007), diese Programme nur bei Kindern mit einem besonderen Förderbedarf einzusetzen. Wählt man die zweite Möglichkeit, dann ist flächendeckend ein diagnostischer Schritt erforderlich, bei dem Kinder mit einem Schriftspracherwerbsrisiko frühzeitig erkannt werden können. Die prognostische Validität eines solchen Verfahrens ist daher seine relevanteste Güte, daneben sollte es wegen des flächendeckenden Einsatzes zeitökonomisch sein und möglichst wenig Durchführungs- und Auswertungsfehler zulassen.

Derzeit liegen nur zwei Screenings vor, bei denen die prognostische Validität geprüft und sich als zufrieden stellend erwiesen hat: Das Bielefelder Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (*BISC*; Jansen, Mannhaupt, H. Marx & Skowronek, 1999) und das Heidelberger Auditive Screening in der Einschulungsuntersuchung (*HASE*; Schöler & Brunner, 2008). Die Wiederholungsprüfung der prognostischen Validität beim *BISC* (P. Marx & Weber, 2006) deutet an, dass einmal gefundene gute Prädiktionen nicht zwingend replizierbar sein müssen. Daher soll im Folgenden auch *HASE* einer Kreuzklassifikationsstudie unterzogen werden.

1.1 Das Heidelberger Auditive Screening in der Einschulungsuntersuchung

Für den Einsatz von *HASE* im Rahmen der Schuleingangsdiagnostik war vom Öffentlichen Gesundheitsdienst in Baden-Württemberg eine zeitökonomische (10 bis maximal 15 Minuten) und gleichzeitig objektive Durchführung gefordert. Nach Erprobungen im flächendeckenden Einsatz bei der Einschulungsuntersuchung in Mannheim (Schäfer, Schöler, Roos, Grün-Nolz & Engler-Thümmel, 2003; Schöler et al., 2002) konnte eine erste Version von *HASE* (Brunner & Schöler, 2001/02) vorgelegt werden, bei der sowohl die Durchführung als auch die Auswertung vollständig computerge-

stützt möglich ist. Die Kinder des Mannheimer Einschulungsjahrgangs 2003 (Schöler et al., 2004) wurden dann erstmals mit *HASE* untersucht, das seit 2009 flächendeckend und obligatorisch in der neu konzipierten ESU in Baden-Württemberg eineinhalb bis zwei Jahre vor der Einschulung eingesetzt wird. Dieser erste, in Mannheim vollständig mit *HASE* untersuchte Einschulungsjahrgang 2003 diente auch der Bestimmung der prognostischen Validität des Screenings *HASE*, die sich als zufriedenstellend erwies (Treutlein, Roos & Schöler, 2007; s. auch Schöler & Brunner, 2008).

HASE enthält vier Aufgabengruppen: *Nachsprechen von Sätzen (NS)*, *Wiedergeben von Zahlen-Folgen (WZ)*, *Nachsprechen von Kunstwörtern (NK)* und *Erkennen von Wortfamilien (EW)*.¹

Die drei Aufgaben *NS*, *WZ* und *NK* haben sich als differenziell valide bei der Diagnostik von Sprachentwicklungsstörungen (Schöler, 1999) und als valide Indikatoren für Funktionen der phonologischen Schleife des Arbeitsgedächtnisses sensu Baddeley (2000) erwiesen. Defizite in diesen basalen informationsverarbeitenden Prozessen bzw. Strukturen gelten als zugrunde liegend für Sprach- und Schriftspracherwerbsstörungen (Schöler, 2011). Mit *EW* soll die Fähigkeit erfasst werden, „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., 2001) übernommen.

Die Entscheidung, ob ein Kind ein Risiko für Schriftspracherwerbsprobleme aufweist und daher eine genauere diagnostische Abklärung erforderlich wird, erfolgt mit Hilfe empirisch gewonnener Risikowerte und -muster (vgl. Treutlein et al., 2007).

Objektivität und Reliabilität. Falls die Durchführung PC-gestützt und damit standardisiert in Durchführung und Auswertung erfolgt, ist Objektivität gewährleistet. Die internen Konsistenzen der vier Aufgaben variieren zwischen $\alpha = .62$ (*NK*) und $\alpha = .83$ (*NS*) (s. Schöler & Brunner, 2008, S. 40). Angesichts der geringen Itemzahlen scheinen diese Koeffizienten zufriedenstellend. Die split-half-Koeffizienten liegen ein wenig höher, die Werte variieren zwischen $r = .63$ und $r = .90$.

Zur internen Validität. Alle Interkorrelationen zwischen den vier Aufgaben sind substantiell und statistisch bedeutsam. Da die Koeffizienten zwischen $r = .29$ und $r = .55$ variieren und somit nur geringe bis mittlere Höhe erreichen (s. Schöler & Brunner, 2008, Tab. 8.5), kann man davon ausgehen, dass die vier Aufgaben ein ausreichend unterschiedliches Leistungsspektrum erfassen.

Zur prognostischen Validität. Die prognostische Validität von *HASE* wurde in der Längsschnittstudie EVER (Entwicklung eines Vorschulscreenings zur Erfassung von Risikokindern von Sprach- und Schriftspracherwerbsproblemen) bestimmt (vgl. dazu Treutlein et al., 2007). Dabei erwies sich die prognostische Validität insbesondere bei der Vorhersage der Leseleistungen als sehr gut (RATZ-Indizes > 66 ; zur Erläuterung s. u.; s. auch Tab. 2). Tabelle 1 zeigt die Klassifikationsgüte aufgrund der Prädiktion durch *HASE* am Ende des zweiten und dritten Grundschuljahres².

¹ Das Screening war zunächst für den Einsatz bei Kindern im letzten Vorschuljahr konzipiert; durch die Neukonzeption der Einschulungsuntersuchung in Baden-Württemberg wird das Screening nun mit Kindern ab vier Jahren und zwar in einer um *EW* reduzierten Form durchgeführt, da diese Aufgabenstellung für die jüngeren Kinder zu schwierig ist.

² Zu beachten ist, dass eine solch gute Vorhersage nur dann erzielt wird, wenn die angepassten Grenzwerte zur Bestimmung der Risikokinder verwendet werden. Bei einer Verwendung der Altersnormen fallen die Gütekriterien deutlich geringer aus.

Tabelle 1 Güteindizes für die Vorhersage der Leistungen in der 3. Klasse (Daten aus der EVER-Studie; Treutlein et al., 2007)

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

1.2 Fragestellung

Kreuzvalidierungen, d. h. Replikationen an einer anderen Stichprobe, bieten eine gute Möglichkeit zur Überprüfung der Gütekriterien von Verfahren, insbesondere der prognostischen Validität eines Screenings. Zwar hat sich *HASE* als ein prognostisch valides Screening in der Längsschnittstudie, mit der gleichzeitig die Risikowerte und -muster bestimmt worden sind, erwiesen. Wie aber das Beispiel des *BISC* zeigt (P. Marx & Weber, 2006), ergibt sich bei Wiederholungsstudien nicht immer eine vergleichbar hohe Messgüte, hier vor allem für die prognostische Validität, wie sie bei der Entwicklung eines Verfahrens festgestellt wurde. Eine Erklärung könnte sein, dass sich die Rahmenbedingungen epochal geändert haben, wie man dies beim *BISC* vermuten kann. Ein Training der phonologischen Bewusstheit ist zum Standard vieler Kindertageseinrichtungen geworden, insofern sind viele der im *BISC* geprüften Leistungen bereits trainiert, sodass sich Risiken nicht mehr so gut vorhersagen lassen (Schöler, 2011). Vielleicht liegt hier ein spezifisches Problem des *BISC*, das sich bei *HASE* weniger zeigt, weil dessen Aufgaben nicht auf Fertigkeiten oder Fähigkeiten fokussieren, die durch Lernen deutlich geändert werden können, sondern eher auf *Entwicklungsvoraussetzungen* und damit auf Fähigkeiten, die sich durch Übungen o. Ä. weniger gut bzw. kaum verändern lassen und für eine Prognose durch ihre Stabilität als valider erweisen.

Da *HASE* in einer weiteren Längsschnittstudie, die zur Evaluation von Sprachförderprogrammen³ erfolgte, eingesetzt wurde, besteht die Möglichkeit, die prognostische Validität des Screenings anhand einer anderen Stichprobe erneut zu überprüfen.

2 Methode

2.1 Untersuchungsgruppe und Durchführung von HASE

Die Daten für die Kreuzvalidierung wurden im Rahmen der Längsschnittstudie EVAS erhoben, einem anderen Längsschnitt als bei der Validierung und Normierung von *HASE*. Zwischen Mai und Juli 2005 vor Beginn des letzten Kindergartenjahres wurde

³ *HASE* wurde im Rahmen der EVAS-Studie durchgeführt, mit der im Auftrag der Landesstiftung Baden-Württemberg Sprachfördermaßnahmen evaluiert werden sollten (s. Baden-Württemberg Stiftung, 2009).

HASE mit 453 Kindern durchgeführt. Von 237 Kindern konnten die Sprach- und Schriftsprachleistungen bis ans Ende der 2. Klasse (Sommer 2008) beobachtet werden. Die Durchführung des Screenings erfolgte entweder im jeweiligen Kindergarten von geschulten Hilfskräften ($N = 218$), oder im Rahmen der üblichen Einschulungsuntersuchung (ESU) vom kinderärztlichen Gesundheitsdienst bzw. ebenfalls von geschulten Hilfskräften. Alle Hilfskräfte führten *HASE* instruktionsgemäß mit Hilfe der CD durch. Von den 237 Kindern sind 56 % Jungen und 44 % Mädchen, 50 % von ihnen sind deutschsprachig, knapp 14 % türkischsprachig und etwa 36 % anderssprachig. Bei der Untersuchung mit *HASE* waren sie zwischen 4;6 und 6;6 Jahre alt (im Durchschnitt 5;5 Jahre).

Schulleistungen und Deutschkenntnisse – Einschätzungen durch die Lehrkräfte

Am Ende der 1. und 2. Klasse wurden die Lehrkräfte gebeten, die Deutschnote, und darüber hinaus eine Note für die Lese- und Rechtschreibleistung anzugeben. Außerdem sollten sie am Ende der 2. Klasse die Deutschkenntnisse in den Bereichen Grammatik, Sprachverstehen, Wortschatz und Ausdrucksfähigkeit einschätzen und anhand der üblichen Notenskala einschätzen.

Prüfung des Lesens und Rechtschreibens durch Schulleistungstests

Am Ende der 2. Klasse (Sommer 2008) wurden die Lese- und Rechtschreibleistungen der Kinder mit Hilfe von Schulleistungstests erfasst. Die Untersuchungen wurden in Gruppen durchgeführt und an zwei verschiedenen Tagen von geschulten Hilfskräften mit den folgenden drei Tests:

(1) Die *Rechtschreibleistung* wurde mit Hilfe der Hamburger Schreibprobe (*HSP 2*, May, 2002) erfasst. Die *HSP 2* besteht aus 15 Wörtern und 3 Sätzen, die dem Kind diktiert und im Testheft mit Bildern veranschaulicht werden. Neben der Anzahl richtig geschriebener Wörter (*HSP* Gesamtwert) kann zusätzlich die Anzahl richtig geschriebener Grapheme ermittelt werden (Graphemtreffer).

(2) Das *Leseverständnis* wurde durch *ELFE 1-6* (Ein Leseverstehenstest für Erst- bis Sechstklässler, Lenhard & Schneider, 2006) erfasst. *ELFE* überprüft zum einen basale Lesestrategien, zum anderen die Fähigkeit zum Verstehen von Wörtern, Sätzen und Texten.

(3) Die *Lesegeschwindigkeit* wurde mittels der Würzburger Leise Leseprobe (*WLLP*, Küspert & Schneider, 1998) erhoben. Aufgabe des Kindes ist es, innerhalb von fünf Minuten zu möglichst vielen von 140 vorgegebenen Wörtern das richtige aus vier Bildalternativen herauszusuchen.

2.2 Zur Bestimmung der Validität von HASE

Kennwerte für die prognostische Validität.

Ein Screening soll das zuverlässige Auffinden entwicklungsgefährdeter Kinder ermöglichen. Zusammenhangsmaße zwischen Prädiktorwerten und Kriteriumsleistungen sind daher für die individuelle Risikobestimmung unzureichend. Für die Bestimmung der prognostischen Validität eines Screenings sind notwendig klassifikatorische Gütekriterien zu ermitteln. 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. Tabelle 2; s. dazu auch Jansen, 2007).

Tabelle 2 Vierfelderschema der möglichen Vorhersageergebnisse als Grundlage zur Berechnung der Güteindizes eines Screenings

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

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 ($a / (a + c) \times 100$) entspricht dem prozentualen Anteil der Problemkinder in der Schule, die durch das Screening auch als Risikokinder vorhergesagt waren (richtig positiv).

Spezifität: Mit Spezifität ($d / (b + d) \times 100$) 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).

Positive Korrektheit: Positive Korrektheit ($a / (a + b) \times 100$) 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.

Negative Korrektheit: Die Negative Korrektheit ($d / (c + d) \times 100$) 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.

RAZ-Index: Der relative Steigerungsbetrag gegenüber dem Zufall (RAZ-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).

Selektions- und Grundrate beeinflussen diese Maße. Eine hohe Grundrate erhöht die Wahrscheinlichkeit für gute Werte bei positiver und negativer Korrektheit. Die Selektionsrate wirkt sich auf Sensibilität und Spezifität aus.

Dichotomisierung des Prädiktors: Zur Bestimmung von Risikokindern.

Die Dichotomisierung des Prädiktors erfolgt anhand empirisch adjustierten Grenzwerten und Risikoprofilen (s. Treutlein et al., 2007). Diese entsprechen nicht den im *HASE*-Manual der 1. Auflage (Brunner & Schöler, 2001) als vorläufig gekennzeichneten Grenzwerten und Risikoprofilen. Im Untertest *NS* gilt die Leistung als problematisch, wenn weniger als 8 Punkte erzielt werden, im Untertest *WZ* weniger als 6 Punkte, bei *EW* weniger als 6 Punkte und bei *NK* weniger als 9 Punkte. Als risikobehaftet wird ein Kind eingestuft, das problematische Leistungen in mindestens drei Untertests zeigt, oder Schwierigkeiten sowohl in *NS* und *WZ* bzw. *EW* hat. Mit diesen Kriterien wurden in der Normierungsstichprobe 44 % der Kinder als Risikokinder ein-

geschätzt (bei Grundraten von 12-15 % - je nach gewähltem Kriterium). In der für die Kreuzvalidierung verwendeten Stichprobe müssen 73 % als Risikokinder eingestuft werden (bei Grundraten von bis zu 40 %; s. u.).

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, deren Schriftsprachleistungen in der 2. Klasse mit der Schulnote 4 oder schlechter bewertet wurden oder die im Leistungstest unterdurchschnittliche Lese- bzw. Rechtschreibtestleistungen ($T < 40$) zeigen. Mit der Note 4 oder schlechter werden zwischen 8 % (Sprachverständnis) und 20 % (Rechtschreibung) bewertet. Eine schwache Rechtschreibnote erhielten 15 % der Kinder ohne und 21 % der Kinder mit Risiko. Eine unterdurchschnittliche Leistung in den Schulleistungstests zeigen zwischen 3 % (alphabetische Strategie in der HSP) und 40 % (Gesamtwert ELFE, 31 % von den Kindern ohne Risiko, 44 % der Kinder mit Risiko).

3 Ergebnisse

Werden die Schulnoten in Deutsch, Lesen, Rechtschreiben und der Deutschkenntnisse (Grammatik, Sprachverstehen, Wortschatz, Ausdrucksfähigkeit) als Kriteriumsleistungen berücksichtigt, resultieren eine hohe Negative Korrektheit und hohe Sensitivität für die Prädiktion durch *HASE* (vgl. Tabelle 3). Die aufgrund der Prädiktion durch das Screening erfolgte Klassifikation der Kinder hinsichtlich ihrer sprachlichen Leistungsfähigkeit im Deutschen am Ende der 2. Klasse gilt gemäß der RATZ-Indizes als sehr gut und spezifisch. Nicht zufrieden stellend gelingt die Klassifikation hinsichtlich der Beurteilung der Lese- und Rechtschreibleistung.

Tabelle 3 Güteindizes für die prognostische Validität am Ende der 2. Klasse bei der Kreuzvalidierung (Schulnoten*) (in Klammern die Angaben)

Note in ...	Sensitivität	Spezifität	Positive Korrektheit	Negative Korrektheit	RATZ-Index
Deutsch	87	31	15	95	55
Lesen	78	30	12	92	25
Rechtschreiben	78	30	22	85	24
Grammatik	95	35	26	97	83
Sprachverstehen	100	32	12	100	100
Wortschatz	92	32	15	97	71
Ausdrucksfähigkeit	93	32	18	97	75

Anmerkung. Die Lehrkräfte waren gebeten, neben den Noten im Fach Deutsch auch die verschiedenen sprachlichen Leistungsbereiche auf einer Notenskala zu bewerten.

Auch bei Berücksichtigung der Leistungen in den Schulleistungstests als Kriterium sind negative Korrektheit und Sensitivität sehr zufrieden stellend (s. Tabelle 4). Die RATZ-Indizes sind dagegen nur beim Textverstehen (*ELFE*) und bei der Rechtschreibleistung (*HSP*) zufriedenstellend, d. h. nur hier kann die Klassifikation als gut, aber unspezifisch bewertet werden.

Die Werte für die positive Korrektheit und die Spezifität sind bei allen Kriterien unbefriedigend.

Tabelle 4 Güteindizes für die prognostische Validität am Ende der 2. Klasse bei der Kreuzvalidierung (Schulleistungstests)

	Sensitivität	Spezifität	Positive Korrektheit	Negative Korrektheit	RATZ-Index
<i>ELFE</i> -Wortverstehen	80	36	38	78	33
<i>ELFE</i> -Satzverstehen	80	36	38	78	33
<i>ELFE</i> -Textverstehen	85	37	36	85	48
<i>ELFE</i> -Gesamt	77	36	44	69	23
<i>WLLP</i>	77	34	35	76	24
<i>HSP</i> -Gesamtwert	84	32	13	94	48
<i>HSP</i> -Graphemtreffer	84	32	10	96	49

4 Diskussion

Mit dem *Heidelberger Auditiven Screening in der Einschulungsuntersuchung (HASE)* sollen Kinder mit einem Risiko für Schriftspracherwerbsprobleme bereits im Vorschulalter aufgefunden werden. In einem ersten Längsschnitt vom letzten Vorschuljahr bis zum Ende der dritten Grundschulklasse (*EVER*-Studie; Treutlein et al., 2007) hatte sich die Klassifikationsgüte des Screenings als zufrieden stellend erwiesen. Eine weitere Längsschnittstudie, in der *HASE* im Vorschulalter eingesetzt wurde und die Kinder bis zum Ende der 2. Klasse untersucht wurden, ermöglichte eine Kreuzvalidierung, mit der die zufrieden stellende prognostische Validität teilweise bestätigt werden kann.

Bei der Kreuzvalidierung (*EVAS*-Studie; s. Baden-Württemberg Stiftung, 2009) ergeben sich Güteindizes, die ebenfalls als gut zu bewerten sind, auch wenn sie etwas geringer ausfallen als in der *EVER*-Studie. Es kann demnach davon ausgegangen werden, dass durch die empirisch erstellten Risikowerte und -muster zutreffende Vorhersagen getroffen werden können. Die mit der Untersuchungsgruppe zur Bestimmung der prognostischen Validität erhaltenen Güteindizes scheinen demnach nicht durch eine „Überanpassung“ an die Daten der Untersuchungsgruppe so positiv ausgefallen zu sein – die prognostische Validität von *HASE* ist zufrieden stellend. Die Grenzwerte und Risikoprofile gelten nicht nur für die Normierungsstichprobe, sondern erweisen sich auch in einer anderen Stichprobe als gute Prädiktoren. Dies gilt unter der Voraussetzung einer ordnungsgemäßen Durchführung und der Verwendung der im Rahmen von *EVER* hergeleiteten Risikoprofile (vgl. Treutlein et al., 2007).

Bei einem Screening muss der Anteil der Kinder, bei denen kein Risiko festgestellt wird, die aber später zu Problemfällen werden, möglichst gering gehalten werden. Bei der Früherkennung von Risikokindern kommt es in erster Linie darauf an, keine gefährdeten Kinder zu übersehen – falsch negativ klassifizierte Kinder zu vermeiden. Dieser Klassifikationsfehler ist bei *HASE* ausreichend klein. Weniger relevant ist bei einem Screening die Zahl der Kinder, die als Risikokinder klassifiziert werden, später aber keine Probleme aufweisen und unauffällig bleiben. Diese so genannte Positive Korrektheit ist bei *HASE* gering ausgeprägt – es werden mehr Kinder als Risikokinder eingestuft als später tatsächlich auffällig werden. Hierdurch wird einmal mehr die Notwendigkeit verdeutlicht, bei Vorliegen eines auffälligen Screening-Befundes weiterführende ausführlichere diagnostische Maßnahmen in die Wege zu leiten. Eventuelle Stigmatisierungsprozesse können durch einen verantwortungsbewussten Umgang mit den Screeningergebnissen eingedämmt werden. Die Vorhersage eines Risikos nach *HASE* wird sowohl bei den Bewertungen der Leserechtschreibfertigkeiten durch die Lehrkräfte als auch bei den Untersuchungen mit den Schulleistungstests bestätigt. Dabei sind die Klassifikationen dann besonders gut, wenn als Kriterien die Beurteilung sprachlicher Leistungsbereiche durch die Lehrkräfte definiert waren, d. h. Ausdrucksfähigkeit, Grammatik, Sprachverständnis und Wortschatz eines Kindes am Ende des zweiten Grundschuljahres. Für den individuellen Bildungsweg dürften die Urteile der Lehrkräfte relevanter sein als die in Schulleistungstests erbrachten Leistungen, zumal diese nur aufgrund einer punktuellen Gruppenuntersuchung erfasst werden konnten.

Die Vorhersage aufgrund der Klassifikation nach *HASE* ist ähnlich gut wie die Vorhersage mit dem zehn Monate vor der Einschulung durchgeführten *BISC* (Jansen et al., 1999), für das am Ende der 2. Klasse RAZ-Indizes zwischen 33 und 65 berichtet werden (H. Marx, Jansen & Skowronek, 2000). Für *HASE* liegen die Werte am Ende der 2. Klasse zwischen 23 und 74 – zu beachten ist dabei, dass die Werte des *BISC* nicht über eine Kreuzvalidierung entstanden, sondern mit der Stichprobe gewonnen wurden, die zur Bestimmung der Risikowerte herangezogen wurde. Die Kreuzklassifikation des *BISC* in der Studie von P. Marx und Weber (2006) erbrachte eine deutlich niedrigere Klassifikationsgüte. Es ist zu vermuten, dass ein Screening, wie das *BISC*, mit dem u. a. auf eine Vorläuferfertigkeit fokussiert wird, die sich gut trainieren lässt und nicht zuletzt deshalb auch häufig trainiert wird, seine Klassifikationsgüte durch eine solche populäre Präventionsmaßnahme einbüßen kann. In *HASE* werden hingegen Aufgaben vorgegeben, mit denen individuelle Fähigkeiten geprüft werden, die sich nicht oder nur geringfügig durch Interventionen verändern lassen.

Zusammenfassend kann festgestellt werden, dass *HASE* zur Früherkennung von Schwierigkeiten beim Lesen und Rechtschreibenlernen tauglich ist. Ein Screening ersetzt allerdings nicht eine differenzierte und differenzierende Diagnostik, die sich notwendig anschließen muss.

Die prognostische und inhaltliche Validität von *HASE* kann allerdings nur gewährleistet werden, wenn die Vorgabe des Screenings instruktionsgemäß zumindest mit CD, möglichst PC-gestützt erfolgt. Allzu oft wird noch auf diese Durchführungs- und Auswertungsstandardisierung und -ökonomie verzichtet, dann werden nur die einzelnen Items über ein CD-Abspielgerät vorgegeben, und leider ist auch zu beobachten, dass noch nicht einmal eine standardisierte Vorgabe von *WZ* und *NK* mittels CD erfolgt. Wie bereits mehrfach gezeigt werden konnte (vgl. Schöler & Schäfer, 2004, S. 16; Treutlein et al., 2007), ist eine standardisierte Vorgabe der Aufgaben, insbesondere der Aufgaben Wiedergabe von Zahlen-Folgen (*WZ*) und Nachsprechen von Kunstwörtern (*NK*), unverzichtbar, um die Objektivität und Reliabilität, und damit auch die prognostische und inhaltliche Validität zu garantieren. Die Leistungs-diffe-

renzen, 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/oder Clusterungen der vorzugebenden Zahlenfolgen ergeben. Durch diese „Hilfe“ werden Risikokinder möglicherweise übersehen.

5 Literatur

- Brunner, M., Pfeiffer, B., Schlüter, K., Steller, F., Möhring, L., Heinrich, I. & Pröschel, U. (2001). *Heidelberger Vorschulscreening zur auditiv-kinästhetischen Wahrnehmung und Sprachverarbeitung, (HVS)*. CD-Rom und Testhandbuch zur Testanweisung und Auswertung. Wertingen: Westra.
- Brunner, M. & Schöler, H. (2001/2002). *HASE – Heidelberger Auditives Screening in der Einschulungsuntersuchung*. Wertingen: Westra.
- Jansen, H. (2007). Screenings zur Früherkennung von Leserechtschreibschwierigkeiten. 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 Leserechtschreibschwierigkeiten (BISC)*. Göttingen: Hogrefe.
- Küspert, P. & Schneider, W. (1998). *Würzburger Leise Leseprobe (WLLP): Ein Gruppenlesetest für die Grundschule*. Göttingen: Hogrefe.
- Landesstiftung Baden-Württemberg. (2009). „EVAS“ – Evaluation von Programmen zur Förderung von Vorschulkindern mit unzureichenden Sprachkenntnissen im Deutschen. Zugriff am 28.12.2009 unter <http://www.sagmalwas-bw.de/sprachfoerderung-fuer-vorschulkinder/wissenschaftliche-begleitung2/ph-heidelberg.html>
- Lenhard, W. & Schneider, W. (2006). *Ein Leseverständnistest für Erst- bis Sechstklässler (ELFE 1-6)*. 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 Leserechtschreibschwierigkeiten. *Diagnostica*, 38, 249-268.
- Marx, H., Jansen, H. & Skowronek, H. (2000). Prognostische, differentielle und konkurrente Validität des Bielefelder Screenings zur Früherkennung von Leserechtschreibschwierigkeiten (BISC). In M. Hasselhorn, W. Schneider & H. Marx (Hrsg.), *Diagnostik von Leserechtschreibschwierigkeiten, Band 1* (S. 9-34). Göttingen: Hogrefe.
- Marx, P. & Weber, J. (2006). Vorschulische Vorhersage von Leserechtschreibschwierigkeiten. Neue Befunde zur prognostischen Validität des Bielefelder Screenings (BISC). *Zeitschrift für Pädagogische Psychologie*, 20, 251-259.
- May, P. (2002). *Hamburger SchreibProbe für zweite Klassen (HSP 2)*. Göttingen: Hogrefe.

- Nation, K., Clarke, P., Marshall, C. M. & Durand, M. (2004). Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language, and Hearing Research*, 47, 199-211.
- Olofsson, A. & Lundberg, I. (1985). Evaluation of long term effects of phonemic awareness training in kindergarten: Illustrations of some methodological problems in evaluation research. *Scandinavian Journal of Psychology*, 26, 21-34.
- Schäfer, P., Schöler, H., Roos, J., Grün-Nolz, P. & Engler-Thümmel, H. (2003). Einschulungsuntersuchung 2002 in Mannheim – Sprachentwicklungsstand bei Schulbeginn. *Gesundheitswesen*, 65, 676-682.
- Schneider, W., Roth, E., Küspert, P. & Ennemoser, M. (1998). Kurz- und langfristige Effekte eines Trainings der sprachlichen (phonologischen) Bewusstheit bei unterschiedlichen Leistungsgruppen: Befunde einer Sekundäranalyse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 30, 26-39.
- Schöler, H. (1999). *IDIS – Inventar diagnostischer Informationen bei Sprachentwicklungsauffälligkeiten*. Heidelberg: Edition S im Universitätsverlag C. Winter.
- Schöler, H. (2001). *Zur Früherkennung von Schriftspracherwerbsproblemen im Rahmen der Einschulungsuntersuchungen* (Arbeitsberichte aus dem Forschungsprojekt Differenzialdiagnostik Nr. 10). Heidelberg: Pädagogische Hochschule, Erziehungs- und Sozialwissenschaftliche Fakultät.
- Schöler, H. (2011). Prognose schriftsprachlicher Leistungen und Risiken im Vorschulalter am Beispiel des Heidelberger Auditiven Screenings in der Einschulungsuntersuchung (HASE). In M. Hasselhorn & W. Schneider (Hrsg.), *Tests & Trends, N. F. Bd. 9 Frühprognose schulischer Kompetenzen* (S. 13-31). Göttingen: Hogrefe.
- Schöler, H. & Brunner, M. (2008). *HASE - Heidelberger Auditives Screening in der Einschulungsuntersuchung* (2., überarb. u. erw. Aufl.). Wertingen: Westra.
- Schöler, H., Dutzi, I., Roos, J., Schäfer, P., Grün-Nolz, P. & Engler-Thümmel, H. (2004). *Einschulungsuntersuchungen 2003 in Mannheim* (Arbeitsberichte aus dem Forschungsprojekt Differenzialdiagnostik 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 Differenzialdiagnostik Nr. 13). Heidelberg: Pädagogische Hochschule, Institut für Sonderpädagogik, Abt. Psychologie in sonderpädagogischen Handlungsfeldern.
- Schöler, H. & Schäfer, P. (2004). *HASE Heidelberger Auditives Screening in der Einschulungsuntersuchung – Itemanalysen und Normen* (Arbeitsberichte aus dem Forschungsprojekt Differenzialdiagnostik, Nr. 17). Heidelberg: Pädagogische Hochschule, Institut für Sonderpädagogik, Abt. Psychologie in sonderpädagogischen Handlungsfeldern.
- Snowling, M., Bishop, D. V. M. & Stothard, S. E. (2000). Is preschool language impairment a risk factor for dyslexia in adolescence? *Journal of Child Psychology and Psychiatry*, 41, 587-600.
- Treutlein, A., Roos, J. & Schöler, H. (2007). *Zur prognostischen Validität des Heidelberger Auditiven Screenings in der Einschulungsdiagnostik HASE. Abschlussbericht des Projekts EVER*. Verfügbar unter: www.ph-heidelberg.de/wp/schoeler/Dateien/Abschlussbericht_EVER-HASE_Feb-2007.pdf [18.12.2008].

- Weber, J. & Marx, P. (2008). Lese-Rechtschreibschwierigkeiten. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (S. 631-641). Göttingen: Hogrefe.
- Weber, J., Marx, P. & Lenhard, W. (2006). Die Prognose und Prävention von schulischen Lernproblemen. *PÄD Forum: unterrichten erziehen*, 34, 289-292.
- 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.
- Weinert, S., Doil, H. & Frevert, S. (2008). Kompetenzmessungen im Vorschulalter: eine Analyse vorliegender Verfahren. In Bundesministerium für Bildung und Forschung (BMBF) (Hrsg.), *Bildungsforschung, Band 24 Kindliche Kompetenzen im Elementarbereich: Förderbarkeit, Bedeutung und Messung* (S. 89-236). Berlin: BMBF.