

## Testrezension

<b>Titel</b>	Aktiver Wortschatztest für 3- bis 5-jährige Kinder (AWST-R)
<b>Autoren</b>	Christiane Kiese-Himmel
<b>Quelle</b>	Kiese-Himmel, Christiane: Aktiver Wortschatztest für 3- bis 5-jährige Kinder (AWST-R). Erste Auflage. Hogrefe Verlagsgruppe (2005): Göttingen. <a href="http://www.hogrefe.de">www.hogrefe.de</a>
<b>Rezensent</b>	Meike Karrasch, Johanna Keim
<b>Datum</b>	16.01.2020

## Angaben aus dem Handbuch

<b>Zielsetzung (gem. Ziele AWST)</b>	<p>(1) Selektion für sprachliche Frühförderung</p> <ul style="list-style-type: none"> <li>- Lexikalische Sprachförderung</li> <li>- Individuelles Testergebnis Ausgangspunkt für pädagogische Sprachförderung</li> <li>- Fürsorgliche Pflicht bzgl. Sprachaufbaumaßnahmen bei Migrantenkinder ohne Kenntnisse des Deutschen muss beachtet werden</li> </ul> <p>(2) Diagnose sowie Indikation für Therapiemaßnahmen</p> <ul style="list-style-type: none"> <li>- Erlaubt Einschätzung der Wortschatzhöhe im Vergleich zu gleichaltrigen</li> <li>- Fester Platz in der Sprachentwicklungsdiagnostik</li> <li>- Beurteilung wichtig, da hinreichend große Lexikonkapazität Grundlage für Entwicklung der Grammatik bildet</li> <li>- Wortschatz = verbale Intelligenz</li> <li>- Bildbenennung komplexe Leistung</li> <li>- Sprachentwicklungsgestörte Kinder im Benennen langsamer</li> </ul> <p>(3) Evaluation für sprachtherapeutische Interventionen</p> <ul style="list-style-type: none"> <li>- Dient Messung intraindividuelle Entwicklungsveränderungen</li> <li>- Messung der Wirksamkeit von Maßnahmen d.h. Evaluation von Sprachtherapie/-förderung durch Messung von Leistungsfortschritten</li> </ul>
<b>Theoretischer Hintergrund</b>	<p>Kapitel über</p> <ul style="list-style-type: none"> <li>- Wortkategorien, ihre Bedeutung und Differenzierung</li> <li>- Skizze der frühen aktiven Wortschatzentwicklung</li> <li>- Quantitative Entwicklung des aktiven Wortschatzes</li> <li>- Skizze der frühen Wortbedeutungsentwicklung</li> </ul>
<b>Operationalisierung / Aufgabenstellung</b>	<p>Benennen von Bildkarten mit der Fragestellung:</p> <ul style="list-style-type: none"> <li>- Bei Objekten: Was ist das?</li> <li>- Bei Tätigkeiten: Was macht die/der?</li> </ul>
<b>Material</b>	<p>Test komplett bestehend aus:</p> <ul style="list-style-type: none"> <li>- Manual</li> <li>- Bildvorlagenmappe 1 und 2</li> <li>- 10 Testprotokollen</li> <li>- 10 Auswertungsprotokollen</li> <li>- Koffer für Aufbewahrung und Transport</li> </ul>
<b>Itemauswahl</b>	<p><i>Itemgenerierung Erstentwurf</i></p> <ul style="list-style-type: none"> <li>- Aufgabenwahl soll repräsentative Teilmenge aus lebensnahem Wortschatz der 3- bis 5-jährigen Kinder darstellen</li> <li>- Übernahme von 47 Items aus dem AWST 3-6</li> </ul>

- Restliche Items aus PLAKKS (Fox, 2002), vergleichenden Wortlisten der 50 häufigsten Inhaltswörter für jedes Kind (in Wagner, 1987), K-ABC-Subtest Wortschatz (Melchers & Preuß, 1994), FTF-W für 5-Jährige (Raatz & Möhling, 1971), Sprachkorpus Rosenkranz (1957), Sprachkorpus für 6-Jährige (in Wagner, 1987), Sandhoffer et. al (2000), Yoshida & Smith (2001), Wortliste „Wortproduktion-Verben (in Kauschke & Siegmüller, 2002, Patholinguist. Diagnostik, Bd. 2)
- Ergänzung um zeitgemäße und weitere, als geeignet eingeschätzte Begriffe
- Kategorisierung der Items in semantische Felder gemäß Gipper (1985)

#### *Pilotierungsphase*

##### *Ergebnisse*

- 66 monolingual deutschsprachige Kinder (33 J. und 33 M.) in Eisenach und Göttingen mit Testvorform untersucht
  - Negativer Schiefekennwert von -0,4 sprach für nichtsymmetrische Verteilung → Testvorform zu leicht
  - Exzess-Kennwert von -1,0 nicht mit statistischer Normalität vereinbar
  - Unpassende Schwierigkeitsgraduierung schien vorzuliegen
  - Konsistenzschätzung belegte hohe interne Konsistenz (.93)
  - Gesamtpunktwert und Lebensalter korrelierten mit  $r=0,59$  ( $p<0,0001$ )
- Erneute Anpassung der Testform

#### *Empirische Erprobung der vorläufigen Testform*

##### *Stichprobe*

- 259 monolingual deutschsprachige Kinder aus Regelkindergärten und Vorschulen (6-jährige Kinder) in Göttingen, Salzgitter (Niedersachsen 86%), Rheine (NRW 14%), Bremen (Bremen 0,4%)
- Auf jeder Altersstufe Mädchen und Jungen gleichhäufig vertreten (außer 6;0-6;6 Jahre → 21 J. vs. 18 M.)

##### *Itemselektion/-revision*

- Erneutes Aussortieren zu leichter Items bzw. Items mit zu niedriger/negativer Trennschärfe (eher Substantive als Verben)
- Itemklassen mit identischer Lösungsrate wurden reduziert

##### *Konsequenzen*

- Verbleib von 71 Items
- Min. Gesamtpunktwert 4 (vorher 7), max. 67 (vorher 77)

	<ul style="list-style-type: none"> <li>- Für ältere Kinder Testform dennoch zu leicht, daher Anwendungsbereich auf bis zu 5;5-Jährige begrenzt</li> <li>- Mittlerer Gesamtpunktwert 40,0 (SD 13,9)</li> <li>- Korrelation Gesamtpunktwert mit Lebensalter <math>r = 0,72</math> (<math>p &lt; 0,0001</math>) → Entwicklungssensitiv</li> <li>- Schwierigkeitsindex streuts zwischen <math>P = 0,08</math> bis <math>P = 0,95</math></li> <li>- Mittlerer Trennschärfekoeffizient 0,44 (min 0,25, max. 0,66)</li> <li>- Konsistenzschätzung (Kuder-Richardson): .96</li> <li>- Vier neue Verb-Items ergänzt</li> </ul>
<b>Durchführungsdauer</b>	15-20 Minuten
<b>Auswertung</b>	<p>a) Rohwertbestimmung: Welche Werte gehen in die Auswertung ein?</p> <ul style="list-style-type: none"> <li>- Dichotome Aufgabenbewertung: Item korrekt ja/nein; mit alternativen Antwortmöglichkeiten, wie <i>Kamera</i> statt <i>Fotoapparat</i></li> <li>- Ein Punkt für ein korrektes Item</li> <li>- Kein Punkt für falsches Item</li> </ul> <p>Korrekte und inkorrekte Items im Handbuch definiert:</p> <ul style="list-style-type: none"> <li>- Inkorrekte Nomen</li> <li>- Inkorrekte Verben</li> </ul> <p>Welche Berechnungen sind notwendig?</p> <ul style="list-style-type: none"> <li>- Summe richtiger Verben</li> <li>- Summe richtiger Substantive</li> <li>- Bildung des Gesamtpunktwertes</li> </ul> <p>Alles pen &amp; paper</p> <p>b) Vergleich mit Normgruppe:</p> <ul style="list-style-type: none"> <li>- Rohwerte lassen sich mit Hilfe von Tabelle in Prozentränge transformieren, diese wiederum in T-Werte</li> </ul>
<b>Normierung</b>	<ul style="list-style-type: none"> <li>- Gesamtpunktwert wird in Normwert übertragen</li> <li>- Normierungsstichprobe stützt sich auf Untersuchung von 551 „sinnesgesunden und vermeintlich normal begabten Kindern im Alter von 3;0-5;5“</li> <li>- Aus 10 Bundesländern in Einzelsitzungen</li> <li>- Mittleres Alter lag bei 51,9 Monaten (SD 8,1)</li> <li>- Zusammensetzung der Normstichprobe in Tabelle geschildert (S. 63) <ul style="list-style-type: none"> <li>- Ges. 551 (m:284, w:267)</li> <li>- 3;0-3;5: 69 Kinder (m:37, w:32)</li> <li>- 3;6-3;11: 122 Kinder (m:63, w:59)</li> <li>- 4;0-4;5: 113 Kinder (m: 58, w: 55)</li> <li>- 4;6-4;11: 124 Kinder (m:65, w: 59)</li> <li>- 5;0-5;5: 123 Kinder (m:61, w: 62)</li> </ul> </li> <li>- Kein geschlechtsspezifischer Unterschied feststellbar</li> <li>- Differenzierung zwischen Kindern aus der Stadt und aus ländlicher Umgebung, allerdings Unterschied als nicht signifikant erkannt</li> </ul>

<b>Objektivität</b>	<p><i>Durchführung</i></p> <ul style="list-style-type: none"> <li>- Durch Standardisierung des Bildsatzes, des Protokollbogens, vorgeschriebene Testinstruktion gewährleistet (solange sich Untersucher daran hält)</li> </ul> <p><i>Auswertung</i></p> <ul style="list-style-type: none"> <li>- Konkrete Richtlinien: Vorgeschriebene mögliche Items, alternative richtige und falsche Antworten angegeben (Beispielliste)</li> </ul> <p><i>Interpretation</i></p> <ul style="list-style-type: none"> <li>- Aufgrund der Normierung lässt sich die individuelle Testleistung vor dem Hintergrund des Konfidenzintervalls interpretieren</li> </ul>
<b>Reliabilität</b>	<p><b>Halbierungs-Reliabilität (Split-Half-Methode)</b></p> <ul style="list-style-type: none"> <li>- Zwischen beiden Testhälften lag ein numerischer Zusammenhang von <math>r_{12} = .72</math></li> <li>- Korrektur mittels Spearman-Brown-Formel belegte Halbierungs-Reliabilität von <math>r_{12} = .83</math></li> <li>- Split-Half wurde jedoch für den Test als unangemessen eingeschätzt, AWST-R Niveautest (weshalb i.d. zweiten Hälfte schlechtere Benennungen)</li> </ul> <p><b>Stattdessen Odd-Even-Methode (Gradzahlige vs. Ungradzahlige)</b></p> <ul style="list-style-type: none"> <li>- Korrelation von <math>r_{12} = .76</math></li> <li>- Unterschätzt aber auch die wahre Reliabilität, da nur die Hälfte der Items beurteilt wird</li> <li>- Daher Verwendung der Spearman-Brown-Formel Reliabilitätskoeffizient von .86</li> </ul> <p><b>Schätzung der internen Konsistenz</b></p> <ul style="list-style-type: none"> <li>- Verallgemeinerung der Testhalbierungsmethode</li> <li>- Bei 0-1-Aufgaben-Codierung wird die Konsistenzschätzung mit Kuder-Richardson-Formel vorgenommen (im Gegensatz zum Crombachs <math>\alpha</math> wird hier Schwierigkeit und Trennschärfe der Items berücksichtigt)</li> <li>- Damit globaler Reliabilitätswert von .88</li> <li>- Angaben zu interner Konsistenz: 3-Jährige .82; 4-Jährige .86; 5-Jährige .80</li> </ul> <p><b>Retest-Reliabilität</b></p> <ul style="list-style-type: none"> <li>- Bei Entwicklungstests nur bedingt sinnvoll, da zwischen Testzeitpunkten Entwicklung voranschreitet (insbesondere Wortschatz ist sehr variabel)</li> <li>- Trotzdem Testwiederholung mit kleiner Stichprobe von 19 monolingual-sprachigen Kindern (7 Jungen, 12 Mädchen; <math>n = 3</math> Dreijährige, <math>n = 10</math> Vierjährige, <math>n = 6</math> Fünfjährige) am selben Ort durch dieselbe Untersucherin</li> <li>- Kurzer Zeitintervall (10-14 Tage), weil bei Kindern ab 3 mit hoher Wahrscheinlichkeit entwicklungsbedingte Verbesserungen</li> <li>- 1. Testung im Mittel 45,8 Bilder (SD 10,9) richtig</li> <li>- 2. Testung im Mittel 48,5 Bilder (SD 12,6) richtig</li> <li>- Unterschied in beiden Testzeiträumen lag im Zufallsbereich</li> <li>- Verbesserungen können auf Erinnerungseffekt beruhen, Vertrautheit mit Messsituation,...</li> <li>- Stabilitätskoeffizient <math>r_{tt} = .87</math></li> </ul>

<p><b>Validität</b></p>	<p><b>Inhaltsvalidität</b></p> <ul style="list-style-type: none"> <li>- Im Sinne inhaltlicher Repräsentativität laut Autoren durch Itemauswahl, Pilotierungsphase, empirische Erprobung der vorläufigen Testform, empirisch begründete Darstellung der Zusammenstellung der Items in der Testendform gegeben</li> </ul> <p><b>Kriteriumsbezogene Validität</b></p> <ul style="list-style-type: none"> <li>- An 113 monolingual deutschsprachig aufgewachsenen im Bundesland Niedersachsen (Städtisch: Göttingen [n=68], ländlich: Friedland [n=30], Wulften [n=15]) gemessen</li> <li>- Äußere Bedingungen sehr ähnlich (im gesonderten Raum in Einzelsituation von Psychologin)</li> <li>- Teilstichproben in Tabelle</li> </ul> <p><i>Schätzurteil</i></p> <ul style="list-style-type: none"> <li>- Vertraute Person schätzt Wortschatzumfang des Kindes ein (6-stufige bipolare Likert-Skala)</li> <li>- Hatte hohen Bezug</li> <li>- Statistisch signifikanter Bezug zum Erzieherinnenurteil für 99 Kinder</li> <li>- Bei Mädchen größere Kovariation des Schätzurteils</li> <li>- Bei 3-jährigen weder für Jungen noch Mädchen statistisch signifikante Zusammenhänge</li> <li>- Erst in Gesamtgruppe wird Korrelationskoeffizient von 0,46 statistisch bedeutsam</li> <li>- Genauso bei 5-jährigen Kindern</li> <li>- Bei 4-Jährigen durchgängig signifikanten Zusammenhang sowie statistische Validität am höchsten 0.59</li> <li>- Bei Jungen im mittleren, bei Mädchen im hohen Ausmaß</li> </ul> <p><i>Korrelation mit Testkriterien</i></p> <ul style="list-style-type: none"> <li>- Problematik: In Deutschland kein produktiver Wortschatztest für 3- bis 5-Jährige vorhanden → entsprechender Benenntest fehlt als Validitätskriterium</li> <li>- Validität mithilfe von Vergleichsmessung mit zwei anderen Subtests ermittelt (lautsprachliches Vokabular ist Voraussetzung für erfolgreiche Bearbeitung dieser Tests): „Sätze ergänzen“ und „Wörter ergänzen“ aus PET von Angermaier (1977)</li> <li>- Keine Angaben zur Validität der Subtests bekannt</li> <li>- Zusammenhang zwischen Gesamtpunktwert von AWST-R und PET-Subtests mit Person'schem Produkt-Moment-Korrelationskoeffizienten auf Rohwertbasis berechnet</li> <li>- Zusätzlich Übereinstimmung mit Summenwert der Substantive u. Verben ermittelt</li> </ul> <p><i>Ergebnisse Kriterium „Sätze Ergänzen“:</i></p> <ul style="list-style-type: none"> <li>- Validitätskoeffizienten: alle Kinder: <math>r_{tc} = 0,78</math>; 3-Jährige: <math>r_{tc} = 0,74</math>; 4-Jährige: <math>r_{tc} = 0,69</math>; 5-Jährige <math>r_{tc} = 0,66</math></li> <li>- Validitätskoeffizient für Verben: <math>r_{tc} = 0,64</math>; <math>p = 0,0001</math> (PET-SE enthält lediglich 4 Aufgaben (11%), die Verb verlangen)</li> </ul>
-------------------------	--

- Validitätskoeffizient für Substantive:  $r_{tc} = 0,76$ ;  $p = 0,000$

*Ergebnisse Kriterium „Wörter Ergänzen“*

- Validitätskoeffizienten:  
alle Kinder:  $r_{tc} = 0,58$ ; 3-Jährige  $r_{tc} = 0,24$ ;  
4-Jährige:  $r_{tc} = 0,46$ ; 5-Jährige  $r_{tc} = 0,40$
- Höhe des korrelativen Zusammenhangs von PET-WE und Benennleistung Substantive im AWST-R entspricht der des AWST-R insgesamt ( $r = 0,58$ ,  $p < 0,0001$ )
- Statistische Beziehung zwischen Verben in AWST-R und PET-WE niedriger → WE-Items nur Substantive
- Validitätskoeffizient wird trotzdem als akzeptabel angesehen, auch wenn es sich um konstruktferneres Kriterium handelt

*Multiple Validität*

- Multipler Validitätskoeffizient beträgt  $0,78$  → nach Fisseni (1997) als „hoch“ einzuordnen

**Konstruktvalidität**

*Altersabhängigkeit Gesamtpunktwert*

- AWST-R muss als Entwicklungstest Veränderungen im Wortschatzumfang über die Zeit abbilden
- Mittlerer Gesamtpunktwert steigt bei Jungen kontinuierlich mit Lebensalter an
- Mittlerer Gesamtpunktwert der Mädchen steigt ebenfalls kontinuierlich mit Lebensalter an, aber nicht von erstem zu zweitem Halbjahr bei 3-Jährigen (allerdings nur 2 Mädchen in dieser Altersstufe)
- Mittlere Wortschatzleistungen von Jungen und Mädchen unterscheiden sich auf keiner Halbjahresstufe signifikant
- Differenzen gemäß ANOVA und nachgeschaltetem Scheffé-Test:  $p < 0,01$ :
- 3-Jährige konnten im Durchschnitt signifikant weniger Wörter benennen als 4-Jährige, diese signifikant weniger als 5-Jährige, 3- und 5-Jährige unterschieden sich ebenfalls statistisch signifikant
- Leistungsdifferenzen von Jungen u. Mädchen zufällig

*Gesamtpunktwert und Wohnumfeld (Stadt vs. Land)*

- Kinder der Validierungsstichprobe wuchsen in Dörfern, als auch in Städten auf (>10 000 Einwohner)
- Tendenziell mittlerer Gesamtpunktwert bezogen auf gesamte Stichprobe, als auch für jeweilige Altersgruppen bei Stadtkindern etwas höher  
→ statistisch jedoch unbedeutend  
→ kein Effekt vom lokalen Wohnumfeld auf Wortschatzumfang zu erwarten, da Kindern über das Fernsehen gleichermaßen überregionale Sprachmodelle zugänglich sind und in Dörfern lebende Bezugspersonen/Kommunikationspartner hinsichtlich ihres Bildungsstandes und sozioökonomischen Status ebenso heterogen sind wie Stadtbewohner

#### *Korrelation zwischen Gesamtpunktwert und lexikalisch-semantischen Kategorien*

- Wortfelder, die Konstrukt „aktiver Wortschatzumfang“ repräsentieren, wurden statistisch und semantisch analysiert
- Vergleichbar den Trennschärfekoeffizienten sind Korrelationen ( $r$ ) von Gesamtpunktwert mit den 13 lexikalisch-semantischen Kategorien angegeben (durchgängig  $p < 0,0001$ )
- Höchster Wert:  $r = 0,71$  („Draußen und Natur“); niedrigster Wert:  $r = 0,36$  („Verbale Ausdruckstätigkeiten“)
- Durchgängig signifikanter Zusammenhang ( $p < 0,0001$ )  
→ alle Kategorien tragen zum Gesamtergebnis bei
- 50% der Varianz des Gesamtpunktwerts wird durch Varianz der größten Kategorie „Draußen u. Natur“ bestimmt

#### *Altersabhängigkeit Summenwert in den lexikalisch-semantischen Kategorien*

- Mittlerer Summenwert für alle lexikalisch-semantischen Kategorien bezogen auf Lebensalter in Tabelle angegeben
- Auf jeder Altersstufe prozentuale Lösungsrate für Verben höher als Substantive → Summenwerte von substantivischen u. Verbkategorien korrelieren mit .68
- Beste Kategorien bei...  
3-Jährigen: „Tätigkeiten Spiel/Sport/Freizeit“ (73%)  
4-Jährigen: „verbale Ausdruckstätigkeiten“ (90%)  
5-Jährigen: „verbale Ausdruckstätigkeiten“ (95%)  
→ diese Kategorien verkörpern laut Autorin dementsprechend altersübergreifend lexikalischen Ausschnitt, der für Kindergartenkind von 3 und 5 typisch ist
- Durchschnittlicher Summenwert jeder sem. Kategorie steigt mit Lebensalter an (Ausnahme: „Symbole“ □ Leistung von 4- und 5-Jährigen gleich gut)
- Durchschnittlicher Summenwert Substantive, sowie Summenwert Verben nehmen ebenfalls mit Lebensalter zu
- Mittelwertsunterschiede wurden varianzanalytisch mit anschließendem Scheffé-Test überprüft

#### *Interkorrelationen der lexikalisch-semantischen Kategorien*

- Lexikalisch-semantische Kategorien des AWST-R sind alle dem Konstrukt „aktiver Wortschatzumfang“ zugeordnet und sollten positiv interkorreliert sein
- Produkt-Moment-Korrelationen sind positiv, unterschiedlich stark ausgeprägt und bis auf drei signifikant
- Keine Korrelation höher  $r > 0,70$
- Stärkster statistischer Zusammenhang zwischen substantivischen Wortfeldern „Gebrauchsgegenstände Haushalt“ und „Körperpflege/Gesundheit/Körper/Krankheit“ ( $r = 0,58$ )
- Numerisch unterschiedliche Besetzung der Kategorien muss berücksichtigt werden

- Interkorrelation der 9 substantivischen Wortfelder liegt in mittlerer Höhe (statistische Zusammenhänge mit Wortfeld „Kleidung“ (3 Items) sind alle deutlich niedriger)
- Interkorrelation der 4 Verb-Wortfelder sind mit Wert von höchstens  $r=0,35$  eher niedrig
- Verb-Wortfelder „Tätigkeiten Spiel/Sport/Freizeit“ und „Verbale Ausdruckstätigkeiten“ korrelieren so gut wie nicht miteinander
- Substantivische Wortfelder korrelieren mit Verb-Wortfeldern in mittlerem bis niedrigem Ausmaß
  - stärkster statistischer Zusammenhang zwischen subst. Wortfeld „Draußen und Natur“ und Verb-Wortfeld „Tätigkeiten im häuslichen Umfeld“ ( $r = 0,51$ )
  - niedrigster u. statistisch insignifikanter Zusammenhang zwischen subst. Wortfeld „Kleidung“ und Verb-Wortfeld „Nonverbale Ausdruckstätigkeiten“ ( $r = 0,14$ ) (Wortfeld „Kleidung hinsichtlich der Items „Bademantel“ u. „Gürtel“ wahrscheinlich eher „untypisch“ besetzt)
- Möglicherweise schlägt sich nicht nur Bestehen/ Fehlen semantischer Ähnlichkeiten in Korrelationen wieder, sondern auch individuelle Unterschiede im Sprachinput der Kinder
- sehr starker positiver linearer Zusammenhang zwischen Gesamtpunktwert und Summenwert der substantivischen Wortfelder ( $r = 0,97$ ,  $p < 0,0001$ ) und dem Summenwert der Verb-Wortfelder ( $r = 0,83$ ,  $p < 0,0001$ )

#### *Extremgruppenvalidierung*

- untersuchte Extremgruppen: bilingual aufwachsende Kinder und Kinder mit Sprachentwicklungsstörungen
- Testvalidierung i.S quasi-experimentellen Zwei-Gruppen- Plans organisiert

#### *Mono- vs. Bilingualität*

Annahme: Bilingual aufwachsende Kinder weisen für das Deutsche einen niedrigeren Wortschatzumfang auf als monolinguale Kinder

- Bilingual aufwachsende Kinder in KiTas untersucht ( $n=67$ )
- Untersuchungssituation entsprach der, der monolingual aufwachsenden Kinder ( $n=113$ )
- Ähnliches mittleres Alter beider Gruppen (54,1 biling. vs. 53,3 Monate monoling.)
- Extremgruppen wurden parallelisiert ☐ jedem bilingualen Kind ein monolinguales Kind gleichen Alters und Geschlechts zugeordnet ☐ 26 Kinder ausgeschlossen, da kein entsprechender „Partner“ vorlag (41 übrige Paare)
- Mono- und bilinguale Kinder unterschieden sich signifikant in der Benennleistung im AWST-R (Gesamtpunktwert) (keine Signifikanzprüfung für 3-Jährige wegen geringer Gruppengröße ( $n=16$ ))

*Nicht-sprachentwicklungsgestörte vs. Sprachentwicklungsgestörte Kinder*

	<ul style="list-style-type: none"> <li>- Annahme: Kinder mit rez. oder prod. Sprachentwicklungsstörungen sollten niedrigeren Gesamtwert im AWST-R erreichen, Kinder mit ausschl. Artikulationsstörungen sollten ähnlich großen Wortschatzumfang wie Kinder mit unauffälliger Sprachentwicklung zeigen</li> <li>- Rekrutierung Kinder mit umschriebenen Entwicklungsstörungen im Sprechen und der Sprache in Göttingen (logopädische Praxen, Schule für Logopädie, Abteilung Phoniatrie/Pädaudiologie des Uniklinikums)</li> <li>- Untersuchung der Kinder mit AWST-R jeweils vor Ort (n= 27, 18 J. und 9 M., durchschnittliches Alter 57,0 Monate)</li>   <li>- Niedrigster Gesamtpunktwert: Kinder mit rezeptiven Störungen (27,0 Punkte)</li> <li>- Kinder mit ausschl. Störung der Artikulation erreichten im Mittel beste Benennleistung (48,7 Punkte)</li> <li>- Kinder mit expressiven Sprachstörungen lagen dazwischen (31,3)</li> <li>- Kinder mit Artikulationsstörung unterschieden sich in durchschnittlicher Benennleistung nicht signifikant von parallelisierter Kontrollgruppe ohne Artikulationsauffälligkeiten (Ergebnis geringfügig schlechter)</li> <li>- Kinder mit rez. und expr. Sprachentwicklungsstörungen erreichten durchschnittlich signifikant weniger Pkte. (30,3) als Kinder mit unauffälliger Sprachentwicklung (41,4)</li> <li>➔ Benennleistung im AWST-R vermag zw. Sprachentwicklungsgestörten und in der Sprachentwicklung unauffälligen Kindern zu trennen</li> </ul>
<b>Preis</b>	241,00 €

## Testentwicklung

Im Vorwort des Testhandbuchs berichtet die Autorin, dass sich der Wortschatz in den vergangenen Jahrzehnten als „eine zentrale Dimension in der Sprachentwicklungspsychologie“ (Handbuch S.6) herausgestellt hat. Demnach kritisiert sie, dass bis zur Entwicklung des AWST-R ausschließlich der „Aktive Wortschatztest für 3- bis 6-jährige Kinder“ (Kiese und Kozielski 1979/1996) als etabliertes Prüfverfahren für diesen wichtigen Aspekt der Sprachentwicklung erhältlich war. Da der AWST 3-6, an dessen Entwicklung die Autorin ebenfalls beteiligt war, allerdings in vielerlei Hinsicht bereits veraltet und somit nicht mehr als valide zu betrachten war, hat sich die Autorin der Aufgabe angenommen eine aktualisierte, valide Testform zu erstellen. Hierbei wurde sie finanziell von der Ravensburger Stiftung unterstützt, die eine umfangreiche Erläuterung des theoretischen Hintergrundes voraussetzte. Abgesehen von den teilweise nicht mehr zeitgemäßen Items und deren Phänomenologie im AWST- 3 bis 6 hatte sich laut Autorin auch die Art der Itemdarstellung (Schwarz-Weiß-Zeichnungen) als für Kinder wenig ansprechend erwiesen, weshalb sich Christiane Kiese-Himmel im neuen Wortschatztest für die Darstellung mittels digitaler, farbiger Photographien entschied.

Bei der Entwicklung des AWST-R wurden dieselben Ziele wie beim AWST 3 bis 6 verfolgt. Aufgrund der Ergebnisse der Pilotierungsphase und der empirischen Erprobung der Testvorform entschied sich die Autorin jedoch dafür, den AWST-R im Gegensatz zum AWST auf die Zielgruppe 3 bis 5;5-jähriger Kinder zu begrenzen.

## Testbewertung

<b>Theoretischer Hintergrund</b>	<ul style="list-style-type: none"><li>- Grundlagenwissen geschildert</li><li>- Relevant für Testentwicklung und -ziel</li><li>- Wird in späteren Kapiteln immer wieder aufgegriffen</li></ul>
<b>Operationalisierung / Aufgabenstellung</b>	<ul style="list-style-type: none"><li>- Sehr klare Anweisungen zur Aufgabenstellung bis hin zu Nachfragen (auf den Ringbuchblockseiten)</li><li>- Setting eindeutig definiert</li></ul>
<b>Material</b>	<ul style="list-style-type: none"><li>- Digitale Fotografien für die Items, allerdings geringe Qualität und veraltete Aufbereitung</li><li>- „Tor“ und „Küken“ uneindeutig abgebildet</li><li>- Verständliche Begründung für Materialauswahl, mehrfache und nachvollziehbare Überarbeitung der Itemauswahl</li></ul> <p><i>Handbuch</i></p> <ul style="list-style-type: none"><li>- Ausführlich</li><li>- Nachvollziehbar geschrieben, mit Erläuterungen zu allen statistischen sowie linguistischen Termini und Vorgehensweisen</li></ul>
<b>Itemauswahl</b>	<ul style="list-style-type: none"><li>- Ausführliche Kapitel zu „Testaufbau und Itemauswahl“, „Pilotierungsphase“ und der „empirischen Erprobung der Testvorform“</li><li>- Beruhend auf Erfahrungen mit AWST, bspw. schwarz-weiße Zeichnungen für Items ersetzt durch digitale Fotografien +</li><li>- Schwierigkeitsgrade der Items anhand linguistischer Parameter eingeschätzt (Geläufigkeit, Wortlänge (Silbenzahl), Artikulationsanforderung)</li><li>- Annäherung an Normalverteilung mit Hilfe von Schiefe-Kennwert und Exzess-Kennwert nachvollziehbar (Ersetzung und Wiederaufnahme von (zu leichten und schweren) Items)</li><li>- Zu leichte Items wurden als Eisbrecher aufgenommen, waren somit schon evaluiert</li><li>- Verben und Substantive nach semantischen Kategorien aufgeführt</li></ul>

	<ul style="list-style-type: none"> <li>- Altersgruppe kritisch reflektiert und korrigiert (da vorläufige Testform für ältere Kinder zu leicht Altersbegrenzung auf 5;5)</li> <li>- Item „rauchen“ inadäquat und tendenziell verherrlichend dargestellt</li> </ul>
<b>Durchführungsdauer</b>	<ul style="list-style-type: none"> <li>- Angemessene Länge</li> </ul>
<b>Auswertung</b>	<ul style="list-style-type: none"> <li>- Objektivität gegeben, durch klare Auswertungskriterien</li> <li>- Durch dichotome Punktwerte simpel und wenig Zeitaufwand</li> <li>- geringe Fehleranfälligkeit</li> </ul>
<b>Normierung</b>	<ul style="list-style-type: none"> <li>- Wie die „Sinnesgesundheit“ und „Normalbegabung“ der Kinder evaluiert wurde, wird nicht näher beschrieben. → ist allerdings von hohem Wert für den Ausschluss von Kindern mit diversen Pathologien und für die Sicherung einer definitiv homogenen Stichprobe</li> <li>- Im Rahmen der Gesamtstrichprobe eine Probandenzahl von hohem repräsentativem Wert (551 Kinder)</li> <li>- Allerdings sind die Altersgruppen im Einzelnen, insbesondere die der 3;0-3;5-Jährigen nur mäßig vertreten</li> <li>- Kinder aus 10 Bundesländern</li> <li>- Bildungshintergrund des häuslichen Umfelds nicht näher erhoben -</li> </ul>
<b>Objektivität</b>	<ul style="list-style-type: none"> <li>- Durchführungsobjektivität gegeben (klare und zugängliche Anweisungen, durch Vorgaben auf Testmaterial)</li> <li>- Auswertungsobjektivität gegeben (eindeutige Definitionen korrekter/inkorrekt Items)</li> <li>- Interpretationsobjektivität gegeben (Rohwerte lassen sich in objektiv evaluierte Prozentränge und T-Werte übertragen und somit geringerer Wortschatzumfang identifizierbar)</li> </ul>
<b>Reliabilität</b>	<ul style="list-style-type: none"> <li>- Mittels Split-Half-/Odd-Even-Methode gegeben</li> <li>- Korrelation von <math>r_{12} = .76</math></li> <li>- Damit gut einsetzbar für Verlaufsdiagnostik/Therapieevaluation</li> </ul>
<b>Validität</b>	<ul style="list-style-type: none"> <li>- Validitätskriterien umfassend und transparent untersucht</li> <li>- Vergleich mit nicht validierten Tests wenig aussagekräftig</li> <li>- Insbesondere Extremgruppenvergleich überzeugt und belegt Validität des AWST-R</li> <li>→ Kinder mit Artikulationsstörung unterschieden sich in durchschnittlicher Benennleistung nicht signifikant von parallelisierter Kontrollgruppe ohne Artikulationsauffälligkeiten (Ergebnis geringfügig schlechter)</li> <li>→ Kinder mit rez. und expr. Sprachentwicklungsstörungen erreichten durchschnittlich signifikant weniger Pkte. (30,3) als Kinder mit unauffälliger Sprachentwicklung (41,4)</li> </ul> <p>Benennleistung im AWST-R vermag zw. Sprachentwicklungsgestörten und in der Sprachentwicklung unauffälligen Kindern zu trennen</p>
<b>Nebengütekriterien</b>	<ul style="list-style-type: none"> <li>- Transparenz gegeben</li> <li>- Äußere Gestaltung durch Itemselektion ausreichend und gut evaluiert</li> </ul>
<b>Theoretischer Rahmen</b>	<ul style="list-style-type: none"> <li>- Ausführlich geschildert</li> </ul>
<b>Klinische Anwendbarkeit</b>	<ul style="list-style-type: none"> <li>- Anwendung einfach und schnell</li> <li>- Jedoch veraltet, Alleinstellungsmerkmal nicht mehr gegeben</li> <li>→ Klinische Anwendbarkeit daher fragwürdig</li> </ul>

**Abschließende  
Bewertung**

Insgesamt handelt es sich bei dem AWST-R um einen umfangreich evaluierten Entwicklungstest für den aktiven Wortschatz. Frau Kiese-Himmel ist durchaus dem Anspruch gerecht geworden, den theoretischen Hintergrund ausführlich und verständlich darzustellen. Die gesamte Testentwicklung lässt sich sehr gut nachvollziehen, insbesondere durch die Schilderung der Pilotierungsphase und Itemauswahl bzw. -selektion. Positiv zu betonen sind die Erläuterungen der einzelnen statistischen Termini, welche das Manual auch für Personen transparent machen, die sich in ihrem Arbeitsalltag nicht intensiv mit Statistik befassen. Wie auf den vorhergegangenen Seiten geschildert, sind Objektivität und Reliabilität gegeben.

Erste kritische Punkte finden sich innerhalb der Normierung, da die Selektion von „sinnesgesunden und vermeintlich normalbegabten Kindern“ nicht näher erläutert wird und zudem für die einzelnen Altersklassen relativ kleine Stichprobengrößen vorliegen. Bezüglich der Validität wurden Testungen zu vielen verschiedenen Validitätsmaßen durchgeführt, welche jedoch nur teilweise als sinnvoll zu betrachten sind (Erheben des Schätzurteils oder aber Hinzuziehen des PET nach Angermaier, trotz nicht vorhandener Werte zur Validität). Insbesondere die Darstellung der Items ist nicht in jedem Fall adäquat und ist den Rezensentinnen teils sehr negativ aufgefallen. Jedoch ist die Validität mit den Extremgruppenvergleichen gegeben.

Die Zielsetzung des AWST-R ist es, Kinder mit einem Bedarf für sprachliche Frühförderung oder sprachtherapeutische Interventionen identifiziert zu können. Außerdem soll nicht nur die Diagnose für eine Sprachentwicklungsstörung gestellt werden können, sondern es sollen auch Indikationen für erste Therapiemaßnahmen evaluiert werden.

Wie sich im Laufe der Analyse des Diagnostikinstrumentes herausstellte, ermöglicht der AWST-R eine Differenzierung von sprachentwicklungsgestörten Kindern und ungestörten Kindern in Bezug auf den Wortschatz. Im Manual ist jedoch keine Angabe dazu zu finden, ob der Test auch für solche Kinder sensitiv ist, bei denen sich die Störung ausschließlich auf den morphosyntaktischen Bereich bezieht oder ob er nicht lediglich Kinder mit einem verminderten Wortschatz identifiziert. Zudem fehlen genauere Angaben zu einem Grenzwert, der zwischen förder- und therapiebedürftigen Kindern unterscheiden lässt. Rückschlüsse in Bezug auf Indikationen für Therapiemaßnahmen sind für die Wortschatztherapie gegeben.

Zusammenfassend war die klinische Anwendung des AWST-R in den vergangenen Jahren sicherlich vertretbar und zu empfehlen. Es handelt sich hier um ein ausführlich evaluiertes Testmaterial, welches jedoch nicht mehr zeitgemäß ist. Mittlerweile gibt es aktuellere Diagnostikinstrumente, welche dem AWST-R das 2005 noch vorhandene Alleinstellungsmerkmal als Wortschatztest für 3- bis 5-jährige Kinder absprechen. Demnach stellen die zu investierenden 241,00€ bereits eine

	hohe Geldsumme dar, welche aufgrund der genannten Punkte nicht mehr als gerechtfertigt zu beurteilen sind. Dennoch hebt sich der AWST-R nach wie vor durch die große Zahl an überprüften Items von anderen Materialien ab. Folglich wäre eine Neuauflage mit zeitgemäßem Bild- und Itemmaterial begrüßenswert.
--	--

## Quellen

Angermaier, Michael J.W.: Psycholinguistischer Entwicklungstest. 2., korrigierte Auflage. Weinheim: Beltz Test 1977.

Bortz, J. & Döring, N.: Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler. Heidelberg: Springer Verlag 2006.

Kauschke, Christina; Siegmüller, Julia: Patholinguistische Diagnostik bei Sprachentwicklungsstörungen (PDSS). 2. Auflage. München: Elsevier GmbH 2010.

Kiese, Christiane und Kozielski, Peter-Michael: AWST 3-6. Aktiver Wortschatztest für 3-6jährige Kinder. Zweite Auflage. Göttingen: Beltz Test GmbH 1979.

Kiese-Himmel, Christiane: Aktiver Wortschatztest für 3- bis 5-jährige Kinder (AWST-R). Erste Auflage. Göttingen: Hogrefe Verlagsgruppe 2005.

Kiese-Himmel, Christiane und Kozielski, Peter-Michael: AWST 3-6. Aktiver Wortschatztest für 3 bis 6-jährige Kinder. Zweite Auflage. Göttingen: Beltz Test GmbH 1996.