



Mehrdimensionale Diagnostik von Textverständnis im Vorschulalter mit der Timogeschichte – eine explorative Untersuchung

Simone Janda & Marina Heel

Zusammenfassung

Hintergrund: Aufgrund fehlender Diagnostikverfahren im Vorschulalter entwickelten Heel und Janda (2011) einen Geschichtentest – die Timogeschichte – zur mehrdimensionalen Überprüfung des Textverständnisses in Anlehnung an die im angloamerikanischen Raum vorhandene „*splish splash story*“ (u.a. Skarakis-Doyle & Dempsey 2008a,b). Die Timogeschichte umfasst drei Untertests *Verständnisfragen (VF)*, *Gemeinsames Nacherzählen (GN)* und *Entdeckung von Erwartungsverletzungen (EV)*.

Fragestellung/Zielsetzung: Im Rahmen einer explorativen Studie wurde dieses Verfahren mit der Timogeschichte auf seine Anwendbarkeit erprobt, u.a. hinsichtlich Sensitivität, Spezifität und Itemanalysen sowie der Tauglichkeit der einzelnen Untertests. Daneben wurden Erkenntnisse zum Textverstehen erhoben, u.a. wie verschiedene Dimensionen des Textverständnisses miteinander bzw. mit anderen sprachlichen Ebenen zusammenhängen.

Methodik: Im Rahmen der Studie wurde das neue Verfahren mit 88 monolingualen Kindern im Alter von 4;0 – 6;11 Jahren durchgeführt. Zusätzlich wurden die Kinder mit zwei Vergleichstests - dem TROG-D zur Überprüfung des Satzverständnisses und dem SETK 3-5 (bis auf den Untertest Verstehen von Sätzen) zur Einschätzung der allgemeinen sprachlichen Fähigkeiten – getestet.

Ergebnisse: Das Verfahren erzielt gute Werte im Bereich Spezifität, aber nicht ausreichende Werte im Bereich Sensitivität. Alle Untertests erwiesen sich jedoch für die Sechsjährigen als nicht mehr geeignet, um auffällige und unauffällige Kinder zu differenzieren. Eine Beeinträchtigung der Leistung in den VF und im GN zieht meist auch geringere Leistungen im Bereich Monitoring des Sprachverstehens (EV) nach sich. Im Gegensatz dazu können Monitoringprozesse auch isoliert beeinträchtigt sein. Mit anderen sprachlichen Bereichen gibt es nur geringe Zusammenhänge zur Textverständnisleistung.

Diskussion: Die Sensitivität muss mittels einer Überarbeitung des Verfahrens noch weiter verbessert werden. Die Überprüfungsmethode *Gemeinsames Nacherzählen* ist vor allem für jüngere Kinder bis 5;6 Jahre geeignet, wobei dies in einer größeren und ausgewogeneren Stichprobe erneut überprüft werden sollte. Weitere Erkenntnisse weisen darauf hin, dass Sprachverständnisdefizite auf Textebene früh entdeckt werden können und sich in der Entwicklung manifestieren, weshalb ein früher Therapiebeginn indiziert ist.

Schlüsselwörter

Sprachverständnis, Textverstehen, Monitoring des Sprachverstehens (MSV), Diagnostik, informelles Testverfahren, Vorschulkinder

Abstract

Background: As there are no valid diagnostic instruments to evaluate text comprehension in preschoolers, Heel and Janda (2011) developed a story test – the *Timostory*. The test was constructed on the basis of the “*splish splash story*” (among others: Skarakis-Doyle & Dempsey 2008 a,b) and the practicality was tested in an explorative study, during which new results regarding text comprehension were obtained.

It consists of three subtests: *comprehension questions (VF)*, *Joint Story Retell (GN)* and the *Expectancy Violation Detection Task (EV)*.

Aims: In an explorative study the diagnostic procedure was evaluated regarding among others sensitivity, specificity, item analyses and the suitability of the individual subtests. Furthermore, information about text comprehension were gained, e.g. how different dimensions of text comprehension and different levels of language profession are connected.

Methods: The new method was carried out with 88 children aged 4;0 – 6;11 years. Additionally we used to tests to compare the results with: TROG-D to evaluate the comprehension of sentences and SETK 3-5 (without the subtest for sentence comprehension) to get an impression of the children's general language abilities.

Results: The specificity of the *Timostory* showed positive results whereas the sensitivity was not acceptable. Yet all subtests were not suitable to differentiate normal developed six year olds from children with problems in text comprehension. If children achieve bad results in *VF* and *GN*, they tend to have problems in the *EV* as well. By contrast, it is possible that only comprehension monitoring is reduced. Other language levels show little discernible connections with text comprehension.

Discussion: Sensitivity has to be improved with the help of a general revision of the test. The Joint Story Retell is appropriate for young children up to age 5;6, but this result should be verified with a larger and more adequate sample of children. The results show that difficulties in text comprehension can be detected in very young children and manifest themselves in the further development. Therefore early therapy is indicated.

Keywords

Language comprehension, text comprehension, comprehension monitoring, diagnostics, informal test, preschoolers

Hintergrund

Die Diagnostik des Textverstehens beschränkte sich im deutschsprachigen Raum oftmals auf das Schulalter, häufig kombiniert mit Schriftsprache. Dabei verfügen bereits Kinder im Vorschulalter bei altersgemäßen Texten über ein grundlegendes Geschichtenverständnis (Hachul & Schönauer-Schneider 2012, Skarakis-Doyle & Dempsey 2008a). Da eine frühzeitige Erfassung und Intervention der Textverständnisseleistungen zur Vorbeugung von sich manifestierenden Einschränkungen, wie Lesesinnverständnisschwierigkeiten (u.a. Skarakis-Doyle & Dempsey 2008, Van den Broek et al. 2005) bis ins Schul- und Jugendalter jedoch dringend indiziert ist, entwickelten Heel und Janda (2011) im Rahmen einer Masterarbeit ein Verfahren, welches die Kompetenzen von Vorschulkindern auf mehreren Ebenen überprüft. Während das Diagnostikinstrument an anderer Stelle (Praxis Sprache) später ausführlich vorgestellt wird, steht im vorliegenden Artikel die Darstellung der zur Erprobung des Verfahrens durchgeführten explorativen Studie im Vordergrund.

Überblick über das Verfahren

Die Timogesichte ist ein Verfahren zur Überprüfung des Textverständnisses von Kindern im Alter von 4;0 bis 6;11 Jahren und stellt eine Adaption der im englischen Sprachraum evaluierten *splash splash story* (Skarakis-Doyle & Dempsey 2008, Dempsey 1999) dar. Das Testinstrument basiert auf einer Bilderbuchgeschichte, die das Routineereignis des Zubettgehens aufgreift und sowohl inhaltlich als auch sprachlich unter Berücksichtigung wissenschaftlicher Erkenntnisse und Theorien auf die Fähigkeiten von Vorschulkindern abgestimmt wurde. Das Textverständnis wird mehrdimensional anhand von drei Untertests überprüft:

- Untertest 1 *Verständnisfragen (VF)*: 11 Fragen zur Geschichte zur Überprüfung der Fähigkeit zur faktischen Informationsentnahme (z. B. „Wer hat vergessen die Zähne zu putzen?“ – Timo) und zur Inferenzbildung (z. B. „Warum sucht Papa den Teddy?“ – Weil Timo den Teddy zum Schlafen braucht)
- Untertest 2 *Gemeinsames Nacherzählen (GN)*: Auslassen relevanter Informationen in einem „mündlichen Lückentext“. Die inhaltliche Genauigkeit der kindlichen Antworten reflektiert die Verständnisseleistung (z. B. „Draußen ist es schon _____ (dunkel).“)
- Untertest 3 *Entdeckung von Erwartungsverletzungen (EV)*: Überprüfung des Monitoring des Sprachverstehens durch das Einfügen von Inkonsistenzen oder Ersetzen bzw. Austauschen bestimmter Elemente der Geschichte (z. B. Personen, Objekte, Handlungen), die das Kind erkennen muss (z. B. „Gut, dass meine Zähne jetzt *schmutzig* (sauber) sind“)

Fragestellungen

Die der Studie zugrunde gelegten Fragestellungen beziehen sich einerseits auf die Erprobung und Anwendbarkeit des entwickelten Verfahrens sowie der einzelnen Untertests, andererseits sollten Erkenntnisse zum Textverständnis im Vorschulalter gewonnen werden. Relevante Fragestellungen werden im Folgenden dargestellt.

- **Sensitivität und Spezifität:** Zur Überprüfung, inwiefern anhand der Timogesichte Kinder mit altersgemäßen Textverständnisleistungen von Kindern mit beeinträchtigtem Textverständnis differenziert werden könnten, wurden Sensitivität und Spezifität des Verfahrens ermittelt. Die Sensitivität bezeichnet dabei die Wahrscheinlichkeit, mit der Kinder mit Beeinträchtigungen im Textverstehen durch die Timogesichte auch als solche erkannt werden. Die Spezifität zeigt die Wahrscheinlichkeit auf, mit der unauffällige Kinder korrekt als unauffällig identifiziert werden.
- **Analyse der Untertests:** Im Rahmen einer Itemanalyse wurden Reliabilitätskoeffizienten, Trennschärfe und Itemschwierigkeiten analysiert.
- **Ausschluss expressiver Sprachleistungen auf die Verstehensleistungen:** Dempsey & Skarakis-Doyle (2001) gehen davon aus, dass Kinder mit expressiven Sprachstörungen im Untertest *Gemeinsames Nacherzählen* besser abschneiden als bei den Verständnisfragen, die auf dieselben Inhalte abzielen, weil die vollständig eigenständige Produktion von Inhalten für diese Kinder schwieriger sein kann, als das Ergänzen einer spezifischen Information innerhalb eines Lückensatzes. Der Einfluss expressiver Sprachleistungen auf die Testleistung soll mit Hilfe der neu entwickelten Überprüfungsmethode, die nur die Ergänzung von Sätzen erfordert und nicht die vollständige Satzproduktion, minimiert werden, wodurch die tatsächliche Verständnisleistung besser abgebildet werden soll. In der Studie wurde überprüft, ob diese Annahme zutreffend ist.
- **Entwicklungssensitivität:** Zugrundeliegend ist die Annahme, dass sich das im Vorschulalter entwickelnde Textverständnis bei einem Vergleich der Altersgruppen in einer Verbesserung der Leistung in den verschiedenen Untertests der Timogesichte abzeichnet. Eine Leistungssteigerung wird in Anlehnung an Mathieu (1998, 2009) insbesondere auch für logische Verknüpfungsfragen erwartet, welche die Fähigkeit zur Inferenzbildung erfassen. Darüber hinaus interessieren mögliche Unterschiede in der Entwicklung der Textverständnisfähigkeiten von Kindern mit rezeptiver Sprachauffälligkeit im Vergleich zur Gesamtstichprobe, welche ebenfalls ermittelt werden sollten.
- **Zusammenhänge zwischen Teilleistungsbereichen im Textverständnis:** Die drei Untertests der Timogesichte untersuchen verschiedene Aspekte des Textverstehens: Speicherung und Abruf von Informationen, Bilden von logischen Verknüpfungen und Monitoring des Sprachverstehens. Von Interesse ist, inwieweit diese Fähigkeiten zusammenhängen und ob die Kinder das gleiche Leistungsniveau in allen Bereichen aufweisen oder ob die Teilleistungsbereiche auch selektiv beeinträchtigt sein können.
- **Zusammenhänge zwischen Leistungen im Textverständnis und anderen sprachlichen Bereichen:** Untersucht wurde außerdem, ob die in den Untertests gezeigten Textverständnisleistungen mit anderen sprachlichen Bereichen zusammenhängen. Florit et al. (2009) und Montgomery et al. (2009) beschreiben das phonologische Arbeitsgedächtnis als Voraussetzung für gutes Textverstehen, weshalb der Untertest *Phonologisches Arbeitsgedächtnis für Nichtwörter* aus dem SETK 3-5 (Sprachentwicklungstest für drei bis fünfjährige Kinder, Grimm et al. 2010) herangezogen wurde, um diese Annahme zu überprüfen. Das Satzverstehen sollte laut Dempsey & Skarakis-Doyle (2001) keine Rückschlüsse auf das Textverstehen ermöglichen, was aus dem Vergleich mit den Leistungen im TROG-D (Fox 2006) hervorgehen sollte.

Erprobung

Studiendesign

Vor der Erprobung des neuen Verfahrens an einer größeren Stichprobe wurde eine Pilotstudie mit neun sprachlich unauffälligen Kindern durchgeführt, woraufhin ein Item und der Aufbau der Protokollbögen verändert wurde. Für die Hauptuntersuchung wurden drei verschiedene Tests verwendet, die mit Pausen und teilweise an verschiedenen Tagen durchgeführt wurden, um eine Überanstrengung der Kinder und einen Konzentrations- oder Motivationsabfall zu vermeiden. Dabei handelte es sich um die zu evaluierende Timogesichte sowie zwei standardisierte und normierte Sprachtests als Vergleichstests. Zum einen diente der TROG-D (Fox 2006) zur Feststel-

lung einer möglichen Sprachverständnisstörung auf Satzebene. Zum anderen wurden die Untertests *Morphologische Regelbildung (MR)*, *Phonologisches Arbeitsgedächtnis für Nichtwörter (PGN)*, *Satzgedächtnis (SG)* sowie die *Gedächtnisspanne für Wortfolgen (GW)* des SETK 3-5 (*Sprachentwicklungstest für drei bis fünfjährige Kinder*: Grimm et al. 2010) zur Erfassung der Fähigkeiten im Bereich Sprachproduktion und Sprachgedächtnis durchgeführt. Von der zusätzlichen Durchführung des Untertests *Verstehen von Sätzen* im SETK 3-5 wurde abgesehen, da dieser Teilbereich bereits durch den TROG-D abgedeckt wurde.

Die Testungen erfolgten in einem ruhigen Raum. Die Vergleichstests wurden von einer Untersucherin durchgeführt, während bei der Durchführung des Geschichtenverfahrens zwei Testleiterinnen anwesend waren. Wegen der fehlenden Möglichkeit, die Testung auf Video aufzuzeichnen, sollte so eine sichere Erfassung der nonverbalen Reaktionen der Kinder beim Untertest *Entdeckung von Erwartungsverletzungen* gewährleistet sein.

Die erhobenen Daten wurden anhand verschiedener statistischer Verfahren (SPSS Statistics 19) analysiert.

Beschreibung der Stichprobe

Die Stichprobe umfasst 88 monolingual deutsche Kinder im Alter von 4;0 - 6;11 Jahren. Voraussetzung für die Aufnahme in die Stichprobe war eine altersgemäße allgemeine Entwicklung, die durch ein Eltern- und Erzieherurteil bestätigt wurde. Sprachliche Schwierigkeiten stellten dabei kein Ausschlusskriterium dar. Die Probanden stammten aus dem Raum Kempten/Allgäu und München. Testungen wurden in Kindergärten, sprachtherapeutischen Praxen und einer schulvorbereitenden Einrichtung durchgeführt.

Die Probanden wurden in fünf Altersgruppen und vier Auffälligkeitsgruppen je nach Leistung in den Vergleichstests unterteilt:

Table 1: Stichprobe

Alter	Geschlecht		Auffälligkeit				Gesamt
	Männlich	Weiblich	keine	SETK 3-5	TROG-D	SETK 3-5 & TROG-D	
4;0-4;11	8	7	7	6	1	1	15
5;0-5;5	7	9	10	4	1	1	16
5;6-5;11	14	12	16	8	0	2	26
6;0-6;5	11	9	11	6	1	2	20
6;6-6;11	7	4	4	3	3	1	11
Gesamt	47	41	48	27	6	7	88
In %	53,4 %	46,6 %	54,5 %	30,7 %	6,8 %	8 %	100 %

Die vier Auffälligkeitsgruppen sind:

- *sprachlich normal entwickelte Kinder* (Testwerte in allen durchgeführten Untertests des SETK 3-5 und im TROG-D im Bereich durchschnittlicher Leistung oder darüber),
- *rezeptiv und produktiv auffällige Kinder* (Testwerte in mindestens einem Untertest des SETK 3-5 und im TROG-D unterhalb des Bereichs durchschnittlicher Leistung),
- *rezeptiv auffällige Kinder* (Testwert unterhalb des Bereichs durchschnittlicher Leistung im TROG-D)
- *produktiv auffällige Kinder* (Testwert in mindestens einem Untertest des SETK 3-5 unterhalb des Bereichs durchschnittlicher Leistung)

Nicht in allen Fällen konnten sämtliche Vergleichsdaten erhoben werden, so dass zwar mit allen Kindern der TROG-D durchgeführt wurde, krankheitsbedingt aber nur von 82 Kindern die Vergleichsdaten des SETK 3-5 vorliegen.

Darstellung und Erläuterung der Ergebnisse

Sensitivität und Spezifität der Untertests

Der Berechnung von Sensitivität und Spezifität eines Diagnostikinstrumentes wird grundsätzlich ein sogenannter Goldstandard zugrunde gelegt, der hinsichtlich des zu messenden Konstrukts als zuverlässig eingestuft wird und daher als Bezugsgröße für einen Vergleich mit dem zu überprüfenden Verfahren herangezogen werden kann. Da im deutschsprachigen Raum bisher jedoch kein evaluiertes Diagnostikinstrument zur Überprüfung der Textverständnisleistungen im Vorschulalter existiert, erwies sich die Ermittlung der Sensitivität und der Spezifität der Untertests der Timogeschichte als problematisch. Die Leistungen anderer sprachlicher Bereiche können nur bedingt als Bezugsgröße dienen, da trotz einer gewissen Abhängigkeit des allgemeinen Sprachentwicklungsstandes sowie der verschiedenen Ebenen des Sprachverständnisses eine Gleichsetzung der unterschiedlichen Konstrukte im Grunde nicht möglich ist (zur Notwendigkeit der diagnostischen Unterscheidung von Wort-, Satz- und Textverstehen siehe auch Hachul & Schönauer-Schneider 2012).

Aufgrund fehlender Alternativen wurde der Berechnung von Sensitivität und Spezifität in einem ersten Schritt die Bezugsgröße Sprachauffälligkeit, in einem weiteren Schritt die Bezugsgröße Rezeptive Sprachauffälligkeit zugrunde gelegt. Zu berücksichtigen ist, dass diese nur bedingt passenden, eigentlich zu unspezifischen Vergleichsdaten jedoch lediglich eine ungefähre Einschätzung erlauben.

Wie in Diagnostikstudien üblich wird zur Ermittlung von Sensitivität und Spezifität eine Vierfeldertafel herangezogen. Die Berechnung der Sensitivität erfolgt, indem die Anzahl aller vom zu überprüfenden Verfahren richtig als auffällig eingestuften Kinder durch die Anzahl aller hinsichtlich der Bezugsgröße auffälligen Kinder (positiv und negativ) dividiert wird. Daraus ergibt sich die Formel $a / (a + c)$. Die Spezifität wird über die Formel $d / (b + d)$ ermittelt. Die Anzahl aller als richtig unauffällig eingestuften Kinder wird dabei durch die Anzahl aller Unauffälligen (falsch positiv und richtig negativ) geteilt. Werte zwischen 80 und 90 % werden dabei im Allgemeinen als akzeptabel betrachtet.

		Bezugsgröße	
		(Rezeptive) Sprachauffälligkeit	
		Auffällig	Unauffällig
Untertest der Timogeschichte	Auffällig	Richtig positiv a	Falsch positiv b
	Unauffällig	Falsch negativ c	Richtig negativ d
		a + c alle Auffälligen	b + d alle Unauffälligen

Unter Heranziehung der Bezugsgröße *Sprachauffälligkeit* konnten hinsichtlich der Sensitivität keine zufriedenstellenden Werte erreicht werden (Untertest VF 30 %, Untertest GN 35 %, Untertest EV 30 %). Hinsichtlich der Spezifität wurden dagegen gute Werte ermittelt (Untertest VF 91,6 %, Untertest GN 89,6 %, Untertest EV 89,6 %). Unter Ausschluss der Kinder mit rein expressiven sprachlichen Auffälligkeiten kann für die Bezugsgröße Rezeptive Sprachauffälligkeit eine Verbesserung der Sensitivität (Untertest VF 38,4 %, Untertest GN 38,4 %, Untertest EV 46,1 %) bei gleichbleibender Spezifität erreicht werden. Ausgehend von diesen Werten und der erläuterten Definition würde der Untertest Verständnisfragen beispielsweise also 38,4 % der tatsächlich textverständnisgestörten Kinder auch als auffällig erkennen. 91,6 % der hinsichtlich ihrer Textverständnisleistung unauffälligen Kinder werden dagegen auch als unauffällig eingestuft.

Trotz der Verbesserung der Werte unter Heranziehung der Bezugsgröße *Rezeptive Sprachauffälligkeit* im Vergleich zur noch ungenaueren Bezugsgröße Sprachauffälligkeit sind die ermittelten Werte im Bereich Sensitivität weiterhin nicht ausreichend. Die Ergebnisse verdeutlichen jedoch eindrücklich die Abhängigkeit der Sensitivität von der Bezugsgröße. Die rezeptive Sprachauffälligkeit unter Ausschluss der Kinder mit rein expressiven Störungen scheint zwar als Bezugsgröße besser geeignet, jedoch wurden wie unter dem Punkt Stichprobe beschrieben rezeptiv sprachauffällige Kinder anhand ihrer Leistungen im Satzverständnis klassifiziert, was nur bedingt Rückschlüsse auf ihre Textverständnisleistung zulässt und daher nicht in ausreichendem Maße den erhobenen Konstrukten, nämlich verschiedenen Aspekten des Textverständnisses,

entspricht. Unter Berücksichtigung dieses Sachverhaltes müsste das oben ausformulierte Beispiel demnach korrekterweise lauten: 38,4% der im Satzverständnis auffälligen Kinder werden auch vom Untertest *Verständnisfragen* als (im Textverständnis) auffällig eingestuft. 91,6% der im Satzverständnis unauffälligen Kinder zeigen dagegen im Untertest *Verständnisfragen* auch auf Textebene keine Auffälligkeiten.

Die Aussagekraft der Werte ist somit eingeschränkt. Vielmehr sind die ermittelten Daten als weiterer Hinweis zu werten, dass die Leistungen im Satz- und Textverständnis nicht gleichzusetzen sind.

Analyse der Untertests

Die drei Untertests *Verständnisfragen*, *Gemeinsames Nacherzählen* und *Entdeckung von Erwartungsverletzungen* wurden hinsichtlich der Aufgabenkennwerte Itemschwierigkeit, Trennschärfe und Reliabilität untersucht. Dabei besteht ein Zusammenhang zwischen den drei genannten Größen. Bei mittlerer Itemschwierigkeit steigt die Wahrscheinlichkeit für hohe Trennschärfen an, womit auch eine höhere Reliabilität erwartet werden kann (Bühner 2011). Die *Itemschwierigkeit* bezeichnet den prozentualen Anteil der richtigen Antworten für ein Item (Lienert & Raatz 1998), wobei Werte zwischen 20 und 80% als akzeptabel gelten. Mittlere Itemschwierigkeiten differenzieren dabei am besten zwischen Probanden mit guten bzw. schlechten Leistungen. Um auch Randbereiche erfassen zu können, sollte ein Testverfahren jedoch auch Items enthalten, die extremere Schwierigkeitsbereiche abdecken. Für vier- und fünfjährige Kinder konnten bei den Untertests *Verständnisfragen* (Tab. 2) und *Entdeckung von Erwartungsverletzungen* (Tab. 4) insgesamt adäquate sowie ausreichend breit gestreute Schwierigkeiten erzielt werden, wohingegen die Items des Untertests *Gemeinsames Nacherzählen* (Tab. 3) durchweg geringe Itemschwierigkeiten aufwiesen und somit nur von wenigen Kindern nicht korrekt gelöst werden konnten. Für die Sechsjährigen erwiesen sich alle Untertests hinsichtlich der Schwierigkeit ihrer Items als zu leicht.

Fehlende Werte weisen darauf hin, dass das jeweilige Item in der entsprechenden Altersgruppe von allen Kindern beantwortet werden konnte und deshalb nicht weiter berücksichtigt wurde.

Table 2: Itemschwierigkeiten Verständnisfragen

Item	Itemschwierigkeiten der Verständnisfragen					Gesamt
	4;0-4;11	5;0-5;5	5;6-5;11	6;0-6;5	6;6-6;11	
VF-1-Info	,60	,63	,69	,95	,55	,70
VF-2-Info	,93	,88	,96			,95
VF-3-Info	,53	,56	,73	,70	,91	,68
VF-4-Info	,73	,69	,65	,80	,64	,70
VF-5-Log	,73	,94	,88	,95	,91	,90
VF-6-Info	,73	,69	,88	,95		,90
VF-7-Info	,73	,50	,92	,95	,91	,85
VF-8-Log	,40	,44	,81	,80	,91	,69
VF-9-Info	,87	,56	,96			,97
VF-10-Info	,33	,81	,69	,70	,55	,57
VF-11-Log	,53	,50	,50	,80	,73	,61
VF-12-Log	,73	,63	,96	,90		,89
VF-13-Log	,27	,88	,65	,70	,82	,59

Table 3: Itemschwierigkeiten des Gemeinsamen Nacherzählens

Item	Itemschwierigkeiten des Gemeinsamen Nacherzählens				Gesamt
	4;0-4;11	5;0-5;5	5;6-5;11	6;0-6;11	
GN-1	,67	,75	,77	,84	,77
GN-2	,20	,19	,19	,19	,19
GN-3	,87	,75	,92	,84	,85
GN-4	,93				,99
GN-5					
GN-6	,60	,69	,54	,87	,69
GN-7	,67	,63	,69	,90	,75
GN-8			,92		,98

Tabelle 4: Itemschwierigkeiten der Entdeckung von Erwartungsverletzungen

Item	Itemschwierigkeiten der Entdeckung von Erwartungsverletzungen					Gesamt
	4;0-4;11	5;0-5;5	5;6-5;11	6;0-6;5	6;6-6;11	
EV-1	,40	,25	,42	,65	,73	,48
EV-2	,47	,69	,81	,90		,77
EV-3	,73	,88	,96			,92
EV-4	,80	,88	,88	,95		,90
EV-5	,73	,75	,81	,95	,73	,81
EV-6	,73	,81	,96			,91
EV-7	,47	,69	,73	,75	,73	,68
EV-8	,53	,69	,65	,80	,55	,66

Die *Trennschärfe* wird durch die Korrelation eines einzelnen Items mit der Gesamtskala eines Tests angezeigt. Dadurch wird eine Einschätzung ermöglicht, wie gut ein Item zwischen Personen mit niedriger und hoher Merkmalsausprägung im Sinne des Gesamttests differenziert. Probanden, die im Gesamttest eine gute Leistung erzielen, müssten ein Item also korrekt beantworten, Probanden mit einem schlechteren Gesamtergebnis eher nicht. Die Trennschärfe umfasst einen Wertebereich von -1 bis 1. Je höher der Wert, desto genauer erfasst das Item dasselbe Kriterium wie der Gesamttest. Eine Trennschärfe von 0 besagt, dass die Aufgabe von guten und schlechten Kindern ungefähr gleich häufig korrekt beantwortet wird (Lienert & Raatz 1998). Werte unter 0,2 sind dabei nicht mehr akzeptabel.

Auch bei der Analyse der Trennschärfe der Items wurde deutlich, dass keiner der Untertests geeignet ist, um im Alter von sechs Jahren im Textverständnis auffällige Kinder von unauffälligen zu differenzieren. Für die vier- und fünfjährigen Kinder erwiesen sich die Trennschärfen für den Untertest *Entdeckung von Erwartungsverletzungen* (Tab. 7) als ausreichend, ebenso für den Untertest *Verständnisfragen* (Tab. 5) mit Ausnahme einzelner Items. Für den Untertest *Gemeinsames Nacherzählen* (Tab. 6) lagen die ermittelten Trennschärfen dagegen nicht im zufriedenstellenden Bereich.

Bei einigen Items existieren für manche Altersgruppen Trennschärfen im negativen Wertebereich. Diese sind jedoch immer sehr nahe am Wert 0, was auf eine zufällige Entstehung des Wertes aufgrund der Stichprobe hindeutet.

Tabelle 5: Trennschärfen der Verständnisfragen

Item	Trennschärfen der Verständnisfragen					Gesamt
	4;0-4;11	5;0-5;5	5;6-5;11	6;0-6;5	6;6-6;11	
VF-1-Info	,171	,689	,424	,046	,104	,370
VF-2-Info	-,177	-,300	,569	,000	,000	,052
VF-3-Info	,453	,612	-,086	,000	,704	,318
VF-4-Info	,469	,297	-,005	,062	,292	,178
VF-5-Log	,469	,000	,142	-,300	,704	,249
VF-6-Info	,275	,123	,627	-,300	,000	,318
VF-7-Info	,606	,473	,057	-,132	,000	,380
VF-8-Log	,365	,497	,406	-,137	,000	,412
VF-9-Info	,515	,000	-,065	,000	,000	,248
VF-10-Info	,282	,635	,477	,195	,104	,435
VF-11-Log	,453	,380	,325	,292	,225	,366
VF-12-Log	,093	,325	,569	,225	,000	,324
VF-13-Log	,148	,555	,037	,426	,028	,335

Tabelle 6: Trennschärfen des Gemeinsamen Nacherzählens

Item	Trennschärfen des Gemeinsamen Nacherzählens					Gesamt
	4;0-4;11	5;0-5;5	5;6-5;11	6;0-6;5	6;6-6;11	
GN-1	,073	,126	,066	-,134	,361	,057
GN-2	,254	,563	-,025	-,102	,327	,161
GN-3	,547	,420	-,114	,250	-,069	,192
GN-4	,000	,000	,000	,000	,000	,023
GN-5	,000	,000	,000	,000	,000	,000
GN-6	,538	,191	,000	-,128	-,054	,170
GN-7	,444	-,355	,077	,093	,361	,094
GN-8	,000	,500	,036	,000	,000	,179

Tabelle 7: Trennschärfen der Entdeckung von Erwartungsverletzungen

Item	Trennschärfen der Entdeckung von Erwartungsverletzungen					Gesamt
	4;0-4;11	5;0-5;5	5;6-5;11	6;0-6;5	6;6-6;11	
EV-1	,508	,376	,239	,072	,000	,320
EV-2	,663	,870	,627	-,338	,000	,644
EV-3	,337	,737	,612	,000	,000	,565
EV-4	,430	,737	,355	,014	,000	,475
EV-5	,589	,842	,627	-,252	,000	,554
EV-6	,502	,847	,612	,000	,000	,660
EV-7	,663	,870	,461	,041	-,479	,507
EV-8	-,101	,734	,472	,147	-,278	,309

Die *Reliabilität* misst die Zuverlässigkeit eines Verfahrens, das heißt die Wahrscheinlichkeit, mit der bei einer Wiederholung der Testung das gleiche Ergebnis (auffällig/unauffällig) erzielt werden würde. Verwendet wurde hierbei einerseits Cronbachs Alpha, welches ein Maß für die interne Konsistenz einer Skala darstellt, also die durchschnittliche Korrelation zwischen den Items überprüft. Der Wertebereich dieser Größe liegt zwischen $-\infty$ und 1, jedoch sind nur Werte über .65 akzeptabel. Zudem wurde die Split-Half-Reliabilität berechnet, bei welcher der Test in zwei Hälften geteilt und Rohwertpaare korreliert werden (Bühner 2011). Hier liegt der mögliche Wertebereich zwischen -1 und 1, wobei Werte möglichst nahe am Wert 1 wünschenswert sind.

Sowohl Cronbachs Alpha (Tab. 8) als auch die Split-Half-Reliabilität (Tab. 9) lagen für die Untertests *Verständnisfragen* und *Entdeckung von Erwartungsverletzungen* in einem für eine erste Testversion zufriedenstellendem Bereich. Der Untertest *Gemeinsames Nacherzählen* erreichte dagegen keinen ausreichenden Wert.

Tabelle 8: Cronbachs Alpha

Untertest	Cronbachs Alpha
Verständnisfragen	.682
Gemeinsames Nacherzählen	.288
Entdeckung von Erwartungsverletzungen	.733

Tabelle 9: Split-Half-Reliabilität

Untertest	Spearman-Brown
Verständnisfragen	.663
Gemeinsames Nacherzählen	.266
Entdeckung von Erwartungsverletzungen	.773

Die Analyse der Untertests zeigte somit auf, dass sich die Untertests *Verständnisfragen* und *Entdeckung von Erwartungsverletzungen* für die Gruppe der vier- und fünfjährigen Kinder zur Erfassung der Textverständnisleistung insgesamt gut eignen. Der Untertest *Gemeinsames Nacherzählen* dagegen hatte nur wenig Aussagekraft. Darüber hinaus erwiesen sich für die Gruppe der Sechsjährigen alle drei Untertests als nicht ausreichend hinsichtlich der Differenzierung von im Textverständnis auffälligen und unauffälligen Kindern. Zudem konnten konkrete Überarbeitungsimpulse bezüglich der einzelnen Items der Untertests gewonnen werden.

Entwicklungssensitivität

Anhand deskriptiver Analysen sowie einer einfaktoriellen Varianzanalyse (ANOVA) wurde die Leistungssteigerung in den drei Untertests der Timogesichte unter Berücksichtigung des Alters untersucht. Darüber hinaus wurde eine Untergruppe gebildet, die ausschließlich Kinder mit rezeptiven Sprachauffälligkeiten einschloss, um eine eventuell diskrepante Entwicklung der Text-

verständnisleistungen von Kindern mit Sprachverständnisstörung im Vergleich zur Gesamtstichprobe festhalten zu können.

Entwicklungsfortschritte der Gesamtstichprobe

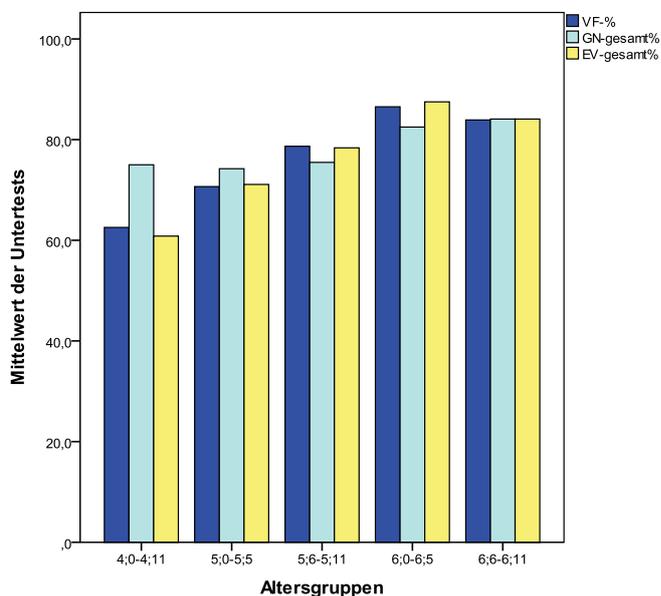


Abbildung 1: Entwicklungssensitivität bei der Gesamtstichprobe

Wie in Abbildung 1 ersichtlich, ergab sich für die Untertests *Verständnisfragen* und *Entdeckung von Erwartungsverletzungen* mit zunehmendem Alter der Probanden ein deutlicher Anstieg der Testleistung. Für den Untertest *Verständnisfragen* konnte ein hochsignifikanter Unterschied der Testmittelwerte in den fünf Altersgruppen ($F(4,83) = 1,875, p = .122$) belegt werden, ebenso für den Untertest *Entdeckung von Erwartungsverletzungen* ($F(4,83) = 3,424, p = .012$). Im Gegensatz dazu zeigte sich im Untertest *Gemeinsames Nacherzählen* eine weitaus geringere Leistungssteigerung der älteren Kinder im Vergleich zu den jüngeren Testteilnehmern. Auch statistisch zeigte sich kein signifikanter Unterschied der Mittelwerte zwischen den Altersgruppen ($F(4,83) = 1,875, p = .122$).

Insbesondere die Untertests *Verständnisfragen* und *Entdeckung von Erwartungsverletzungen* erwiesen sich somit als entwicklungs sensitiv, die sich verbessernden Textverständnisleistungen im Vorschulalter zeichnen sich in den beiden Aufgabentypen statistisch nachweisbar ab.

Entwicklungsfortschritte der rezeptiv auffälligen Kinder

Für die Untergruppe der rezeptiv auffälligen Kinder dagegen konnte trotz einer in der deskriptiven Analyse ebenfalls sichtbaren Steigerung der Leistungen in allen drei Untertests lediglich eine vergleichsweise geringere Verbesserung der Testergebnisse mit zunehmendem Alter gemessen werden.

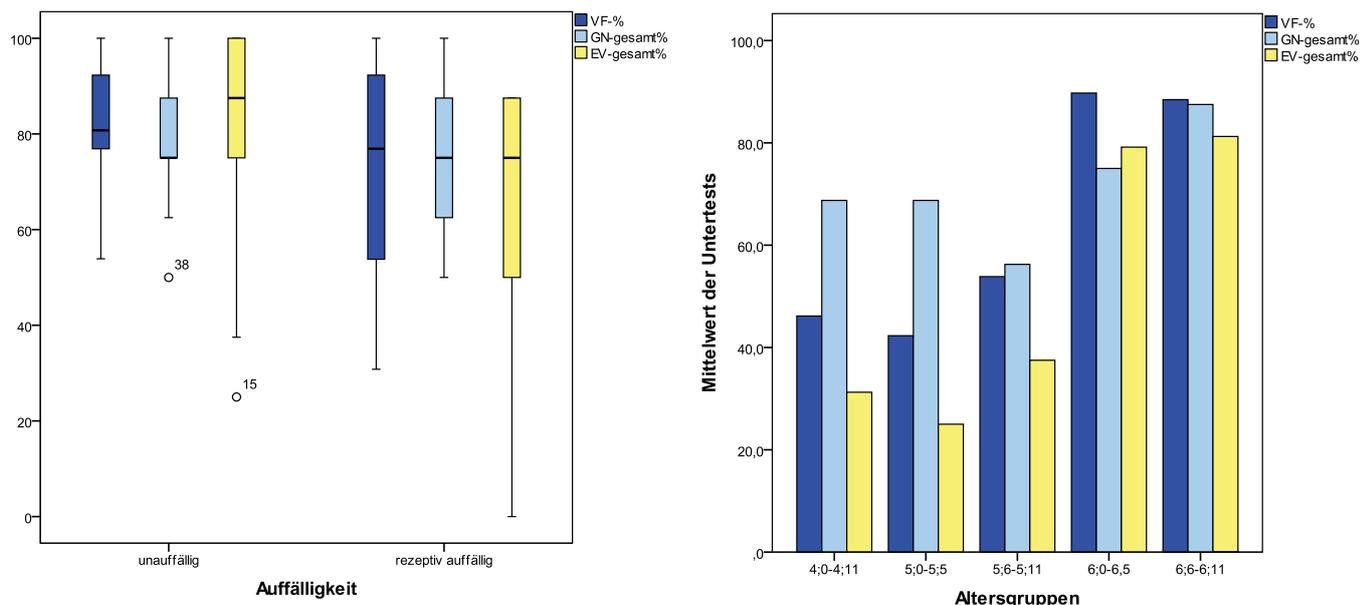


Abbildung 2: Entwicklungssensitivität für rezeptiv auffällige Kinder

Für den Untertest *Verständnisfragen* konnte anhand einer einfaktoriellem Varianzanalyse ein signifikanter Unterschied zwischen den Mittelwerten der Gesamtgruppe und der Gruppe rezeptiv auffälliger Kinder belegt werden ($F(1,59) = 5.950, p = .018$). Auch rein auf die Gruppe der Kinder mit Sprachverständnisstörung bezogen erweist sich die Mittelwertdifferenz zwischen den Altersgruppen für diesen Untertest als signifikant ($F(4,8) = 4.827, p = .028$). Der Untertest *Entdeckung von Erwartungsverletzungen* weist ebenfalls einen hochsignifikanten Mittelwertunterschied zwischen der Gesamtgruppe und den rezeptiv auffälligen Kindern auf ($F(1,59) = 12.796, p = .001$). Zwischen den Altersgruppen gibt es jedoch keinen signifikanten Unterschied ($F(4,8) = 2.980, p = .088$). Eine Entwicklung im Sinne einer signifikanten Verbesserung der Leistungen ist somit bei rezeptiv auffälligen Kindern für den Untertest *Entdeckung von Erwartungsverletzungen* nicht gegeben. Beim Untertest *Gemeinsames Nacherzählen* konnte weder ein signifikanter Mittelwertunterschied zwischen der Gesamtgruppe und der Gruppe rezeptiv auffälliger Kinder ($F(1,59) = 2.559, p = .115$), noch zwischen den Altersgruppen der Kinder mit rezeptiver Auffälligkeit ($F(4,8) = 3.427, p = .065$) festgestellt werden.

Wie bereits ausgehend von den auf die Gesamtstichprobe bezogenen Analysen zu erwarten, erwies sich der Untertest *Gemeinsames Nacherzählen* somit als wenig sensitiv hinsichtlich der Einschätzung des Entwicklungsstandes rezeptiv auffälliger Kinder und ließ darüber hinaus kaum Rückschlüsse auf den Entwicklungsverlauf der Einschränkungen im Textverständnis zu. Die Ergebnisse der Analysen zu den Untertests *Verständnisfragen* und *Entdeckung von Erwartungsverletzungen* dagegen sprechen dafür, dass rezeptiv auffällige Kinder wie erwartet auch im Bereich des Textverständnisses geringere Fähigkeiten aufweisen als Kinder ohne rezeptive Sprachauffälligkeiten. Darüber hinaus ergaben sich deutliche Hinweise darauf, dass bei Kindern mit Sprachverständnisstörung vor allem die Leistungen im Monitoring des Sprachverstehens (überprüft anhand des Untertests *Entdeckung von Erwartungsverletzungen*) vergleichsweise stärker beeinträchtigt sind und sich zudem auch langsamer verbessern als bei rezeptiv unauffälligen Kindern.

Entwicklung der Inferenzbildung

Nach Mathieu (1998, 2009) ist die Überprüfung der Inferenzbildung bzw. der Fähigkeit, aus den explizit im Text genannten Inhalten Schlussfolgerungen auf implizite Informationen ziehen zu können, für die Einschätzung der Textverständnisleistung wesentlich, da nur darüber Rückschlüsse auf tiefgreifendes Verständnis gezogen werden können. Die Beantwortung von faktischen Informationsfragen, für deren Beantwortung lediglich explizit genannte Inhalte abgerufen werden müssen, fällt insbesondere sprachverständnisgestörten Kindern leichter.

Für die im Untertest *Verständnisfragen* überprüften logischen Verknüpfungsfragen zeigten sich in der statistischen Analyse signifikante Mittelwertsunterschiede hinsichtlich der Altersgruppen ($F(4,83) = 5.845, p < .001$). Mit steigendem Alter fiel den teilnehmenden Probanden die Beantwor-

tung der Inferenzfragen demnach zunehmend leichter. Bei der Betrachtung der Gruppe rezeptiv auffälliger Kinder im Vergleich zur Gesamtstichprobe konnte dagegen keine derartige Leistungssteigerung festgestellt werden: Die ermittelte Mittelwertdifferenz zwischen den Altersgruppen erwies sich für diese Untergruppe als nicht signifikant ($F(4,8) = 1.420, p = .311$).

Die Entwicklungssensitivität der zur Überprüfung herangezogenen logischen Verknüpfungsfragen des Untertests *Verständnisfragen* ebenso wie eine deutliche Leistungssteigerung bei der Inferenzbildung mit zunehmendem Alter konnte demnach für die Gesamtstichprobe bestätigt werden. Die Ergebnisse für die Untergruppe der rezeptiv auffälligen Kinder dagegen deuten klar darauf hin, dass die Fähigkeit zur Inferenzziehung bei Kindern mit Sprachverständnisstörungen in besonderem Maße und längerfristig beeinträchtigt ist.

Ausschluss expressiver Sprachleistungen auf die Testleistung

Dempsey & Skarakis-Doyle (2001) gehen davon aus, dass die expressiven Sprachfähigkeiten nur wenig zur Leistung beim *Gemeinsamen Nacherzählen* beitragen. Diese Annahme konnte von ihnen für sprachlich unauffällige Kinder im Alter von 30–50 Monaten bestätigt werden. Um diese Hypothese auch für deutschsprachige unauffällige Kinder zu untersuchen und darüber hinaus auch mit produktiv eingeschränkten Kindern zu vergleichen, inwiefern der expressive Sprachleistungsstand der Kinder die Ergebnisse in den zwei verschiedenen Testmodalitäten *Verständnisfragen* (faktische Informationsfragen) und *Gemeinsames Nacherzählen* beeinflusst, wurden die Ergebnisse zweier Stichprobengruppen miteinander verglichen:

- rezeptiv und produktiv unauffällige Kinder
- produktiv auffällige Kinder, die keine rezeptiven Einschränkungen aufweisen.

Produktiv unauffällige Kinder

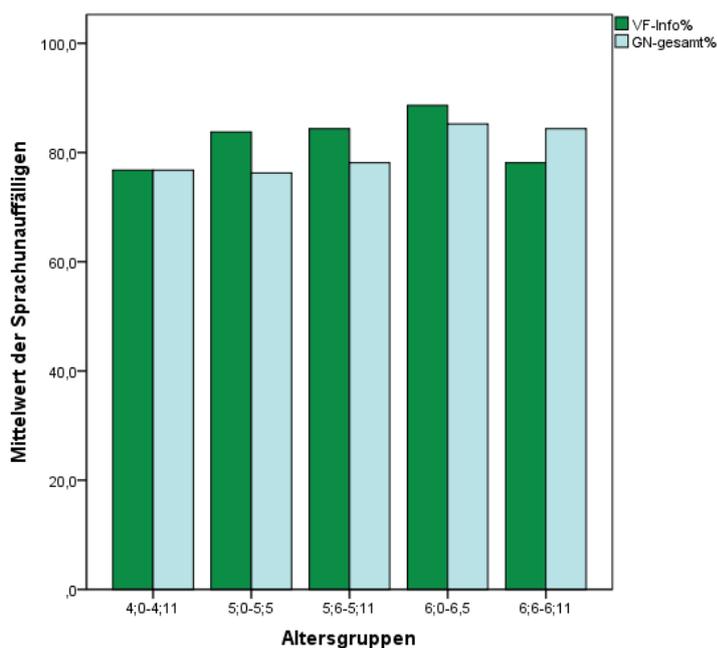


Abbildung 3: Vergleich VF Info – Gemeinsames Nacherzählen für unauffällige Kinder und sprachlich auffällige Kinder

Die Leistungen der sprachlich unauffälligen Kinder steigen bei den *faktischen Informationsfragen* zunächst leicht graduell an. Lediglich in der Gruppe der 6;6 – 6;11-Jährigen ist ein Leistungsabfall zu verzeichnen. Beim *Gemeinsamen Nacherzählen* bleiben die Werte in den Altersgruppen von 4;0 – 5;11 Jahren fast gleich. Ein Anstieg erfolgt erst bei den Sechsjährigen. Nur bei den 6;6-6;11-Jährigen sind die Ergebnisse im Untertest *Gemeinsames Nacherzählen* besser als bei den *Verständnisfragen*. Ein Mittelwertvergleich durch eine Varianzanalyse der Leistungen in den einzelnen Altersgruppen ergab, dass die Leistungsschwankungen nicht signifikant sind ($F(4,43) = 1,274, p = .295$ für *VF-Info* und $F(4,43) = 1,515, p = .316$ für *GN*).

Das *Gemeinsame Nacherzählen* bietet für sprachlich unauffällige Kinder gegenüber den *Verständnisfragen* keinen signifikanten Vorteil, was dafür spricht, dass beide Testmethoden die Leistungen im Textverstehen gleich gut messen. Lediglich in der Gruppe der 6;6-6;11-Jährigen

wurde ein besseres Ergebnis im Untertest *Gemeinsames Nacherzählen* erreicht als bei den *Verständnisfragen*. Ein Grund dafür könnte sein, dass diese Kinder, die aufgrund ihrer allgemeinen Entwicklung noch nicht eingeschult wurden, auch sprachlich nicht so weit entwickelt sein könnten wie ihre Altersgenossen und ihnen daher das *Gemeinsame Nacherzählen* leichter fällt. Da die Entwicklung der Aufgabenstellung den Einfluss von expressiven Sprachfähigkeiten reduzieren sollte, könnten gerade diese schon älteren Kinder davon profitieren und somit bessere Ergebnisse im *Gemeinsamen Nacherzählen* erreichen als bei den *faktischen Informationsfragen*.

Sprachproduktiv auffällige Kinder

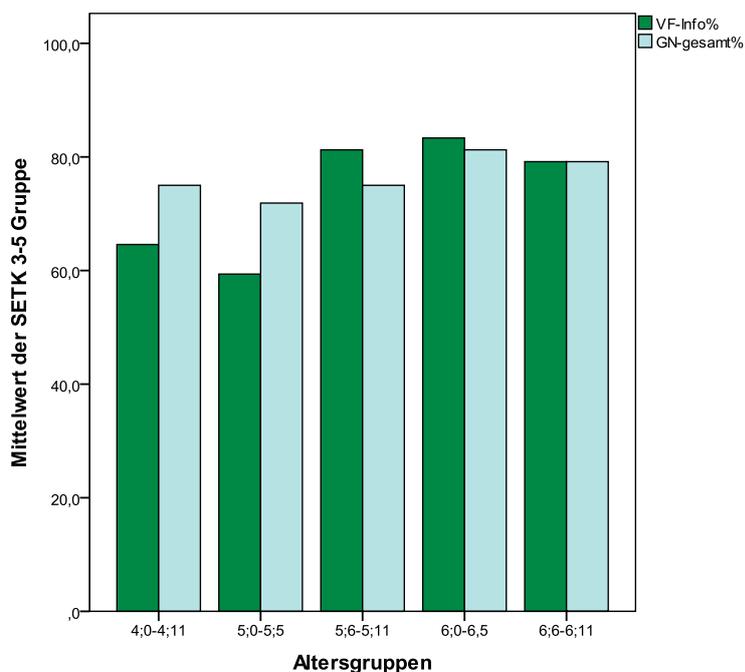


Abbildung 4: Vergleich VF-Info – Gemeinsames Nacherzählen: produktiv auffällige Kinder

In der Gruppe der Kinder, die im SETK 3-5 eine unterdurchschnittliche Leistung erzielten, ergab der Leistungsvergleich ein anderes Bild. Der Mittelwertunterschied ist jedoch auch hier nicht signifikant ($T(4,22) = 1,150, p = .360$ für *VF-Info* und $T(4,22) = .189, p = .942$ für *GN*). Der Leistungsanstieg bei den *Informationsfragen* ist weniger graduell als bei den unauffälligen Kindern. Ein deutlicher Leistungssprung ist zwischen den 5;0-5;5 und 5;6-5;11-Jährigen erkennbar. Erst im höheren Alter schneiden die Kinder bei den *Verständnisfragen* besser ab als bei der Satzergänzung, bis schließlich kein Unterschied mehr feststellbar ist. Die 4;0-5;5-Jährigen sind deutlich besser beim *Gemeinsamen Nacherzählen* als bei den *faktischen Informationsfragen*. Dieser Unterschied spricht dafür, dass für diesen Altersbereich die expressiven Sprachfähigkeiten weniger Einfluss beim *Gemeinsamen Nacherzählen* besitzen und bestätigt somit die Annahme, dass das *Gemeinsame Nacherzählen* die Textverständnisleistung von diesen unabhängiger überprüft. Der Test scheint für jüngere Kinder, in diesem Fall sogar bis zum Alter von 5;5 Jahren, zur Überprüfung der Textverstehensleistung gut geeignet zu sein. Ab 5;6-6;11 wird der Einfluss der expressiven Sprachleistungen allerdings relativiert. Die Kinder schnitten entweder gleich gut oder sogar besser bei *faktischen Informationsfragen* ab als beim *Gemeinsamen Nacherzählen*. Die expressiv sprachlichen Leistungen scheinen nun soweit entwickelt zu sein, dass der Einfluss des Überprüfungsformats aufgehoben wird. Die Entwicklung der expressiven Sprachfähigkeiten in diesem Alter würde auch für den sprunghaften Leistungsanstieg bei den *faktischen Informationsfragen* ab 5;6 Jahren sprechen.

Direktvergleich

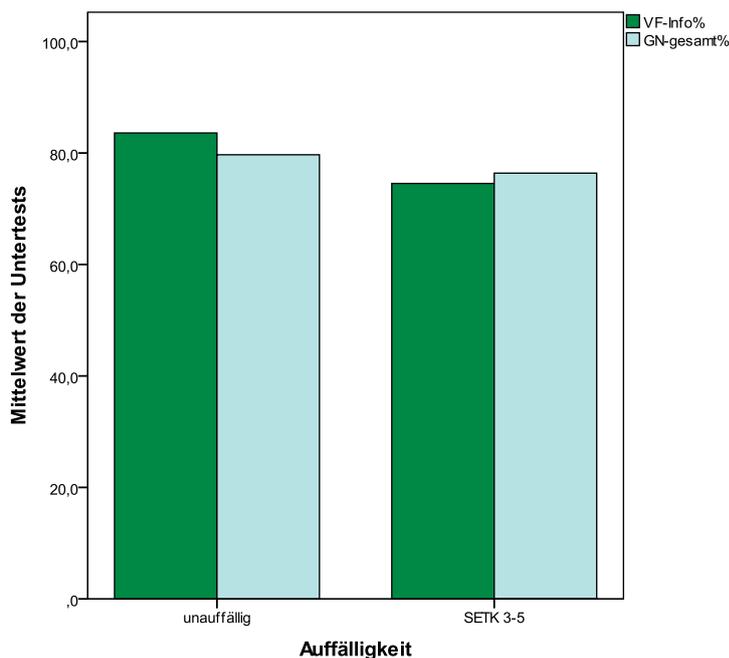


Abbildung 5: Vergleich VF-Info – Gemeinsames Nacherzählen

Der Mittelwert aller unauffälligen Probanden und aller im SETK 3-5 auffälligen Probanden ist in etwa gleich, da sich der Effekt der expressiven Sprachleistungen mit zunehmendem Alter aufhebt. Die Leistungen im Untertest *Gemeinsames Nacherzählen* schwankten sowohl in der Gruppe der unauffälligen als auch der auffälligen Kinder kaum. Die Mittelwerte sind zwar altersbedingt signifikant verschieden, jedoch ergab der Direktvergleich der Mittelwerte aller Probanden lediglich für die *faktischen Informationsfragen* ein signifikant besseres Ergebnis der sprachlich unauffälligen Kinder ($T(73) = 2,229, p = .029$). Für das *Gemeinsame Nacherzählen* war dies nicht der Fall, so dass eine gute Differenzierungsleistung aufgrund des Testwerts nicht mehr gewährleistet werden kann.

Der Untertest *Gemeinsames Nacherzählen* scheint demnach nur so lange dafür geeignet, das Textverstehen weniger von expressiven Sprachfähigkeiten beeinträchtigt zu überprüfen, bis diese weit genug entwickelt sind, dass die Kinder auch bei den *Verständnisfragen* gute Leistungen erbringen können. Dies scheint ab 5;6 Jahren der Fall zu sein. Zu bedenken ist zudem, dass sich der Untertest *Gemeinsames Nacherzählen* ausgehend von der Datenanalyse als zu leicht erwies (siehe Punkte *Entwicklungssensitivität* und *Analyse der Untertests*), weshalb die besseren Leistungen im Untertest GN im Vergleich zu den *Verständnisfragen* zum Teil auch darauf zurückführbar scheinen. Weder diese Tatsache noch der Einfluss der sprachlichen Fähigkeiten können allerdings ausgeschlossen werden, so dass ein eindeutiger Rückschluss auf den tatsächlichen Grund der teilweise besseren Leistungen im *Gemeinsamen Nacherzählen* nicht möglich ist.

Zusammenhänge zwischen Teilleistungsbereichen im Textverständnis

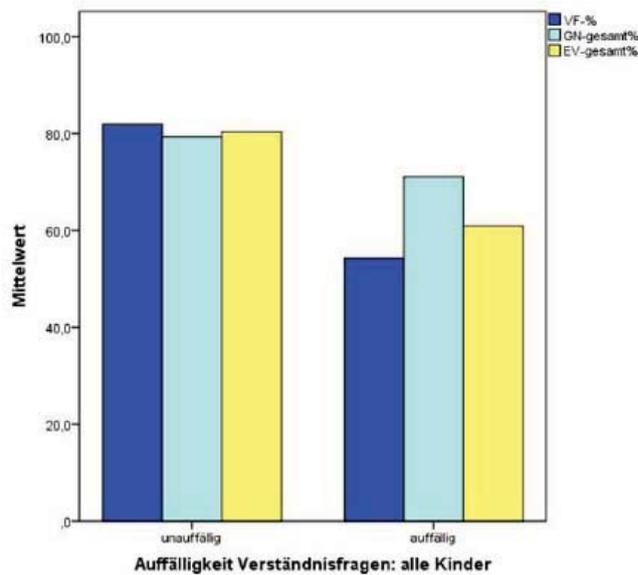
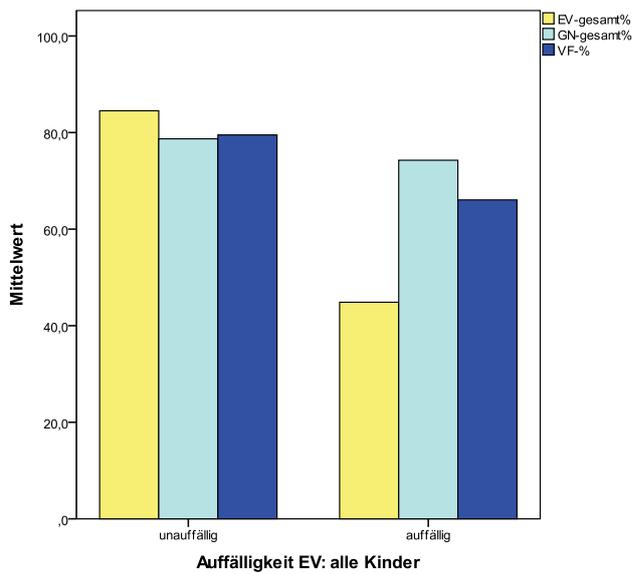
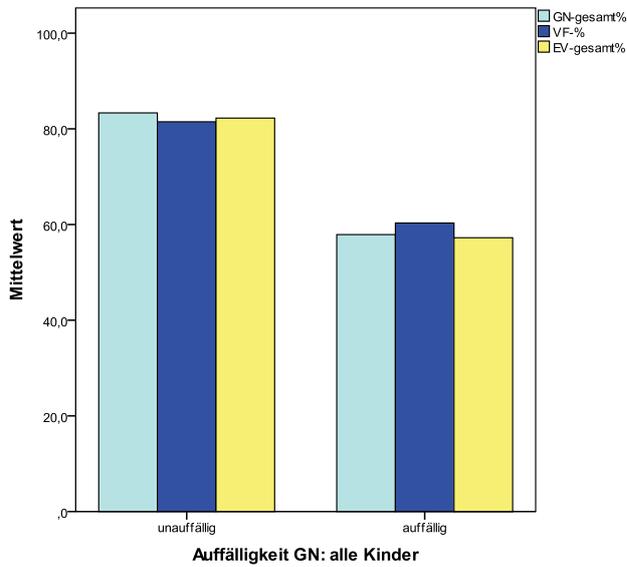


Abbildung 6: Teilleistungsbereiche im Textverständnis – Vergleich von VF, GN und EV

Beim Vergleich der Leistungen in den drei Untertests der Timogesichte konnte festgestellt werden, dass Kinder, die im Untertest *Verständnisfragen* oder *Gemeinsames Nacherzählen* auffällig abschnitten, auch in den anderen beiden Untertests ähnlich schlechte Leistungen erzielten. Der Mittelwertunterschied ist hierbei signifikant zwischen den auffälligen und unauffälligen Leistungen:

- Vergleich des Leistungsniveaus bei den Verständnisfragen bei allen Kindern: $T(86) = 6,720$, $p < .001$ für VF, $T(86) = 3,027$, $p = .003$ für EV und $T(86) = 1,903$, $p = .028$ für GN
- Vergleich des Leistungsniveaus im Gemeinsamen Nacherzählen bei allen Kindern: $T(86) = 5,059$, $p < .001$ für VF, $T(86) = 11,425$, $p < .001$ für GN, $T(86) = 4,363$, $p < .001$ für EV

Lediglich bei der *Entdeckung von Erwartungsverletzungen* ergibt sich ein anderes Bild: Die Kinder, die hier auffällig niedrige Punktzahlen erreichten, haben in den beiden anderen Untertests deutlich bessere Leistungen erbracht. Der Mittelwertvergleich zwischen den auffälligen und unauffälligen Leistungen ergibt nur für die Untertests VF und EV signifikante Unterschiede ($T(86) = 2,840$, $p = .006$ für VF, $T(86) = 7,883$, $p < .001$ für EV). Dieses Bild verstärkt sich auch, wenn nur die Gruppe der rezeptiv auffälligen Kinder betrachtet wird. Kinder mit Satzverständnisstörungen sind vor allem im Untertest EV deutlich schlechter als die Kinder mit regelrechtem Sprachverständnis.

Interessant ist dabei die Beobachtung, dass die Verstehensleistung in den *Verständnisfragen* und dem *Gemeinsamen Nacherzählen* auch einen Einfluss auf das Monitoring des Sprachverstehens zu haben scheint. Schneiden die Kinder schlecht in den beiden ersten Untertests ab, sind auch die Monitoring-Kompetenzen beeinträchtigt. Ist also das Textverstehen an sich defizitär, scheint als Konsequenz auch die Sprachverstehenskontrolle nicht zu funktionieren.

Im Gegensatz dazu kann es durchaus sein, dass Kinder im Bereich Monitoring schlecht abschnitten, jedoch in den beiden anderen Untertests keine deutlich schlechteren Ergebnisse erzielen als die Kinder mit guter Sprachverstehenskontrolle. Das Monitoring des Sprachverstehens kann demnach isoliert beeinträchtigt sein, ohne dass gleichzeitig ein gravierender Einfluss auf das allgemeine Textverstehen zurückzuführen ist. Eine Auffälligkeit im Bereich Satzverständnis scheint diese Tendenz noch zu erhöhen.

Zusammenhänge zwischen Leistungen im Textverständnis und anderen sprachlichen Bereichen

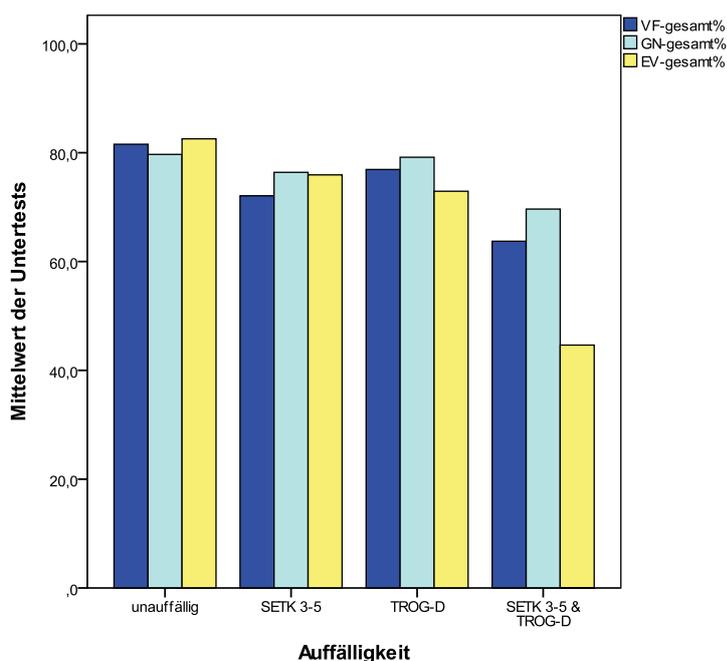


Abbildung 7: Zusammenhänge von Textverständnis und anderen sprachlichen Bereichen

Der deskriptive Vergleich der Mittelwerte zeigt, dass Kinder, die in einem oder beiden Vergleichstests auffällige Ergebnisse erreichten auch in den Tests der Timogeschichte schlechter abschnitten. Dabei sind Ergebnisse in den Untertests bei den entweder produktiv oder rezeptiv auffälligen Kindern in etwa auf einem Level. Die Kinder, die sowohl im TROG-D als auch im SETK 3-5 unterhalb des Normbereichs liegen, schneiden am schlechtesten ab, besonders im Untertest *Entdeckung von Erwartungsverletzungen*. Die Timogeschichte scheint für sprachlich auffällige Kinder schwieriger zu sein als für sprachlich unauffällige. Je stärker die Kinder in den Vergleichstests beeinträchtigt sind, desto schlechter sind auch die Leistungen in den Untertests der Timogeschichte, was für einen Zusammenhang zwischen den Leistungen in den verschiedenen Verfahren spricht.

Ein Mittelwertvergleich mittels einer univariaten Varianzanalyse ergab, dass sich die Mittelwerte der einzelnen Gruppen im Untertest *Entdeckung von Erwartungsverletzungen* signifikant unterscheiden ($F(3,84) = 5,891, p = .001$). Auch für die *Verständnisfragen* ist dies der Fall ($F(3,84) = 3,094, p = .031$). Die Leistungen in diesen beiden Untertests scheinen demnach von sprachlichen Einschränkungen beeinflusst zu sein, die Untertests somit gut zwischen sprachlich unauffälligen und auffälligen Kindern zu differenzieren. Lediglich der Vergleich der Mittelwerte im Untertest *Gemeinsames Nacherzählen* erbrachte keinen signifikanten Unterschied, was dafür spricht, dass die Aufgaben für alle Kinder gleich leicht zu beantworten sind.

Korrelationsanalysen bestätigten Zusammenhänge zwischen den Leistungen in der Timogeschichte und den Vergleichstests. Die höchste aufgeklärte Varianz (14,6%) und den stärksten Zusammenhang mit den Leistungen im TROG-D besitzt der Untertest *Entdeckung von Erwartungsverletzungen*. Bei den Untertests *VF* und *GN* ist die aufgeklärte Varianz bis zu 8,4% niedriger. Die Kapazität des Arbeitsgedächtnisses (abgeprüft im Untertest *Phonologisches Gedächtnis für Nichtwörter* aus dem SETK 3-5) hängt am stärksten mit dem *Gemeinsamen Nacherzählen* zusammen (13,8% aufgeklärte Varianz). Von beiden Vergleichstests am unabhängigsten zeigte sich der Untertest *Verständnisfragen*, wohingegen die *Entdeckung von Erwartungsverletzungen* die stärksten Zusammenhänge aufweist. Insgesamt sind jedoch alle Korrelationen in einem niedrigen Bereich einzuordnen, woraus sich schließen lässt, dass die Timogeschichte und TROG-D sowie SETK 3-5 verschiedene sprachliche Konstrukte zu messen scheinen. Ziel war es, ein Überprüfungsinstrument zu erstellen, welches das Textverständnis möglichst unabhängig von anderen sprachlichen Domänen misst. Aufgrund der geringen Korrelationen und der geringen aufgeklärten Varianz zu den anderen sprachlichen Bereichen, scheint die Leistung in der Timogeschichte nur wenig von den untersuchten anderen sprachlichen Fähigkeiten abzuhängen und so eine unverzerrte Messung der Textverständniskompetenz zu liefern. Weder unterdurchschnittliche Leistungen im TROG-D noch im SETK 3-5 lassen also auf Schwierigkeiten beim Textverstehen schließen, weshalb eine separate Überprüfung dieser Fähigkeit indiziert ist.

Diskussion

Methodenkritik

Weder die Stichprobe noch die verwendeten Vergleichstest waren zur Erprobung des Verfahrens optimal geeignet. Die Stichprobe ist in ihren Untergruppen nicht ausgewogen: Zwar konnte eine in etwa ausgeglichene Geschlechteraufteilung erreicht werden, jedoch war es nicht möglich, aus jeder Altersstufe gleich viele Kinder zu akquirieren. Vor allem eine größere Gruppe vierjähriger Kinder wäre sicher interessant gewesen, da sich das Textverstehen in diesem Entwicklungsabschnitt rasant entwickelt. Außerdem waren mit zunehmendem Alter prozentual mehr Kinder von einer Sprachauffälligkeit betroffen, was daraus resultieren könnte, dass viele der sechsjährigen Teilnehmer vom Schuleintritt zurückgestellt wurden und dafür nicht selten sprachliche Probleme ausschlaggebend sind.

Die Ergebnisse der Vergleichstests führten zu einer Einteilung in vier Gruppen, die jedoch besonders im Falle der produktiv auffälligen Kinder schwierig war. Die Auswertung des SETK 3-5 liefert kein Gesamtergebnis, sondern lediglich untertestbezogene Normwerte. Der Zuordnung zur sprachproduktiv auffälligen Gruppe wurde ein recht strenges Kriterium zugrunde gelegt, nachdem bereits ein auffälliges Ergebnis in *einem* der durchgeführten Untertests des SETK 3-5 für die Zuordnung in diese Gruppe ausreichend war. Kinder, die sich im Grenzbereich zwischen Auffälligkeit und normaler Entwicklung befinden, könnten daher eventuell teilweise falsch zugeordnet worden sein. Darüber hinaus ist der SETK 3-5 nur für Kinder bis zum Alter von 5;11 Jahren normiert. In unserer Zielgruppe befanden sich jedoch auch sechsjährige Kinder. Wegen

der Vergleichbarkeit in der statistischen Auswertung erschien es nicht sinnvoll, diese mit einem anderen Verfahren zu testen. Mit den Sechsjährigen wurde deshalb lediglich der Untertest *Phonologisches Arbeitsgedächtnis für Nichtwörter (PGN)* durchgeführt, um die Leistungen der Kinder trotz fehlender Normwerte wenigstens grob einschätzen zu können. Dadurch musste für die ältesten Kinder ein Kriterium festgelegt werden, ab wann eine Leistung im Untertest *PGN* als auffällig zu werten ist. Ausgehend von den T-Werten der 5;0-5;11-jährigen Kinder wurde ein „Bereich durchschnittlicher Leistung“ für die 6;0-6;11-Jährigen festgelegt. Dieses Vorgehen ist nicht ganz unproblematisch, wurde aber nur für drei Probanden im Grenzbereich angewendet. Die restlichen sechsjährigen Kinder erzielten entweder ein selbst für Fünfjährige auffälliges Ergebnis, womit sie klar im auffälligen Bereich lagen, oder erzielten eine so deutlich überdurchschnittliche Leistung, dass sie eindeutig der Gruppe der sprachlich unauffälligen Kinder zugeordnet werden konnten. Aus organisatorischen Gründen konnten die Vergleichstests außerdem nicht durchgehend in der gleichen Reihenfolge durchgeführt werden. Auch die Pausen zwischen den einzelnen Testungen waren unterschiedlich lang. Durch Krankheitsfälle konnten zudem nicht alle Vergleichsdaten erhoben werden. Eine fehlende adäquate Bezugsgröße erschwerte darüber hinaus die korrekte Berechnung von Sensitivität und Spezifität, die erst dann möglich wird, wenn eine solche vorliegt.

Aufgrund der beschriebenen methodischen Mängel und zu geringen Größe der Stichprobe beschränkt sich die Gültigkeit der Ergebnisse somit auf die vorliegende Studie, zur Generalisierung sind weitere Erhebungen notwendig.

Veränderungen bei der Testentwicklung im Vergleich zur *splash splash story*

Ausgehend von einer genauen Analyse der *splash splash story* wurden bei der Übertragung ins Deutsche einige Veränderungen am Verfahren vorgenommen, die unserer Ansicht nach die Durchführbarkeit verbessert haben. Der Aufbau der Geschichtenstruktur blieb in etwa gleich, jedoch wurden bei den Untertests einige Details verändert. Zu Beginn jedes Untertests wurden zwei Übungsaufgaben eingefügt, die zur Sicherung des Aufgabenverständnisses dienen. Im Untertest *Verständnisfragen* wurde außerdem die Anzahl der Informationsfragen reduziert, dafür aber mehr Inferenzfragen in Anlehnung an Mathieu (2009) integriert. Diese haben zur gleichzeitigen Reduzierung der Ratewahrscheinlichkeit vor allem Ja/Nein-Fragen ersetzt. Darüber hinaus wurde der Test nicht wie bei der Forschungsgruppe um Skarakis-Doyle und Dempsey an 2;6-4;0-Jährigen durchgeführt, sondern bei passender sprachlicher Aufbereitung der Geschichte an einer Gruppe von Vorschulkindern im Alter von 4;0-6;11 Jahren. Da sich die Textverständnisleistungen in diesem Zeitraum rasch ausdifferenzieren und ein ausreichendes Textverständnis dann insbesondere zum Schuleintritt relevant ist, erschien dies sinnvoller.

Gütekriterien

Bei der Evaluation der *splash splash story* wurden durchwegs sehr gute Werte in Bezug auf die Gütekriterien erreicht. Wie bereits Dempsey (1999) und Dempsey & Skarakis-Doyle (2001) konnten wir ebenfalls eine Entwicklungssensitivität in den Tests *Verständnisfragen* und *Entdeckung von Erwartungsverletzungen* nachweisen, jedoch im Gegensatz zu dieser Forschungsgruppe nicht im Untertest *Gemeinsames Nacherzählen*. Spezifität und Sensitivität waren in der Untersuchung von Skarakis-Doyle et al. (2008) (VF: Spezifität 92 %, Sensitivität 92 %; GN: Spezifität 94 %, Sensitivität 73 %; EV: 92 % Spezifität, Sensitivität 91 %) durchwegs deutlich höher als in unserer Studie (VF: Sensitivität 38,4 %, Spezifität 91,6 %; GN: Sensitivität 38,4 %, Spezifität 89,6 %; EV: Sensitivität: 46,1 %, Spezifität 89,6 %). Auch das englischsprachige Verfahren wurde bisher nicht durch unabhängige Forschungsgruppen evaluiert. Eine Replizierung dieser hohen Werte durch unabhängige Gruppen wäre wünschenswert. Auch für den deutschsprachigen Raum besteht Bedarf nach weiteren Untersuchungen mit einer modifizierten Version der Timogeschichte.

Bessere Eignung des Untertests *Gemeinsames Nacherzählen* für expressiv sprachlich eingeschränkte Kinder

Unsere Ergebnisse können die Annahme, dass sprachlich expressiv beeinträchtigte Kinder in diesem Testformat ihre Verständnisleistungen besser wiedergeben, nur bedingt bestätigen. Jüngere Kinder bis zum Alter von 5;5 Jahren scheinen von der Methode *Gemeinsames Nacherzählen* zu profitieren. Für die Kinder höheren Alters scheint sich der Einfluss zu relativieren, was aber auch an der zu geringen Schwierigkeit der Items liegen könnte. Die neue Testmethode wurde von Dempsey & Skarakis-Doyle (2001) für Kinder ab 2 ½ Jahren entwickelt. In einer weiteren Über-

prüfung an Kindern ab 3;0 Jahren bis zum Alter von 5;5 Jahren sollte demnach erneut überprüft werden, ob sich die Annahmen wie bei Dempsey & Skarakis-Doyle (2001) für jüngere Kinder bestätigen. Hier sollten allerdings nicht wie in dieser Studie nur sprachunauffällige Kinder näher untersucht werden, sondern auch Kinder mit Sprachstörungen, um die Zusammenhänge zwischen sprachlichen Auffälligkeiten und der Testleistung beim *Gemeinsamen Nacherzählen* genauer zu evaluieren.

Inferenzbildung

Nach Mathieu (1995, 2009) lassen im Bereich der *Verständnisfragen* solche Fragen, die auf im Text nicht explizit genannte Inhalte abzielen und daher ein Verständnis von übergreifenden Zusammenhängen und logischen Verknüpfungen erfordern, im Vergleich zu faktischen Informationsfragen eindeutigere Rückschlüsse auf die Verständnisleistung zu. In der vorliegenden Untersuchung konnten Mathieus Ergebnisse bei ihrer Evaluierung der *Michi-Geschichte* (1995) sowie der *Anna-Geschichte* (2009) weitgehend bestätigt werden: Inferenzfragen scheinen eine größere Anforderung an die Verständnisleistung von Kindern darzustellen, da sie einerseits von den älteren Kindern besser beantwortet werden können als von den jüngeren Testteilnehmern, andererseits für Kinder mit rezeptiven Auffälligkeiten schwieriger zu beantworten sind als Fragen, die lediglich das Abrufen faktischer Informationen verlangen. Für die Differenzierung von hinsichtlich ihrer Textverständnisleistung auffälligen und unauffälligen Kindern scheinen Inferenzfragen daher eine geeignete Methode darzustellen.

Zusammenfassung, Fazit und Ausblick

Verfahrensbesprechung

Die Ergebnisse der statistischen Auswertung deuten darauf hin, dass bereits der bestehende erste Verfahrensentwurf eine informelle Einschätzung der Textverständnisleistungen von Vorschulkindern erlaubt. Dafür sprechen nicht nur die belegte Entwicklungssensitivität des Verfahrens, sondern auch die guten Kennwerte hinsichtlich der Reliabilität, insbesondere in den Untertests *Verständnisfragen* und *Entdeckung von Erwartungsverletzungen*. Die Bestätigung der Ergebnisse an einer umfangreicheren Stichprobe, die außerdem eine größere Anzahl an Vierjährigen enthalten sollte, ist dennoch erforderlich. Darüber hinaus ist eine Erweiterung des Altersbereichs auf dreijährige Kinder denkbar, wohingegen das Verfahren bei sechsjährigen Kindern nicht mehr verlässlich zwischen auffälligen und unauffälligen Leistungen differenzieren kann. Die Resultate der Itemanalyse lieferten jedoch erste Impulse für die Überarbeitung und Verbesserung des Verfahrens.

Die Art des Verfahrens, insbesondere das Bilderbuchformat, erwies sich in der Erprobung nicht nur als sehr ansprechend und motivierend für die teilnehmenden Kinder, sondern war auch für den Untersuchungsleiter mit nur geringer Einarbeitungszeit leicht durchführbar. Mit einem Gesamtaufwand von ungefähr 30 Minuten für Durchführung und Auswertung kann die Timogeschichte zudem als praktikabel und zeitökonomisch im sprachtherapeutischen Praxisalltag bezeichnet werden.

Erkenntnisse zur Entwicklung und Diagnostik des Textverständnisses

In den durchgeführten Analysen zeigte sich die Notwendigkeit, die Fähigkeiten im Textverständnis separat von anderen rezeptiven Ebenen sowie der Sprachproduktion zu erfassen, da die Leistungen in anderen sprachlichen Bereichen nur bedingt Rückschlüsse auf die Textverständnisleistung zulassen. Ebenso scheint eine Differenzierung der verschiedenen Teilbereiche des Textverständnisses erforderlich. So können insbesondere Defizite im Monitoring des Sprachverstehens isoliert von anderen Einschränkungen auftreten. Die Untersuchungsergebnisse deuten zudem darauf hin, dass Kinder mit eingeschränktem Sprachverständnis insbesondere bei der Inferenzbildung sowie im Monitoring des Sprachverstehens nicht nur stärkere Beeinträchtigungen, sondern auch langsamere Entwicklungsfortschritte aufweisen. Eine frühzeitige Erfassung und Intervention noch im Vorschulalter ist demnach empfehlenswert, um eine Manifestierung der Einschränkungen zu vermeiden.

Literatur

- Bühner, M. (2011): *Einführung in die Test- und Fragebogenkonstruktion*. 3. aktual. Aufl. München: Pearson Studium
- Dempsey, L.F. (1999): *The validity of the Joint Retell as a Measure of Young children's Comprehension of Familiar Stories*. Masterarbeit an der School of Communication Sciences and Disorders, London, Ontario
- Dempsey, L.F. & Skarakis-Doyle, E. (2001): The Validity of the Joint Story Retell as a Measure of Young Children's Comprehension of Familiar Stories. In: *Journal of Speech-Language Pathology and Audiology* (25, 4), 201-211
- Florit, E., Roch, M., Altoè, G. & Levorato, M.C. (2009): Listening comprehension in preschoolers: the role of memory. In: *British Journal of Developmental Psychology* (27), 935-951
- Fox, A. (2006): *TROG-D. Test zur Überprüfung des Grammatikverständnisses*. Idstein: Schulz-Kirchner Verlag
- Grimm, H., Aktas, M., Frevert, S. (2010): *SETK 3-5. Sprachentwicklungstest für drei- bis fünfjährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiven Gedächtnisleistungen*, Hogrefe, 2. Auflage
- Hachul, C. & Schönauer-Schneider, W. (2012): *Sprachverstehen bei Kindern: Grundlagen, Diagnostik und Therapie*. München, Elsevier-Verlag
- Heel, M. & Janda, S. (2011): Diagnostik von Sprachverständnis im Vorschulalter. Explorative Untersuchung zum Textverständnis. Unveröffentlichte Masterarbeit im Studiengang Sprachtherapie an der LMU München
- Mathieu, S. (1998): Entwicklung und Abklärung des Sprachverständnisses In: Zollinger, B. (Hrsg.): *Kinder im Vorschulalter. Erkenntnisse, Beobachtungen und Ideen zur Welt der Drei- bis Siebenjährigen*. Bern, Stuttgart, Wien: Verlag Paul Haupt, 83-137
- Mathieu, S. (2009): Entwicklung und Abklärung des Sprachverständnisses. In: Zollinger, B. (Hrsg.): *Kinder im Vorschulalter. Erkenntnisse, Beobachtungen und Ideen zur Welt der Drei- bis Siebenjährigen*. 3. Korrigierte Aufl. Bern, Stuttgart, Wien: Verlag Paul Haupt, 83-137
- Montgomery, J.W., Polunenko, A. & Marinellie, S.A. (2009): Role of working memory in children's understanding spoken narrative: A preliminary investigation. In: *Applied Psycholinguistics* (30), 485-509
- Lienert, G. & Ratz, U. (1998): *Testaufbau und Testanalyse*. 6. Aufl. Weinheim: Psychologie Verlags Union
- Skarakis-Doyle, E. & Dempsey, L. (2008a): Assessing story comprehension in Preschool Children. In: *Topics on Language Disorders* (28, 2), 131-148
- Skarakis-Doyle, E. & Dempsey, L. (2008b): The Detection and Monitoring of Comprehension Errors by Preschool Children With and Without Language Impairment. In: *Journal of Speech, Language and Hearing Research* (51), 1227-1243
- Van den Broek, P., Kendeou, P., Kremer, K., Lynch, J.S., Butler, J., White, M.J. & Lorch, E.P. (2005): Assessment of comprehension abilities in Young children. In S. Stahl & S. Paris (Hrsg.): *Children's Reading Comprehension and Assessment*, Mahwah, NJ: Erlbaum, 107-130

Zu den Autoren

Simone Janda (M.A.) absolvierte den Masterstudiengang Sprachtherapie an der Ludwig-Maximilians-Universität München und schloss diesen 2011 ab. Im Rahmen der Abschlussarbeit entstand die Timogesichte. Seit 2010 ist sie als akademische Sprachtherapeutin in München und Hamburg tätig, derzeit in einem interdisziplinären Frühförderzentrum.

Marina Heel (M.A.) beendete 2011 mit der Masterarbeit zur beschriebenen Timogesichte ihr Studium der Sprachtherapie an der Ludwig-Maximilians-Universität München. Nach Tätigkeit in sprachtherapeutischen Praxen arbeitet sie seit 2012 als Sprachtherapeutin in der Spezialambulanz für Entwicklungsstörungen des Heckscher-Klinikums in München.

Korrespondenzadresse

Marina Heel
kbo-Heckscher-Klinikum gGmbH
Deisenhofener Str. 28
81539 München
E-Mail: marina.heel@kbo.de

DOI dieses Beitrags: 10.2443/skv-s-2014-57020140102

