

Sprachtherapeutische Diagnostik mit dem Mottier-Test

Viele Normierungen und viele Fragezeichen

Tanja Ulrich

ZUSAMMENFASSUNG. Als verursachender bzw. aufrechterhaltender Faktor für laut- und schriftsprachliche Defizite bei Kindern kommt der Kapazität des phonologischen Kurzzeitgedächtnisses eine wichtige Rolle zu. Im Rahmen der Sprachdiagnostik wird diese meist anhand von Aufgaben zum Nachsprechen von Pseudowörtern erfasst. Bei der Anwendung des weit verbreiteten Mottier-Tests stehen DiagnostikerInnen in Schule und Praxis oftmals vor dem Problem, sich begründet für eine der aktuellen Normierungen entscheiden zu müssen. Der vorliegende Beitrag stellt die Ergebnisse einer Pilotuntersuchung dar, in der die Leistungen von 30 fünfjährigen Kindern beim Nachsprechen von Pseudowörtern erfasst wurden. Die im Mottier-Test anhand dreier unterschiedlicher Normierungen erreichten Werte wurden miteinander verglichen sowie die Ergebnisse aus zwei standardisierten und normierten Untertests zum verbalen Kurzzeitgedächtnis als Außenkriterium herangezogen. Es zeigte sich, dass die verschiedenen Normierungen des Mottier-Tests in mindestens 10% der Fälle zu unterschiedlichen Entscheidungen hinsichtlich einer Therapieindikation führten. Zudem wurden die Leistungen des verbalen Kurzzeitgedächtnisses durch den Mottier-Test oftmals anders eingeschätzt als durch Untertests mit standardisierter Durchführung. Implikationen aus den Ergebnissen für die sprachtherapeutische Diagnostik werden aufgezeigt.

Schlüsselwörter: Diagnostik – verbales Kurzzeitgedächtnis – Sprachentwicklungsstörung

Dr. Tanja Ulrich schloss ihre Ausbildung zur Logopädin im Jahr 2002 ab und studierte anschließend Lehr- und Forschungslgopädie (Diplom) an der RWTH Aachen. 2012 promovierte sie an der Universität zu Köln. Aktuell arbeitet sie als wissenschaftliche Mitarbeiterin am Lehrstuhl für Sprachbehindertenpädagogik (Prof. Motsch) der Universität zu Köln und hat sich in Lehre, Forschung und Therapie auf den ungestörten und gestörten Spracherwerb bei Kindern spezialisiert.



Einleitung

Auf der Grundlage ihrer Normierungsstudie zum Mottier-Test stellt *Welte* (1981, 124) fest: „Mottier-Testergebnisse von <11 bei den 4- und 5-Jährigen und <12 bei den 6-Jährigen sind unterdurchschnittlich.“

Seit Erscheinen der ersten Normwerte für den Mottier-Test im Jahr 1981 wurden zwar einige weitere Referenz- und Vergleichswerte veröffentlicht (Tab. 1); die ursprünglichen Normen von *Welte* (1981) und *Bohny* (1981) wurden jedoch auch weiterhin herangezogen, um Leistungen von Kindern beim Nachsprechen von Pseudowörtern zu beurteilen.

Aus dem Bestreben heraus, aktuelle Normwerte für den Mottier-Test zu ermitteln, wurden in den letzten sechs Jahren drei größere Normierungsstudien in Deutschland und der Schweiz durchgeführt (*Kiese-Himmel & Risse* 2009, *Gamper et al.* 2012, *Wild & Fleck* 2013).

Erreicht ein fünfjähriges Kind bei der Diagnostik mit dem Mottier-Test nun einen Rohwert von 10 Punkten, so entspricht dies der Normierung von *Kiese-Himmel & Risse* (2009) zufolge einer grenzwertigen Leistung um den T-Wert 40 (T-Wert 42 für 5;0-5;5 J., T-Wert 39

für 5;6- 5;11 J.), in den beiden Schweizer Normierungen dagegen einer völlig unauffälligen Leistung im mittleren Prozentrangbereich (PR 51-83 bei *Gamper et al.* 2012, PR 53/T-Wert 50,8 bei *Wild & Fleck* 2013).

Für SprachtherapeutInnen und SprachheilpädagogInnen stellt sich daher berechtigterweise die Frage, ob der Mottier-Test als Diagnostikverfahren in Praxis und Schule weiterhin sinnvoll eingesetzt werden kann und welche der aktuellen Normierungen zu bevorzugen ist, um die Testergebnisse von Kindern zuverlässig beurteilen zu können.

Theoretischer Hintergrund

Der Mottier-Test

Der Mottier-Test wurde 1951 von der Zürcher Logopädin Grete Mottier entwickelt (*Mottier* 1951). Er sollte die Diagnostik lese-rechtschreibschwacher Kinder ergänzen und wurde als Zusatzverfahren „zur Prüfung der phonematischen Speicherung der sprechmotorischen Koordination und der Artikulation“ in den Zürcher Lesetest (ZLT) mit aufgenommen (*Linder & Grisseman* 1996, 15).

In der Weiterentwicklung dieses Verfahrens, dem Zürcher Lesetest II (ZLT II, *Petermann & Daseking* 2012) wurden die Silben des ursprünglichen Mottier-Tests verändert und als Untertest „Pseudowörter nachsprechen“ in die Testbatterie integriert.

Der Mottier-Test erfuhr in den letzten Jahrzehnten rege Verbreitung und dürfte mittlerweile den meisten in der Sprachdiagnostik tätigen Praktikern bekannt sein. Die Aufgabe des Kindes besteht darin, vorgespochene Pseudowörter in der korrekten Reihenfolge unmittelbar zu reproduzieren.

Die Pseudowörter des Mottier-Tests bestehen aus einfachen Konsonant-Vokal-Verbindungen, deren Länge von Zweisilbern (z.B. „re-la“) hin zu Sechssilbern (z.B. „bi-ga-dona-fe-ra“) zunimmt. Der Test besteht aus insgesamt 30 Items.

Mottier (1951) empfahl, die Pseudowörter mit verdecktem Mundbild und gleichmäßiger Betonung bei einer Sprechrage von einer Silbe pro Sekunde vorzusprechen (*Linder & Grisseman* 1996). Die Anzahl der vollständig korrekt nachgesprochenen Pseudowörter entspricht dem erreichten Rohwert.

Was wird mit dem Mottier-Test überprüft?

Aufgaben zum Pseudowörter-Nachsprechen werden in der Regel herangezogen, um die Kapazität der phonologischen Schleife des Arbeitsgedächtnisses (Baddeley 2003) bzw. des phonologischen Kurzzeitspeichers zu erfassen (Boudreau & Costanza-Smith 2011). Das Arbeitsgedächtnis ist ein „kognitives System [...], das es ermöglicht, mehrere Informationen vorübergehend zu speichern und miteinander in Beziehung zu setzen“ (Hasselhorn & Grube 2003, 32).

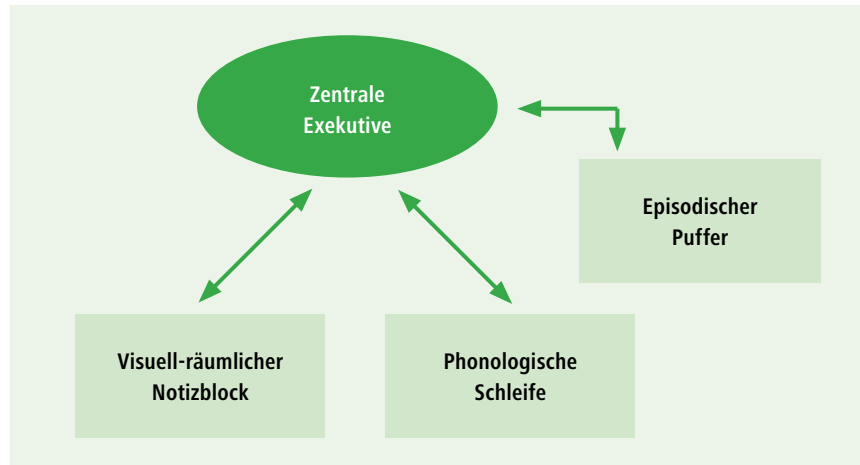
Im häufig verwendeten Modell von Baddeley (z.B. 2003) besteht dieses System aus einer übergeordneten, modalitätsunabhängigen Kontrolleinheit, der zentralen Exekutive, die für die Aufmerksamkeitskontrolle verantwortlich ist, sowie zwei modalitätsspezifischen Hilfssystemen, dem visuell-räumlichen Notizblock und der phonologischen Schleife. Der episodische Buffer bringt schließlich modalitätsspezifische Informationen zusammen und verknüpft sie zu Episoden (Abb. 1).

Die phonologische Schleife ist der Teil des Arbeitsgedächtnisses, der für die Speicherung von sprachlichen und schriftsprachlichen Informationen zuständig ist. Sie besteht aus einem phonologischen „buffer“ (auch: phonologischer oder phonetischer Speicher), in dem verbale Informationen für etwa 1,8 Sekunden gehalten werden können. Ein subvokaler, artikulatorischer Kontrollprozess, das „rehearsal“, ermöglicht es, die Informationen über dieses Zeitfenster hinaus aufrecht zu erhalten, indem das zu analysierende Material mehrfach laut wiederholt, geflüstert oder innerlich vorgesprochen wird.

Die auditive Merk- oder Gedächtnisspanne wird in der Regel als „Maß für die funktionale Gesamtkapazität des phonologischen Arbeitsgedächtnisses“ herangezogen (Hasselhorn & Grube 2003, 33). Sie wird ermittelt über die Anzahl der Zahlenfolgen oder Pseudowörter, die unmittelbar nachgesprochen werden können. Das Nachsprechen von bedeutungsfreien Pseudowörtern ermöglicht es dabei besonders gut, die Kapazität der phonologischen Schleife unabhängig vom lexikalischen Vorwissen zu erfassen (Baddeley 2003, Graf Estes et al. 2007).

In aktuelleren Veröffentlichungen wird in der Regel darauf hingewiesen, dass für die reine Wiedergabe von Informationen ausschließlich Kurzzeitgedächtnisleitungen erbracht werden müssen. Daher wird die Fähigkeit, Pseudowörter (oder Zahlenfolgen) unmittelbar nachzusprechen, spezifischer als eine Leistung des „phonologischen oder verbalen Kurzzeitgedächtnisses“ beschrieben und abgegrenzt von Leistungen des Arbeitsgedäch-

■ Abb. 1: Modell des Arbeitsgedächtnisses nach Baddeley (2003)



nisses selbst, die unter Einbezug der zentralen Exekutive eine Bearbeitung des Gedächtnismaterials erfordern (z.B. Zahlenfolgen in umgekehrter Reihenfolge wiedergeben; Graf Estes et al. 2007, Archibald & Joanisse 2009, Boudreau & Costanza-Smith 2011).

Neben der Kapazität des phonologischen Kurzzeitgedächtnisses erfasst das Nachsprechen der Mottier-Silben jedoch noch weitere sprachverarbeitende Fähigkeiten des Kindes:

Analyse- und Verarbeitungsschritte beim Nachsprechen von Pseudowörtern

- „Einscannen“: Auditive und anschließende phonologische Analyse des Pseudowortes, Phonemdifferenzierung und -identifikation
- Reihenfolgenregetreues Memorieren der Lautkette (subvokales artikulatorisches Rehearsal) zum Aufrechterhalten in der phonologischen Schleife des Arbeitsgedächtnisses
- Wiedergabe der Lautkette: Aktivieren der entsprechenden motorischen Programme, Sprechplanung, Ausführen der Artikulationsbewegungen

Im Gegensatz zu anderen, ähnlichen Diagnostikverfahren (z.B. „Pseudowörter nachsprechen“ im SET 5-10, Petermann 2010) handelt es sich bei den Mottier-Silben um einfache Konsonant-Vokal-Verbindungen ohne Konsonantencluster oder geschlossene Silben, wodurch die Anforderungen an die artikulationsmotorische Umsetzung bei der Wiedergabe möglichst gering gehalten werden sollen (Kiese-Himmel & Risse 2009).

Dennoch kommen in vielen Silbensequenzen des Mottier-Tests beispielsweise die velaren

Plosive /k,g/ vor, die einigen aussprachege-störten Kindern Schwierigkeiten bereiten und so möglicherweise erhöhte Anforderung an die phonetischen und phonologischen Fähigkeiten dieser Kinder stellen können. In einigen Items erfordert die Wiedergabe darüber hinaus diadochokinetische – also schnelle, wechselhafte – Artikulationsbewegungen („ka-pe-to“), die für Kinder mit eingeschränkten artikulationsmotorischen Fähigkeiten ebenfalls erhöhte Anforderungen darstellen können.

Wozu eine Überprüfung des phonologischen Kurzzeitgedächtnisses im Rahmen der Sprachdiagnostik?

Zusammenhänge zwischen der Kapazität der phonologischen Schleife und dem lexikalischen Lernen bei Kindern sind mittlerweile umfassend belegt, wobei mit zunehmendem Alter auch der Umfang des bereits vorhandenen lexikalischen Wissens dazu beiträgt, wie gut Kinder neue Wörter erlernen können (Gathercole et al. 1997, Baddeley et al. 1998).

Darüber hinaus spielt das phonologische Arbeitsgedächtnis auch für andere Bereiche der Sprachentwicklung eine zentrale Rolle. Einschränkungen in der Kapazität der phonologischen Schleife können sich auf die phonologische, die lexikalische, die morphologische und auch die syntaktische Sprachebene auswirken (Archibald & Gathercole 2006, Montgomery & Evans 2009). Zudem stellt sowohl das phonologische Kurzzeit-, in stärkerem Maße jedoch das phonologische Arbeitsgedächtnis, einen signifikanten Prädiktor für spätere Leseleistungen dar (Carette et al. 2009).

Sprachentwicklungsgestörte Kinder erreichen in Aufgaben zum Nachsprechen von Pseudowörtern signifikant schlechtere Werte als sprachunauffällige Kinder. Graf Estes et

al. (2007) ermittelten im Rahmen einer Meta-Analyse, dass die Leistungen von sprachergestörten Kindern im Mittel 1,27 Standardabweichungen unterhalb der Leistungen sprachunauffälliger Kinder beim Pseudowörter-Nachsprechen liegen (vgl. auch *Baddeley et al.* 1998, *Baddeley* 2003).

Einige Autoren diskutieren, ob sich das Nachsprechen von Pseudowörtern als diagnostischer Marker zur Identifizierung von spezifisch sprachergestörten Kindern eigne (*Ellis Weismer et al.* 2000). Dies erscheint insbesondere vor dem Hintergrund zunehmender Zwei- und Mehrsprachigkeit der zu diagnostizierenden Kinder eine lohnenswerte Fragestellung (*Archibald & Joanisse* 2009).

Phonologisches Kurzzeitgedächtnis und Mehrsprachigkeit

So wird das Nachsprechen von Pseudowörtern als diagnostischer Indikator diskutiert, der gerade auch bei Kindern mit nicht-deutscher Erstsprache Hinweise auf das Vorliegen einer therapiebedürftigen Sprachentwicklungsstörung geben kann (*Kohnert & Windsor* 2004).

Da es sich bei den Mottier-Silben um einfache Konsonant-Vokal-Verbindungen handelt, gehen *Kiese-Himmel & Risse* (2009) davon aus, dass das Nachsprechen der Zielitems einsprachig deutsch aufwachsende und mehrsprachig aufwachsende Kinder vor vergleichbar hohe Anforderungen stellt (*Kiese-Himmel & Risse* 2009, 945). Die Ergebnisse der Schweizer Normierungsstudien der letzten Jahre bestätigen dies im Großen und Ganzen:

Für die Altersgruppen der 6-17-Jährigen konnten beim Nachsprechen der Pseudowörter keine Gruppenunterschiede zwischen mono- und bilingualen Kindern festgestellt werden (*Gamper et al.* 2012, *Wild & Fleck* 2013). Nur in der Kohorte der fünfjährigen Kinder bei *Wild & Fleck* (2013) zeigte sich ein signifikanter Mittelwertunterschied zugunsten der mehrsprachigen Probanden.

Eine Unterschätzung der realen Leistungen mehrsprachiger Kinder aufgrund der Items im Mottier-Test kann somit weitgehend ausgeschlossen werden. Allerdings weisen die Autorinnen darauf hin, dass dies möglicherweise nur für Kinder mit einer dem Deutschen verwandten, ebenfalls indoeuropäischen Erstsprache gilt. Da der Anteil

an Kindern mit nicht-indoeuropäischer Erstsprache (z.B. Türkisch) in ihrer Normierungsstichprobe sehr gering gewesen sei, müsse der Nachweis der Vergleichbarkeit für diese Erstsprachen noch erbracht werden (*Wild & Fleck* 2013, 153f.; vgl. aber die Ergebnisse von *Kruse* 2010 für Dreijährige).

Pilotuntersuchung zum Vergleich unterschiedlicher Normierungen

Ausgangspunkt und Zielsetzung

Da die ursprünglichen Normwerte (*Bohny* 1981 in der Schweiz, *Welte* 1981 in Deutschland) mittlerweile als veraltet gelten, wurden in den letzten Jahren in Deutschland und in der Schweiz mehrere Normierungsstudien mit dem Ziel durchgeführt, aktuelle Vergleichswerte bereitzustellen (Tab. 1).

Wie aus Tabelle 1 ersichtlich ist, wurden dabei teils unterschiedliche Altersgruppen in den Blick genommen. Auch die Größe der Normierungsstichproben sowie die Anzahl der Kinder in den einzelnen Alterskohorten unterschieden sich deutlich. Während in einigen Normierungen nur sprachunauffällige Kinder untersucht wurden (z.B. *Kiese-*

■ Tab. 1: Übersicht über Charakteristika der Normierungsstudien zum Mottier-Test 1981-2013

Autor/Jahr	Altersbereich; Größe der Normierungsstichprobe	Ein- und mehrsprachige Kinder?	Sprachauffällige Kinder dabei?	Standardisierte Durchführung (Audio-CD)?
<i>Bohny</i> (1981)	5;0-16;11 Jahre; n = 415	keine Angabe	Ja, zusätzlich getrennte Erhebung für Kinder mit LRS-Verdacht	Nein
<i>Welte</i> (1981)	3;11-6;11 Jahre; n = 122 sprachunauffällige Kinder, n = 107 sprachauffällige Kinder	keine Angabe	Ja, getrennte Werte für beide Gruppen; zudem befanden sich auch in der sprachunauffälligen Gruppe einige Kinder mit sprachlichen Auffälligkeiten	Nein
<i>Wagner</i> (1994)	6-15 Jahre; n = 149	keine Angabe	Ja, klinische Stichprobe	keine Angabe
<i>Seibert et al.</i> (2001)	7-9 Jahre, 9-11 Jahre; n = 272	keine Angabe	keine Angabe	Ja
<i>Dockter et al.</i> (2005)	6;1-8;3 Jahre; n = 67	keine Angabe	keine Angabe	keine Angabe
<i>Renner et al.</i> (2008)	5-10 Jahre; n = 312	Ja, keine getrennten Werte	Ja, klinische Stichprobe	Nein
<i>Kiese-Himmel & Risse</i> (2009)	4-6 Jahre; n = 308	Ja, keine getrennten Werte, kein Gruppenvergleich	Nein: Logopädische Therapie Ausschlusskriterium	Ja, aber Darbietung hier 1 Silbe pro 0,5 Sek.
„Berner Normen“ <i>Gamper et al.</i> (2012)	4-12 Jahre; n = 1 549 (1 145 monolinguale, 418 bilinguale)	Ja, da keine signifikanten Unterschiede keine getrennten Normwerte	Ausschluss von Kindern mit aktueller logopädischer Therapie	Ja, ggf. qualitativ schlechte Aufnahme? (<i>Taubert</i> 2014)
„St. Galler-Normen“ <i>Wild & Fleck</i> (2013)	5-17 Jahre; n = 1 242 (797 monolinguale, 445 bilinguale)	Ja, getrennte Normwerte bei den fünfjährigen Kindern, für alle anderen Altersgruppen gemeinsame Werte	Ja, ehemalige oder aktuelle logopädische Therapie kein Ausschlusskriterium	Keine CD, standardisierte Instruktion der Testleiter, Sprechrates 1 Silbe/Sek. (<i>Wild & Fleck</i> 2013, 154)

Himmel & Risse 2009, Gamper et al. 2012), inkludierten andere Studien auch Kinder mit Sprach- und/oder anderen Entwicklungsauffälligkeiten (z.B. Wild & Fleck 2013).

In den meisten Untersuchungen wurden den Kindern die Mottier-Silben von der UntersucherIn vorgesprochen. In drei Studien wurden die Testitems per Audiodatei präsentiert. Dabei wich die Sprechrate bei Kiese-Himmel & Risse (2009) mit 2 Silben/Sekunde von der sonst üblichen Sprechrate (1 Silbe/Sekunde) ab.

Die Aufnahme, die in der Normierung von Gamper et al. (2012) verwendet wurde, wird von Taubert (2014) als qualitativ unzureichend sowie bei der Darbietung der Pseudowörter mehrfach vom Original abweichend beschrieben. Da die Audio-Datei mittlerweile nicht mehr öffentlich zugänglich ist, kann dies jedoch nicht verifiziert werden.

In den meisten Normierungsstudien wurden sowohl mono- als auch bilinguale Kinder berücksichtigt. Getrennte Werte für diese beiden Gruppen wurden jedoch nur in den beiden aktuellen Schweizer Normierungen berechnet (Gamper et al. 2012, Wild & Fleck 2013).

Angesichts der Unterschiede in der methodischen Konzeption der Normierungsstudien überrascht eine gewisse Abweichung in den Ergebnissen der Normierung nicht. Wie deutlich die verschiedenen Normierungen voneinander abweichen und ob eine Beurteilung von Diagnostikergebnissen tatsächlich zu unterschiedlichen Einschätzungen hinsichtlich auffälliger/unauffälliger Leistungen führen kann, war die Fragestellung einer Pilotuntersuchung, die an der Universität zu Köln durchgeführt wurde (Hagel 2015).

Methode

Verglichen wurden drei verschiedene Normierungen: die deutsche Normierung von Kiese-Himmel & Risse (2009) sowie die beiden Schweizer Normierungen aus Bern bzw.

St. Gallen (Gamper et al. 2012, Wild & Fleck 2013).

Die Auswahl dieser drei Normierungen erfolgte aufgrund ihrer Aktualität sowie der Tatsache, dass Normwerte für die Zielgruppe dieser Untersuchung, fünfjährige Kinder, vorhanden waren. Die Pilotuntersuchung verfolgte das Ziel, die anhand der drei unterschiedlichen Normierungen erreichten Werte im Mottier-Test zu vergleichen sowie diese Werte mit den Testwerten in zwei standardisierten und normierten Verfahren, dem Untertest „Kunstwörter nachsprechen, KN“ aus dem SET 5-10 (Petermann 2010) und dem Untertest „Phonologisches Arbeitsgedächtnis für Nichtwörter, PGN“ aus dem SETK 3-5 (Grimm 2001) in Beziehung zu setzen.

Die Untersuchungsstichprobe umfasste 30 Kinder im Alter von 5;0-5;11 Jahren (19 Jungen, 11 Mädchen). Um eine möglichst große Bandbreite der Leistungen im Pseudowörter-Nachsprechen abbilden zu können, sollte die Stichprobe sowohl sprachunauffällige als auch spracherwerbsgestörte Kinder umfassen. Daher wurden die Kinder zur Hälfte aus Regelkindergärten und zur anderen Hälfte aus logopädischen Praxen rekrutiert. 23 Kinder sind monolingual deutschsprachig, 7 haben einen mehrsprachigen Hintergrund.

Die Testungen der Kinder fanden im Einzelsetting in ihren Kitas bzw. logopädischen Praxen statt. Um Reihenfolgeneffekte zu vermeiden, wurde die Reihenfolge der Durchführung der verschiedenen Diagnostikverfahren systematisch variiert.

Die Testitems des Mottier-Tests wurden entsprechend der ursprünglichen Durchführungsanleitung im Sekundentakt mit verdecktem Mundbild und gleichbleibender Betonung von der Testleiterin vorgesprochen (Linder & Grisseman 1996). Die Testitems der beiden standardisierten Testverfahren wurden von der Audio-Kassette (SETK 3-5) bzw. Audio-CD (SET 5-10) vorgespielt, um die standardisierte Durchführung zu gewährleisten. Die Antwort-

ten der Kinder wurden mit einem Diktiergerät aufgenommen und anschließend von der Testleiterin protokolliert.

Die Testergebnisse der Kinder wurden mit der Software IBM SPSS 22 ausgewertet und inferenzstatistisch analysiert. Da für die Normierung von Gamper et al. (2012) keine T-Werte vorliegen, erfolgte der Vergleich der Testergebnisse der Kinder vorrangig anhand der erreichten Prozentränge (bei Gamper et al. 2012 wurde jeweils der mittlere Wert des Prozentrangbands berücksichtigt); dementsprechend wurden statistische Methoden für ordinalskalierte Daten eingesetzt (Kendall-Tau-b, Friedman's ANOVA, anschließende paarweise Vergleiche mit Wilcoxon-Tests).

Ergebnisse

Es zeigten sich signifikante Korrelationen sowohl zwischen den Ergebnissen der verschiedenen Mottier-Normierungen als auch zwischen den Ergebnissen des Mottier-Tests und den beiden externen Verfahren (Tab. 2). Die Korrelationskoeffizienten entsprechen großen Effektstärken nach Cohen (1988). Insgesamt korrelieren die Werte, die in den drei Normierungen des Mottier-Tests gewonnen wurden, gut mit den Ergebnissen der beiden externen Verfahren, sodass die externe Validität insgesamt als gegeben angesehen werden kann.

In einem nächsten Schritt wurden die mittleren erreichten Testwerte der Kinder anhand der verschiedenen Normierungen des Mottier-Tests miteinander verglichen (Tab. 3).

Tabelle 3 macht deutlich, dass der mittlere Prozentrangwert bei Zugrundelegung der beiden Schweizer Normierungen (Gamper et al. 2012, Wild & Fleck 2013) deutlich höher ausfällt als bei Berücksichtigung der Normierung von Kiese-Himmel & Risse (2009). Ein Vergleich mittels des Chi-Quadrat-Tests belegt, dass es sich um einen statistisch höchst signifikanten Unterschied handelt, $\chi^2(2) = 24.17, p < .001$.

■ Tab. 2: Ergebnisse der Korrelationsanalysen

n = 30	Normierungen Mottier-Test			Untertests aus standardisierten Testbatterien	
	Kiese-Himmel & Risse (2009)	Blöchliger et al. (2011)	Gamper et al. (2012)	PGN (SETK 3-5, Grimm 2001)	KN (SET 5-10, Petermann 2010)
Kiese-Himmel & Risse (2009)		.86**	.83**	.65**	.68**
Wild & Fleck (2013)	.86**		.94**	.63**	.70**
Gamper et al. (2012)	.83**	.94**		.63**	.71**
PGN (SETK 3-5, Grimm 2001)	.65**	.63**	.63**		.63**
KN (SET 5-10, Petermann 2010)	.68**	.70**	.71**	.63**	

Korrelationskoeffizienten nach Kendall-Taub-b, **): signifikant mit $p < .01$ (zweiseitig)

■ Tab. 3: Deskriptive Statistik – Prozentrangwerte und T-Werte der Untersuchungsstichprobe in den verschiedenen Verfahren

n = 30	Prozentrangwerte			T-Werte	
	Median	Mittelwert	Standardabweichung	Mittelwert	Standardabweichung
Normierungen Mottier-Test					
<i>Kiese-Himmel & Risse (2009)</i>	88.0	67.3	34.8	58.4	14.4
<i>Wild & Fleck (2013)</i>	94.0	81.3	27.3	67.6	15.4
<i>Gamper et al. (2012)</i> Mittlerer Prozentrang	94.0	85.8	19.3	---	---
Untertests aus standardisierten Testbatterien					
PGN (SETK 3-5, <i>Grimm 2001</i>)	53.9	52.6	28.5	51.1	9.6
KN (SET 5-10, <i>Petermann 2010</i>)	55.5	54.7	31.0	51.5	10.6

Anschließend paarweise Vergleiche mit dem Wilcoxon-Test zeigen einen statistisch signifikanten Unterschied zwischen den Werten der Normierung von *Kiese-Himmel & Risse (2009)* und denen der beiden Schweizer Normierungen:

- *Gamper et al. (2012) – Kiese-Himmel & Risse (2009)*: $z=3.5$, $p=.001$, $r=.45$, mittlerer bis großer Effekt,
- *Wild & Fleck (2013) – Kiese-Himmel & Risse (2009)*: $z=4.5$, $p<.001$, $r=.57$, großer Effekt.

Im Mittel erreichen die fünfjährigen Kinder der Untersuchungsstichprobe somit signifikant niedrigere Prozentrangwerte anhand der Normierung von *Kiese-Himmel & Risse (2009)* gegenüber den beiden Schweizer Normierungen.

Zwischen den Ergebnissen der beiden Schweizer Normierungen (*Wild & Fleck 2013* und *Gamper et al. 2012*) besteht hingegen kein signifikanter Gruppenunterschied ($z=.97$, $p=.999$).

Die Einzelfallbetrachtung der Kinder stützt das Bild, das oben durch den Gruppenvergleich gewonnen wurde. Legt man als einheitliches Kriterium für ein auffälliges Testergebnis einen T-Wert <40 bzw. einen $PR < 16$ zugrunde, so werden 3 der 30 untersuchten Kinder (10%) anhand der Normierung von *Kiese-Himmel & Risse (2009)* als unterdurchschnittlich eingeschätzt, in den anderen beiden Normierungen erreichen sie hingegen durchschnittliche Werte.

Folgt man dem Vorschlag von *Kiese-Himmel & Risse (2009)*, alle Leistungen unterhalb eines Prozentranges von 25 als auffällig einzuschätzen, betrifft die unterschiedliche Klassifikation sogar 4 der 30 Kinder (13,3% der Stichprobe).

Der Vergleich mit den beiden externen Verfahren zeigt, dass alle drei Normierungen des Mottier-Tests die Leistungen der Kinder

im Nachsprechen von Kunstwörtern signifikant besser bewerten als die Untertests aus den Testbatterien SET 5-10 und SETK 3-5 (vgl. Tab. 3).

Besonders groß ist dieser Unterschied zwischen den beiden Schweizer Normierungen und den beiden standardisierten Verfahren. Hier liegen jeweils höchst signifikante Gruppenunterschiede vor, die einem großen Effekt nach *Cohen (1988)* entsprechen:

- PGN – Mottier (*Gamper et al. 2012*): $z=-6.1$, $p<.001$, $r=-.79$, großer Effekt
- PGN – Mottier (*Wild & Fleck 2013*): $z=-6.37$, $p<.001$, $r=-.82$, großer Effekt
- KN – Mottier (*Gamper et al. 2012*): $z=-5.5$, $p<.001$, $r=-.72$, großer Effekt
- KN – Mottier (*Wild & Fleck 2013*): $z=-5.84$, $p<.001$, $r=-.75$, großer Effekt.

Wie aus Tabelle 3 ersichtlich ist, liegen die im Mittel erreichten Prozenträge auf der Grundlage der Normierung von *Kiese-Himmel & Risse (2009)* recht nah an den mittleren Prozentrangwerten der beiden standardisierten Verfahren.

Paarweise Vergleiche zeigen einen signifikanten Unterschied mittlerer Effektgröße zwischen der Normierung von *Kiese-Himmel & Risse (2009)* und dem Untertest PGN (SETK 3-5: $z=-2.94$, $p=.033$, $r=-.38$); der Unterschied zwischen *Kiese-Himmel & Risse* und den Werten im KN (SET 5-10) erreicht bei adjustiertem Signifikanzniveau zur Kontrolle der Alpha-Fehler-Kumulierung keine statistische Signifikanz, obwohl auch hier ein Effekt mittlerer Größe vorliegt ($z=-2.41$, $p=.16$, $r=-.31$).

Die beiden Untertests aus dem SET 5-10 und dem SETK 3-5 kommen zu sehr ähnlichen Ergebnissen, für die kein signifikanter Mittelwertunterschied gefunden wurde. Vergleichbare Ergebnisse zeigen sich beim Mittelwertvergleich anhand der T-Werte, die für vier der fünf Normierungen zur Verfügung stehen.

In der Einzelfallbetrachtung der Klassifikation der Kinder (auffällige Leistung: T-Wert <40 , $PR < 16$) werden die Nachsprecheleistungen von 5 der 30 Kinder (16,7%) im Mottier-Test unabhängig von der verwendeten Normierung als unauffällig eingeschätzt, in mindestens einem der beiden standardisierten Verfahren jedoch als auffällig.

Diskussion der Ergebnisse

Die beiden Schweizer Normierungen bewerten die Nachsprecheleistungen der Kinder signifikant besser als die Normierung von *Kiese-Himmel & Risse (2009)*. Dies kann zu einer unterschiedlichen Einschätzung hinsichtlich auffälliger bzw. unauffälliger Leistungen der Kinder führen; in der vorliegenden Stichprobe traf dies auf 10% der Fälle zu.

Im Vergleich mit den beiden Untertests aus dem SETK 3-5 und dem SET 5-10 beurteilt die Normierung von *Kiese-Himmel & Risse (2009)* die Nachsprecheleistungen für Kunstwörter am ähnlichsten. Dennoch wurden 16,7% der Kinder, die anhand der Normierungen des Mottier-Tests unauffällige Werte erreichten, von mindestens einem der beiden externen Verfahren als auffällig eingeschätzt.

Dies wirft die Frage auf, ob die verwendeten Diagnostikinstrumente vergleichbare Fähigkeiten der Kinder messen. Während die Mottier-Silben einfache Konsonant-Vokal-Verbindungen sind, enthalten die Testitems der Untertests PGN und KN eine Vielzahl von Konsonantenclustern; neben offenen werden auch geschlossene Silben verwendet.

Bezüglich der Binnenstruktur der Testitems stellen somit die Mottier-Silben geringere Anforderungen an die phonetischen und phonologischen Fähigkeiten der Kinder. Möglicherweise trägt dies dazu bei, dass die Kinder im Mottier-Test durchgängig bessere Werte erreichen als in den beiden anderen Verfahren.

Für Kinder mit Aussprachestörungen im phonetischen und/oder phonologischen Bereich könnte es somit sinnvoller sein, den Mottier-Test einzusetzen, da dieser die geringsten Anforderungen an die phonologische Analyse und die sprechmotorische Umsetzung der Items stellt und somit möglicherweise ein „reineres Maß“ für die Kapazität der phonologischen Schleife liefern kann.

Andererseits entsprechen Pseudowörter mit Mehrfachkonsonanz eher der typischen Struktur vieler deutscher Wörter als einfache KV-Verbindungen, weshalb das Nachsprechen „anspruchsvollerer“ Pseudowörter möglicherweise die für den Spracherwerb zentrale Fähigkeit, neue Wörter analysieren und einspeichern zu können, besser abbilden kann.

Das entscheidende Argument gegen den Einsatz des Mottier-Tests ist seine immer noch unzureichende Standardisierung, insbesondere in Bezug auf die Durchführungsobjektivität. Wie bereits beschrieben, wurden in den Normierungsstudien die Testitems teilweise mündlich durch die Testleiter vorgegeben (Wild & Fleck 2013), teilweise von einer Audioaufnahme, deren Qualität offenbar nicht in allen Fällen optimal war (Gamper et al. 2012, Taubert 2014, Kiese-Himmel & Risse 2009, Kiese-Himmel 2014).

Um die Leistungen von Kindern im Mottier-Test mit Normwerten beurteilen zu können, die anhand einer standardisierten Audioaufnahme gewonnen wurden, muss diese dem Diagnostiker jedoch vorliegen, damit er sie auch für seine Testung einsetzen kann. Nur so können die Durchführungs- und Auswertungsobjektivität des Messinstruments gewährleistet werden. Es wäre daher wünschenswert, wenn die entsprechende Audio-datei, die der Normierung zugrunde liegt, den Anwendern gemeinsam mit den Normwerten zur Verfügung gestellt würde.

Die mündliche Vorgabe der Testitems durch den Testleiter erscheint zum einen hinsichtlich der Durchführungsobjektivität unzureichend; zum anderen sind auch die Normwerte nur dann sinnvoll anwendbar, wenn die Rohwerte des Kindes genau durch die Präsentation der Items gewonnen wurden wie in der Normierungsstudie.

Die Sprechrate, mit der die Silben präsentiert werden, stellt vermutlich einen weiteren Einflussfaktor auf die Ergebnisse dar. Werden die Silben des Mottier-Tests wie ursprünglich empfohlen mit einer Sprechrate von einer Silbe pro Sekunde vorgesprochen (Linder & Grisse-mann 1996), überschreitet das zu ver-

arbeitende Material die zeitliche Kapazität des phonologischen Buffers von maximal 1,8 Sekunden. Somit wird nicht nur der Speicher, sondern darüber hinaus auch die Funktionsfähigkeit des artikulatorischen Rehearsals überprüft.

Eine erhöhte Sprechrate bei der Präsentation der Pseudowörter oder Silben verändert hingegen die Anforderungen an die Verarbeitung: So könnten bei einer Präsentationsrate von einer Silbe pro 0,5 Sekunden (vgl. Risse & Kiese-Himmel 2009) Pseudowörter bis zu einer Länge von drei Silben (1,5 Sekunden) allein über das Nutzen des phonetischen Speichers wiedergegeben werden, ohne dass ein Rehearsal-Prozess aktiviert werden muss. Die Frage nach der Präsentationsgeschwindigkeit der Items ist somit entscheidend, wenn vergleichbare theoretische Konstrukte gemessen werden sollen.

Fazit und Ausblick

Der Mottier-Test stellt ein zeitökonomisches und im deutschsprachigen Raum weit verbreitetes Diagnostikverfahren dar. In seiner jetzigen Form erfüllt er jedoch die Testgütekriterien nur in unzureichendem Maße. Insbesondere die Durchführungsobjektivität ist als mangelhaft zu bewerten. Wünschenswert wäre es, dieses Diagnostikverfahren einer umfassenden Standardisierung zu unterziehen, die neben einer Audiodatei zur einheitlichen Präsentation der Items auch Informationen über Testgütekriterien und Itemkennwerte zur Verfügung stellt, wie dies teilweise bereits geschehen ist (z.B. Risse & Kiese-Himmel 2009, Renner et al. 2008, Kruse 2010).

Solange ein solches Verfahren nicht zur Verfügung steht, sollte der Mottier-Test als orientierendes Screening-Verfahren betrachtet werden. Für fünfjährige Kinder konnte in einer Pilotuntersuchung anhand der Normierung von Kiese-Himmel & Risse (2009) die größte Übereinstimmung mit den Ergebnissen anderer Verfahren zum Pseudowörter-Nachsprechen erzielt werden. Im Hinblick auf die Durchführungsobjektivität erscheint es jedoch sinnvoller, das phonologische Kurzzeitgedächtnis mithilfe eines Untertests aus standardisierten Testbatterien unter Verwendung der Original-Audiodatei zu erfassen.

Für aussprachegestörte Kinder empfiehlt sich ggf. der Rückgriff auf andere Aufgabenformate wie bspw. das Nachsprechen von Zahlenfolgen. Dabei sollte dann allerdings der mögliche Einfluss des lexikalischen Vorwissens der Kinder auf die Ergebnisse berücksichtigt werden.

LITERATUR

- Archibald, L.M.D. & Gathercole, S.E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders* 41 (6), 675-693
- Archibald, L.M.D. & Joanisse, M.F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language and Hearing Research* 52, 899-914
- Baddeley, A. (2003). Working memory and language: an overview. *Journal of Communication Disorders* 36 (3), 189-208
- Baddeley, A., Gathercole, S. & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review* 105 (1), 158-173

- Bohny, A. (1981). Verbal auditive Dysgnosie. *Der Sprachheilpädagogie* 13, 50-59
- Boudreau, D. & Costanza-Smith, A. (2011). Assessment and treatment of working memory deficits in school-age children: the role of the speech-language pathologist. *Language, Speech, and Hearing Services in Schools* 42 (2), 152-166
- Carretti, B., Borella, E., Cornoldi, C. & de Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: a meta-analysis. *Learning and Individual Differences* 19 (2), 246-251
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale: Erlbaum.
- Dockter, S., Feldhusen, F., Brunner, M. & Pröschel, U. (2005). *Auditive Wahrnehmung: Normwerte für Klassenstufe 1*. 22. Wissenschaftliche Jahrestagung der Deutschen Gesellschaft für Phoniatrie und Pädaudiologie e.V., 16.-18.9.2005 in Berlin. www.egms.de/static/de/meetings/dgpp2005/05dgpp017.shtml (13.11.2015)
- Ellis Weismer, S., Tomblin, J.B., Zhang, X., Buckwalter, P., Chynoweth, J.G. & Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech, Language, and Hearing Research* 43 (4), 865-868
- Gamper, H., Keller, U., Messerli, N., Moser, M. & Wüst, J. (2012). *Normen für den Mottier-Test bei 4- bis 12-jährigen Kindern*. Praxis Forschung der Erziehungsberatung des Kantons Bern. www.erz.be.ch/erz/de/index/erziehungsberatung/erziehungsberatung/praxisforschung/projekte.assetref/content/dam/documents/ERZ/AKVB/de/Erziehungsberatung/Praxisforschung/Schriften/EB_PF_Band%2012%20Normen%20Mottier.pdf (13.11.2015)
- Gathercole, S.E., Hitch, G.J., Service, E. & Martin, A.J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology* 33 (6), 966-979
- Graf Estes, K., Evans, J.L. & Else-Quest, N.M. (2007). Differences in the nonword repetition performance of children with and without specific language impairment: a meta-analysis. *Journal of Speech, Language, and Hearing Research* 50 (1), 177-195
- Grimm, H. (2001). *SETK 3-5. Sprachentwicklungstest für drei- bis fünfjährige Kinder. Diagnose von Sprachverarbeitungsfähigkeiten und auditiver Gedächtnisleistungen*. Göttingen: Hogrefe
- Hagel, S. (2015). *Der Mottier-Test – Vergleich unterschiedlicher Normierungen*. Schriftliche Hausarbeit im Rahmen der Ersten Staatsprüfung im Lehramt Sonderpädagogik. Universität zu Köln, Köln.
- Hasselhorn, M. & Grube, D. (2003). Das Arbeitsgedächtnis: Funktionsweise, Entwicklung und Bedeutung für kognitive Leistungsstörungen. *Sprache Stimme Gehör* 27 (1), 31-37
- Kiese-Himmel, C. (2014). Mottier-Test: Kontroverse um die Normierung. *Sprache Stimme Gehör* 38 (2), 57
- Kiese-Himmel, C. & Risse, T. (2009). Normen für den Mottier-Test bei 4- bis 6-jährigen Kindern. *HNO* 57 (9), 943-948
- Kohnert, K. & Windsor, J. (2004). The search for common ground: part II. Nonlinguistic performance by linguistically diverse learners. *Journal of Speech, Language, and Hearing Research* 47, 891-903
- Kruse, E.S.U. (2010). *Die phonologische und auditive Verarbeitungskapazität bei dreijährigen Kindern mit ein- und zweisprachigem Lebenskontext. Eine Analyse des Mottier-Tests im Vergleich zum PGN*. Inauguraldissertation an der LMU München. https://edoc.ub.uni-muenchen.de/11949/1/Krauser_Elke.pdf (13.11.2015)
- Linder, M. & Grisseemann, H. (1996). *ZLT – Zürcher Lesetest. Förderdiagnostik bei gestörtem Schriftspracherwerb*. Bern: Hans Huber
- Montgomery, J.W. & Evans, J.L. (2009). Complex sentence comprehension and working memory in children with specific language impairment. *Journal of Speech, Language, and Hearing Research* 52 (2), 269-288
- Mottier, G. (1951). Mottier-Test. Über Untersuchungen der Sprache lesegestörter Kinder. *Folia Phoniatria* 3 (3), 170-177
- Petermann, F. (2010). *SET 5-10. Sprachstandserhebung für Kinder im Alter zwischen 5 und 10 Jahren*. Göttingen: Hogrefe

- Petermann, F. & Daseking, M. (2012). *ZLT-II: Zürcher Lesetest-II. Weiterentwicklung des Zürcher Lesetests (ZLT) von Maria Linder und Hans Grissemann*. Göttingen: Hans Huber
- Renner, G., Rothermel, C. & Krampen, G. (2008): Befunde zur Reliabilität und Validität des Mottier-Tests in einer klinisch-sozialpädiatrischen Stichprobe. *Sprache Stimme Gehör* 32 (1), 30-35
- Risse, T. & Kiese-Himmel, C. (2009). *Der Mottier-Test. Teststatistische Überprüfung an 4-bis 6-jährigen Kindern*. *HNO* 57 (5), 523-528
- Seibert, A., Dierks, A., Strehlow, U., Haffner, J., Parzer, P. & Resch, F. (2001). Der Mottier-Test als computergestütztes Screeningsverfahren bei der Legastheniediagnostik. *Zeitschrift für Differentielle und Diagnostische Psychologie* 22 (2), 118-126
- Taubert, G. (2014). *Gravierende methodische Mängel entwerfen die Ergebnisse der Berner Neunormierung des Mottier-Tests von 2012 durch Gamper et al.* www.praxis-taubert.de/images/Mottier-Test-Studie%20Bern%202012%20Kritik%20Taubert%2002.05.2014.pdf (13.11.2015)
- Wagner, H. (1994). Auditive Wahrnehmungsprobleme und verbale und nonverbale Intelligenzleistungen. *Praxis der Kinderpsychologie und Kinderpsychiatrie* 43 (3), 106-109
- Welte, V. (1981). Der Mottier-Test, ein Prüfmittel für die Lautdifferenzierungsfähigkeit und die auditive Merkfähigkeit. *Sprache Stimme Gehör* 5, 121-125
- Wild, N. & Fleck, C. (2013). Neunormierung des Mottier-Tests für 5- bis 17-jährige Kinder mit Deutsch als Erst- oder als Zweitsprache. *Praxis Sprache* 58 (3), 152-158

SUMMARY. Language assessment using the Mottier-Test: different standardizations and some open questions

The capacity of the verbal short term memory is discussed as an important factor to cause or to maintain language deficits in children. It is usually measured using nonword repetition tasks. Using one of the most used assessment tools in German-speaking countries, the Mottier-Test, practitioners are faced with the challenge to reasonably decide for one of the standardizations available. This paper presents the results of a pilot study which assessed nonword repetition abilities of 30 five-year-old children. It aimed at a comparison of the scores gained by using three recent standardizations of the Mottier-Test. In addition, the Mottier-scores were compared to the results of nonword repetition in two standardized subtests from test batteries which served as external criterion. The results show that the interpretation of nonword repetition skills was highly influenced by the choice of the standardization: for 10% of the children observed, the differences in the standardization led to divergent decisions regarding therapy indication. Furthermore, the subtests of the test batteries in some cases led to a different evaluation of verbal short term memory capacity compared to the Mottier-Test. Implications of these findings for language assessments are being discussed.

KEY WORDS: Diagnosis – verbal short-term memory – developmental language disorders

DOI dieses Beitrags (www.doi.org)

10.2443/skv-s-2016-53020160203

Autorin

Dr. Tanja Ulrich
 Lehrstuhl für Sprachbehindertenpädagogik
 in schulischen und außerschulischen Bereichen
 Department Heilpädagogik und Rehabilitation
 Universität zu Köln
 Klosterstr. 79b
 50931 Köln
tanja.ulrich@uni-koeln.de