



„Halt!“ Ein Diagnoseverfahren zum Textverstehen bei siebenjährigen Kindern*

„Halt!“ A diagnostic tool to evaluate text comprehension in seven year old children

Corinne Weyda, Wilma Schönauer-Schneider

Zusammenfassung

Hörverstehen von Texten ist eine wichtige Kompetenz für Lern- und Bildungserfolg, die Kindern mit Sprachverständnisstörungen oftmals Probleme bereitet und deshalb überprüft werden sollte. Skarakis-Doyle (2002) entwickelte für Kinder im Kindergartenalter eine mehrdimensionale Diagnostik, die neben Fragen zu einer Geschichte auch das Entdecken von Erwartungsverletzungen umfasst. Für das Schulalter fehlen hier jedoch insbesondere im deutschsprachigen Raum noch derartige diagnostische Instrumentarien. In der vorliegenden Studie wurde deshalb ein Diagnoseverfahren zum Hörtextverstehen im Grundschulalter entwickelt. Als Grundlage diente die Adaption der Geschichte „Halt!“, rief der Bär von Manfred Mai (1995). Nach dem Hören der Geschichte müssen Kinder Inhalts- und Inferenzfragen im „Multiple Choice-Format“ beantworten sowie in einer Version mit Inkonsistenzen diese Fehler entdecken und korrigieren.

Das Verfahren sowie weitere Vergleichstests, u. a. zum Satz- und Textverstehen wurden 2012 an 58 monolingual deutschsprachigen Kindern im Alter von 7;0–7;11 Jahren überprüft. 13 Kinder davon wiesen rezeptive Auffälligkeiten auf. Im Vordergrund stand die Frage, ob sich Kinder mit rezeptiven Auffälligkeiten von sprachlich unauffälligen Kindern bei dem vorliegenden Verfahren unterscheiden. Zusätzlich wurden die Items hinsichtlich Schwierigkeit, Trennschärfe und Reliabilität untersucht.

Die Ergebnisse zeigen, dass Kinder mit rezeptiven Auffälligkeiten sowohl beim Beantworten von Fragen als auch beim Erkennen und Korrigieren von Erwartungsverletzungen signifikant schlechter abschneiden als sprachlich unauffällige Kinder. Die Itemanalyse zeigte für die meisten Items einen ausreichenden Schwierigkeitsgrad bzw. Trennschärfe und eine zufriedenstellende Reliabilität. Insgesamt scheint somit das Verfahren geeignet, Kinder mit nicht ausreichendem Verstehen von Hörtexten zu erfassen. Da jedoch nur 13 Kinder mit rezeptiven Auffälligkeiten überprüft wurden, sind diese Ergebnisse in weiteren, größer angelegten Studien zu evaluieren.

Schlüsselwörter

Sprachverstehen, Monitoring des Sprachverstehens, Textverstehen, Diagnostik, Sprachentwicklungsstörungen

Abstract

Text comprehension is a crucial skill for successful learning and education. As children with receptive language disorders often struggle with understanding texts, it is important to assess listening comprehension skills. Skarakis-Doyle (2002) developed a multidimensional diagnostic instrument for children at the age of kindergarten, which consists of comprehension questions and the detection of expectancy violations. There are no corresponding diagnostic procedures for school age children in the German-speaking area. Therefore, in this study, a diagnostic instrument was developed in order to evaluate text comprehension for elementary school. It is based on an adaption of the story „Halt!“, rief der Bär by

* Dieser Beitrag hat das Peer-Review-Verfahren durchlaufen.

Manfred Mai (1995). After listening to the story, children had to answer several multiple choice questions. Some of the questions asked about literal information, others required the listener to draw an inference. Additionally the children had to detect and correct expectancy violations. This procedure and other comparative tests were carried out with 58 monolingual German speaking children at the age of 7;0–7;11 years. 13 of these children showed receptive language disorders. The diagnostic procedure was evaluated regarding item analyses and the suitability of the two subtests.

Results demonstrated that children with receptive language disorders revealed significant poorer skills in both subtests. The item analyses for most of the items indicated sufficient difficulty and selectivity indices and good reliabilities. Overall, „Halt!“ seems to be appropriate to detect children with difficulties in text comprehension, but this result should be verified with a larger sample of children.

Keywords

language comprehension, comprehension monitoring, assessment, language impairment

1 Einleitung

Der Wissenserwerb im Kindesalter hängt stark vom Hörverständnis für Texte ab, da bereits im Kindergarten und dann besonders in der Grundschule viel Wissen in Erzählform vermittelt und daher über das Sprachverstehen angeeignet wird (Zollinger, 2010). In Längsschnittstudien konnte zudem gezeigt werden, dass die Kompetenzen im Hörverständnis für Geschichten und Schlussfolgern mit späteren Leistungen im Leseverstehen zusammenhängen (Kim, 2015; van den Broek, White, Kendeou & Carlson, 2009; Oakhill & Cain, 2007). Daher haben Textverständnisstörungen gravierende Auswirkungen auf die gesamte Entwicklung des Kindes. Defizite im Text- und Leseverstehen können bis ins Erwachsenenalter persistieren und stehen mit Problemen im sozialen Bereich im Zusammenhang (Rißling, Ronniger, Petermann & Melzer, 2016; Clegg, Hollis, Mawhood & Rutter, 2005). Die Diagnostik sprachlicher Leistungen konzentriert sich aber oft auf produktive Fähigkeiten und vernachlässigt rezeptive Leistungen, insbesondere die Erfassung des Textverstehens. Da das Sprachverständnis als Fähigkeit nicht direkt beobachtbar und nur durch sichtbar werdende Reaktionen auf Sprache zu beurteilen ist, sind Interpretationsfehler bei dessen Erfassung häufig (Hachul & Schönauer-Schneider, 2016). Die rezeptiven Leistungen von Kindern werden dadurch oftmals überschätzt und Störungen werden übersehen.

Im deutschen Sprachraum bestehen nur wenige Verfahren zur Erfassung des Hörtextverstehens im Grundschulalter, z. B. die Mäusegeschichte von Gebhard (2008) oder der Untertest Fragen zum Text des SET 5-10 (Petermann, Metz & Fröhlich, 2010), die wichtige Aspekte des Textverstehens wie das Monitoring des Sprachverstehens, Ziel-Plan-Strukturen oder Perspektivenübernahme vernachlässigen. Ziel ist es aus diesen Gründen, ein geeignetes Verfahren zur Erfassung des Textverstehens im Grundschulalter zu entwickeln, das zwischen auffälligem und nicht-auffälligem Textverstehen unterscheidet, und dieses an siebenjährigen Kindern zu überprüfen.

2 Theoretischer Hintergrund

2.1 Textverstehen

Das Verstehen von Texten stellt die komplexeste Form des Sprachverstehens dar und beinhaltet sowohl Aspekte des Wort-, als auch des Satzverstehens, geht aber noch weit darüber hinaus. Aus kognitionspsychologischer Perspektive werden beim Textverstehen im Sinne eines Konstruktions-Integrations-Modells (van Dijk & Kintsch, 1983) mentale Repräsentationen auf drei hierarchischen Ebenen angenommen, wobei die ersten beiden Ebenen textbasiert sind (Richter & Schnotz, 2018; Kintsch & Rawson, 2007).

Auf der ersten Ebene wird eine **Oberflächenstruktur** erstellt. Es werden Wörter und Phrasen, die im Text enthalten sind, v.a. linguistisch verarbeitet. Das Verstehen von Wörtern, wörtlichen Formulierungen und syntaktischen Strukturen liefert alle sprachlichen Details für die Textoberfläche. Die zweite Ebene bildet textbasierte Repräsentationen als **Textbasis** ab. Sie stellt den semantischen Gehalt bzw. die semantische Struktur eines Textes in Form von Propositionen dar. Darin sind Wortbedeutungen enthalten, die kombiniert und zu komplexen Netzwerken zusammengefügt werden, z. B. Prädikat-Argument-Strukturen (GRABEN [agent: HASE; objekt: LOCH]). Lediglich eine kleine Anzahl an Inferenzen ist enthalten, die für die Herstellung von lokaler Textkohärenz notwendig ist. Zum Beispiel werden pronominale Referenten mit der passen-

den Person, auf das sich das Pronomen bezieht, innerhalb und zwischen Sätzen verbunden (Snyder & Caccamise, 2010). Diese Textbasis auf Grundlage expliziter Informationen reicht jedoch für ein tieferes und umfassenderes Verstehen nicht aus. Auf einer dritten, nicht textbasierten Repräsentationsebene muss der Rezipient das im Text enthaltene Wissen mit sachbezogenem Weltwissen, eigenen Erfahrungen bzw. Erwartungen verknüpfen und ein **mentales Modell** (Johnson-Laird, 1983) oder **Situationsmodell** (van Dijk & Kintsch, 1983) konstruieren. Erst nach der Erstellung dieses Situationsmodells hat der Rezipient eine wirkliche Vorstellung des Textes konstruiert.

Dazu werden verschiedene zugrundeliegende kognitive Fähigkeiten bzw. Prozesse angenommen. Der Hörer muss u. a. implizite Bedeutungen ableiten oder fehlende Informationen ergänzen, denn in Texten sind Informationen oftmals nicht explizit formuliert oder fehlen. Beispielsweise müssen zwischen verschiedenen Propositionen neue Verknüpfungen erstellt werden oder müssen durch Weltwissen zusätzliche Informationen integriert werden. Diese Verknüpfungen bzw. Schlussfolgerungen werden auch als **Inferenzen** bezeichnet, die vorhandene Textinformationen anreichern, strukturieren oder verdichten (Richter & Schnotz, 2018). Man unterscheidet zwischen lokalen und globalen Inferenzen. Lokale Inferenzen sind textverbindend bzw. referentiell, z. B. Pronomen mit dem jeweiligen Nomen oder Ableiten von Wortbedeutungen aus dem Kontext, während globale Inferenzen wissensbasiert die Verknüpfung von Informationen im Text mit Erfahrungen oder Weltwissen des Hörers erfordern (Hall & Barnes, 2017). In der hier untersuchten Geschichte muss der Hörer z. B. eine kausale Inferenz ziehen, warum ein Fuchs, der einen Bären weglocken will, dem Bären von einem Honig in der Nähe erzählt.

Neben Inferenzen spielen Prozesse der Validierung von Textinformationen eine große Rolle. Hörer bzw. Leser nutzen ihr Vorwissen auch zur Überprüfung der verstandenen Inhalte im Hinblick auf Erwartungsverletzungen und mangelnde Plausibilität. Bei inkonsistenten Textinformationen oder nicht zutreffenden Erwartungen reagieren sie beispielsweise mit Stirnrunzeln bzw. „Hä?“ oder beim Lesen mit verlängerten Blickbewegungen und Lesezeiten (Richter & Schnotz, 2018; Skarakis-Doyle & Dempsey, 2008). Diese metakognitive Strategie wird auch Comprehension Monitoring oder **Monitoring des Sprachverstehens** genannt (Marks, Mayer & Schönauer-Schneider, 2018; Wassenburg, Beker, van den Broek & van der Schoot, 2015).

In vielen Texten ist es erforderlich, andere Perspektiven einzunehmen und sich in die Gedanken und Verhaltensweisen anderer hineinzusetzen, die möglicherweise von der eigenen Sichtweise abweichen. Rezipienten müssen somit die Fähigkeit der **Theory of Mind (ToM)** nutzen. Insbesondere das Erkennen mentaler und emotionaler Zustände bestimmter Personen ist für das vertiefte Verstehen von Texten unerlässlich. Studien konnten belegen, dass Messungen der ToM für das Verstehen von Texten prädiktiv sind (Dore, Amendum, Golinkoff & Hirsh-Pasek, 2018; Kim, 2015).

Das Bilden von Inferenzen, das Monitoring des Sprachverstehens und auch ToM sind kognitiv anspruchsvolle Fähigkeiten, die an hierarchieniedrigere kognitive Kompetenzen wie dem phonologischen Arbeitsgedächtnis ansetzen. Es wird vermutet, dass das phonologische Arbeitsgedächtnis für die Konstruktion des Situationsmodells und die dafür erforderlichen Fähigkeiten grundlegend bedeutsam ist, indem Informationen aufrechterhalten und vernetzt werden, die für eine lokale und globale Kohärenzbildung erforderlich sind (u. a. Kim, 2015).

Somit erfordert das Textverstehen multiple Repräsentationen, die auf jeder Ebene unterschiedliche sprachliche und kognitive Prozesse erfordern. Dazu gehören u. a. Wortverstehen, syntaktisches Verstehen, phonologisches Arbeitsgedächtnis, Comprehension Monitoring, Inferenzen ziehen oder auch Theory of Mind (Kim, 2015). Für eine Überprüfung des Textverstehens wäre es deshalb hilfreich, nicht nur das erstellte Situationsmodell durch Fragen zum Text abzu prüfen, sondern auch Prozesse der Erstellung wie die Evaluation des Textes (Monitoring des Sprachverstehens) oder Fähigkeiten der Perspektivenübernahme/ToM in der Diagnostik mit zu berücksichtigen.

2.2 Methoden zur Erfassung des Textverstehens

Es gibt verschiedene Methoden, um das Textverstehen eines Kindes zu überprüfen. Im deutschsprachigen Raum dominiert vor allem die Methode der Fragenbeantwortung. Selten lassen sich die Methoden des (gemeinsamen) Nacherzählens und des Entdeckens von Erwartungsverletzungen finden (Hachul & Schönauer-Schneider, 2016). Im Folgenden wird im Hinblick auf das erarbeitete Verfahren auf die Methoden der Fragenbeantwortung und des Entdeckens von Erwartungsverletzungen eingegangen.

Bei den Verständnisfragen können offene Fragen (W-Fragen), „Fragen“ im Multiple-Choice-Format bzw. Mehrfachwahlaufgaben oder Entscheidungsfragen gestellt werden. Vorteile der Multiple Choice-Aufgaben sind ihre hohe Auswertungsobjektivität, die einfache ökonomische Auswertung und die geringen Anforderungen an die Sprachproduktion im Gegensatz zu den offenen Fragen. Nachteilig wirkt sich jedoch die höhere Ratewahrscheinlichkeit aus (Hachul & Schönauer-Schneider, 2016).

Bei der Fragestellung kann man zwischen faktischen Informationsfragen und schlussfolgernden bzw. Inferenzfragen unterscheiden. Informationsfragen können in der Regel mit einem Wort beantwortet werden und betreffen meist explizit im Text genannte Informationen (Gebhard, 2008). Bei den Inferenzfragen müssen Schlussfolgerungen gezogen werden, um verschiedene im Text genannte Informationen sinnvoll zu verknüpfen. Diese Frageart bezieht sich eher auf den Verstehensprozess und ist somit eine geeignete Methode, um kohärentes Verstehen bzw. Inferenzen und somit das Situationsmodell zu erfassen (Kim, 2016; Mathieu, 1995). In zahlreichen Studien mit sprachunauffälligen und sprachauffälligen Kindern konnte gezeigt werden, dass die sprachauffälligen Kinder meist größere Probleme mit Inferenzfragen aufweisen (u. a. Dodwell & Bavin, 2008, Karasinski & Ellis Weismer, 2010) und daher besondere Schwierigkeiten beim Verstehen von Texten haben.

Um die Fähigkeiten des Monitorings des Spracherstehens beim Verstehen von Texten zu erfassen, entwickelte Skarakis-Doyle (2002) vorrangig für Kinder im Kindergartenalter die Methode der *Expectancy Violation Detection Task*, d. h. das *Entdecken von Erwartungsverletzungen*. Hierbei wird die Leistung der Kinder überprüft, Unstimmigkeiten zwischen Elementen in der Geschichte und ihren Erwartungen zu entdecken und zu korrigieren. In bekannten und thematisch vertrauten Geschichten werden verschiedene inhaltliche Verletzungen, wie Zielunterbrechungen, prosodische Veränderungen oder Ersetzungen innerhalb der Geschichte vorgenommen (Skarakis-Doyle & Dempsey, 2008). Bei der Dokumentation der Reaktionen auf die Erwartungsverletzungen werden sowohl verbale als auch nonverbale Reaktionen, die innerhalb von fünf Sekunden oder am Ende der Phrase auftreten, erfasst und in die Auswertung einbezogen. Bei den verbalen Antworten wird zwischen unspezifischen Protesten („Nein“) und Protesten mit Korrekturen („Nein, das war ...“) unterschieden. Korrekturen beziehen sich auf die verletzten Textstellen und den richtigen Inhalt (Skarakis-Doyle, 2002, Skarakis-Doyle & Dempsey, 2008). Bei diesem Test handelt es sich um eine online-Methode zur Überprüfung der frühen Monitoring-Leistungen, da die Kinder während des Verstehensprozesses auf Unstimmigkeiten reagieren müssen. Dadurch sind sowohl die Anforderungen an die expressiven Leistungen als auch an den Arbeitsspeicher gering. Skarakis-Doyle, Dempsey und Lee (2008) berichten moderate Korrelationen der Leistungen im primären Verstehen und dem Entdecken von Erwartungsverletzungen. Jüngere Kinder im Alter von 30 bis 48 Monaten können Ersetzungen von Objekten und Personen besser entdecken als Ersetzungen von Handlungen (Skarakis-Doyle, 2002). Auch bei Heel und Janda (2011, 2014), die Kinder im Alter von vier bis sechs Jahren untersuchten, stellte die Ersetzung der Handlungen eines Protagonisten die schwierigste Erwartungsverletzung dar.

2.3 Konzeption des Verfahrens

Als Grundlage der Überprüfung dient die kurze Geschichte *„Halt!“, rief der Bär* von Manfred Mai (1995). Sie handelt von einem Bären, der den Eingang eines sonst rundherum eingezäunten Waldes bewacht. Er ist der Meinung, dass ihm der Wald gehöre und er daher kein Tier in den Wald lassen dürfe. Als Erstes kommt ein Hase und möchte in den Wald, was ihm der Bär verbietet. Der Hase hoppelt allerdings ein Stück weiter, gräbt eine Mulde unter den Zaun und erhält so unbemerkt Zutritt. Als Nächstes kommt ein Stinktier, das ebenfalls um Einlass bittet. Der Bär verweigert dies abermals und das Stinktier lässt daraufhin eine Stinkbombe los, sodass der Bär ohnmächtig wird und das Stinktier in den Wald gehen kann. Als der Bär aus seiner Ohnmacht erwacht, kann er sich an das Geschehene nicht erinnern und fühlt sich noch etwas schwach. Diesen Umstand nutzt ein Fuchs aus, der dem Bären rät, etwas Honig essen zu gehen. Dieser übergibt die Verantwortung den Wald zu bewachen dem Fuchs, der allerdings alle vorbeikommenden Tiere in den Wald lässt. Als der Bär gestärkt zurückkehrt, gewährt er dem Fuchs zur Belohnung Eintritt in den Wald. Der Bär bewacht den Wald weiter und wundert sich darüber, dass kein Tier mehr vorbeikommt.

Die Geschichte ist linear aufgebaut, folgt dem gängigen Geschichtenschema und enthält viele schlussfolgernde Elemente sowie komplexe Ziel-Plan-Strukturen mit drei Episoden und vier Protagonisten. Die drei Episoden bestehen jeweils aus einem verursachenden Geschehen, bei

dem die Tiere nacheinander in den Wald gehen wollen und der Bär ihnen das untersagt. Darauf folgt eine Aktion bzw. ein Lösungsversuch der Tiere, um in den Wald zu kommen. Diese Aktionen beinhalten jedes Mal eine Täuschung des Bären und gelingen, ohne dass der Bär dies bemerkt. Am Ende der Geschichte wird daher noch die Verwunderung als Reaktion des Bären beschrieben. Nach Mathieu (1998) kann ein Kind ab sechs Jahren zwei Ereignisse mit etwa fünf Protagonisten innerhalb einer Erzählung verstehen und Verknüpfungsfragen dazu beantworten. Für Grundschul Kinder sollte die Komplexität der Geschichte daher adäquat sein. Die Hauptfiguren weisen miteinander konkurrierende Zielstrukturen auf: Der Bär bewacht seinen Wald, damit niemand hineingehen kann. Alle anderen Tiere hingegen verfolgen das Ziel in den Wald zu kommen. Daher denkt sich jedes Tier einen anderen Plan aus, um Zutritt in den Wald zu erhalten. Der Bär erreicht sein Ziel nicht, da er nicht bemerkt, wie ihn die anderen Tiere täuschen. Der Hase, das Stinktier und der Fuchs erreichen jedoch ihr Ziel ohne dass der Bär die Täuschung bemerkt. Die Wahl der Geschichte hängt mit dieser recht komplexen Ziel-Plan-Struktur der Protagonisten zusammen, da in vielen Studien (u. a. Graesser, Olde & Klettke, 2002) die Bedeutsamkeit von Zielstrukturen in Geschichten für das Verständnis betont wurde. Nur wenn die Ziele eines Charakters entdeckt und die dazugehörigen Motive und Pläne durch Perspektivenübernahme bzw. ToM geschlussfolgert werden können, ist es möglich, dessen Handlungen zu interpretieren (Kim, 2015, 2016; Schank & Abelson, 1977).

Sprachlich ist die Geschichte im Präteritum erzählt, beinhaltet jedoch keine unbekanntenen oder niedrigfrequenten Vergangenheitsformen. Bezüglich des verwendeten Wortschatzes und der Grammatik wurden kleine Veränderungen an der Geschichte vorgenommen. Zudem wurde nach Behrens (2010) darauf geachtet, dass die Geschichte nicht größeren Schulgruppen bekannt ist und die angesprochenen Themen nicht bestimmte Gruppen, wie beispielsweise Jungen oder Mädchen, systematisch bevorzugen oder benachteiligen. Außerdem sind das Setting *Wald* und die vorkommenden Tiere wohl den meisten Vorschul- und Grundschulkindern bekannt.

2.3.1 Untertest Verständnisfragen

Dieser Untertest wurde in Anlehnung an die Mäusegeschichte von Gebhard (2008) konzipiert. Bei den Verständnisfragen wurde das Multiple Choice-Format mit drei Antwortmöglichkeiten gewählt, da die produktiven Sprachfähigkeiten weitgehend ausgeschlossen werden sollen (Weyda, 2012). Bei den Antwortmöglichkeiten wurde darauf geachtet, dass bei Unkenntnis der Geschichte alle drei Antworten gleich wahrscheinlich sind. Außerdem wurde ein ausbalanciertes Antwortformat verwendet, indem die Positionen der richtigen Antwortmöglichkeiten variiert wurden. Die 25 Verständnisfragen bestehen in relativ ausgewogenem Verhältnis aus elf Informationsfragen (z. B. Der Fuchs ließ ... a) kein Tier in den Wald; b) alle Tiere in den Wald; c) ein Tier in den Wald) und vierzehn schlussfolgernden Fragen (z. B. Das Stinktier wollte ... a) den Wald bewachen; b) in den Wald gehen; c) mit dem Bären spielen). Letztere setzen sich aus verschiedenen Arten der Inferenzen zusammen (referentielle Inferenzen, Verstehen der Ziel-Plan-Struktur bzw. der Täuschung, Perspektivenübernahme, kausale Inferenzen, Schlussfolgerungen zum übergeordneten Ziel). Die schlussfolgernden Fragen können in lokale und globale Kohärenzbildung eingeteilt werden. Dabei handelt es sich um neun schlussfolgernde Fragen, die die Herstellung von lokaler Kohärenz verlangen und fünf Inferenzfragen, die die globale Kohärenzbildung erfordern. Somit wurden zum einen hierarchieniedrige bzw. einfache Schlussfolgerungen, bei denen Informationen aus benachbarten Sätzen verknüpft werden müssen, miteinbezogen. Zum anderen wurden hierarchiehohe bzw. komplexe Inferenzen, bei denen Informationen aus mehreren Sätzen und mehreren Abschnitten mit Weltwissen zusammengeführt werden müssen, integriert.

2.3.2 Untertest Erwartungsverletzungen

Der Untertest zur Überprüfung der Fähigkeiten im Monitoring des Sprachverstehens wurde in Anlehnung an den Untertest *Expectancy Violation Detection Task* von Skarakis-Doyle (2002) konzipiert. Zu Beginn wurden zwei Übungsaufgaben eingefügt. Die in die Geschichte integrierten Erwartungsverletzungen bestehen aus drei Zielunterbrechungen, einer Änderung des Refrains mit Zielunterbrechung, vier Ersetzungen innerhalb der Geschichte und einer widersprüchlichen Ersetzung. Die Ersetzungen innerhalb der Geschichte im Verfahren *Halt* wurden in Anlehnung an Skarakis-Doyle (2002) aufgeteilt in Ersetzungen von Protagonisten, Objekten, Handlungen eines anderen Protagonisten und Handlungen desselben Protagonisten. Die Auftretenshäufigkeit der einzelnen Kategorien wurde nicht übernommen, da die vorgenommenen Ersetzungen der Geschichte angepasst wurden. Um den Schwierigkeitsgrad für das Alter der in diesem Verfahren

getesteten Kinder anzupassen, wurde nur eine Verletzung bezüglich des Protagonisten (z. B. „Das kann ich doch für dich tun, solange du weg bist“, bot *das Stinktier* an (anstatt: *der Fuchs*) und des Objekts (z. B. „Dagegen hilft am besten *eine Stinkbombe* (anstatt *Portion Honig*) in den Untertest einbezogen. Da dem Verstehen von Zielstrukturen bei der Rezeption von Geschichten ein besonderer Stellenwert zukommt, wurden vier Zielunterbrechungen integriert. Das Ausmaß, in dem ein Rezipient eine Geschichte versteht, hängt mit dem Verfolgen der Ziele und Pläne der Charaktere zusammen (Kim, 2015; Magliano, Taylor & Kim, 2005). Daher ist es besonders wichtig, dass Kinder überwachen, ob die Handlungen der Protagonisten deren Plänen folgen und in einer erfolgreichen Problemlösung resultieren. In den eingefügten Zielunterbrechungen verhalten sich die Protagonisten jedoch nicht ihrem Ziel entsprechend, sondern handeln entgegen ihres Plans (z. B. Der Fuchs legte sich vor den Eingang und ließ *kein Tier in den Wald* (anstatt *ließ jedoch alle Tiere in den Wald, die hineinwollten*).

Für eine möglichst hohe Durchführungs- und Auswertungsobjektivität wurden vor der Erprobung der Tests Instruktionen für die genaue Vorgehensweise sowie Auswertungsregeln ausgearbeitet. Bei den Verständnisfragen erhält das Kind für jede korrekt gelöste Frage einen Punkt, bei einer falschen Antwort sind null Punkte zu vergeben. Der Untertest Entdecken von Erwartungsverletzungen ist in der Auswertung zweigeteilt. Zuerst wird ein Punkt für die Entdeckung der Erwartungsverletzung (durch einen eindeutigen verbalen Kommentar des Kindes, z. B. „Nein!“ oder „Stimmt nicht!“) gegeben. Kann das Kind dem Inhalt weiter durch eine darauffolgende Spezifizierung des entdeckten Fehlers oder durch eine Korrektur widersprechen, werden zwei Punkte vergeben (z. B. „Nein, keine Stinkbombe, Honig!“). Eine nonverbale Reaktion wird zwar protokolliert, da sie für die qualitative Einschätzung von Bedeutung ist, wird aber nicht bewertet.

3 Fragestellungen

Im Rahmen der Studie wurde das Verfahren Halt zur Überprüfung des Textverstehens erprobt (Weyda, 2012). Dazu wurden die einzelnen Untertests und Items des zu überprüfenden Verfahrens auf ihre Eignung hin untersucht und die Ergebnisse in den Untertests von rezeptiv auffälligen Kinder mit denen unauffälliger Kinder verglichen, um mögliche Unterschiede feststellen zu können. Es wurden explizit rezeptiv auffällige Kinder als Subgruppe ausgewählt, da Kinder mit überwiegend expressiven Sprachentwicklungsstörungen nicht zwingend auch Auffälligkeiten im Hörtextverstehen aufweisen. Vorstudien mit ähnlich konzipierten Verfahren bei Vorschulkindern belegen bessere Leistungen beim Beantworten von Fragen und beim Entdecken von Erwartungsverletzungen bei Kindern mit expressiven Sprachentwicklungsstörungen (Schulze, 2010; Hagemester, 2011). Im Folgenden werden vorrangig Fragestellungen und Ergebnisse zum neu konzipierten Verfahren im Vergleich der beiden Gruppen berichtet, während genauere Analysen zu Zusammenhängen mit dem phonologischen Arbeitsgedächtnis und der Perspektivenübernahme hier nicht näher dargestellt werden.

■ Fragestellung 1: Finden sich Zusammenhänge zwischen den Untertests von Halt?

Die beiden Untertests untersuchen verschiedene Aspekte des Textverstehens, die sich allerdings hinsichtlich ihrer Anforderungen auch überschneiden. Skarakis-Doyle und Dempsey (2008), Skarakis-Doyle und Kollegen (2008) sowie Heel und Janda (2011, 2014) fanden Zusammenhänge zwischen den Leistungen beim Beantworten von Verständnisfragen und Entdecken von Erwartungsverletzungen für jüngere Kinder. Daher wird untersucht, inwieweit die Leistungen in den beiden Untertests bei siebenjährigen Kindern zusammenhängen.

■ Fragestellung 2: Finden sich Unterschiede zwischen den rezeptiv auffälligen und unauffälligen Kindern in den beiden Untertests von Halt?

Andere Geschichtentests fanden unterdurchschnittliche Leistungen von sprachgestörten Kindern beim Beantworten von Verständnisfragen (z. B. Gebhard, 2008). Weitere Studien (u. a. Skarakis-Doyle & Dempsey, 2008; Heel & Janda, 2011, 2014; Hagemester, 2011) konnten darüber hinaus zeigen, dass Vorschulkinder mit rezeptiven Sprachstörungen beim Entdecken von Erwartungsverletzungen schlechter abschneiden als Gleichaltrige mit normaler Sprachentwicklung bzw. auch Kinder mit expressiven Sprachentwicklungsstörungen.

■ **Fragestellung 3: Finden sich Unterschiede zwischen den rezeptiv auffälligen und unauffälligen Kindern beim Beantworten der verschiedenen Fragearten?**

Es ist von Interesse, ob sich Gruppenunterschiede hinsichtlich der Beantwortung sowohl von beiden Arten der Inferenz- als auch Informationsfragen zeigen. Nach dem Konstruktions-Integrations-Modell nach van Dijk und Kintsch (1983) wird angenommen, dass die für die Beantwortung globaler Inferenzfragen benötigten Verarbeitungsprozesse höher und komplexer einzustufen sind. Karasinski und Ellis Weismer (2010) fanden signifikant geringere Leistungen sprachauffälliger Jugendlicher bei der Bildung von lokaler und globaler Kohärenz im Vergleich zu sprachunauffälligen. Dies wird auch für die jüngeren Kinder dieser Studie untersucht.

■ **Fragestellung 4: Finden sich Unterschiede zwischen den rezeptiv auffälligen und unauffälligen Kindern in den Reaktionsmustern im Untertest *Entdecken von Erwartungsverletzungen*?**

Bei Skarakis-Doyle und Dempsey (2008) reagierten Kinder mit einer Sprachentwicklungsstörung im Gegensatz zu den elaborierten Antworten der unauffälligen Kinder hauptsächlich nonverbal. Daher ist von Interesse, ob auch rezeptiv auffällige Kinder im Alter von sieben Jahren ein anderes Muster hinsichtlich der Reaktionen auf die Erwartungsverletzungen zeigen als gleichaltrige Kinder mit altersgemäßen Fähigkeiten im Sprachverständnis.

■ **Fragestellung 5: Sind die Items der beiden Untertests hinsichtlich ihrer Schwierigkeit, Trennschärfe und Reliabilität geeignet?**

Im Rahmen der Itemanalyse werden Itemschwierigkeiten, Trennschärfen und Reliabilitätskoeffizienten analysiert.

4 Methode

4.1 Untersuchungsablauf und Beschreibung der Untersuchungsinstrumente

In der Untersuchung (Weyda, 2012) wurden fünf verschiedene Tests verwendet, die mit jedem Kind einzeln durchgeführt wurden. Das Verfahren *Halt* wurde zu Beginn jeder Testung durchgeführt. Um die teilnehmenden Kinder hinsichtlich ihrer rezeptiven Leistungen einschätzen und differenzieren zu können, wurden zwei Vergleichstests herangezogen. Ziel war es, in Kinder mit unauffälligem bzw. auffälligem Sprachverstehen einzuteilen. Hier wurden Verfahren zum grammatischen Verstehen und Textverstehen verwendet, um auffälliges Sprachverstehen auf mehreren Ebenen zu erfassen. Zum Wortverstehen waren zum Testzeitpunkt für die Altersgruppe keine ausreichend normierten Verfahren im deutschsprachigen Raum vorhanden. Der TROG-D (Test zur Überprüfung des Grammatikverständnisses) von Fox (2006) wurde zur Erfassung des Sprachverstehens auf Satzebene verwendet, da dieser etablierte und standardisierte Test Normwerte für den gewünschten Altersbereich zwischen 7;0 und 7;11 Jahre bietet und grammatisches Verstehen in enger Verbindung mit Textverstehen steht (u. a. Kim, 2015). Als zweites Verfahren wurde die Mäusegeschichte von Gebhard (2008) eingesetzt, um die Textverstehensleistung zu überprüfen. Dabei handelt es sich um einen informellen Test zum Hörverständnis für Geschichten bei Kindern zwischen 7 und 9 Jahren, bei dem das Textverstehen mit 12 Fragen im Multiple Choice-Format mit jeweils drei Antwortmöglichkeiten und einer zusätzlich offenen Frage erfasst wird. Das Verfahren beinhaltet drei schlussfolgernde Fragen (z. B. Gefährlich für das Mäuslein war in Wirklichkeit ... a) das erste Tier; b) das zweite Tier; c) gar kein Tier) und neun Informationsfragen (z. B. Das erste Tier war auf dem Kopf ... a) schneeweiß; b) mausgrau; c) feuerrot). Alle korrekt beantworteten Fragen werden jeweils mit einem Punkt bewertet, während die offene Frage der qualitativen Einschätzung dient. Der Test liefert ein quantitatives Maß, ob ein Kind zu den schwächsten 15 % seiner Altersgruppe gehört (Gebhard, 2008). Die Mäusegeschichte wurde als Vergleichstest herangezogen, da der erste Untertest im Verfahren *Halt* in Anlehnung an die Mäusegeschichte entwickelt wurde und beide Verfahren für das Grundschulalter konzipiert wurden.

Ferner wurden zwei weitere Fähigkeiten im Zusammenhang mit Textverstehen untersucht: das phonologische Arbeitsgedächtnis und ToM. Für die Erfassung des phonologischen Arbeitsgedächtnisses wurde der Mottier-Test, ein Bestandteil des Zürcher Lesetests (Linder & Grisseemann, 2000), zur Einschätzung der auditiven Merkspanne durch das Nachsprechen von Sinnlossilben eingesetzt. Ein Wert unter 12 Rohpunkten galt als auffällig. Da die Studie 2012 durchgeführt wurde, stand die 2. Auflage des Tests noch nicht zur Verfügung.

ToM wurde durch eine Aufgabe zur falschen Überzeugung überprüft, da die Geschichte *Halt* komplexe Ziel-Plan-Strukturen und Täuschungen beinhaltet. Dazu wurde die Testung zum falschen Inhalt nach Gopnik und Astington (1988) verwendet, um die Anforderungen an expressive und rezeptive Sprachfähigkeiten relativ gering zu halten. In der Untersuchung wurde den teilnehmenden Kindern eine Smarties-Schachtel präsentiert, in der sich entgegen der Erwartungen Stifte befanden. Es sollte nun bestimmt werden, welchen Inhalt ein Stofftier in der verschlossenen Schachtel vermutete und was sich tatsächlich darin verbarg („Was denkt die Ratte ist in der Schachtel?“ und „Was ist wirklich in der Schachtel?“). Konnte das Kind die zweiteilige Testfrage richtig beantworten, erhielt es einen Punkt. Bei inkorrekt beantworteter Frage wurde kein Punkt vergeben, so dass lediglich die Unterscheidung ein oder kein Punkt im Sinne einer vorhandenen bzw. nicht vorhandenen ToM möglich war.

Die Tests wurden in 35 bis 45 Minuten in der jeweiligen Schule bzw. der sprachtherapeutischen Praxis in den Regionen München bzw. Stuttgart in einem separaten Raum durchgeführt und anschließend anhand einer statistischen Analyse mit SPSS (SPSS Statistics 19) ausgewertet.

4.2 Stichprobenbeschreibung

Die Stichprobe der Untersuchung setzte sich aus $N = 58$ Probanden zusammen. Davon waren 22 (37,9 %) Kinder weiblich und 36 (62,1 %) männlich. Das Durchschnittsalter betrug 7;57 Jahre. Voraussetzungen für die Teilnahme war ein Alter von 7;0 bis 7;11 Jahre, eine altersgemäße allgemeine Entwicklung, unauffällige Hörfähigkeit, durchschnittliche kognitive Fähigkeiten und ein monolingual deutscher Spracherwerb. 31 Kinder besuchten eine Regelschule (53,4 %) und 27 Kinder eine Sprachheilschule (46,6 %). Der hohe Anteil von Kindern aus Sprachheilschulen sollte möglichst viele rezeptiv auffällige Kinder sicherstellen.

Für die Studie sollten die Kinder in eine Gruppe mit rezeptiv unauffälligen und rezeptiv auffälligen Kindern eingeteilt werden. Kinder wurden als rezeptiv auffällig eingestuft, wenn sie im TROG-D und/oder der Mäusegeschichte (MG) unterdurchschnittlich waren, d. h. bei einem T-Wert unter 40 beim TROG-D und bei einem Rohwert unter 7 bei der Mäusegeschichte. Da das grammatische Verstehen eine bedeutsame Rolle für das Textverstehen spielt, wurden auch Kinder, die nur im TROG-D auffielen, als rezeptiv auffällig eingestuft. Zudem wurde die Mäusegeschichte zwar an 117 sieben-, acht- und neunjährigen Kindern überprüft, die Untergruppe der Siebenjährigen ist jedoch nicht ausreichend untersucht, um lediglich von diesem Test auf rezeptiv auffällig zu schließen.

Die Testung mit den Vergleichstests ergab, dass 45 Kinder keine rezeptiven Defizite aufwiesen, während 13 Kinder durch unterdurchschnittliche Leistungen im TROG-D und/oder der Mäusegeschichte (MG) als rezeptiv auffällig eingestuft wurden (vgl. Tab. 1). In der Mäusegeschichte fielen insgesamt nur fünf Kinder rezeptiv auf, wobei davon 3 auch im TROG-D unterdurchschnittlich waren. Acht Kinder zeigten nur im TROG-D auffällige Ergebnisse. Somit stuft der TROG-D elf Kinder als rezeptiv auffällig ein, während die Mäusegeschichte nur fünf Kinder als rezeptiv auffällig erkannte. Für die weiteren Analysen wurden alle rezeptiv auffälligen Kinder als Untersuchungsgruppe ($N = 13$) gewertet.

Tab. 1: Einteilung der Probanden in rezeptiv auffällig bzw. unauffällig

	unauffällig	nur Mäusegeschichte auffällig	nur TROG-D auffällig	Mäusegeschichte u. TROG-D auffällig
Anzahl der Kinder	45	2	8	3

5 Ergebnisse

Der Schwerpunkt der Ergebnisdarstellung liegt auf der Beantwortung der Fragestellungen im Hinblick auf die Halt-Geschichte im Vergleich der beiden Untersuchungsgruppen (rezeptiv auffällig und rezeptiv unauffällig). Zur Vollständigkeit sind in der folgenden Tabelle (vgl. Tab. 2) jedoch auch die Ergebnisse in den Vergleichstests TROG-D bzw. Mäusegeschichte und den Zusatztests Mottier-Test bzw. Aufgabe zur Perspektivenübernahme/ToM (falscher Inhalt) dargestellt. Dabei wurden beim TROG-D die Rohwerte verwendet, da einige Kinder so gut abschnitten, dass ihnen kein T-Wert zugewiesen werden konnte.

Tab. 2: Deskriptive Statistik der Untersuchungsgruppen in den Vergleichs- und Zusatztestverfahren

Test	TROG-D		Mäusegeschichte		Mottier		Falsch. Inhalt (ToM)	
	unauff.	auffällig	unauff.	auffällig	unauff.	auffällig	unauff.	auffällig
M	16,31	8,92	10,69	7,62	16,33	12,08	0,89	0,38
SD	2,37	2,75	1,59	2,06	4,49	2,78	0,32	0,51
Max	21	14	12	11	26	18	1	1
Min	12	5	7	4	9	8	0	0
Anzahl (%) auffälliger Kinder	s.o.	s.o.	s.o.	s.o.	6 (13,3%)	6 (46,2%)	8 (11,1%)	5 (61,5%)

Die rezeptiv auffälligen Kinder sind erwartungsgemäß im TROG-D und in der Mäusegeschichte deutlich schwächer als die Kontrollgruppe, wobei in der Mäusegeschichte auch einige rezeptiv auffällige Kinder höhere Werte erreichten, was der Subgruppeneinteilung (s. 4.2) geschuldet ist.

Im phonologischen Arbeitsgedächtnis liegen der Mittelwert und der Maximalwert der rezeptiv auffälligen Kinder deutlich unter den Werten der Vergleichsgruppe. Vergleicht man den prozentualen Anteil auffälliger Kinder in den Untersuchungsgruppen, so weisen 46,2% der rezeptiv auffälligen Kinder und nur 13,3% der rezeptiv unauffälligen Kinder unterdurchschnittliche Leistungen im auditiven Kurzzeitgedächtnis auf.

Bei der Aufgabe zur ToM (Erkennen des falschen Inhalts) schneiden die rezeptiv auffälligen Kinder im Mittelwert deutlich schlechter ab als die unauffälligen Kinder, wobei die Einteilung in 0 oder 1 Punkt sehr wenig differenziert. Betrachtet man den prozentualen Anteil auffälliger Kinder im Bereich ToM, so haben 61,5% der rezeptiv auffälligen Kinder Schwierigkeiten bei der Perspektivenübernahme, jedoch nur 11,1% der unauffälligen. Diese Ergebnisse sprechen für größere Schwierigkeiten rezeptiv auffälliger Kinder im phonologischen Arbeitsgedächtnis und vor allem in der ToM, die das Textverstehen negativ beeinflussen könnten. Diese wichtigen Aspekte werden hier jedoch nicht weiter vertieft.

Im Folgenden sind die Ergebnisse in Bezug auf die Fragestellungen zur Halt-Geschichte beschrieben (vgl. 3). Die rezeptiv auffälligen Kinder schneiden in beiden Untertests im Mittelwert deutlich schlechter ab als die rezeptiv unauffälligen Kinder (vgl. Tab. 3). Dabei zeigt die Standardabweichung jeweils eine größere Streuung bei den rezeptiv auffälligen Kindern.

Die rezeptiv unauffälligen Kinder konnten bei den Verständnisfragen mindestens 14 Fragen korrekt beantworten, die rezeptiv auffälligen zeigten einen Minimalwert von 5, so dass insbesondere die schwächeren der jeweiligen Gruppe sich deutlich unterscheiden. Allerdings liegt der Maximalwert bei beiden Gruppen bei der Gesamtzahl von 25, so dass auch in der rezeptiv auffälligen Gruppe die Beantwortung altersgemäß gelang. Die Mittelwerte belegen jedoch deutliche Unterschiede in der Anzahl der korrekt beantworteten Fragen.

Beim Erkennen der Erwartungsverletzungen liegt der Mittelwert der rezeptiv auffälligen Kinder wiederum erkennbar unter dem der Vergleichsgruppe. Während die rezeptiv unauffälligen Kinder den höchstmöglichen Wert von 18 erreichen, gelingt dies in der rezeptiv auffälligen Gruppe nicht. Der Minimalwert von 0 in der rezeptiv auffälligen Gruppe belegt zudem, dass hier teilweise keine einzige Erwartungsverletzung erkannt wurde. Der hohe Anteil an expressiven Auffälligkeiten in der Gruppe der rezeptiv unauffälligen Kinder kann möglicherweise die extrem niedrigen Werte der rezeptiv unauffälligen Kinder beim Entdecken von Erwartungsverletzungen erklären.

Tab. 3: Ergebnisse der Untersuchungsgruppen in den Untertests der Halt-Geschichte

Untertest Halt	Verständnisfragen		Erwartungsverletzungen	
	unauffällig	auffällig	unauffällig	auffällig
M	22	15,54	15,64	9,15
SD	2,71	5,74	3,59	5,79
Max	25	25	18	16
Min	14	5	3	0

Genauere Analysen zu den Subgruppen finden sich nun in der Beantwortung der Fragestellungen. Da für die rezeptiv unauffälligen Kinder in beiden Untertests keine Normalverteilung der Messwerte angenommen werden kann und die Gruppengröße der rezeptiv auffälligen Kinder mit $N=13$ sehr gering ist, werden zur inferenzstatistischen Überprüfung der Fragestellungen nicht-parametrische Verfahren verwendet.

■ Fragestellung 1: Zusammenhänge zwischen den Untertests von *Halt*

Um die Zusammenhänge der Leistungen aller getesteten Kinder in den beiden Untertests von *Halt* überprüfen zu können, wurden die prozentualen Anteile der richtig gelösten Aufgaben in den beiden Untertests miteinander korreliert. Dazu wurde der Rangkorrelationskoeffizient nach Spearman verwendet. Im folgenden Streudiagramm (vgl. Abb. 1) werden die Ergebnisse aller Kinder zwischen den beiden Untertests von *Halt* verglichen.

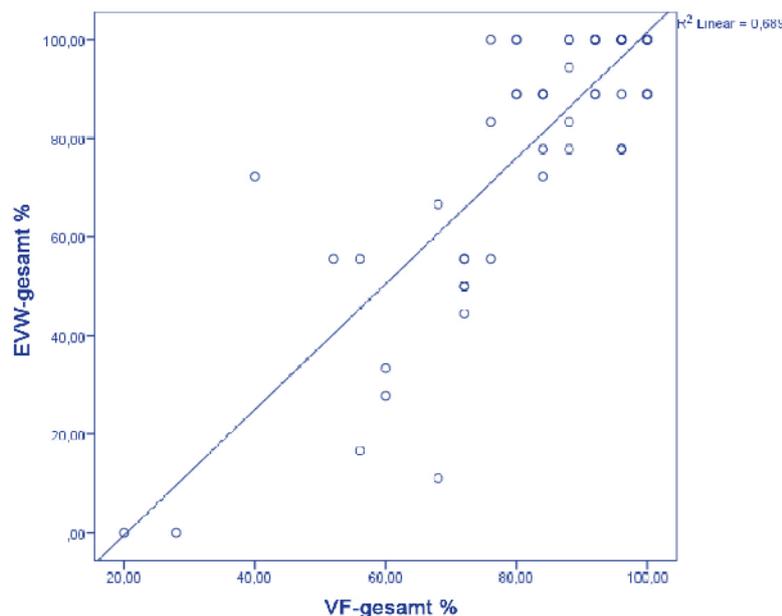


Abb. 1: Zusammenhang der Untertests von *Halt*

Durch die Ergebnisse in den Erwartungsverletzungen werden 68,9% der Varianz bei den Verständnisfragen aufgeklärt. Es zeigt sich eine hohe positive Korrelation von $r(56)=.724$ ($p<.001$) zwischen den Leistungen in den beiden Untertests von *Halt*.

■ Fragestellung 2: Unterschiede zwischen den Untersuchungsgruppen in den Untertests von *Halt*

Im Untertest *Verständnisfragen* zeigen die rezeptiv unauffälligen Kinder einen Mittelwert von 22.0 ($SD=2.71$). Die rezeptiv auffälligen hingegen weisen einen Mittelwert von 15.5 ($SD=5.74$) auf. Ein Mittelwertvergleich mittels des Mann-Whitney-U-Tests ergab einen signifikanten Unterschied ($U=85$, $p<.001$). Im Untertest *Entdecken von Erwartungsverletzungen* zeigen die rezeptiv unauffälligen Kinder einen Mittelwert von 15.6 ($SD=3.59$), während die rezeptiv auffälligen einen Mittelwert von 9.2 ($SD=5.79$) aufweisen. Ein Mittelwertvergleich mittels der Mann-Whitney-U-Tests für zwei unabhängige Stichproben ergab ebenfalls einen signifikanten Unterschied ($U=83$, $p<.001$). Das folgende Boxplotdiagramm verdeutlicht diese Gruppenunterschiede für beide Untertests (vgl. Abb. 2).

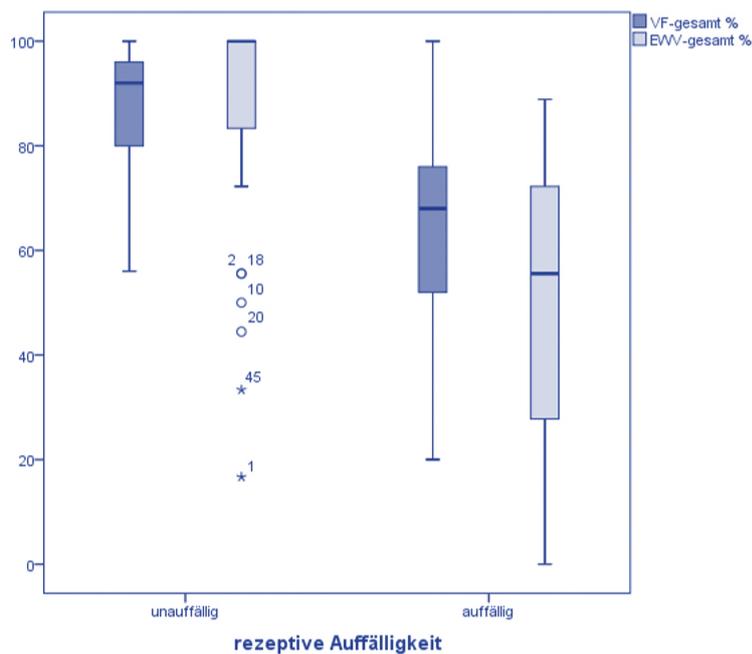


Abb. 2: Gruppenunterschiede für die Untertests von Halt

Die rezeptiv auffälligen Kinder erzielen in beiden Untertests niedrigere Werte als die unauffälligen. Die Mediane der sprachverstehensauffälligen Kinder liegen in einem geringeren Prozentbereich. Die Streuung der Rohwerte ist bei den rezeptiv auffälligen Probanden besonders im Untertest *Entdecken von Erwartungsverletzungen* deutlich größer. Allerdings sind in der Gruppe der unauffälligen Kinder bei den Erwartungsverletzungen Ausreißer und Extremwerte zu finden.

■ Fragestellung 3: Unterschiede zwischen den Untersuchungsgruppen bei den unterschiedlichen Fragearten im Untertest *Verständnisfragen*

Des Weiteren wurden die Leistungen bei den verschiedenen Fragearten zwischen den Untersuchungsgruppen der rezeptiv unauffälligen und auffälligen Kinder verglichen. Beide Gruppen weisen bessere Leistungen bei den Informationsfragen als bei den Inferenzfragen auf. Die rezeptiv unauffälligen Kinder schneiden sowohl bei den Informations- als auch bei den Inferenzfragen besser ab als die Kinder mit rezeptiver Auffälligkeit. Ein Mittelwertvergleich zwischen den Gruppen mittels des Mann-Whitney-U-Tests für zwei unabhängige Stichproben ergab sowohl für Informationsfragen einen signifikanten Unterschied ($U=104$, $p<.001$), als auch für Inferenzfragen ($U=86$, $p<.001$). Beim Vergleich der lokalen und globalen Inferenzfragen fällt auf, dass beide Gruppen bessere Leistungen bei den lokalen Inferenzfragen zeigen. Die rezeptiv unauffälligen Kinder weisen bei beiden Fragearten bessere Leistungen auf. Die niedrigsten Prozentwerte weisen die rezeptiv auffälligen Kinder bei den globalen Inferenzfragen auf. Ein Mittelwertvergleich zwischen den Gruppen mittels der Mann-Whitney-U-Tests für zwei unabhängige Stichproben ergab bei lokalen Inferenzfragen einen signifikanten Unterschied ($U=95$, $p<.001$), ebenso bei globalen Inferenzfragen ($U=121$, $p<.001$).

■ Fragestellung 4: Unterschiede zwischen den Untersuchungsgruppen in den Reaktionsmustern im Untertest *Entdecken von Erwartungsverletzungen*

Es wurde überprüft, ob sich die Reaktionen der rezeptiv auffälligen von den unauffälligen Kindern im Hinblick auf die verschiedenen Erwartungsverletzungen unterscheiden (vgl. Tab. 4). Der prozentuale Vergleich der Reaktionen der rezeptiv unauffälligen und auffälligen Kinder auf die Erwartungsverletzungen ergibt, dass beide Gruppen am häufigsten mit einer Spezifizierung bzw. Korrektur der Erwartungsverletzung reagieren, die rezeptiv unauffälligen Kinder diese Reaktion jedoch häufiger (85,7 %) zeigen als die rezeptiv auffälligen (46,2 %). Die Kinder mit Verstehensdefiziten zeigen etwa gleich oft (45,3 %) keine Reaktion, während die Kinder ohne Verstehensdefizite nur in 11,4 % der Fälle nicht reagieren. In beiden Gruppen kommt der isolierte Widerspruch relativ selten vor.

Tab. 4: Vergleich der Häufigkeitsverteilungen der Reaktionen auf die Erwartungsverletzungen

	Keine Reaktion	Widerspruch (Nein)	Spezifizierung/Korrektur
Rezeptiv unauffällig	11,4 %	2,9 %	85,7 %
Rezeptiv auffällig	45,3 %	8,5 %	46,2 %

■ Fragestellung 5: Itemanalyse

Zur Bestimmung und Verbesserung der psychometrischen Güte des entwickelten Verfahrens wurde eine Itemanalyse durchgeführt. Dabei sollen Items mit zu geringer oder hoher Schwierigkeit und zu geringer Trennschärfe ermittelt werden, um nicht genügend geeignete Items zu eliminieren oder zu verbessern.

Schwierigkeitsindizes

Der Schwierigkeitsindex einer Aufgabe ist nach Lienert und Raatz (1998) der prozentuale Anteil P der auf diese Aufgabe entfallenden richtigen Antworten in einer Analysestichprobe von der Größe N . Der Schwierigkeitsindex liegt somit bei schwierigen Aufgaben niedrig und bei leichten hoch. Ein Wert zwischen 20 und 80 % gilt dabei als akzeptabel (Lienert & Raatz, 1998). Allerdings soll das Verfahren *Halt* rezeptiv auffällige Kinder herausfiltern und erhebt nicht den Anspruch, zwischen guten Verstehern zu differenzieren. Die Schwierigkeitsindizes sollten daher in einem Bereich liegen, der diese beiden Gruppen besonders gut differenziert, was in diesem Fall einem hohen Schwierigkeitsindex entspricht. Deswegen sind auch Werte von ca. 85 % erwünscht (Gebhard, 2008).

Im Untertest Verständnisfragen erfolgt die Bestimmung der Itemschwierigkeiten über eine zufallskorrigierte Formel, da es sich hierbei um Mehrfachwahlaufgaben mit zwei Distraktoren handelt. Sieben Items in diesem Untertest zeigen zu geringe Schwierigkeiten ($P > 85$), eine globale Inferenzfrage weist einen negativen Schwierigkeitsindex und damit eine zu hohe Schwierigkeit auf. Die durchschnittliche Itemschwierigkeit liegt unter Ausschluss des negativen Schwierigkeitsindex dieses Items bei $P=77$. Es sind sowohl relativ schwierige Aufgaben als auch mittlere Schwierigkeiten und leichte Items enthalten. Im Hinblick auf den Verwendungszweck dieses Verfahrens, die rezeptiv schwächsten Kinder der Gesamtgruppe herauszufiltern, sind auch Items mit Schwierigkeiten über $P=80$ integriert.

Im Untertest Erwartungsverletzungen geschieht die Berechnung der Itemschwierigkeiten nicht über die Zufallskorrektur, da hier Raten eine eher untergeordnete Rolle spielt. Außerdem erfolgt keine dichotome, sondern eine dreistufige Bewertung (0/1/2 Punkte). Die Berechnung der Itemschwierigkeit wird über das Verhältnis des Itemmittelwerts zu dem maximal erreichbaren Wert eines Items berechnet. Die neun Items des zweiten Untertests weisen einen geeigneten Schwierigkeitsgrad zwischen 60-89 % auf. Die durchschnittliche Itemschwierigkeit liegt bei $P=79$. Damit erweist sich dieser Untertest im Gesamten als etwas leichter als der erste Untertest.

Trennschärfe

Die Trennschärfe bezeichnet die Korrelation eines Items mit dem Gesamtergebnis eines Tests. Sie gibt also an, wie gut ein einzelnes Item das gesamte Testergebnis voraussagt. Der mögliche Wertebereich liegt dabei zwischen -1 und 1. Bei einer hohen positiven Trennschärfe erfasst das Item dasselbe Kriterium wie der Gesamttest, während bei einer Trennschärfe nahe 0 das Item den Gesamttest nicht ausreichend widerspiegelt. Die Trennschärfe gibt an, wie gut ein Item zwischen Personen mit niedriger und hoher Merkmalsausprägung unterscheidet. Ein Item sollte mindestens eine Trennschärfe von 0,3 aufweisen. Zwischen Itemschwierigkeit, Trennschärfe und Reliabilität besteht ein enger Zusammenhang, da eine Variation von Itemschwierigkeiten zu einer Abnahme der Homogenität und damit zu einer Abnahme der Trennschärfe führt. Eine mittlere Itemschwierigkeit begünstigt daher eine hohe Trennschärfe und eine höhere Reliabilität (Bühner, 2011).

Sechs Items des Untertests Verständnisfragen weisen eine zu geringe Trennschärfe auf ($< 0,3$), das Item mit der höchsten Trennschärfe stellt eine globale Inferenzfrage nach dem übergeordneten Ziel des Stinktiers dar.

Im Untertest Erwartungsverletzungen weisen alle Items eine hohe Trennschärfe über 0,5 auf und sind daher dazu geeignet, um zwischen Personen mit niedriger und hoher Merkmalsausprä-

gung zu trennen. Die Erwartungsverletzung mit besonders hoher Trennschärfe (0,74) beinhaltet eine Zielunterbrechung.

Reliabilität

Die Reliabilität bezeichnet die Messgenauigkeit eines Tests. Für beide Untertests wurden jeweils Cronbachs Alpha als Maß für die interne Konsistenz einer Skala und die Split-Half-Reliabilität berechnet. Bei der Berechnung der internen Konsistenz wird der Test in so viele Untertests zerlegt, wie er Items besitzt. Dabei erfolgt eine Überprüfung des inneren Zusammenhangs der Items unter Berücksichtigung der Testlänge. Die interne Konsistenz bezieht sich daher auf die Korrelation der Items untereinander. Es ist zu beachten, dass Cronbachs Alpha ein kombinierter Index ist, der die Höhe der Itemkorrelationen und gleichzeitig die Itemanzahl berücksichtigt (Bühner, 2011). Für die Untertests des Verfahrens *Halt* ergaben sich dabei die folgenden Werte (vgl. Tab. 5):

Tab. 5: interne Konsistenz der Untertests von Halt

Untertest	Cronbachs Alpha
Verständnisfragen	.876
Entdecken von Erwartungsverletzungen	.878

Zur Feststellung der Split-Half Reliabilität wurden die Items ihrer Reihenfolge nach in eine erste und zweite Hälfte geteilt und miteinander korreliert. Mithilfe der Spearman-Brown-Formel wird die Korrelation der beiden Hälften auf die ursprüngliche Testlänge hochgerechnet. Für den Untertest *Verständnisfragen* ergaben sich dabei folgende Werte (vgl. Tab. 6):

Tab. 6: Split-Half-Reliabilität Untertest Verständnisfragen

Reliabilitätsstatistiken Untertest VF		
Cronbachs Alpha	Teil 1	.834
	Teil 2	.724
Korrelation zwischen Formen		.744
Spearman-Brown-Koeffizient		.853

Die Ergebnisse für den Untertest *Verständnisfragen* zeigen, dass beide Testhälften jeweils bereits zufriedenstellende Reliabilitäten (Cronbachs Alpha) aufweisen. Allerdings sind die beiden Testhälften nicht exakt parallel: Die erste Testhälfte weist eine höhere interne Konsistenz (.834) auf als die zweite Testhälfte (.724). Die Korrelation zwischen den Testhälften

ohne Spearman-Brown-Korrektur ist bereits zufriedenstellend (.744). Die Schätzung der Reliabilität der Gesamtskala unter Verwendung der Spearman-Brown-Korrektur, die eine Aufwertung der Reliabilität der Testhälften auf die Reliabilität des Gesamttests vornimmt, zeigt mit einem Wert von .853 einen etwas geringeren Wert an als die Schätzung der Reliabilität der Gesamtskala über Cronbachs Alpha.

Für den Untertest *Entdecken von Erwartungsverletzungen* wurden folgende Werte erzielt (vgl. Tab. 7):

Tab. 7: Split-Half-Reliabilität Untertest Entdecken von Erwartungsverletzungen

Reliabilitätsstatistiken Untertest Erwartungsverletzungen		
Cronbachs Alpha	Teil 1	.800
	Teil 2	.759
Korrelation zwischen Formen		.787
Spearman-Brown-Koeffizient		.881

Die beiden Testhälften zeigen jeweils bereits zufriedenstellende Reliabilitäten (Cronbachs Alpha). Wiederum sind die beiden Testhälften nicht exakt parallel: Die erste Testhälfte weist eine etwas höhere interne Konsistenz (.800) auf als die zweite Testhälfte (.759). Die Korrelation zwischen den Testhälften ohne Spearman-Brown-Korrektur ist bereits zufriedenstellend (.787). Die Schätzung der Reliabilität der Gesamtskala unter Verwendung der Spearman-Brown-Korrektur zeigt mit einem Wert von .881 einen etwas höheren Wert an als die Schätzung der Reliabilität der Gesamtskala über Cronbachs Alpha.

6 Diskussion

6.1 Untertest Verständnisfragen

Da das Ziel des Verfahrens *Halt* darin liegt, Kinder mit Verstehensproblemen auf Textebene herauszufiltern, wurde untersucht, ob rezeptiv auffällige Kinder beim Beantworten der Verständnisfragen schlechtere Leistungen zeigen als rezeptiv unauffällige. Dazu wurden die Mittelwerte der beiden Gruppen statistisch verglichen. Die Mittelwerte der rezeptiv auffälligen und unauffälligen Kinder unterschieden sich signifikant, sodass angenommen werden kann, dass Kinder mit Verstehensdefiziten bei der Bearbeitung des Untertests *Verständnisfragen* besondere Schwierigkeiten aufweisen. In anderen Geschichtentests konnten ebenfalls unterdurchschnittliche Leistungen von sprachgestörten Kindern beim Beantworten von Verständnisfragen zu Geschichten festgestellt werden (u. a. Heel & Janda, 2011, Skarakis-Doyle et al., 2008a, Gebhard, 2008).

In mehreren Studien wurde gezeigt, dass Kinder mit Sprachstörungen bei Aufgaben zur Bildung von Inferenzen schlechter abschneiden als gleichaltrige Kinder mit normaler Sprachentwicklung (u. a. Dodwell & Bavin, 2008, Karasinski & Ellis Weismer, 2010). Es lassen sich jedoch keine ganz übereinstimmenden Ergebnisse darüber finden, ob sprachlich auffällige Kinder ein isoliertes Defizit beim Beantworten von schlussfolgernden Fragen zu Geschichten aufweisen oder ob ebenso Schwierigkeiten beim Beantworten von Informationsfragen bestehen. Im Verfahren *Halt* schnitten sowohl die rezeptiv unauffälligen als auch die rezeptiv auffälligen Kinder besser bei den Informationsfragen als bei den Inferenzfragen ab, was für eine allgemein erhöhte Schwierigkeit der Inferenzfragen spricht. Beim statistischen Vergleich der beiden Gruppen konnte allerdings festgestellt werden, dass die rezeptiv unauffälligen Kinder bei beiden Fragearten signifikant höhere Mittelwerte erzielten. Diese Ergebnisse sprechen gegen isolierte Schwierigkeiten beim Beantworten von schlussfolgernden Fragen, die in einigen Studien berichtet werden (Dodwell & Bavin, 2008, Cain & Oakhill, 1999). Vielmehr zeigen die in dieser Studie getesteten rezeptiv auffälligen Kinder neben unterdurchschnittlichen Leistungen im Bilden von Inferenzen auch schlechtere Leistungen beim Beantworten von Informationsfragen. Verminderte Leistungen beim Lösen von Informationsfragen sprachauffälliger Kinder konnten ebenso in weiteren Studien festgestellt werden (Ellis Weismer, 1985, Bishop & Adams, 1992). Nach Bishop und Adams (1992) könnten u. a. Defizite, der Geschichte eine Struktur zu verleihen, die Schwierigkeiten bei Informations- und schlussfolgernden Fragen der rezeptiv auffälligen Kinder erklären. Ebenso könnten Probleme im phonologischen Arbeitsgedächtnis für das schlechte Beantworten von Informationsfragen verantwortlich sein, denn insbesondere die Kinder mit rezeptiven Auffälligkeiten hatten hier geringere Fähigkeiten.

Da im Verfahren *Halt* die unterschiedliche Verstehenstiefe einbezogen wird und deswegen auch die schlussfolgernden Fragen hinsichtlich ihrer Komplexität unterschieden werden, sollen diese getrennt voneinander betrachtet werden. In Anlehnung an das Konstruktions-Integrationsmodell nach van Dijk und Kintsch (1983) wird zwischen eher hierarchieniedrigen Verarbeitungsprozessen, die für die Herstellung lokaler Kohärenz verantwortlich sind, und eher hierarchiehöheren Verarbeitungsprozessen, die die Herstellung globaler Kohärenz ermöglichen, differenziert. Diese unterscheiden sich dadurch in der Distanz der zu integrierenden Informationen. Der Vergleich der beiden Gruppen ergab, dass sich signifikante Gruppenunterschiede sowohl bei den lokalen als auch den globalen Inferenzen feststellen ließen, was sich mit den Ergebnissen von Karasinski und Ellis Weismer (2010) für sprachauffällige Jugendliche deckt. Daher stellt nicht nur die globale Kohärenzbildung, die von höheren Anforderungen an das Arbeitsgedächtnis abhängt, eine besondere Schwierigkeit für sprachverstehensgestörte Kinder dar. Das Ziehen von lokalen Schlussfolgerungen ist zwar den hierarchieniedrigen Verarbeitungsprozessen zuzuordnen, bereitet den rezeptiv auffälligen Kindern ebenfalls Probleme.

Allerdings maskiert der Vergleich von Gruppenmittelwerten mögliche Variationen innerhalb der Gruppen. Weitere Faktoren wie ToM oder personbezogene Faktoren wie Motivation und Erfahrung könnten Einfluss nehmen. Denn es handelt sich beim Textverstehen um interagierende Verarbeitungsprozesse. Somit können auch inhaltliche Präferenzen eine Rolle bei der Fragenbeantwortung spielen. Beispielsweise fiel beim Vorlesen der Geschichte auf, dass besonders das Loslassen der Stinkbombe häufig Lachen oder Kommentare der Kinder hervorrief. Diese Stelle der Geschichte scheint also für die Kinder besonders interessant bzw. witzig zu sein und sollte daher auch besser verstanden und erinnert werden. Daher ist es wichtig, die vorgenommene Einteilung der Fragen als hilfreich für die qualitative Auswertung, aber nicht als einzige Betrachtungsmöglichkeit anzusehen und in weiteren Studien zusätzliche Faktoren mit zu berücksichtigen.

6.2 Untertest Entdecken von Erwartungsverletzungen

Der zweite Untertest dient der Erfassung der Leistungen im Monitoring des Sprachverstehens auf Textebene. Hierbei wurde eine dreistufige Bewertung gewählt, welche sich von Skarakis-Doyle (2002) und Heel und Janda (2011, 2014) unterscheidet, die die Methode allerdings bei jüngeren Kindern einsetzten. Im Einklang mit Skarakis-Doyle und Dempsey (2008b) sowie Heel und Janda (2011, 2014) wiesen die rezeptiv auffälligen Kinder signifikant geringere Mittelwerte auf als die rezeptiv unauffälligen Kinder. Bei den rezeptiv unauffälligen Kindern gab es jedoch einige Ausreißer mit geringen Werten im Monitoring des Sprachverstehens. Diese könnten möglicherweise durch geringe Fähigkeiten im phonologischen Arbeitsgedächtnis oder geringe expressive Fähigkeiten für die Korrektur der Erwartungsverletzung, aber auch durch geringe Aufmerksamkeitsleistungen erklärt werden. Studien belegen, dass Kinder mit Aufmerksamkeitsstörungen besondere Probleme im Monitoring des Sprachverstehens haben (Berthiaume, Lorch & Milich, 2010). In zukünftigen Studien sollte dieser Aspekt deshalb zusätzlich erfasst werden.

Insgesamt ist für beide Gruppen feststellbar, dass ein isolierter Widerspruch („Nein“) relativ selten vorkam. Die rezeptiv unauffälligen Kinder reagierten eindeutig am häufigsten mit einer zusätzlichen Spezifizierung bzw. einer Korrektur (85,7%). Nur in 11,4% der Fälle reagierten sie gar nicht. Die Kinder mit Verstehensdefiziten hingegen zeigten in den Bewertungskategorien *keine Reaktion* und *Spezifizierung bzw. Korrektur* etwa gleich häufige Verteilungen. Expressive Sprachprobleme könnten für eine fehlende Korrektur verantwortlich sein. Da allerdings bei den rezeptiv auffälligen Kindern häufig gar nicht reagiert wurde, spricht dies für ein Nicht-Erkennen von Unstimmigkeiten, möglicherweise aufgrund von Problemen in der Aufmerksamkeit oder aufgrund einer generellen Nichtevaluation von Inhalten in Bezug auf Erwartungsverletzungen oder Inkonsistenzen und damit einem mangelhaften Monitoring des Sprachverstehens. Dies bleibt in weiteren Studien zu klären.

6.3 Itemanalyse

Die meisten Items des ersten Untertests *Verständnisfragen* weisen einen geeigneten Schwierigkeitsgrad auf. Es sind sowohl relativ schwierige Aufgaben als auch mittlere Schwierigkeiten und leichte Items enthalten. Die trennscharfen Items bestehen ungefähr zu gleichen Teilen aus Informations- und schlussfolgernden Fragen. Die Itemanalyse lieferte wichtige Hinweise für die Überarbeitung des Untertests *Verständnisfragen*. Da die Methode *Entdecken von Erwartungsverletzungen* zuvor noch nicht an siebenjährigen Kindern durchgeführt wurde, stellte sich bei der Konzeption des Untertests die Frage nach einem geeigneten Bekanntheitsgrad der Geschichte und Schwierigkeitsgrad für Kinder dieses Alters. Durch den Ausschluss des zweiten Untertests *Gemeinsames Nacherzählen* und das dreistufige Bewertungssystem ohne Einbezug nonverbaler Reaktionen sollten die Anforderungen den Fähigkeiten von Schulkindern angepasst werden. Die Itemschwierigkeiten weisen darauf hin, dass dies gelungen ist. Somit stellte sich die Geschichte als ausreichend bekannt für die Durchführung der Überprüfung des Monitoring des Sprachverstehens heraus. Alle Items dieses Untertests weisen eine hohe Trennschärfe über 0,5 auf und sind daher dazu geeignet, um zwischen Personen mit niedriger und hoher Merkmalsausprägung zu trennen. Das trennschärfste Item stellte dabei eine Zielunterbrechung dar.

Für beide Untertests konnten sowohl hinsichtlich der internen Konsistenz als auch der Split-Half Reliabilität zufriedenstellende Werte festgestellt werden.

6.4 Methodenkritik

Für die Untersuchung wurden die Kinder anhand einer Auffälligkeit im TROG-D und/oder der Mäusegeschichte in rezeptiv auffällig vs. unauffällig eingeteilt. Diese Einteilung birgt Probleme, da die Mehrzahl der Kinder nur in einem Testverfahren und nicht in beiden auffällig war. Auch wenn das grammatische Verstehen mit dem Textverstehen korreliert, könnten einige der Kinder isolierte grammatische Schwierigkeiten aufweisen, was die durchaus auch guten Verstehensleistungen bei der Halt-Geschichte einiger Kinder der rezeptiv auffälligen Gruppe belegt. Es ist auch fraglich, ob der geforderte Wert von 7 in der Mäusegeschichte eindeutig im Hinblick auf die Gruppen differenziert, da in der gewählten Altersgruppe nur fünf Kinder unterhalb der geforderten Punkte in der Mäusegeschichte lagen. Bei 58 Kindern wären bei der Grenze von 15% der Schlechtesten jedoch acht bis neun rezeptiv auffällige Kinder zu erwarten bzw. durch die hohe Anzahl von Kindern aus Sprachheilschulen eher mehr. In der Gruppe der rezeptiv nicht auffälligen Kinder waren u. a. Kinder, die exakt den geforderten Wert von 7 erreichten, so dass auch dieser Wert noch Schwierigkeiten belegen könnte. Dafür sprechen auch die geringen Leistungen

in den Subtests der Haltgeschichte einiger Kinder der rezeptiv nicht auffälligen Gruppe. In zukünftigen Untersuchungen sind deshalb weitere Parameter mit in die Gruppenauswahl miteinzubeziehen. Beispielsweise könnten expressive Sprachentwicklungsstörungen, Auffälligkeiten im phonologischen Arbeitsgedächtnis und in der ToM als weitere Marker oder für die Einteilung in weitere Subgruppen herangezogen werden.

Für die Überprüfung der ToM wurde das Stofftier einer Ratte herangezogen. Einige Kinder reagierten hier verunsichert oder nannten bei der Frage nach dem Inhalt der Smartiesschachtel eine ihrer Meinung nach typische Nahrung einer Ratte oder antworteten „Weiß ich nicht!“. Diese Reaktionen hätten möglicherweise umgangen werden können, wenn eine zweite Person bei der Testung anwesend gewesen wäre, die erst nach dem Besprechen des tatsächlichen Inhalts der Smartiesschachtel den Raum betreten hätte. Da vermutlich für alle Menschen eine Smarties-Schachtel den Inhalt „Smarties“ impliziert, während Tieren Smarties-Schachteln normalerweise unbekannt sind, wären die Kinder somit aufgrund der Aufgabenstellung wahrscheinlich nicht gleichermaßen verunsichert gewesen. Zudem wird diese Aufgabe normalerweise mit jüngeren Kindern durchgeführt, sodass eine Anpassung an das Alter der im Rahmen dieser Arbeit getesteten Kinder eventuell von Vorteil gewesen wäre. Ebenso würde eine Differenzierung der Überprüfung mit mehreren Fragen wie bei Kim (2016) zu genaueren Ergebnissen führen.

7 Fazit/Ausblick

Die Ergebnisse dieser ersten Erprobung des Verfahrens *Halt* zeigen geeignete Schwierigkeiten sowie Trennschärfen der Items und gute Reliabilitäten und sprechen somit dafür, dass mit *Halt* eine mögliche Alternative bzw. Ergänzung zu den wenigen vorhandenen Testverfahren gefunden werden konnte. Der Untertest *Verständnisfragen* sollte nach einer Kürzung und Überarbeitung der Items erneut an einer größeren Stichprobe überprüft werden. Die Wahl der Geschichte im Hinblick auf die Einordbarkeit in das gängige Geschichtenschema und die Konzeption der Verständnisfragen unter Berücksichtigung von Ziel-Plan-Strukturen der Protagonisten und dem Verstehen von Täuschungen stellte sich als wichtig zur Unterscheidung zwischen Kindern mit und ohne Defiziten im Textverstehen heraus. Durch eine qualitative Betrachtung der Verständnisfragen kann außerdem zwischen verschiedenen inhaltlichen Aspekten des Textverstehens sowie der unterschiedlichen Verstehentiefe differenziert werden. Eine quantitative Erfassung der Verstehentiefe wäre allerdings nach einer weiteren Erprobung erstrebenswert. Für den Untertest *Entdecken von Erwartungsverletzungen* konnte gezeigt werden, dass die vorgenommenen Änderungen der ursprünglichen Methode nach Skarakis-Doyle (2002) für siebenjährige Kinder geeignet sind. Neben einer umfangreicheren Stichprobe zur Evaluation wäre es wichtig, das Verfahren auch an Kindern in anderen Altersstufen zu erproben, um eine Entwicklungssensitivität der Überprüfungsmethoden feststellen und eine mögliche Erweiterung des Altersbereichs vornehmen zu können. Bei jüngeren Kindern könnte die Geschichte inhaltlich die Merkfähigkeit übersteigen. Deshalb wird derzeit für diese Kinder der Einsatz des Verfahrens mit zusätzlicher Bildunterstützung durch Metacom-Symbole überprüft. Generell wäre das Verfahren durch die geringe Durchführungs- und Auswertungszeit im sprachtherapeutischen und inklusiven Alltag praktikabel und könnte als wichtige Ergänzung zum TROG-D und zu vorhandenen Geschichtenüberprüfungen herangezogen werden, um eine umfassendere Betrachtung der Leistungen im primären Textverstehen und auch im Monitoring des Sprachverstehens zu erhalten.

8 Literatur

- Behrens, U. (2010): Aspekte eines Kompetenzmodells zum Zuhören und Möglichkeiten ihrer Testung. In V. Bernius & M. Imhof (Hrsg.). *Zuhörkompetenz in Unterricht und Schule. Beiträge aus Wissenschaft und Praxis* (S. 31-51). Göttingen: Vandenhoeck & Ruprecht.
- Berthiaume, K.S., Lorch, E.P. & Milich R. (2010). Inferential processing and comprehension monitoring in boys with ADHD. *Journal of Attention Disorders*, 14 (1), 31-42.
- Bishop, D.V.M. & Adams, C. (1992). Comprehension problems in children with specific language impairment: *Literal and inferential meaning*. *Journal of Speech & Hearing Research*, 35 (1), 119-129.
- Buhl, H. M., Möller, F., Oebser, M., Stein, F. & Noack, P. (2009). Zusammenhänge zwischen der Fähigkeit zur Perspektivenübernahme und dem Textverstehen. *Diskurs Kindheits- und Jugendforschung*, 1, 75-90.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion*. 3. aktualisierte Aufl. München: Pearson.

- Cain, K. & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal*, 11, 489-503.
- Clegg, J., Hollis, C., Mawhood, L. & Rutter, M. (2005). Developmental language disorders – a follow-up in later adult life. Cognitive, language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry*, 46 (2), 128-149.
- Dodwell, K. & Bavin, E. L. (2008). Children with specific language impairment: an investigation of their narratives and memory. *International Journal of Language and Communication Disorders*, 43 (2), 201-218.
- Dore, R.A., Amendum, S.J., Golinkoff, R.M. & Hirsh-Pasek, K. (2018). Theory of mind: a hidden factor in reading comprehension? *Educational Psychology Review* 30 (3), 1067-1089.
- Ellis Weismer, S. (1985). Constructive comprehension abilities exhibited by languagedisordered children. *Journal of Speech and Hearing Research*, 28 (2), 175-184.
- Fox, A. V. (2006). *Test zur Überprüfung des Grammatikverständnisses (TROG-D)*. Idstein: Schulz-Kirchner.
- Gebhard, W. (2008). *Entwicklungsbedingte Sprachverständnisstörungen bei Kindern im Grundschulalter. Status und Diagnostik im klinischen Kontext*. 2. aktualisierte Aufl. München: Herbert Utz.
- Gopnik, A. & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, 59 (1), 26-37.
- Graesser, A. C., Olde, B. A. & Klettke, B. (2002). How does the mind construct and represent stories? In M.C. Green, J.J. Strange & T.C. Brock (Hrsg.). *Narrative impact: Social and cognitive foundations* (S. 213-263). Mahwah, NJ: Lawrence Erlbaum.
- Hachul, C. & Schönauer-Schneider, W. (2016). *Sprachverstehen bei Kindern. Grundlagen Diagnostik Therapie*. München: Elsevier.
- Hagemeister, F. (2011). *Monitoring des Sprachverstehens: Vergleich von Kindern mit spezifischen Sprachentwicklungsstörungen und Kindern mit unterdurchschnittlichen kognitiven Fähigkeiten*. Unveröffentlichte Bachelorarbeit. Ludwig-Maximilians-Universität München.
- Hall, C. & Barnes, M.A. (2017). Inference instruction to support reading comprehension for elementary students with learning disabilities. *Intervention in School and Clinic*, 52 (5), 279-286.
- Heel, M. & Janda, S. (2011). *Diagnostik von Sprachverständnis im Vorschulalter. Explorative Untersuchung zum Textverständnis*. Unveröffentlichte Masterarbeit an der LMU München.
- Heel, M. & Janda, S. (2014). Mehrdimensionale Diagnostik von Textverständnis im Vorschulalter mit der Timogeschichte – eine explorative Untersuchung. *Forschung Sprache*, 2 (1), 14-32.
- Johnson-Laird, P.N. (1983). *Mental models*. Cambridge: Cambridge University Press.
- Karasinski, C. & Ellis Weismer, S. (2010). Comprehension of inferences in discourse processing by adolescents with and without language impairment. *Journal of Speech, Language, and Hearing Research*, 53, 1268-1279.
- Kim, Y.-S. (2016). Direct and mediated effects of language and cognitive skills on comprehension of oral narrative texts (listening comprehension) for children. *Journal of Experimental Child Psychology*, 141, 101-120.
- Kim, Y.-S. (2015). Language and cognitive predictors of text comprehension: Evidence from multivariate analysis. *Child Development*, 86, 128-144.
- Kintsch, W. & Rawson, K.A. (2007). Comprehension. In M.J. Snowling & Ch. Hulme (Eds.). *The Science of Reading: A Handbook* (p. 209-226). Malden, Oxford & Victoria: Blackwell Publishing.
- Lienert, G. & Raatz, U. (1998). *Testaufbau und Testanalyse*. 6. Aufl. Weinheim: Psychologie Verlags Union
- Linder, M. & Grisseemann, H. (2000). *Züricher Lesetest*. 6. Aufl. Bern: Huber
- Magliano, J. P., Taylor, H. A. & Kim, H.-J. J. (2005). When goals collide: Monitoring the goals of multiple characters. *Memory & Cognition*, 33 (8), 1357-1367.
- Mai, M. (1995). *44 Zweiminutengeschichten*. Ravensburg: Ravensburger Buchverlag.
- Marks, D., Mayer, A. & Schönauer-Schneider, W. (2018). Vermittlung von Verstehensstrategien zur Förderung des Textverständnisses. In: T. Jungmann, B. Gierschner, M. Meindl & S. Sallat (Hrsg.). *Sprach- und Bildungshorizonte. Wahrnehmen – Beschreiben – Erweitern* (S. 41-53). Idstein: Schulz-Kirchner Verlag.
- Mathieu, S. (1995). Sprachverständnis im Kindergarten – Vergleichende Untersuchung. *Vierteljahresschrift für Heilpädagogik und ihre Nachbargebiete*, 64, 36-52.
- Mathieu, S. (1998). Entwicklung und Abklärung des Sprachverständnisses. In B. Zollinger (Hrsg.). *Kinder im Vorschulalter. Erkenntnisse, Beobachtungen und Ideen zur Welt der Drei- bis Siebenjährigen* (S. 83-137). Bern, Stuttgart, Wien: Haupt.
- Oakhill, J. & Cain, K. (2007). Issues of causality in children's reading comprehension. In D. Mc Namara (Hrsg.). *Reading comprehension strategies: Theories, interventions, and technologies* (S. 47-72). New York: Erlbaum.
- Petermann, F., Metz, D. & Fröhlich, M. P. (2010). *Sprachstandserhebungstest für Fünf- bis Zehnjährige (SET 5-10)*. Göttingen: Hogrefe.
- Richter, T. & Schnotz, W. (2018). Textverstehen. In S. Buch, D. Rost & J. Sparfeldt (Hrsg.). *Handwörterbuch Pädagogische Psychologie* (4. Aufl. S. 826-837). Weinheim: Beltz.
- Rißling, J.-K., Ronniger, P., Petermann, F. & Melzer, J. (2016). Psychosoziale Belastungen bei Sprachentwicklungsstörungen. *Kindheit und Entwicklung*, 25, 145-152.

- Schank, R. C. & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Schulze, M. (2010). *Monitoring des Sprachverstehens im Kindergartenalter – Untersuchungen zum MSV auf Textebene bei Kindern mit und ohne spezifische Sprachentwicklungsstörungen*. Unveröffentlichte Zulassungsarbeit: Ludwig-Maximilians-Universität München
- Skarakis-Doyle, E. (2002). Young children's detection of violations in familiar stories and emerging comprehension monitoring. *Discourse Processes*, 33 (2), 175-197.
- Skarakis-Doyle, E., Dempsey, L. & Lee, C. (2008a). Identifying language comprehension impairment in preschool children. *Language, Speech, and Hearing Services in Schools*, 39, 54-65
- Skarakis-Doyle, E. & Dempsey, L. (2008). The detection and monitoring of comprehension errors by preschool children with and without language impairment. *Journal of Speech, Language and Hearing Research*, 51, 1227-1243.
- Snyder, L. & Caccamise, D. (2010). Comprehension processes for expository text: Building meaning and making sense. In M. Nippold & C. Scott (Hrsg.). *Expository discourse in children, adolescents, and adults. Development and disorders* (S. 13-39). New York, NY: Psychology Press.
- van den Broek, P.W., White, M. J., Kendeou, P. & Carlson, S. (2009). Reading between the lines. Developmental and individual differences in cognitive processes in reading comprehension. In R.K. Wagner, C. Schatschneider & C. Pythian-Sence, (Hrsg.). *Beyond Decoding. The Behavioral and Biological Foundations of Reading Comprehension* (S. 107-123). New York: The Guilford Press.
- van Dijk, T.A. & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Wassenburg, S.I., Beker, K., van den Broek, P. & van der Schoot, M. (2015). Children's comprehension monitoring relates differently to the multiple dimensions of narrative. *Reading and Writing* 28 (8), 1203-1232.
- Weyda, C. (2012). *Entwicklung und Erprobung eines Verfahrens zur Erfassung von Textverständnis im Grundschulalter*. Unveröffentlichte Masterarbeit im Studiengang Sprachtherapie an der LMU München.
- Zollinger, B. (2010). *Die Entdeckung der Sprache*. Bern: Haupt.

Zu den Autorinnen

Corinne Weyda (M.A.) beendete 2012 mit der Masterarbeit zum beschriebenen Diagnoseverfahren „Halt“ ihr Studium der Sprachtherapie an der Ludwig-Maximilians-Universität München. Nach Tätigkeit in sprachtherapeutischen Praxen sowie einem entwicklungstherapeutischen Kinderheim arbeitet sie seit 2016 als Sprachtherapeutin in der Spezialambulanz für Entwicklungsstörungen des Heckscher-Klinikums in München.

Wilma Schönauer-Schneider ist Professorin für Sprachbehindertenpädagogik an der Pädagogischen Hochschule Heidelberg mit den Schwerpunkten Sprachverstehen, Monitoring des Sprachverstehens und sprachheilpädagogischer Unterricht.

Korrespondenzadresse

Corinne Weyda
kbo-Heckscher-Klinikum Haar
Max-Isserlin-Straße 23
85540 Haar
corinne.veyda@kbo.de

Prof. Dr. Wilma Schönauer-Schneider
Pädagogische Hochschule Heidelberg
Institut für Sonderpädagogik
Keplerstr. 87
69102 Heidelberg
schoenauer@ph-heidelberg.de