

# Psychometrische Kennwerte zum Subtest „Graphästhesie“ aus dem Göttinger Entwicklungstest der Taktil-Kinästhetischen Wahrnehmung (TAKIWA). Eine empirische Untersuchung an einer aktuellen Stichprobe

## Psychometric Parameters of the “Graphesthesia Test” as part of the German Göttinger Developmental Test of Tactile-Kinesthetic Perception (TAKIWA). An empirical analysis with an actual sample

**Autoren:** Christiane Kiese-Himmel, Herbert Poinstingl  
**Institution:** Institut für Medizinische Psychologie und Medizinische Soziologie Universitätsmedizin Göttingen

### Schlüsselwörter

- Entwicklungstest
- Graphästhesie
- psychometrische Kennwerte
- TAKIWA

### Key Words

- Developmental test
- graphesthesia
- psychometric parameters
- TAKIWA

DOI des Beitrags (www.doi.org):  
10.2443/skv-s-2018-54020180203



### Korrespondenzanschrift

Prof. Dr. rer. nat. Christiane Kiese-Himmel, Dipl.-Psych. (Leiterin: Phoniatrie / Pädaudiologische Psychologie), Universitätsmedizin Göttingen, Waldweg 37, D-37073 Göttingen  
Tel.: 0551/39-22844  
E-Mail: ckiese@med.uni-goettingen.de

ergoscience 2018; 13(2): 66-72  
© Schulz-Kirchner, Idstein  
ISSN 1861-6348

### Zusammenfassung

**Einleitung:** Der Subtest „Graphästhesie“ aus dem Göttinger Entwicklungstest der Taktil-Kinästhetischen Wahrnehmung (Kiese-Himmel, 2003) wurde einer Re-Evaluation unterzogen: (1) Ergebnisvergleich auf Rohwertebene mit denen im Testmanual; (2) Ergebnisvergleich von Kindern mit Migrationshintergrund (KMM) und ohne Migrationshintergrund (KOM) unter Beachtung von Lebensalter und Geschlecht; (3) Berechnung der Itemkennwerte „Schwierigkeit“ und „Trennschärfe“; (4) Bestimmung von Reliabilität (interne Konsistenz) und (5) Validität (Testgültigkeit).

**Methode:** Den Analysen lagen die Daten an 3- bis 6-jährigen Kindern im mittl. Alter von 50,04 (SD 9,55) Monaten zu Grunde (N=303; n=144 Jungen), mehrheitlich mit Migrationshintergrund (n=263).

**Ergebnisse:** KOM zeigten mit zunehmendem Alter im Durchschnitt höhere Ergebnisse als die entsprechenden Altersgruppen der Eichstichprobe, das Gleiche traf für KMM zu – ausgenommen die 3-Jährigen, deren mittlere Leistung der der Eichstichprobe entsprach. Die Gesamtleistung von KMM und KOM unterschied sich statistisch nicht signifikant (95 %-KI: -1.25–0.70). Graphästhesie korrelierte nicht mit dem Migrationshintergrund (.04). Der bekannte Alterseffekt (s. Manual) wurde bestätigt ( $\rho=.57$ ;  $p<.01$ ), ebenso der fehlende Geschlechtseffekt ( $r_{pb}=.09$ ). Die Item-Schwierigkeiten wichen von den im Manual ab, die Item-Trennschärfen fielen niedriger aus. Die interne Konsistenz (Cronbachs  $\alpha$ : .72) war noch akzeptabel. Durch eine nahezu fehlende Korrelation mit dem konstruktfernen Gesamtscore des grafomotorischen Tests zum Schöpferischen Denken-Zeichnerisch (Urban & Jellen, 2010) wurde diskriminante Validität ( $\rho=.027$ ; n.s.) belegt. In Ergänzung zu den im Manual genannten konvergenten Validitäten wurden mittlere bis hohe Validitätskennwerte mit konstrukt-nahen, zeitgleich durchgeführten Tests gewonnen.

**Schlussfolgerung:** Die Ergebnisse im Graphästhesie-Test lassen sich auf KMM generalisieren. Insgesamt empfiehlt sich die Abfolge der Items zu verändern, ggf. sogar eine Überarbeitung der Items in Erwägung zu ziehen.

### Abstract

**Introduction:** The graphesthesia subtest of the Göttinger Entwicklungstest der Taktil-Kinästhetischen Wahrnehmung (Kiese-Himmel, 2003) was re-evaluated: (1) comparison of test performances (raw scores) with those in the test manual; (2) of children with (KMM) and without (KOM) migration background considering age and gender; (3) calculation of item parameters like “difficulty” and “discrimination”; (4) reviewing reliability (internal consistency), and (5) validity.

**Methods:** Analyses were made with the test results of children aged 3 to 6 years (N=303; M= 50.04, SD 9.55 months; n=144 boys), mostly with migration background (n=263).

**Results:** KOM showed higher test performance with increasing age compared to the age groups in the test manual, on average. This also applied for KMM – with the exception of 3-year-old children, whose mean raw score was identical with the TAKIWA norm sample. The mean test performance of KMM did not significantly differ from KOM (95 %-KI: -1.25–0.70). Graphesthesia performance did not correlate with the migration background (.04). The known age effect according to the test manual could be replicated ( $\rho=.57$ ;  $p<.01$ ), the missing gender effect too ( $r_{pb}=.09$ ). Item difficulties deviated from those in the manual, and the indices of item discriminatory power were lower as mentioned there. The internal consistency was still satisfactory (Cronbach's  $\alpha$  .72). Discriminant validity could be confirmed ( $\rho=.027$ , n.s.) by almost missing correlation with the German adaption of the graphomotor working Creative Thinking and Drawing Production Test (Urban & Jellen, 2010). In addition to the validity coefficients specified in the test manual, medium to high convergent validity values were obtained with construct-related simultaneous performed standardized tests.

**Conclusion:** The graphesthesia test results are able to generalize to children with migration background. However, the sequence of the ten items should be changed, and if appropriate, an item revision should be considered.

## Einleitung

Standardisierte Testinstrumente erfüllen im Kindesalter eine wichtige Aufgabe in Diagnostik und Differenzialdiagnostik, bspw. zur Entwicklungsstandbestimmung oder zur Dokumentation des Entwicklungsverlaufes. Voraussetzung hierfür ist, dass sie valide und reliabel sind sowie das Hauptgütekriterium der Objektivität erfüllen. Lienert und Raatz schrieben 1994 (S.173), dass „... der Test als wissenschaftliches Prüfinstrument einer dauernden und ins Einzelne gehenden Forschungsarbeit bedarf“. Hierzu ist eine unabhängige empirische Überprüfung seiner Gütekriterien nötig, um dem Qualitätserhalt, insbesondere aber der ggf. erforderlichen Qualitätsverbesserung Rechnung zu tragen. Im vorliegenden Beitrag werden deshalb aktuelle psychometrische Item- und Testkennwerte zum Subtest „Graphästhesie“ eines spezifischen Entwicklungstests präsentiert, des Göttinger Entwicklungstests der Taktil-Kinästhetischen Wahrnehmung (TAKIWA; Kiese-Himmel, 2003). Damit wird auch der im Manual dieses Tests im Kapitel „Ausblick“ (S. 77) gegebenen Empfehlung ergänzender Validierungsstudien für einzelne Prüfdimensionen entsprochen.

Der TAKIWA ist ein nach der klassischen Testtheorie konstruierter standardisierter Test für Kinder im Alter von 3 Jahren und 6 Monaten (3;6) bis sechs Jahren, der u.a. in der ergotherapeutischen und neuropsychologischen Funktionsdiagnostik Anwendung findet. Er besteht aus sieben Subtests (insgesamt 47 Testaufgaben), mit denen sowohl passiv-taktile als auch haptische Leistungen erhoben werden. Mit einer internen Konsistenz (Cronbachs  $\alpha$ ) für den Gesamttest von .85 (in der Eichstichprobe; s. S. 74 im TAKIWA-Testmanual) sowie einer Halbierungsreliabilität von .84 bzw. .87 (je nach Halbierungsmethode; s. S. 48 im Testmanual) darf von einem reliablen Test ausgegangen werden. Durchführungsobjektivität des TAKIWA ist aufgrund der schriftlich festgelegten Instruktionen, Vorgehensweisen und Abbruchkriterien gegeben, wengleich lediglich beim Graphästhesie-Test die Größe der in den Handteller (= Basis der Handfläche innen bis zum Handgelenk) zu zeichnenden Form von der Größe der Kinderhand abhängig ist und demgemäß variieren kann; deswegen wurden im Testmanual keine Angaben zur Größe der Form gemacht. Zur Auswertungsobjektivität tragen eindeutige Auswertungsrichtlinien bei. Interpretationseindeutigkeit ist durch Altersnormen für den Summenwert jedes Subtests sowie für den taktil-kinästhetischen Gesamtscore gewährleistet (Prozentrangplätze; T-Werte). Der TAKIWA verfügt über inhaltlich-logische Gültigkeit und über empirisch geprüfte Validitätsaspekte (faktoriell; kriterienbezogen; entwicklungspsychologische Perspektive), die ihn als valides Untersuchungsinstrument ausweisen. Eine Extremgruppenvalidierung zeigte in Übereinstimmung mit der Modellvorstellung zur frühen Sprach-, insbesondere lexikalischen Entwicklung eines Kindes (Kiese-Himmel, 2001; 2005; 2007; 2008), dass sprachun auffällige Kinder eine signifikant höhere taktil-kinästhetische Gesamtleistung hatten im Vergleich zu sprachentwicklungsge störten Kindern. Hierzu trug stark die Leistung im Graphästhesie-Test bei. Konzepte als „symbolische Repräsentationen“ sind die Schnittstelle von Graphästhesie und Sprache (vgl. Piagets Stufenmodell der kognitiven Entwicklung, 1969).

Mit „Graphästhesie“ wird die passiv-aktive (taktile) Wahrnehmung von Symbolen unterschiedlicher Komplexität (ohne Sicht) bezeichnet. Sie erfordert die Identifizierung einer vom Untersucher aus dem sukzessiven Verlauf seiner Fingerbewegung resultierenden in die Hand der Testperson „gemalten“ Form. Zur Bearbeitung einer graphästhetischen Aufgabe sind intakte tak-

tile Sensibilität sowie spezifische kognitive Fähigkeiten (Raumerfassung, Fähigkeit zur Wahrnehmung der Bewegungsrichtung eines taktilen Reizes, Arbeitsgedächtnis) nötig. Das Wahrnehmungsergebnis muss von der Testperson zentral, d.h. mit seinen mentalen Vorstellungsmustern, abgeglichen und identifiziert werden.

Der TAKIWA-Subtest „Graphästhesie“ enthält zehn Aufgaben (Items) ansteigender Schwierigkeit (mittlerer Schwierigkeitsindex: 0.50), die in einer zumutbaren Testdauer von zehn bis 15 Minuten durchzuführen sind. Dem Kind werden mit einem Watteabchen ein vertikaler Strich sowie geometrische Grundformen (Kreuz, Kreis, gleichseitiges Dreieck und Viereck) die es zu erkennen gilt, jeweils in seine linke bzw. rechte Handfläche „gemalt“. Nur die beiden vorgeschalteten Übungsaufgaben werden einmal mit und hiernach ohne Sicht durchgeführt. Die Lösungsangabe ist nonverbal möglich (Aufzeigen der in die Hand „gemalten“ Form unter insgesamt fünf Formen). Jedes richtig gelöste Item wird mit einem Punkt bewertet (max. Summenwert: 10). Die interne Konsistenz (Cronbachs  $\alpha$  in der Eichstichprobe) beträgt .81. Der Graphästhesie-Test weist die mit Abstand höchste Korrelation eines Subtests mit dem TAKIWA-Gesamtscore auf. Graphästhesie und Lebensalter korrelieren miteinander bedeutsam positiv von .46 ( $p < .0001$ ) in der Eichstichprobe (TAKIWA-Testmanual, 2003, S.73; s. auch Kiese-Himmel & Schiebusch-Reiter, 1995).

Psychologische Tests sollen nach ihrem Erscheinen (weiter) evaluiert werden (und die Normen mindestens alle acht Jahre aktualisiert) – in Anlehnung an die DIN-Norm 33430 [Version 2016] der berufsbezogenen Eignungsdiagnostik. Ziel der vorliegenden Studie war es, (1) die Subtestergebnisse einer aktuellen Stichprobe einem Vergleich mit denen im Testmanual zu unterziehen; (2) sollten die Subtestergebnisse von Kindern mit Migrationshintergrund (KMM) mit denen von Kindern ohne Migrationshintergrund (KOM) unter Berücksichtigung von Lebensalter und Geschlecht verglichen werden; (3) sollten Kennwerte der Testaufgaben (Schwierigkeit und Trennschärfe) berechnet sowie (4) Reliabilität (Testzuverlässigkeit) und (5) Validität (Testgültigkeit) bestimmt werden (in methodischer Anlehnung an Lienert & Raatz, 1994).

## Methode

### Design und Stichprobe

Die Datenerhebung<sup>1</sup> fand im Setting der (von der lokalen Ethikkommission der Universität Göttingen genehmigten) Längsschnittstudie (drei Messzeitpunkte T1 – T3) statt. Die Teilnahme an dieser Studie war freiwillig. Alle Kinder wurden zwischen Mai 2009 und Juli 2014 am Vormittag im Einzelverfahren in einem separaten Raum durch geschulte pädagogische und psychologische Fachkräfte in sieben Kindertagesstätten im Großraum Frankfurt/Main, Darmstadt und Offenbach mit einer umfassenden psychologischen Testbatterie untersucht. Eine Einverständniserklärung seitens der Erziehungsberechtigten lag den Fachkräften vor.

1 Finanziell unterstützt durch das Niedersächsische Institut für frühkindliche Bildung und Entwicklung Nifbe, Hannover (FP 17-09), die Albert und Barbara von Metzler-Stiftung, Frankfurt/Main (1348840), die gemeinnützige Hertie Stiftung, Frankfurt/Main (1.00.1/07 006 H) sowie die Adolf-Messer-Stiftung, Bad Soden (1351110).

Den aktuellen Analysen lagen die Daten von T1 (Baseline-Testung) zugrunde. Die Stichprobe umfasste 303 Kindergartenkinder (47,5 % Jungen; 52,5 % Mädchen) mit einem hohen Anteil an KMM mit Deutsch als Zweitsprache (→ Tab. 1). Diese wuchsen im Gegensatz zu den 40 monolingual deutschsprachigen Kindern ohne Migrationshintergrund mehrsprachig auf. Die Muttersprache von 48 Kindern (15,8 %) war Türkisch, von 44 Kindern (14,5 %) Arabisch, und 14 Kinder (4,6 %) sprachen Kroatisch. Andere Muttersprachen wurden jeweils nur von wenigen Kindern gesprochen, was die Heterogenität des Migrationshintergrundes dokumentiert. Bezogen auf das Lebensalter bestand die Gesamtstichprobe aus 3-Jährigen (n=136), 4-Jährigen (n=105) und 5-Jährigen (n=62) im mittleren Alter von 50,04 (SD = 9,55) Monaten. Alle Kinder waren durchschnittlich intelligent (K-ABC-Skala „Intellektuelle Fähigkeiten“). Die Eltern aller Kinder hatten im Durchschnitt ca. zehn Jahre eine Schule besucht. Mütter bzw. Väter von KOM gingen durchschnittlich ein Jahr länger zur Schule als die Eltern von KMM. Kein Elternteil der KOM war ohne Schulbildung, 0,7 % der Mütter von KMM hatten keine Schulbildung.

#### Testinstrumente zur Untersuchung der Validität

„Die Validität oder Gültigkeit eines Tests gibt den Grad der Genauigkeit an, mit dem dieser Test dasjenige Persönlichkeitsmerkmal oder diejenige Verhaltensweise, das (die) er messen oder vorhersagen soll, tatsächlich mißt oder vorhersagt“ (Lienert & Raatz, 1994, S. 10).

*Diskriminante Validität* (Begrifflichkeit 1959 von Campbell & Fiske eingeführt) ist dann gegeben, wenn aufgrund der Messun-

gen eines Merkmals mit einem bestimmten Messinstrument keine Rückschlüsse auf die Ausprägungen eines anderen Merkmals (hier auf die Graphästhesie), das von dem zu erfassenden Merkmal unabhängig sein soll, gezogen werden können (vgl. Wirtz, 2017: Dorsch Lexikon der Psychologie). Rechnerisch bezieht sich diskriminante Validität auf die Korrelation von zwei Tests, die nur gering oder gar nicht miteinander korrelieren sollten. Um diskriminante Validität zu ermitteln, wurde der (zum taktischen Graphästhesie-Test) konstruktferne, weil grafomotorisch zu bearbeitende Test zum Schöpferischen Denken – Zeichnerisch (TSD-Z; Urban & Jellen, 2010) verwendet, ein standardisiertes Instrument zur Erfassung des individuellen kreativen Potenzials. Auf einem Testbogen werden sechs grafische (figurale) Fragmente angeboten, die frei weiter zu gestalten und abschließend zu benennen sind. Anhand von 14 Kriterien (u.a. Weiterführung, Ergänzung, Hinzufügung neuer Elemente, Perspektive, Herstellen eines zeichnerischen, bzw. thematischen Zusammenhangs), die das Konstrukt „kreatives Denken“ operationalisieren, werden die Zeichnungen ausgewertet. Die Summe der Punkte pro Kriterium bildet den Gesamtscore. Reliabilitätsangaben liegen nur für Schulkinder vor. Aktuelle psychometrische Analysen an Nicht-Schulkindern wurden jüngst von v. Steinbüchel, Meeuwesen, Poinstingl und Kiese-Himmel (2018) durchgeführt; Cronbachs  $\alpha$ -Koeffizienten belaufen sich hiernach auf .69 bis .73.

*Konvergente Validität* (Übereinstimmungsvalidität) bezeichnet die Korrelation von Merkmalsausprägungen, die durch zwei verschiedene Tests, die dasselbe Konstrukt messen, gewonnen werden; sie sollte bei einem validen Test möglichst hoch ausfal-

Tab. 1: Stichprobenbeschreibung

		Gesamtstichprobe (N=303) N (%)	KMM* (n=263) N (%)	KOM** (n=40) N (%)
<b>Geschlecht</b>	Jungen	144 (47,5 %)	124 (47,1 %)	20 (50 %)
	Mädchen	159 (52,5 %)	139 (52,9 %)	20 (50 %)
	<b>Gesamt</b>	<b>303 (100 %)</b>	<b>263 (100 %)</b>	<b>40 (100 %)</b>
<b>Alter</b> (Monate)	M (SD)	50,04 (9,55)	50,33 (9,43)	48,10 (10,19)
<b>Intelligenz</b> (Standardwert)	M (SD)	47,39 (8,76)	46,75 (8,85)	51,66 (6,82)
<b>Schuljahre Mutter***</b>	M (SD)	10,37 (2,51)	10,24 (2,60)	11,28 (1,58)
<b>Schuljahre Vater***</b>	M (SD)	10,56 (2,28)	10,44 (2,34)	11,39 (1,58)
<b>Bildung Mutter***</b>	Keine Schule	2 (0,7 %)	2 (0,8 %)	–
	Keine Grundschule	22 (7,3 %)	22 (8,4 %)	–
	Berufsschulreife	94 (31,0 %)	88 (33,5 %)	6 (15,0 %)
	Berufsausbildung	124 (40,9 %)	104 (39,5 %)	20 (50,0 %)
	Universität	41 (13,5 %)	29 (11,0 %)	12 (30,0 %)
	<b>Gesamt</b>	<b>283 (93,4 %)</b>	<b>245 (93,2 %)</b>	<b>38 (95,0 %)</b>
<b>Bildung Vater***</b>	Keine Schule	–	–	–
	Keine Grundschule	7 (2,3 %)	7 (2,7 %)	–
	Berufsschulreife	86 (28,4 %)	83 (31,6 %)	3 (7,5 %)
	Berufsausbildung	135 (44,6 %)	114 (43,3 %)	21 (52,5 %)
	Universität	42 (13,9 %)	32 (12,2 %)	10 (25,0 %)
	<b>Gesamt</b>	<b>270 (89,2 %)</b>	<b>236 (89,8 %)</b>	<b>34 (85 %)</b>

Anmerkungen: \*KMM = Kinder mit Migrationshintergrund; \*\*KOM = Kinder ohne Migrationshintergrund  
M = Mittelwert; SD = Standardabweichung; \*\*\* = missing data

len (Wirtz, 2017). Konvergente Validität des Subtests „Graphästhesie“ wurden mit Testkriterien berechnet, die zum einen funktionale Fähigkeiten erfassen, die in der Graphästhesieleistung enthalten sind, zum anderen mit sprachgebundener Symbolkompetenz (als Gegenstück zur sprachfreien Symbolkompetenz). Im Einzelnen:

- Subtest „Räumliche Beziehungen“ des Frostigs Entwicklungstest der visuellen Wahrnehmung (FEW-2; Büttner, Dacheneder, Schneider & Weyer, 2008); er prüft eine Wahrnehmungsleistung, die die Analyse von Formen und Mustern in Relation zum eigenen Körper und zum Raum verlangt. Die interne Konsistenz bis zum Alter von 6 Jahren beläuft sich auf (Cronbachs  $\alpha$ ) .77 bis .87.
- Dt. Version der Kaufman Assessment Battery for Children (K-ABC; Melchers & Preuss, 2009), ein Test zur Messung von Intelligenz und spezifischen Fertigkeiten für Kinder im Alter von 2;6 bis 12;5 Jahren. Da es sich bei der aktuellen Stichprobe nicht um Schulkinder handelte, konnten nicht alle Subtests durchgeführt werden. Die K-ABC-Gesamtskala (Skala „Intellektuelle Fähigkeiten“) resultiert aus dem Summenergebnis der Skalen „Einzelheitliches“ und „Ganzheitliches Denken“. „Einzelheitliches Denken“ beinhaltet den folgerichtigen Umgang mit Reizen (z. B. bei seriellen Kurzzeitgedächtnisleistungen), „Ganzheitliches Denken“ meint die gestalthafte und räumliche Organisation von Reizen. Ihre Reliabilitätskoeffizienten belaufen sich in der Eichstichprobe im Alter von 3 bis 6 Jahren auf .90 bis .94.
- K-ABC-Handbewegungen (HB), ein Subtest der Skala „Einzelheitliches Denken“; er verlangt die Nachahmung einer Folge einfacher, vom Testleiter vorgemachter Handbewegungen. Halbierungsreliabilität für 3-Jährige .78; für 4-Jährige .72; für 5-Jährige .69 (Eichstichprobe).
- K-ABC-Zahlen Nachsprechen (ZN); ebenfalls ein Subtest der Skala „Einzelheitliches Denken“, bei dem vorgeschene Zahlenfolgen ansteigender Länge nachzusprechen sind. Split-Half-Reliabilitätskoeffizienten für das Alter 3 bis 6 Jahre: .87 bis .86 (4-Jährige: .83).
- Subskala „Wörter Erklären“ aus dem Wiener Entwicklungstest (WET; Kastner-Koller & Deimann, 2002); sie erfasst die Fähigkeit zur sprachlichen Begriffsbildung, was die Verfügung über Wörter als Symbole voraussetzt. Cronbachs  $\alpha$  und Halbierungsreliabilität (dt. Eichstichprobe): .81.

Durchführungs- und Auswertungsobjektivität war bei allen Tests durch die Anwendung der wörtlichen Testinstruktion laut jeweiligem Testmanual gegeben, ebenso die Einhaltung der dort vorgegebenen Durchführungs- sowie Auswertungsbedingungen.

#### Statistische Auswertung

Die Datenauswertung erfolgte mit der Software IBM SPSS Statistics 24 sowie der Software R (R Core Team, 2017). Deskriptiv wurden Mittelwerte und Standardabweichungen auf Rohwertebene berechnet. Nach Überprüfung der Varianzhomogenität wurde zu Mittelwertvergleichen der t-Test für unabhängige Stichproben durchgeführt bzw. im Falle ungleicher Varianzen der Welch-Test (Signifikanzniveau:  $\alpha_{\text{Irrtum}} = 0.05$ ). Zur Berechnung des Zusammenhangs von Graphästhesie-Summenwert (intervallskaliertes Merkmal) und Geschlecht bzw. Migrationshintergrund (dichotomes Merkmal) wurde der punktbiseriale Korrelationskoeffizient ( $r_{pb}$ ) gewählt.

Eine Testaufgabe (Item) wird u.a. durch ihre Schwierigkeit und ihre Trennschärfe gekennzeichnet. Die Schwierigkeit eines

Items wurde aus der Anzahl der Kinder, die dieses richtig lösten (%-Anteil der Gesamtstichprobe), abgeleitet. Dabei ist der Schwierigkeitsindex eines leichten Items numerisch hoch, der eines schwierigen niedrig. Um etwaige Rateeffekte zu berücksichtigen, wurden die Aufgabenschwierigkeiten der Zufallskorrektur nach Guilford (1936) unterzogen. Die Trennschärfe eines Items erlaubt die Einschätzung, wie gut eine Testaufgabe zwischen Personen mit hoher bzw. mit niedriger Ausprägung im gemessenen Testmerkmal trennt. Rechnerisch wird sie durch die Korrelation der jeweiligen Testaufgabe mit dem Gesamtergebnis im Test ermittelt. Hierzu wurden der biseriale Korrelationskoeffizient ( $r_{pb}$ ) von Aufgabenlösung (dichotome Ausprägung: richtig vs. falsch) und der Summenwert im Graphästhesie-Test berechnet. Ein hoher Trennschärfekoeffizient bedeutet, dass diese Aufgabe bzgl. des Testmerkmals gute von schlechten Probanden zu unterscheiden vermag.

Unter Reliabilität oder Zuverlässigkeit wird der Grad der Genauigkeit eines Tests verstanden, mit dem er ein Merkmal misst. Der Reliabilitätsschätzung über die Methode der Konsistenzanalyse lag die Überlegung zugrunde, dass der Graphästhesie-Test in so viele Teile zu untergliedern ist, wie er Testaufgaben hat. Zur Erfassung der internen Konsistenz wurde Cronbachs  $\alpha$  mit der Kuder-Richardson KR-20-Formel berechnet (s. Lienert & Raatz, 1994; S.193), da das 5-kategorielle nonverbale Antwortformat in „richtig“ vs. „falsch“ dichotomisiert wurde.

Im Rahmen der Untersuchungen zur Validität (Testgültigkeit) wurden bei Verletzung der Normalverteilungsvoraussetzung nicht-parametrische Korrelationen (Spearman's Rangkorrelationskoeffizient rho) zwischen dem Graphästhesie-Summenwert und den Leistungen in den zuvor genannten konstrukt-nahen Tests gerechnet, ansonsten Produkt-Moment-Korrelationskoeffizienten.

Die Höhe der Korrelationskoeffizienten wurde in Anlehnung an Cohen (1988) wie folgt bewertet:  $\geq .10$  = klein;  $\geq .30$  = mittel;  $\geq .50$  = groß.

#### Ergebnisse

##### Mittlere Graphästhesieleistung

KOM hatten mit zunehmendem Alter im Durchschnitt höhere Summenweltergebnisse als die entsprechenden Altersgruppen in der Eichstichprobe (s. TAKIWA-Testmanual, S.73), das Gleiche trifft für KMM zu – mit Ausnahme der 3-Jährigen, deren Durchschnittsleistung genau der der Eichstichprobe entsprach ( $\rightarrow$  Tab. 2). Mädchen (nicht aus Tab. 2 ersichtlich) zeigten mit einem mittleren Summenwert von 4,65 (SD = 2,51) eine geringfügig höhere Leistung als Jungen (M = 4,21; SD = 2,64). Die entsprechenden Werte der Eichstichprobe im Testmanual (S.70) betragen – unabhängig vom Alter – für Mädchen (N=54) M = 4,5 (SD = 2,8) und für Jungen (N=57) M = 4,8 (SD = 2,8). Die Graphästhesieleistung unterlag somit keinem Geschlechtseffekt ( $r_{pb} = .09$ ).

Im Zeitfenster von drei bis fünf Jahren stieg die mittlere Leistung von KOM und KMM an. Entsprechend korrelierte die Graphästhesieleistung mit dem Lebensalter positiv (.57;  $p < .01$ ). In beiden Gruppen fiel die mittlere Leistung in den Altersgruppen 3, 4 sowie 5 Jahre (wie auch in den jeweiligen Gesamtstichproben) ähnlich groß aus, die Gesamtleistung von KMM unterschied sich nicht statistisch signifikant von der der KOM (95 %-KI: -1.25 – 0.70). Graphästhesie korrelierte nicht mit dem Migrationshintergrund ( $r_{pb} = .04$ ).

Tab. 2: Mittlerer Summenwert (Standardabweichung SD) im TAKIWA-Subtest Graphästhesie – angeordnet nach Lebensalter – sowie Referenzwerte aus der Eichstichprobe gem. Testmanual

Kinder	3 Jahre M (SD)	4 Jahre M (SD)	5 Jahre M (SD)	ALLE M (SD)
KMM* (N=263)	3,00 (1,88) (n=113)	5,00 (2,45) (n=95)	6,61 (1,91) (n=55)	<b>4,48 (2,53)</b>
KOM** (N=40)	3,13 (2,70) (n=23)	4,90 (2,77) (n=10)	6,71 (2,06) (n=7)	<b>4,20 (2,91)</b>
Eichstichprobe (Testmanual; N=111)	3,0 (2,4) (n=22)	4,3 (2,8) (n=42)	5,8 (2,6) (n=47)	<b>4,7 (2,8)</b>

Anmerkungen: \*KMM = Kinder mit Migrationshintergrund; \*\*KOM = Kinder ohne Migrationshintergrund.  
M = Mittelwert; SD = Standardabweichung

Tab. 3: Itemschwierigkeiten im TAKIWA-Subtest „Graphästhesie“

Item	Gem. Testmanual (N=146)	Aktuelle Gesamtstichprobe (N=303)	KMM* (n=263)	KOM** (n=40)	Nach Zufalls- korrektur für aktuelle Gesamt- stichprobe***
1 Kreis links	67,1	49,18	48,67	52,50	39,01
2 Strich (vertikal) rechts	69,2	62,38	63,88	52,50	54,85
3 Kreuz rechts	64,4	53,47	55,13	42,50	44,16
4 Kreis rechts	56,8	50,50	51,71	42,50	40,59
5 Strich (vertikal) links	67,1	52,05	63,12	55,00	54,46
6 Kreuz links	47,9	47,86	47,91	47,50	37,43
7 Dreieck links	43,8	31,68	33,46	20,00	18,02
8 Viereck links	32,9	29,70	28,14	40,00	15,64
9 Dreieck rechts	28,8	29,04	28,14	35,00	14,85
10 Viereck rechts	22,6	28,05	27,38	32,50	13,66

Anmerkungen: \*KMM = Kinder mit Migrationshintergrund; \*\*KOM = Kinder ohne Migrationshintergrund.  
\*\*\* In der Erprobungsstichprobe (s. TAKIWA-Manual) wurden die Itemschwierigkeiten keiner Zufallskorrektur unterzogen.

Tab. 4: Itemtrennschärfen ( $r_{it}$ ) und interne Konsistenz (Cronbachs  $\alpha$ ) im TAKIWA-Subtest „Graphästhesie“

Item	$r_{it}$ gem. Testmanual (N=146)	$r_{it}$ Aktuelle Gesamtstichprobe (N=303)	Cronbachs $\alpha$ , wenn Item weggelassen (aktuelle Gesamtstichprobe) (N=303)
1 Kreis links	0.54	0.36	.71
2 Strich (vertikal) rechts	0.47	0.48	.69
3 Kreuz rechts	0.70	0.44	.69
4 Kreis rechts	0.63	0.35	.71
5 Strich (vertikal) links	0.66	0.54	.67
6 Kreuz links	0.62	0.51	.68
7 Dreieck links	0.62	0.30	.71
8 Viereck links	0.55	0.37	.70
9 Dreieck rechts	0.58	0.23	.72
10 Viereck rechts	0.59	0.24	.72

### Itemschwierigkeit und Itemtrennschärfe

→ **Tabelle 5** zeigt die Schwierigkeiten (Anteil der Richtiglösungen) für die zehn Graphästhesie-Items in der Erprobungsstichprobe der Testendform (s. TAKIWA-Testmanual, S. 39/40); sie liegen im mittleren Bereich zwischen knapp 23% und knapp 70%, was einen Bodeneffekt (bei zu großer Schwierigkeit) bzw. einen Deckeneffekt (bei zu geringer Schwierigkeit) ausschließen sollte. In der aktuellen Gesamtstichprobe fielen die Itemschwierigkeiten ähnlich aus und bewegen sich von 28% bis 62%; die ersten fünf Items waren etwas schwieriger, doch lagen in etwa im mittleren Schwierigkeitsbereich. Für KMM betrug die Itemschwierigkeiten 27% bis knapp 64,4%. Das erste Item war für KMM schwieriger als für KOM, während sie die vier folgenden Items häufiger richtig lösten als KOM. Item 6 war in der Testerprobungs- und in der aktuellen Gesamtstichprobe gleich schwierig, die Items 7 und 8 wiederum waren für die aktuelle Stichprobe schwieriger, hingegen die beiden letzten geringfügig leichter im Vergleich zum Testmanual. Bei einem Item mit zwei Antwortmöglichkeiten beträgt die Ratewahrscheinlichkeit 50%, bei einer Aufgabe mit fünf Antwortmöglichkeiten, wie das bei der Lösungsangabe im Graphästhesie-Test durch das Aufzeigen der richtigen Form unter fünf Formen auf einer Bildvorlage der Fall ist, beläuft sie sich auf 20,4%. Um etwaige Rateeffekte zu berücksichtigen, wurden daher die Itemschwierigkeiten der Zufallskorrektur nach Guilford (1936) unterzogen (s. ebenfalls → **Tab. 5**). Die Items werden (bis auf den Strich links) schwieriger; vor allem in der zweiten Testhälfte liegen sie im oberen bis extrem hohen Schwierigkeitsbereich.

Während die Trennschärfen der Items ( $r_{it}$ ) lt. Testmanual (S.39/40) einen Wertebereich von .47 bis .70 haben und als gut zu bezeichnen sind (ausgenommen das Item „Strich rechts“ mit .47), fielen sie in der aktuellen Gesamtstichprobe niedriger aus (Wertebereich .23 bis .54; → **Tab. 4**), doch lagen mehrheitlich im mittleren Wertebereich von .3 bis .5 (Bortz, 2005). Mit Ausnahme von drei Items in der zweiten Testhälfte waren die Werte  $>.32$ , der Konvention für die Untergrenze akzeptabler Trennschärfe. Mit .54 hatte das Item „Strich links“ die höchste Trennschärfe, gefolgt vom Item „Kreuz links“ (.51).

### Reliabilität (Testzuverlässigkeit)

„Unter der Reliabilität oder Zuverlässigkeit eines Tests versteht man den Grad der Genauigkeit, mit dem er ein bestimmtes Persönlichkeits- oder Verhaltensmerkmal mißt“ (Lienert & Raatz, 1994, S. 9). Sie wird in einem Reliabilitätskoeffizienten ausgedrückt, „... der angibt, in welchem Maße unter gleichen Bedingungen gewonnene Meßwerte über ein und denselben Pbd übereinstimmen, in welchem Maße also das Testergebnis reproduzierbar ist (ebd.). Die Schätzung der Reliabilität als interne Konsistenz war für die aktuelle Gesamtstichprobe akzeptabel (.72). Für KMM betrug sie .71, für KOM .80.

Die letzte Spalte in → **Tabelle 4** zeigt Cronbachs  $\alpha$ , wenn das entsprechende Item aus der Berechnung ausgeschlossen wird. Sofern Cronbachs  $\alpha$  nach dem Ausschluss des Items ansteigt, könnte darüber nachgedacht werden, dieses aus dem Test zu nehmen. Hier wird jedoch der  $\alpha$ -Wert für jedes Item, das aus dem Test entfernt wird, niedriger – mit Ausnahme der beiden letzten Items. Allerdings würde Cronbachs  $\alpha$  dennoch nicht über .72 anwachsen, was in der aktuellen Gesamtstichprobe ohnehin erreicht wurde.

Tab. 5: Kriterienbezogene Validität (konvergent und diskriminant; Korrelationskoeffizient rho von TAKIWA-Graphästhesie-Summenwert i.d aktuellen Gesamtstichprobe [N=303] mit kognitiven Testmaßen

Test	rho
Konvergente Validität:	
FEW-2 Räumliche Beziehungen	.43**
K-ABC Handbewegungen	.50**
K-ABC Zahlen Nachsprechen	.41**
K-ABC Intellektuelle Fähigkeiten	.58**
WET-Wörter Erklären	.41**
Diskriminante Validität:	
TSD-Z	.027

Anmerkungen: \*\* =  $p < .01$

### Validität (Testgültigkeit)

Da Graphästhesie eine aus verschiedenen Leistungen resultierende hoch komplexe hohe kognitive Leistung ist, wurden zur Erhebung von konvergenter Validität (Übereinstimmungsvalidität) Zusammenhänge mit ähnlichen Konstrukten ermittelt, wie in Abschnitt „Testinstumentente zur Untersuchung der Validität“ ausgeführt, u. a. mit der Fähigkeit zur Raumerfassung, mit einzelheitlicher Informationsverarbeitung (visuell und auditiv), mit Intelligenz und mit auf Sprachsymbolen basierenden kognitiven Verstehensprozessen. → **Tabelle 5** zeigt die korrelativen Beziehungen zwischen dem Graphästhesie-Summenwert und diesen Variablen (Testinstrumenten). Ausgeprägte positive, signifikante Zusammenhänge (durchweg  $p < .01$ ) bestanden mit der K-ABC-Skala „Intellektuelle Fähigkeiten“ (.58) und mit dem K-ABC-Subtest „Handbewegungen“ (.50). Mittlere Zusammenhänge wurden mit den Leistungen im FEW-2-Subtest „Räumliche Beziehungen“ (.43), WET-Wörter Erklären (.41) sowie K-ABC-Zahlen Nachsprechen (.41) festgestellt. Zur Bestimmung der diskriminanten Validität, die sich auf die rechnerische Korrelation mit einem Test bezieht, der ein ganz anderes Konstrukt misst, wurde der Zusammenhang mit dem TSD-Z-Gesamtscore berechnet. Ein Zusammenhang mit diesem konstruktfernen Instrument bestand erwartungsgemäß nicht ( $\rho = .027$ ; n.s.).

### Diskussion

Der Subtest „Graphästhesie“ aus dem Göttinger Entwicklungstest TAKIWA (Kiese-Himmel, 2003) wurde im Rahmen einer Re-Evaluation an einer aktuellen Stichprobe von Kindergartenkindern überprüft. Dieser Subtest wurde gewählt, weil er als Maß der sprachfreien Symbolkompetenz für eine hohe kognitive Leistung steht und im Rahmen eines Drittmittelprojekts zeitgleich mit anderen kognitiven Maßen erhoben werden konnte. Zudem handelt es sich bei Graphästhesie um eine personale Ressource, die derzeit mit keinem normierten, deutschsprachigen, psychologischen Test bei Nicht-Schulkindern erfasst werden kann, was diesem Subtest a priori eine Existenzberechtigung hinsichtlich seiner Nützlichkeit zuspricht. Söchting (1999) hat in einer österreichischen Studie an einer im Vergleich zu der vorliegenden Studie deutlich kleineren Stichprobe (111 Kinder) die Sensory Integration and Praxis Tests (SIPT; Ayres, 1989) durchgeführt, die

u.a. auch einen Graphästhesie-Test enthalten, aber an US-amerikanischen und kanadischen Kindern von 4 bis 8;11 Jahren in den späten Achtzigern des vergangenen Jahrhunderts normiert wurden. In dieser, inzwischen als überaltert zu bewertenden Studie, fand sie in der Testleistung keinen signifikanten Leistungsunterschied zwischen österreichischen und amerikanischen Kindern. Die Bundesrepublik Deutschland hat sich seit Ende der 1950er Jahre zu einem Einwanderungsland entwickelt (Köckeritz, Klinkhammer & von Salisch, 2010). Inzwischen haben 30 von 100 Kindern im Alter unter 6 Jahren einen Migrationshintergrund (Schneider & Becker-Stoll, 2013), für Ballungsgebiete werden sogar über 40% genannt, weswegen eine Überprüfung des Einflusses des Migrationshintergrunds auf die Testleistung relevant erschien. Zu den meist gesprochenen Muttersprachen von Personen mit Migrationshintergrund in Deutschland gehört Türkisch, was mit knapp 16% in der vorliegenden Stichprobe von KMM abgebildet wird, gefolgt von Arabisch mit 14,5%. Die durchschnittliche Leistung im Graphästhesie-Subtest von KMM wie auch von KOM unterschied sich statistisch nicht signifikant, doch sie war geringfügig niedriger als in der TAKIWA-Eichstichprobe. Die mittlere Leistung für 3-Jährige KMM wie auch KOM entsprach der der Eichstichprobe, für 4- sowie für 5-Jährige fiel sie etwas höher aus. Die Differenzierungsfähigkeit des Graphästhesie-Tests für das Lebensalter, abgeleitet aus der deutlichen Beziehung zur chronologischen Entwicklung durch den beständigen Anstieg der Mittelwerte, wurde auch für KMM bestätigt. Eine durchgängig linear ansteigende Schwierigkeit der zehn Aufgaben konnte in der aktuellen Gesamtstichprobe nicht belegt werden. Aufgabenschwierigkeiten ändern sich in Abhängigkeit von der Stichprobe, was die Abweichung zur Erprobungsstichprobe der Testendform ohne KMM im TAKIWA-Testmanual erklären könnte und besonders die Stichprobenabhängigkeit von Ergebnissen nach der klassischen Testtheorie konstruierten Tests unterstreicht. Die niedrigere Rate an Richtiglösungen des ersten Items (Kreis links) bei KMM kann eventuell auf eine noch nicht abgeschlossene Warming-up-Phase zurückgeführt werden (erlebte Diskrepanz eines KMM aufgrund der minimalen sozialen Interaktion des Testleiters zwecks Einhaltung von Durchführungsobjektivität und der folgenden, möglicherweise als überraschend erlebten manuellen Berührung durch ihn). Mit Ausnahme von drei Items in der zweiten Testhälfte, die aufgrund nachlassender Aufmerksamkeit und Motivation und/oder auf Grund ihrer Schwierigkeit ein zufälliges Lösungsmuster zu zeigen scheinen (leichte Items werden selten durch Raten gelöst), waren die Trennschärfen in der aktuellen Gesamtstichprobe akzeptabel. Die interne Konsistenz für die aktuelle Gesamtstichprobe war ausreichend, wenngleich mit .72 niedriger als die im TAKIWA-Testmanual (S. 74) genannte von .81 in der Eichstichprobe. Das resultiert vermutlich aus der Differenz in den Reliabilitätsschätzungen für KMM (.71) und KOM (.80), wobei grundsätzlich – unabhängig von der Variable „Migrationshintergrund“ – auch Raten die Reliabilität der Ergebnisse beeinträchtigt. Um erfolgreich eine graphästhetische Leistung zu erbringen, sind verschiedene kognitive Leistungen erforderlich, was auf theoretischen Überlegungen basierend durch die Korrelationen mit entsprechenden anderen Kriterien (Tests) belegt wird. Es werden moderate bis tendenziell hohe signifikante positive Korrelationen (Validitätskoeffizienten) mit Merkmalen abgebildet, die konstruktnahe Teilfähigkeiten erfassen (FEW-2-Subtest) oder ebenfalls sequenziell verlaufen (K-ABC-Handbewegungen). Niedriger fiel der Zusammenhang mit dem K-ABC-Zahlen Nachspre-

chen aus, weil dieser Subtest sequenzielle Informationsverarbeitung in einer anderen, der auditiven Sinnesmodalität verlangt. Die moderate Korrelation mit dem WET-Subtest „Wörter Erklären“, der sprachbasierte Symbolfähigkeit erfasst, ist ebenfalls ein Hinweis für konvergente Validität. Evidenz für diskriminante Validität wurde erwartungsgemäß durch den fehlenden substanziellen Zusammenhang mit dem kreativen Potenzial eines Kindes belegt, ausgedrückt im Gesamtscore des Tests zum Schöpferischen Denken – Zeichnerisch (Urban & Jellen, 2010). Offensichtlich stellt die feinmotorische Weiter- bzw. Ausgestaltung eines grafischen Elements ein anderes Konstrukt dar als das Konstrukt „Graphästhesie“, welches die passiv-taktile Wahrnehmung eines bekannten grafischen Elements verlangt.

### Schlussfolgerung

Mit der vorliegenden Studie wurden (a) psychometrische Item- und Testkennwerte zum TAKIWA-Subtest „Graphästhesie“ an einer aktuellen Stichprobe von Kindergartenkindern berechnet, die (b) viele Kinder mit Migrationshintergrund umfasste. Die gemäß der klassischen Testtheorie stets stichprobenabhängigen psychometrischen Kennwerte (was ihre Generalisierbarkeit erschwert) unterschieden sich punktuell von den Kennwerten der TAKIWA-Erprobungs- sowie von der TAKIWA-Eichstichprobe. Für KMM waren sieben der zehn Items schwieriger als im Testmanual angegeben, zwei in etwa gleich schwierig und eines leichter. In etwa traf das auch für KOM zu (6 Items schwieriger, eines in etwa gleich schwierig und 2 Items leichter). Die Items der aktuellen Gesamtstichprobe hatten niedrigere, doch ausreichend hohe Trennschärfekoeffizienten. Insgesamt sprechen die Ergebnisse für ein noch akzeptabel reliables und valides Untersuchungsinstrument, das ein Konstrukt misst wie derzeit kein anderes deutschsprachiges psychodiagnostisches Verfahren im jungen Kindesalter, wodurch das Gütekriterium „Nützlichkeit“ gerechtfertigt wird. Allerdings ist zu überlegen, ob nicht das leichte Item 2 (Strich [vertikal rechts]) an den Testbeginn gestellt oder sogar eliminiert werden sollte, da das etwas weniger leichte Item 6 (Strich [vertikal links]) eine deutlich höhere Trennschärfe hat. Insgesamt empfiehlt sich die Abfolge der zehn Items zu verändern, denn auch die Trennschärfe einer Aufgabe kann letztlich von ihrer Stellung beeinflusst werden. Längerfristig ist ggf. sogar eine Überarbeitung der Items in Erwägung zu ziehen.

### Quintessenz

Nunmehr liegen aktualisierte, ergänzende Informationen zum TAKIWA-Subtest „Graphästhesie“ für den Anwender in pädiatrischer Ergotherapie, Neuropädiatrie, Entwicklungspsychologie und Kinderneuropsychologie vor, zu deren diagnostischen Aufgaben auch die Untersuchung und Bewertung von graphästhetischen Fähigkeiten gehört. Hiernach lassen sich die Ergebnisse zum TAKIWA-Graphästhesie-Test auf mehrsprachig aufwachsende Kinder mit Migrationshintergrund generalisieren.

### Quellenverzeichnis

→ Das Quellenverzeichnis zu diesem Beitrag finden Sie unter [www.download.schulz-kirchner.de](http://www.download.schulz-kirchner.de)