

# Zusätzliche Angaben zur Reliabilität und Validität von „Screening grammatischer Fähigkeiten für die 2. Klasse (SGF 2)“ von Kathrin Mahlau

In diesem Zusatztext zu Kapitel 5 des „Screenings grammatischer Fähigkeiten für die 2. Klasse“ (SGF 2) werden ergänzende Informationen zur Reliabilität sowie zur konvergenten, diskriminanten und prognostischen Validität des Verfahrens dargestellt.

## Korrelationen

Die Reliabilität wurde unter anderem durch die Prüfung der Korrelationen zwischen den Subtests erhoben. Alle Korrelationen sind untereinander hochsignifikant ( $p < .000$ ). Es ergeben sich Korrelationskoeffizienten zwischen  $r = .291$  und  $r = .559$  von moderater Größe. Das bedeutet, dass die Subtests alle eine ähnliche Leistung erheben. Tab. 1 stellt die Korrelationen im Überblick dar.

## Konvergente Validität

Im Folgenden werden Korrelationen mit anderen standardisierten Verfahren, die ebenfalls die Sprachentwicklungsleistung der untersuchten Kinder messen, dargelegt. Dazu werden zunächst die Leistungen einer größeren Kindergruppe ( $N = 96$ ), die sowohl sprachentwicklungsnormale ( $N = 78$ ) als auch sprachentwicklungsauffällige Kinder ( $N = 18$ ) sowie eine ausgewählte Gruppe von Kindern mit Sprachentwicklungsstörungen ( $N = 27$ ) enthält, in die Analysen mit einbezogen.

Die Vergleiche mit anderen standardisierten Sprachdiagnostikverfahren wurden mit vier Referenzverfahren durchgeführt: dem „Marburger Sprachentwicklungstest für Kinder“ (MSVK, Elben/

Lohaus 2000), dem „Münsteraner Screening“ (MÜSC, Mannhaupt 2006) und zwei Teilen des „Elternfragebogens zur Anamnese der Sprachentwicklung“ (Mahlau 2010). Diese Verfahren wurden zu Beginn der Schulzeit und somit ein Jahr vor dem Einsatz des vorliegenden Testverfahrens eingesetzt.

Mit  $N = 27$  Kindern konnten zudem Daten des „Sprachstandserhebungstests für Kinder im Alter zwischen 5 und 10 Jahren“ (SET 5–10, Petermann 2010) und mit  $N = 15$  Kindern Daten des Tests zur Überprüfung des Grammatikverständnisses (TROG-D, Fox 2011) erhoben werden.

Der MSVK und das MÜSC bieten sich als Vergleichstests an, da sie ebenfalls von der Testmethodik her Gruppenscreeningverfahren sind und sprachliche Leistungen erheben. Angaben zum Sprachentwicklungsverlauf und -stand erfolgten durch die Eltern anhand des Elternfragebogens zur Anamnese der Sprachentwicklung und ermöglichen einen Vergleich mit der produktiven Sprachentwicklung, die im SGF 2 ebenfalls erwartet wird.

Aus dem SET 5–10 wird der Subtest 4 Handlungssequenzen dargestellt, der wie das SGF 2 grammatische Leistungen überprüft. Für den Subtest Plural des SGF 2 wird zudem der gleichnamige Untertest Plural des SET 5–10 in Zusammenhang gebracht.

Zunächst werden die Ergebnisse der Tests über die Gesamtgruppe der zu den Vergleichsanalysen untersuchten Kinder korreliert. Es wird der Rangkorrelationskoeffizient nach Spearman (Rho) dargestellt (Bortz/Döring 2006). Der Korrelationskoeffizient nimmt Werte zwischen  $-1$  und  $+1$  an. Je näher er bei  $-1$  oder  $+1$  liegt, desto stärker ist der Zusammenhang:

**Tab. 1:** Korrelationen zwischen den Subtests ( $N = 2.253$ )

Subtests	Genus	Plural	Akkusativ	Dativ	Subjekt-Verb-Kongruenz
Genus	1	.450**	.549**	.360**	.312**
Plural	.450**	1	.530**	.380**	.484**
Akkusativ	.549**	.530**	1	.390**	.476**
Dativ	.360**	.380**	.390**	1	.291**
Subjekt-Verb-Kongruenz	.312**	.484**	.476**	.291**	1

- $0,0 \leq \text{Rho} \leq 0,2 \Rightarrow$  kein bis geringer Zusammenhang
- $0,2 < \text{Rho} \leq 0,5 \Rightarrow$  schwacher bis mäßiger Zusammenhang
- $0,5 < \text{Rho} \leq 0,8 \Rightarrow$  deutlicher Zusammenhang
- $0,8 < \text{Rho} \leq 1,0 \Rightarrow$  hoher bis perfekter Zusammenhang

Die Korrelation nach Spearman liefert ein Maß für eine monotone Beziehung zwischen zwei kontinuierlichen Zufallsvariablen, die nicht normalverteilt sind. Die Spearman-Korrelation ist im Gegensatz zur Pearson-Korrelation robust gegenüber Ausreißern. Es zeigt sich, dass alle Untertests des SGF 2 mit dem MSVK und dem MÜSC hochsignifikant ( $p < .001$ ) korrelieren und meist deutliche Zusammenhänge mit dem SGF 2 bestehen. Der Zusammenhang mit den beiden Bereichen des Elternfragebogens ist etwas geringer ausgeprägt. Hier zeigen sich bei signifikanten ( $p < .05$ ) bis hochsignifikanten Korrelationen ( $p < .001$ ) im Spearman-Koeffizienten eher mäßige Zusammenhänge. Zu beachten ist, dass die negativen Vorzeichen (MÜSC, Eltern-

fragebogen) durch die Verrechnung von Risikopunkten, nicht von Leistungspunkten (MSVK; dort positive Korrelation) zustande kommen (Tab. 2).

Um auch zwischen dem SGF 2 und standardisierten Sprachdiagnostikverfahren in Einzeltestungen Vergleichswerte zu haben, wurden zwei grammatische Subtests des SET 5–10 bei 27 Kindern vergleichend eingesetzt. Die Gruppe dieser Kinder zeigte Sprachentwicklungsauffälligkeiten. Die Probanden wurden als sprachentwicklungsauffällig klassifiziert, wenn im SET 5–10 bei mindestens zwei Subtests unterdurchschnittliche Leistungen ( $T\text{-Wert} < 41$ ) vorlagen.

Da es innerhalb der Testmethodik (Gruppen- vs. Einzeltest) Unterschiede gibt und es sich bei den Probanden um unterdurchschnittlich leistungsfähige Kinder handelt, zeigen sich vermutlich Auswirkungen auf die Korrelationen zwischen den Subtests des SGF 2 und den verwendeten Referenzverfahren. So ergeben sich zwischen dem Subtest Handlungssequenzen des SET 5–10 und den Subtests Plural und Akkusativ deutliche, zwischen den Subtests Genus, Dativ und Subjekt-Verb-Kon-

**Tab. 2:** Korrelationsberechnung zwischen dem SGF 2 und Gruppentest-Referenzverfahren

Sprachentwicklungsscreening Klasse 2	N	Referenzverfahren	Spearman's Rho (Signifikanz p)
Subtest 1 Genus	96	MSVK	.651 (.000)
	87	Elternfragebogen Teil 1	-.279 (.009)
	87	Elternfragebogen Teil 2	-.490 (.000)
	96	MÜSC	-.621 (.000)
Subtest 2 Plural	96	MSVK	.584 (.000)
	87	Elternfragebogen Teil 1	-.245 (.022)
	87	Elternfragebogen Teil 2	-.261 (.015)
	96	MÜSC	-.498 (.000)
Subtest 3 Akkusativ	96	MSVK	.612 (.000)
	87	Elternfragebogen Teil 1	-.309 (.004)
	87	Elternfragebogen Teil 2	-.334 (.002)
	96	MÜSC	-.566 (.000)
Subtest 4 Dativ	96	MSVK	.661 (.000)
	87	Elternfragebogen Teil 1	-.296 (.005)
	87	Elternfragebogen Teil 2	-.463 (.000)
	96	MÜSC	-.593 (.000)
Subtest 5 Subjekt-Verb-Kongruenz	96	MSVK	.516 (.000)
	87	Elternfragebogen Teil 1	-.402 (.000)
	87	Elternfragebogen Teil 2	-.407 (.000)
	96	MÜSC	-.489 (.000)
Gesamttest	96	MSVK	.697 (.000)
	87	Elternfragebogen Teil 1	-.370 (.000)
	87	Elternfragebogen Teil 2	-.513 (.000)
	96	MÜSC	-.529 (.000)

gruenz eher mäßige Zusammenhänge. Zwischen den Subtests Plural sowohl im SET 5–10 als auch im SGF 2 ist der Zusammenhang signifikant, der Korrelationskoeffizient nach Spearman eher mäßig (Tab. 3). Aufgrund der unterschiedlichen Testmethodik und der spezifischen Besonderheiten in der Stichprobe sind die Zusammenhangswerte erwartungsgemäß zufriedenstellend.

Die Ergebnisse der Vergleiche mit den Referenzverfahren verweisen darauf, dass das SGF 2 zufriedenstellend valide die verschiedenen Bereiche der morphologischen Fähigkeiten als Gruppenverfahren untersucht. Sowohl die Gruppenverfahren MSVK und MÜSC als auch die Einzeltestungen mit dem Subtest Handlungssequenzen des SET 5–10 korrelieren auf signifikantem ( $p < .05$ ) bis hochsignifikantem ( $p < .001$ ) Niveau (Ausnahme Handlungssequenzen und Subtests Dativ und Subjekt-Verb-Kongruenz). Die Ergebnisse verweisen auf mindestens substantielle Zusammenhänge zwischen den in den Testverfahren erhobenen Leistungen.

## Diskriminante Validität

Um zu prüfen, ob das SGF 2 mit weiteren Fähigkeiten in Zusammenhang steht, wurden die kognitiven Fähigkeiten (CFT 1, Weiß/Osterland 1997), die mathematischen Fähigkeiten (DEMAT 1+, Krajewski/Küspert/Schneider/Visé 2002) und die schriftsprachlichen Fähigkeiten (WLLP, Küspert/Schneider 1998 sowie DERET 1–2+, Stock/Schneider 2008) untersucht. Da es sich bei dem SGF 2 und den differenzierenden Verfahren um eine ähnliche Testmethodik handelt (Gruppenverfahren) und im Screening auch sehr einfache Fähigkeiten des Schreibens gefordert werden, zeigen sich erwartungsgemäß mit den kriteriums-differenzierenden Verfahren mäßige bis schwache Zusammenhänge. Dabei korrelieren die Verfahren zum Schriftspracherwerb mit dem SGF 2 jedoch nicht höher als die Verfahren zur Überprüfung kognitiver und mathematischer Fähigkeiten (Tab. 4).

## Prognostische Validität

Bei Screeningverfahren spielt die prognostische Validität, also die Frage, wie gut das Verfahren spätere Schwierigkeiten vorauszusagen vermag, eine zentrale Rolle. Risikokinder sollten mit großer Wahrscheinlichkeit tatsächlich Schwierigkeiten im erhobenen Bereich bekommen, Nichtrisikokinder sollten mit hoher Wahrscheinlichkeit keinen Problemen entgegensehen können.

Um einzuschätzen, ob sich auch mit standardisierten Verfahren, die wesentlich später Anwendung fanden, Korrelationen mit dem SGF 2 zeigen, wurde ca. drei Jahre nach dem Einsatz des SGF 2 der Wortschatztest des „Culture Fair Intelligence Test Scale 2 Revision“ (CFT 20-R, Weiß 2008) mit 381 Probanden der Erststichprobe durchgeführt. Die Ergebnisse korrelieren in allen Subtests hochsignifikant mit dem Wortschatztest des CFT 20-R und Spearman's Rho verweist in allen Subtests auf einen deutlichen Zusammenhang zwischen den beiden Testverfahren. Tab. 5 stellt die Ergebnisse im Überblick dar.

Das Ziel des vorliegenden Screeningverfahrens ist es, Schüler der zweiten Klasse zu identifizieren, die ein Risiko im Bereich der Sprachentwicklung tragen. Zur Vorhersagegüte gehören Angaben zur Sensitivität, Spezifität, Prädiktortrefferquote und dem RATZ-Index (Marx/Lenhard 2010). Die nachfolgend berichteten Kennwerte wurden nach den Vorgaben von Lenhard und Lenhard (2014) berechnet. Dazu wurden die Entwicklungsverläufe der 381 Schüler betrachtet, die sowohl zu Beginn der zweiten Klasse am SGF 2 als auch am Ende der vierten Klasse am Wortschatztest des CFT 20-R teilgenommen haben. Es wurden diejenigen Kinder als Risikokinder bezeichnet, die im CFT 20-R einen  $PR \leq 15$  und im SGF 2 einen  $PR \leq 25$  aufwiesen. Der weitere Prozentrang ist dadurch begründet, dass es sich beim vorliegenden Testverfahren um ein Screening handelt. Die Schülerinnen und Schüler mit einem  $PR \geq 15$  im CFT 20-R bzw. mit einem  $PR \geq 25$  im SGF 2 wurden als Nichtrisikokinder bezeichnet.

**Tab. 3:** Korrelationsberechnung zwischen dem SGF 2 und Einzeltest-Referenzverfahren

Sprachentwicklungsscreening Klasse 2	N	Referenzverfahren	Spearman's Rho (Signifikanz p)
Subtest 1 Genus	27	SET 5-10 Subtest 4 Handlungssequenzen	.433 (.022)
Subtest 2 Plural	27	SET 5-10 Subtest 4 Handlungssequenzen	.604 (.001)
	27	SET 5-10 Subtest 8 Plural	.405 (.036)
Subtest 3 Akkusativ	27	SET 5-10 Subtest 4 Handlungssequenzen	.564 (.002)
Subtest 4 Dativ	27	SET 5-10 Subtest 4 Handlungssequenzen	.302 (.125)
Subtest 5 Subjekt-Verb-Kongruenz	27	SET 5-10 Subtest 4 Handlungssequenzen	.341 (.082)
Gesamttest	27	SET 5-10 Subtest 4 Handlungssequenzen	.526 (.005)

**Tab. 4:** Korrelationsberechnung zwischen dem SGF 2 und kriteriumsdifferenziellen Verfahren

Sprachentwicklungsscreening Klasse 2	N	Referenzverfahren	Spearman's Rho (Signifikanz p)
Subtest 1 Genus	410	CFT 1	.352 (.000)
	406	DEMAT 1+	.377 (.000)
	406	WLLP	.307 (.000)
	406	DERET 1-2+	-.373 (.000)
Subtest 2 Plural	410	CFT 1	.320 (.000)
	406	DEMAT 1+	.371 (.000)
	406	WLLP	.319 (.000)
	406	DERET 1-2+	-.460 (.000)
Subtest 3 Akkusativ	410	CFT 1	.249 (.000)
	406	DEMAT 1+	.326 (.000)
	406	WLLP	.427 (.000)
	406	DERET 1-2+	-.453 (.000)
Subtest 4 Dativ	410	CFT 1	.280 (.000)
	406	DEMAT 1+	.360 (.000)
	406	WLLP	.365 (.000)
	406	DERET 1-2+	-.422 (.000)
Subtest 5 Subjekt-Verb-Kongruenz	410	CFT 1	.189 (.000)
	406	DEMAT 1+	.307 (.000)
	406	WLLP	.380 (.000)
	406	DERET 1-2+	-.389 (.000)
Gesamttest	410	CFT 1	.417 (.000)
	406	DEMAT 1+	.562 (.000)
	406	WLLP	.489 (.000)
	406	DERET 1-2+	-.625 (.000)

**Tab. 5:** Korrelationsberechnung zwischen dem SGF 2 und dem Wortschatztest des CFT 20-R zur Einschätzung der prognostischen Validität

Sprachentwicklungsscreening Klasse 2	N	Referenzverfahren	Spearman's Rho (Signifikanz p)
Subtest 1 Genus	381	Wortschatztest CFT 20-R	.448 (.000)
Subtest 2 Plural	381	Wortschatztest CFT 20-R	.408 (.000)
Subtest 3 Akkusativ	381	Wortschatztest CFT 20-R	.423 (.000)
Subtest 4 Dativ	381	Wortschatztest CFT 20-R	.492 (.000)
Subtest 5 Subjekt-Verb-Kongruenz	381	Wortschatztest CFT 20-R	.382 (.000)
Gesamttest	381	Wortschatztest CFT 20-R	.636 (.000)

Im Gesamttest zeigte sich, dass von den 381 Schülerinnen und Schülern 77 im SGF 2-Gesamtwert Risikokinder (18,8%) und 304 Kinder keine Risikokinder waren (81,2%). Davon wurden im CFT 20-R 31 Kinder drei Jahre später im Wortschatztest wiederum als Risikokinder identifiziert (korrekt positiv), 37 der ehemals als Risikokinder identifizierten Kinder fielen im CFT 20-R nicht als Risikokinder auf (falsch positiv). Korrekt negativ

als Nichtrisikokind wurden in beiden Testverfahren übereinstimmend 294 Kinder festgestellt. Nicht erkannt wurden im SGF 2 19 Kinder (falsch negativ).

Im Subtest Genus zeigte sich, dass von den 381 Schülerinnen und Schülern 75 im SGF 2-Genus Risikokinder (19,7%) und 306 Kinder keine Risikokinder waren (80,3%). Davon wurden im

CFT 20-R 28 Kinder drei Jahre später im Wortschatztest wiederum als Risikokinder identifiziert (korrekt positiv), 46 der ehemals als Risikokinder identifizierten Kinder fielen im CFT 20-R nicht als Risikokinder auf (falsch positiv). Korrekt negativ als Nichtrisikokind wurden in beiden Testverfahren übereinstimmend 283 Kinder festgestellt. Nicht erkannt wurden im SGF 2 22 Kinder (falsch negativ).

Im Subtest Plural zeigte sich, dass von den 381 Schülerinnen und Schülern 85 im SGF 2-Plural Risikokinder (22,3%) und 296 Kinder keine Risikokinder waren (77,7%). Davon wurden im CFT 20-R 28 Kinder drei Jahre später im Wortschatztest wiederum als Risikokinder identifiziert (korrekt positiv), 57 der ehemals als Risikokinder identifizierten Kinder fielen im CFT 20-R nicht als Risikokinder auf (falsch positiv). Korrekt negativ als Nichtrisikokind wurden in beiden Testverfahren übereinstimmend 274 Kinder festgestellt. Nicht erkannt wurden im SGF 2 22 Kinder (falsch negativ).

Im Subtest Akkusativ zeigte sich, dass von den 381 Schülerinnen und Schülern 90 im SGF 2-Akkusativ Risikokinder (23,6%) und 291 Kinder keine Risikokinder waren (76,4%). Davon wurden im CFT 20-R 31 Kinder drei Jahre später im Wortschatztests wiederum als Risikokinder identifiziert (korrekt positiv), 59 der ehemals als Risikokinder identifizierten Kinder fielen im CFT 20-R nicht als Risikokinder auf (falsch positiv). Korrekt negativ als Nichtrisikokind wurden in beiden Testverfahren übereinstimmend 271 Kinder festgestellt. Nicht erkannt wurden im SGF 2 19 Kinder (falsch negativ).

Im Subtest Dativ zeigte sich, dass von den 381 Schülerinnen und Schülern 92 im SGF 2-Dativ Risikokinder (24,1%) und 289 Kinder keine Risikokinder waren (75,9%). Davon wurden im CFT 20-R 23 Kinder drei Jahre später im Wortschatztest wiederum als Risikokinder identifiziert (korrekt positiv), 69 der ehemals als Risikokinder identifizierten Kinder fielen im CFT 20-R nicht als Risikokinder auf (falsch positiv). Korrekt negativ als Nichtrisikokind wurden in beiden Testverfahren übereinstimmend 261 Kinder festgestellt. Nicht erkannt wurden im SGF 2 27 Kinder (falsch negativ).

Im Subtest Subjekt-Verb-Kongruenz zeigte sich, dass von den 381 Schülerinnen und Schülern 78 im SGF 2-Subjekt-Verb-Kongruenz Risikokinder (20,5%) und 303 Kinder keine Risikokinder waren (79,5%). Davon wurden im CFT 20-R 21 Kinder

drei Jahre später im Wortschatztests wiederum als Risikokinder identifiziert (korrekt positiv), 57 der ehemals als Risikokinder identifizierten Kinder fielen im CFT 20-R nicht als Risikokinder auf (falsch positiv). Korrekt negativ als Nichtrisikokind wurden in beiden Testverfahren übereinstimmend 273 Kinder festgestellt. Nicht erkannt wurden im SGF 2 29 Kinder (falsch negativ).

Wesentliche Kennwerte, welche Hinweise zur Vorhersagegüte eines Tests liefern, sind wie bereits erwähnt die Sensitivität, die Spezifität, der negative sowie der positive prädiktive Wert und der RAZ-Index (Hasselhorn/Marx et al. 2005; Marx/Lenhard 2010).

Die Sensitivität gibt den Anteil an Schülern an, der durch das Screening korrekt als Risikokind identifiziert wurde.

Die Spezifität gibt den Anteil an nicht betroffenen Personen an, der durch den Test korrekt als nicht betroffen klassifiziert wurde.

Die Prädiktortrefferquote (positiver prädiktiver Wert) gibt den Anteil der Risikokinder an, bei dem später tatsächlich Probleme auftreten.

Der negative prädiktive Wert gibt an, mit welcher Wahrscheinlichkeit ein als Nichtrisikokind klassifiziertes Kind auch später kein Risiko entwickelt.

Die Trefferquote gibt den Anteil der richtigen Klassifikationen an.

Die Zufallstrefferquote verweist auf den Anteil richtiger Vorhersagen, die auch zufällig sein können.

Der RAZ-Index gibt an, wie stark die Trefferquote durch ein Testverfahren im Vergleich zur Zufallstrefferquote ansteigt.

Während die Kennwerte einzeln nur bedingt aussagekräftig sind, gilt der RAZ-Index als der bedeutendste Kennwert für die Beurteilung der prognostischen Validität eines testdiagnostischen Verfahrens. In der Regel liegen die Werte zwischen 0 und 1, können jedoch auch negativ ausfallen. Üblich ist die Angabe des RAZ-Indexes in Prozent. Laut Jansen, Mannhaupt et al. (1999) bedeuten Werte  $< 0$ , dass eine zufällig vorgenommene Klassifikation bessere Ergebnisse als der Test erbringt,  $= 0$ , dass das Testverfahren keine bessere Vorhersage als eine zufällige Klassifikation erbringt und ab 33%, dass ein Testverfahren eine leichte Verbesserung gegenüber einer zufälligen Zuteilung und somit eine gute Klassifikationsleistung aufweist. Testverfahren mit einem RAZ-Index ab 66% stellen eine deutliche Verbesserung gegenüber zufälligen Befunden dar und weisen sehr gute prognostische Klassifikationsleistungen auf (siehe auch Dornheim 2008).

Tab. 6 stellt die Kennwerte der prognostischen Validität des SGF 2 im Überblick dar.



**Tab. 6:** Testgütekriterien prognostische Validität

Kriterium	Kennwert Gesamttest	Kennwert Genus	Kennwert Plural	Kennwert Akkusativ	Kennwert Dativ	Kennwert Subjekt-Verb-Kongruenz
Sensitivität	.62	.56	.56	.62	.46	.42
Spezifität	.888	.86	.828	.821	.791	.827
positiver prädiktiver Wert	.456	.378	.329	.344	.25	.269
negativer prädiktiver Wert	.939	.928	.926	.934	.906	.904
Trefferquote	.853	.821	.793	.795	.747	.774
Zufalls-trefferquote	.737	.724	.704	.694	.69	.717
RATZ-Index	.537	.453	.434	.502	.288	.27

Die Sensitivität ergibt sich aus dem Verhältnis der richtig positiv klassifizierten Kinder bezogen auf die Gesamtheit der 381 Kinder mit niedrigen Wortschatzleistungen ( $PR \leq 15$ ) zum Ende der Klasse 4. Es zeigt sich, dass das SGF 2-Gesamttest 62 % der Kinder mit späteren geringen Wortschatzleistungen korrekt identifizieren konnte (für die Subtests: zwischen .42 und .62). Die  $\beta$ -Fehlerquote ist mit 38 % (für die Subtests: zwischen 58 % und 38 %) nicht gering, erklärt sich jedoch durch den relativ langen Zeitraum von drei Jahren, der zwischen den beiden Erhebungen liegt sowie durch den Unterschied, der zwischen einer morphologischen Leistung (SGF 2) und einer Wortschatzleistung (CFT 20-R) liegen kann. Die Sensitivität ist vor diesem Hintergrund für ein Screeningverfahren für den Gesamttest durchaus zufriedenstellend.

Die Spezifität gibt Aufschluss darüber, wie hoch das Verhältnis von richtig negativ klassifizierten Kindern bezogen auf die Gesamtheit der Kinder ohne schulische Minderleistungen im Wortschatz ( $PR > 15$ ) zum Ende der Klasse 4 ist. Es ergeben sich Spezifitätswerte von .89 im Gesamttest (für die Subtests: zwischen .791 und .86). Das bedeutet, dass der Prozentsatz der Kinder ohne Risiko in der Sprachleistung, die auch als unauffällig eingestuft wurden 89 % (für die Subtests: zwischen 86 % und 79%) beträgt. Die  $\alpha$ -Fehlerquote bildet als Komplement zur Spezifität den prozentualen Anteil falsch positiver Klassifikationen ab. Die  $\alpha$ -Fehlerquote ist mit einem Wert von 11 % relativ gering, für die Subtests (zwischen 21 % und 14 %) jedoch etwas höher.

Der positive prädiktive Wert gibt an, mit welcher Wahrscheinlichkeit ein anhand des Prädiktors (SGF 2) als auffällig klassifiziertes Kind auch tat-

sächlich innerhalb der Kriteriumsmessung mit dem CFT 20-R auffällig wird. Die Prädiktortrefferquote liegt im Gesamttest bei .46 und verweist darauf, dass 46 % der Kinder, welche im SGF 2 zu Beginn der Klasse 2 als auffällig klassifiziert wurden, zum Ende der Klasse 4 tatsächlich unterdurchschnittliche Sprachleistungen aufweisen (für die Subtests: zwischen .25 und .38). Der negative prädiktive Wert gibt an, mit welcher Wahrscheinlichkeit ein anhand des Prädiktors als unauffällig klassifiziertes Kind später auch tatsächlich innerhalb der Kriteriumsmessung unauffällig bleibt. Mit einem Wert von .94 gelingt es mit dem SGF 2 im Gesamttest korrekterweise als unauffällig eingestufte Kinder nach drei Jahren zu 94 % zu klassifizieren (für die Subtests: zwischen .93 und .90).

Als einer der bedeutsamsten Kennwerte zur Vorhersagegüte von Testverfahren wurde abschließend der RATZ-Index ermittelt, welcher den relativen Anstieg gegenüber einer Zufallstrefferquote unter Berücksichtigung der maximalen Trefferquote angibt. Der RATZ-Index liegt für den Gesamttest mit .54 im guten bis zufriedenstellenden Bereich, d.h. dass die Vorhersagen auf Basis des SGF 2 sich gegenüber einer zufälligen Zuordnung verbessern und das Verfahren als Gesamttest eine gute Klassifikationsleistung aufweist. Für die Subtests liegen die RATZ-Indizes zwischen .27 und .50 und sind damit deutlich geringer als im Gesamttest. Dies erklärt sich wiederum durch den relativ langen Zeitraum von drei Jahren, der zwischen den beiden Erhebungen liegt sowie durch den Unterschied, der zwischen einer morphologischen Leistung (SGF 2) und einer Wortschatzleistung (CFT 20-R) liegen kann.

# Literatur

- Bortz, J., Döring, N. (2006): Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. 4. Aufl. Springer, Berlin
- Dornheim, D. (2008): Prädiktion von Rechenleistung und Rechenschwäche: Der Beitrag von Zahlen-Vorwissen und allgemein-kognitiven Fähigkeiten. Logos, Berlin
- Elben, C.E., Lohaus, A. (2000): Marburger Sprachverständnistest (MSVK). Hogrefe, Göttingen
- Fox, A. (2011): TROG-D. Test zur Überprüfung des Grammatikverständnisses. 5. Aufl. Schulz-Kirchner, Idstein
- Hasselhorn, M., Marx, H., Schneider, W. (Hrsg.) (2005): Diagnostik von Mathematikleistungen. Tests und Trends. Bd. 4. Hogrefe, Göttingen
- Jansen, H., Mannhaupt, G., Marx, H., Skowronek, H. (1999): Bielefelder Screening zur Früherkennung von Lese-Rechtschreib-Schwierigkeiten (BISC). Hogrefe, Göttingen
- Krajewski, K., Küspert, P., Schneider, W., Visé, M. (2002): Deutscher Mathematiktest für erste Klassen (DEMAT 1+). Hogrefe, Göttingen
- Küspert, P., Schneider, W. (1998): Würzburger Leise Leseprobe (WLLP). Hogrefe, Göttingen
- Lenhard, W., Lenhard, A. (2014): Berechnung von Testgütekriterien für Screeningverfahren. In: <http://www.psychometrica.de/testkennwerte.html>, 09.09.2015
- Mahlau, K. (2010): Elternfragebogen zur Anamnese der Sprachentwicklung. Material der Universität Rostock. In: [www.lernfortschrittsdokumentation-mv.de/pdf-lounge/Elternfragebogen\\_Sprachentwicklung.pdf](http://www.lernfortschrittsdokumentation-mv.de/pdf-lounge/Elternfragebogen_Sprachentwicklung.pdf), 09.09.2015
- Mannhaupt, G. (2006): Münsteraner Screening zur Früherkennung von Lese-Rechtschreibschwierigkeiten (MÜSC). Cornelsen, Berlin
- Marx, H., Lenhard, W. (2010): Diagnostische Merkmale von Screeningverfahren. In: Hasselhorn, M., Schneider, W. (Hrsg.): Frühprognose schulischer Kompetenzen. Tests und Trends. Bd. 9. Hogrefe, Göttingen, 68–84
- Petermann, F. (2010): Sprachstandserhebungstest für Fünf- bis Zehnjährige (SET 5–10). Hogrefe, Göttingen
- Stock, C., Schneider, W. (2008): Deutscher Rechtschreibtest für das erste und zweite Schuljahr (DERET 1–2+). Hogrefe, Göttingen
- Weiß, R. (2008): CFT 20-R mit WS/ZF-R. Grundintelligenztest Skala 2 – Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgentest – Revision (WS/ZF-R). Hogrefe, Göttingen
- Weiß, R., Osterland, J. (1997): Grundintelligenztest CFT 1 – Skala 1. 5., revidierte Aufl. Hogrefe, Göttingen