

Das KIDS Quiz – Ein Test zur Messung kindlicher Fähigkeiten im Bereich der Stochastik

1. Einleitung

Viele Entscheidungen und Vorhersagen werden aufgrund von statistischen Überlegungen getroffen. Fehlinterpretationen von Daten oder Wahrscheinlichkeiten führen zu falschen Einschätzungen von Situationen oder Gewinnchancen und haben oft weitreichende Folgen. Aufgrund der Flut an Daten und Diagrammen in den Medien sowie auch der mannigfaltigen Wett- und Gewinnspiele wird es für Schülerinnen und Schüler immer wichtiger, Kompetenzen in diesem Bereich aufzubauen. Das Denken in Wahrscheinlichkeiten, das Verstehen von Abbildungen, in welchen Häufigkeiten oder Prozentverteilungen dargestellt werden und die Ausbildung guter Entscheidungsstrategien – das stochastische Grundverständnis – sollte demnach schon im Kindesalter so gut wie möglich geschult werden.

Die Stochastik im Elementarbereich gliedert sich in 3 Teilbereiche: (1) Daten und Häufigkeit, (2) Wahrscheinlichkeit und (3) Kombinatorik (Lehner & Mehlretter, 2009). Der Teilbereich Daten und Häufigkeit ist folgendermaßen definiert: „Daten sammeln, strukturieren und in Tabellen, Schaubildern und Diagrammen darstellen, aus Tabellen, Schaubildern und Diagrammen Informationen entnehmen.“ Der Bereich Wahrscheinlichkeit umfasst „Wahrscheinlichkeiten von Ereignissen in Zufallsexperimenten vergleichen, Grundbegriffe kennen (z. B. sicher, unmöglich, wahrscheinlich) und Gewinnchancen bei einfachen Zufallsexperimenten (z. B. bei Würfelspielen) einschätzen“. Der Bereich Kombinatorik umfasst folgende Kompetenzen: „Sachsituationen mit kombinatorischem Inhalt nachvollziehen, systematisch erschließen und verbalisieren“ (Bildungsstandards im Fach Mathematik, S. 11). Bisher ist uns – obwohl die Stochastik ein nicht zu unterschätzendes Segment der Mathematik ist – kein Instrument bekannt, mit dem die kindlichen Fähigkeiten auf diesem Gebiet verlässlich gemessen werden können. Aufgrund dieser Lücke wurde ein Instrument entwickelt, das die Themenbereiche stochastische Begriffe („möglich“, „wahrscheinlich“, „sicher“ und „unmöglich“), Tabellen, Grafiken, Kombinatorik und Zufallsexperimente enthält. Es wurde für Kinder zwischen 6 und 15 Jahren entwickelt. Im vorliegenden Bericht wird die Testgüte des neu entwickelten Instruments in der Form A analysiert und der Test normiert.

2. Material & Methoden

Stichprobe

Pilotierung: 400 Schülerinnen und Schüler in Österreich, Ungarn, Deutschland und der Schweiz (48.4 % weiblich, Durchschnittsalter 10.9 Jahre, $SD = 2.12$ Jahre) dienten als Pilotstichprobe zur Itemselektion.

Normstichprobe: 974 Schülerinnen und Schüler (49.7 % weiblich, Durchschnittsalter 9.85 Jahre, $SD = 2.49$ Jahre) an 51 Schulen (Volksschulen, Sonderschulen, neue Mittelschulen und Gymnasien in Österreich, Deutschland, der Schweiz und Ungarn) bearbeiteten die Endversion des Tests.

Das KIDS-Quiz

Das Kids Quiz misst die Kenntnisse von Schülerinnen und Schülern der Schulstufen 1 bis 9 im Bereich der Stochastik. Es handelt sich um einen fünfteiligen Test, in dem kindgerechte Aufgaben zu den Themenbereichen Begriffe, Tabellen, Grafiken, Zufallsexperimente und Kombinatorik enthalten sind. Die Zusammenstellung der Themenbereiche orientiert sich an den Bildungsstandards der Kultusministerkonferenz für das Fach Mathematik, die unter anderem den Umgang mit Daten, Häufigkeit und Wahrscheinlichkeit festlegt (Bildungsstandards im Fach Mathematik). Die Aufgaben zu den Themenbereichen wurden alle gleich gewichtet und mit jeweils 12 Punkten im Test berücksichtigt. Pro Aufgabenfeld werden unterschiedlich viele Items vorgegeben, weil diese so gewählt wurden, dass die Bearbeitungszeit für alle Testteile in etwa gleich lang ist. Der Test war zeitlich nicht limitiert, wurde jedoch von fast allen Kindern im Rahmen einer Schulstunde gelöst. Im Folgenden werden die Testteile und die darin enthaltenen Aufgaben kurz beschrieben. Der Test ist weiter hinten in diesem Dokument abgedruckt (Abbildung 1).

Testteil „Begriffe“: Im ersten Teil des Tests finden sich 16 Aufgaben zu den Begriffen „sicher“, „möglich“, „unmöglich“ und „wahrscheinlich“ (Items 1–16). Obwohl der Anwendungsbereich dieser Begriffe durchaus nicht immer eindeutig ist (Patt & Schrag, 2003; Weber & Hilton, 1990), so sind doch die wenigstens die Endpunkte der Skala festgelegt: sicher bzw. unmöglich. Darüber hinaus ist der Komparativ eindeutig (A ist wahrscheinlicher als B wenn mehr Ereignisse vom Typ A als vom Typ B vorliegen). Es wird geprüft, ob ein adäquates Verständnis dieser Begriffe vorliegt. Konkret werden die Kompetenzen „Grundbegriffe bilden

und sichern“ abgefragt (Bildungsstandards im Fach Mathematik, S. 11). Den Kindern werden dazu vier geöffnete Schachteln präsentiert, die mehrere Kugeln unterschiedlicher Farben enthalten. Insgesamt müssen zu jedem der Begriffe vier Fragen beantwortet werden. Die Kinder werden aufgefordert, Aussagen wie „es ist möglich, eine gelbe Kugel zu ziehen“ mit „richtig“ oder „falsch“ zu beantworten. Nur wenn die Kinder alle Fragen zum Begriff richtig beantworten, werden Punkte (3 Punkte pro Begriff) vergeben. Die Ratewahrscheinlichkeit wird somit auf $(\frac{1}{2})^4 = 1/16$ minimiert.

Testteil „Grafiken“: Im zweiten Teil des Tests finden sich 6 Aufgaben zum Thema Grafiken (Items 17–22). Üblicherweise werden Häufigkeitsverteilungen in Form von Balkendiagrammen oder Kuchendiagrammen dargestellt, deshalb wurden diese beiden Darstellungsarten in den Fragenteil integriert. Es geht in diesem Teil um die Kompetenz von Kindern, Informationen aus verschiedenen Darstellungen zu entnehmen, zu verbalisieren und zu interpretieren (Bildungsstandards im Fach Mathematik). Für die Messung des Verständnisses von Balkendiagrammen müssen die Kinder zwei kurze Texte lesen (z. B. „In der Klasse sind 8 Buben und 15 Mädchen“) und dem jeweiligen Text die richtige Grafik aus jeweils 3 Alternativen zuordnen. Für die Messung des Verständnisses von Kuchendiagrammen bekommen die Kinder ein Kuchendiagramm vorgegeben und sollen verschiedene Informationen herauslesen (z. B. „Wie viele Stunden Sport hat die Klasse?“). Hier müssen die Kinder offene Antworten geben. Für die richtige Lösung bei den Balkendiagrammen wurden vier Punkte vergeben. Dies unter der Voraussetzung, dass beide Fragen korrekt gelöst wurden. (Ratewahrscheinlichkeit = $(\frac{1}{3})^2 = 1/9$). Für die richtige Beantwortung der offenen Fragen zum Kuchendiagramm können die Kinder insgesamt 8 Punkte erreichen (2 Punkte pro Aufgabe).

Testteil „Tabellen“: Im dritten Teil des Tests (Items 23–35) finden sich 13 Aufgaben zum Themenbereich Tabellen. Hier wird die Kompetenz, eine Anzahl durch eine Strichliste in einer Tabelle zu erfassen und Daten aus Beobachtungen richtig zu interpretieren, abgefragt (Bildungsstandards im Fach Mathematik). Die Kinder lesen einen kurzen Text und sollen daraus eine Tabelle erstellen. Andererseits bekommen sie eine Tabelle vorgegeben und sollen aus dieser verschiedene Informationen herauslesen (z. B. „Wie viele Radfahrer hat das Kind gesehen?“). Das Antwortformat ist hier offen. Für das Tabellen lesen und das Tabellen interpretieren werden jeweils 6 Punkte vergeben.

Testteil „Zufallsexperimente“: Im vorletzten Teil des Tests (Items 36–39) werden die Kinder mit Zufallsexperimenten konfrontiert (Würfeln, Glücksrad). Hier geht es um die Kompetenz „Gewinnchancen bei einfachen Zufallsexperimenten einschätzen“

(Bildungsstandards im Fach Mathematik, S. 11). Die Kinder sollen bei 4 Fragen vorhersagen, wie ein Zufallsexperiment wahrscheinlich ausgehen wird (z.B. bei der Vorgabe von Glücksrädern: „Es wird schwarz öfter gedreht werden“ oder bei der Vorgabe eines Würfelexperiments „Die ACHT wird öfter gewürfelt werden als die ZEHN“). Auch hier gibt es multiple Choice Format-Fragen („Glücksräder“) und offenes Antwortformat („Würfeln“). Bei den MC-Items zum Glücksrad wird die Ratewahrscheinlichkeit auf $(\frac{1}{3})^3 = (1/27)$ minimiert. Sowohl für die Glücksrad-Aufgabe als auch für die Aufgabe zum Würfeln gibt es 6 Punkte.

Testteil „Kombinatorik“: In diesem Testteil werden die kombinatorischen Probleme „Ziehen mit Zurücklegen“, „Ziehen ohne Zurücklegen“ und „n über k“ in 7 verschiedenen Aufgaben (Items 40–46) behandelt. Geprüft wird hier die Kompetenz, „Sachsituationen mit kombinatorischem Inhalt nachzuvollziehen, systematisch zu erschließen und zu verbalisieren“ (Bildungsstandards im Fach Mathematik, S. 11). Die Kinder sollen Aufgaben lösen wie: „Wie viele 3-stöckige Türme kann ein Kind mit 2 Farben bauen?“ oder: „Wie viele Möglichkeiten haben 3 Tiere, sich in einer Reihe aufzustellen?“ oder: „Wie viele Möglichkeiten gibt es, zwei Eissorten aus vier möglichen auszuwählen wenn keine Sorte doppelt gewählt werden darf?“. Das Antwortformat ist offen. Insgesamt können 12 Punkte erreicht werden.

Im Rahmen des Quiz wurden Geschlecht, Schulstufe, Alter, Note in Mathematik, ob der Schüler das Fach Mathematik gerne mag (dichotom mit ja / nein) und der sozioökonomische Hintergrund abgefragt. Letzterer wurde über die Anzahl der Bücher zu Hause erhoben, da sich in einer Reihe von Studien gezeigt hat, dass der heimische Buchbestand als guter Indikator für den sozio-ökonomischen Status des Elternhauses genutzt werden kann (Rost & Wessel, 1994; Elley, 1994; Lietz, 1996; Schwippert, 2002; Bos et al., 2003). Die jeweiligen Klassenlehrer machten auf fünfstufigen Skalen zusätzlich Angaben zum Interesse an Mathematik, zu den mathematischen Fähigkeiten und zum Leseverständnis des Kindes. Von einem Teil der Kinder (N = 123) lagen Daten aus Intelligenztests (HAWIK IV), welche zur selben Zeit wie das Quiz durchgeführt wurden, vor.

Abbildung 1. Das KIDS Quiz

(farbige Teile wie zum Beispiel die Kugeln aus dem Aufgabenteil Begriffe oder die Legosteine aus dem Aufgabenteil Kombinatorik sind hier in schwarz-weiß wiedergegeben)

KIDS-Quiz

Code:
(2 erste Buchstaben aus dem Vornamen + 2 erste Buchstaben des Familiennamens):

Schule:

Schulstufe:

Bist Du...

- Mädchen
- Bub

Wie alt bist Du?

Wie viele Bücher hast Du zu Hause? (kreuze an!)

- 0 – 10 Bücher
- 11 – 25 Bücher
- 26 – 100 Bücher
- 101 – 200 Bücher
- Mehr als 200 Bücher

Welche Note hattest Du im letzten Zeugnis im Fach Mathematik?

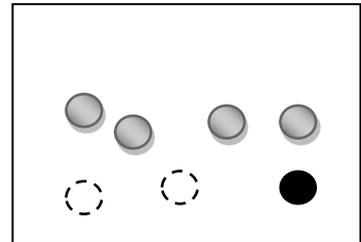
Magst Du Mathematik gerne?

- ja
- nein

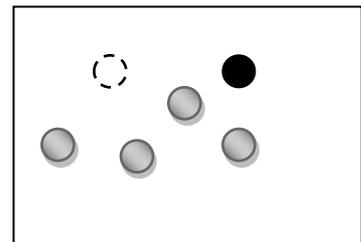
Felix zieht mit verbundenen Augen Kugeln aus einer Kiste.

Kreuze an, welche Aussagen richtig sind!

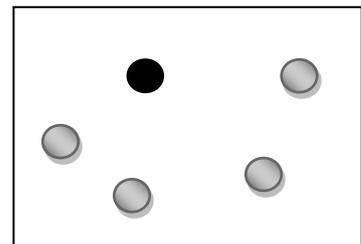
- Es ist möglich, eine rote Kugel zu ziehen!
- Es ist wahrscheinlicher, eine gelbe Kugel zu ziehen als eine rote Kugel zu ziehen!
- Es ist sicher, eine gelbe Kugel zu ziehen!
- Es ist unmöglich, eine schwarze Kugel zu ziehen!



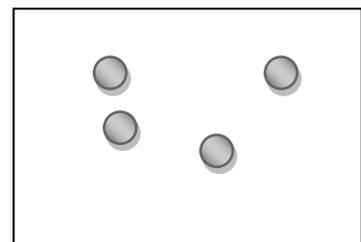
- Es ist wahrscheinlicher, eine rote Kugel zu ziehen als eine schwarze Kugel zu ziehen!
- Es ist sicher, eine gelbe oder eine rote Kugel zu ziehen!
- Es ist unmöglich, eine blaue Kugel zu ziehen!
- Es ist möglich, eine schwarze Kugel zu ziehen!



- Es ist möglich, eine rote Kugel zu ziehen!
- Es ist unmöglich, eine schwarze Kugel zu ziehen!
- Es ist sicher, eine gelbe Kugel zu ziehen!
- Es ist wahrscheinlicher eine gelbe Kugel zu ziehen als eine schwarze Kugel zu ziehen!



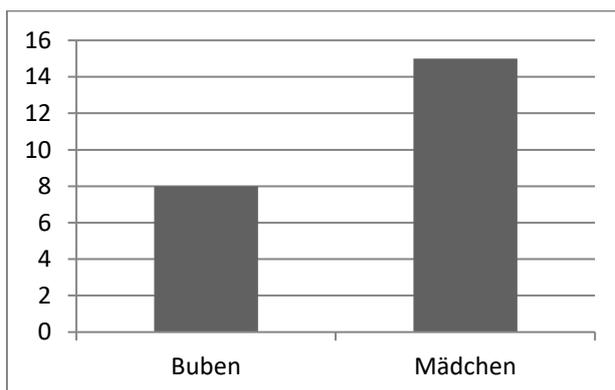
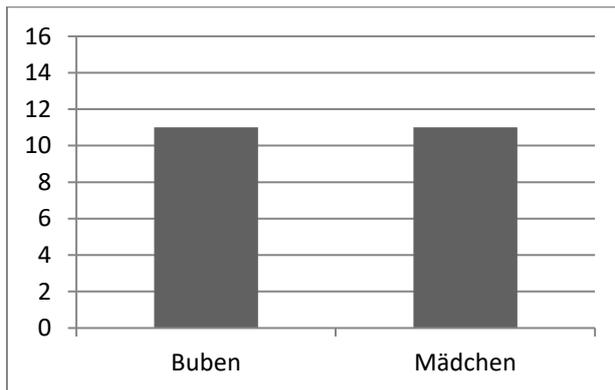
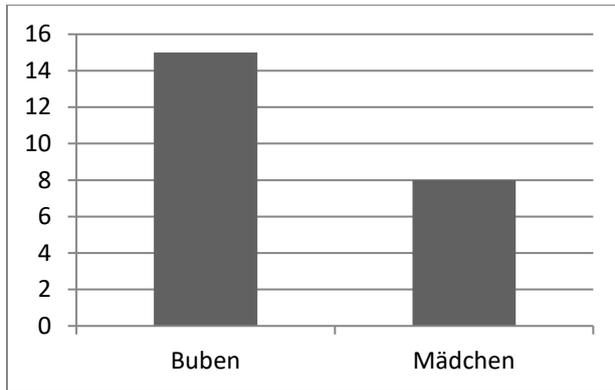
- Es ist möglich, eine rote Kugel zu ziehen
- Es ist sicher, eine gelbe Kugel zu ziehen
- Es ist wahrscheinlicher, eine rote Kugel zu ziehen als eine gelbe Kugel zu ziehen
- Es ist unmöglich, eine gelbe Kugel zu ziehen



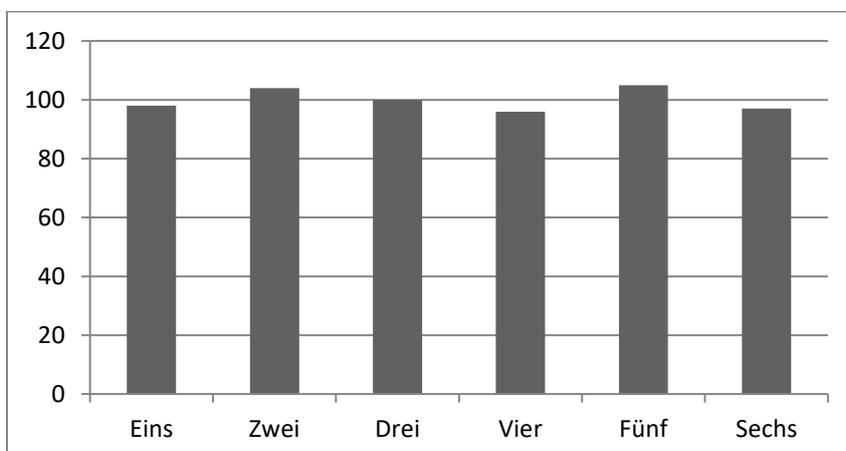
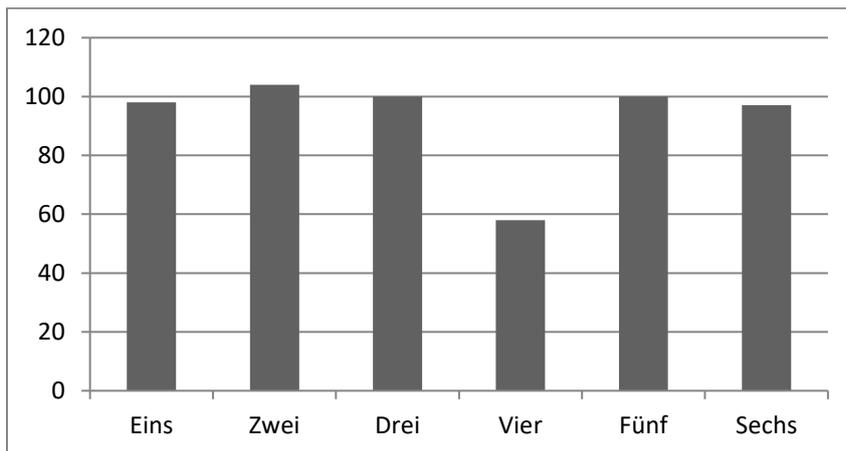
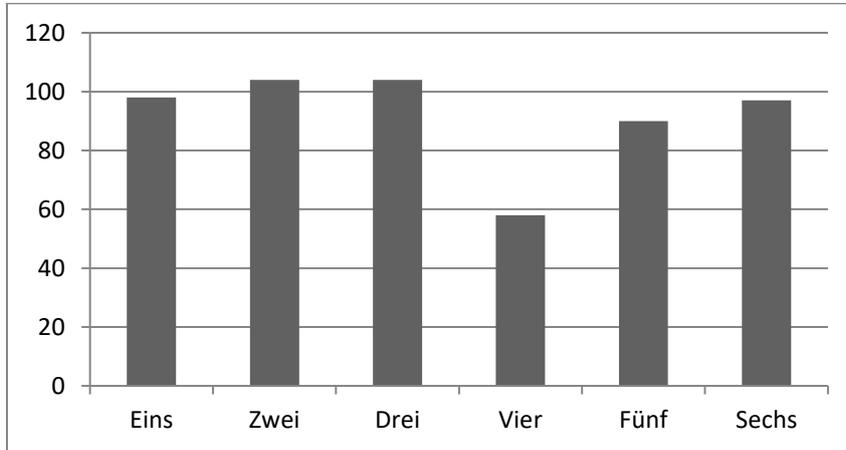
Welche Grafik passt zu den Aussagen?

Mache ein Kreuz neben die Grafik!

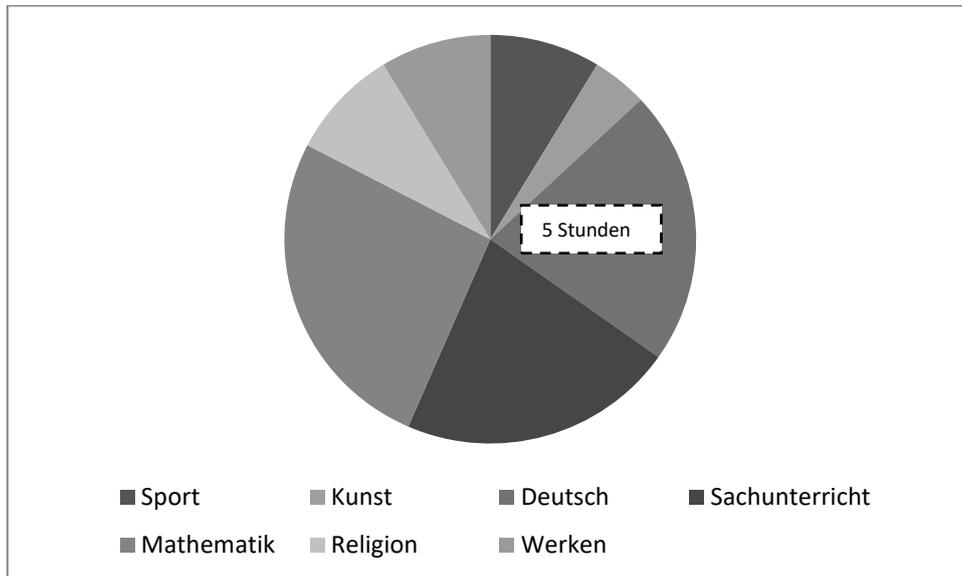
➤ In der Klasse sind 15 Buben und 8 Mädchen



Susi hat 600 mal gewürfelt, sie hat 98 mal eine Eins gewürfelt, 104 mal eine Zwei, 100 mal eine Drei, 96 mal eine Vier, 105 mal eine Fünf und 97 mal eine Sechs!
Mache ein Kreuz neben der Grafik!



Schau dir das Diagramm an, es zeigt die Stunden einer 4. Klasse...



- In welchem Fach hat die Klasse am wenigsten Stunden?
- In welchem Fach am meisten Stunden?
- Wie viele Stunden Sport hat die Klasse?
- Wie viele Stunden hat sie insgesamt in der Woche?

Hans schreibt in seinem Tagebuch:

Am Sonntag habe ich in der Früh den Hasen gefüttert, dann bin ich in den Wald gefahren und habe mich dort mit Franz und Leon getroffen. Wir haben dort bis zum Mittagessen gespielt. Nach dem Mittagessen habe ich zu Hause gelesen und fern gesehen. Vor dem Bettgehen habe ich den Hasen noch einmal gefüttert. Am Montag habe ich Besuch von Mia und Lea bekommen. Nach dem Abendessen habe ich gelesen.

Kannst Du das in dieser Tabelle einzeichnen?

Tätigkeit	Sonntag	Montag	Wie oft?
Lesen	I	I	2
Freunde treffen			
Fernsehen			
Haustier versorgen			
Wald besuchen			

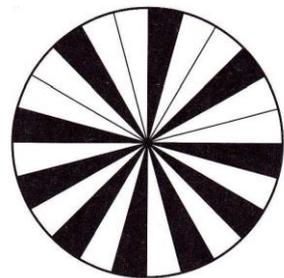
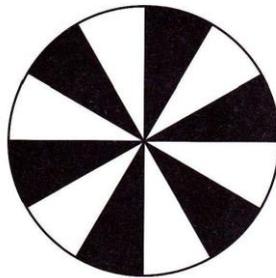
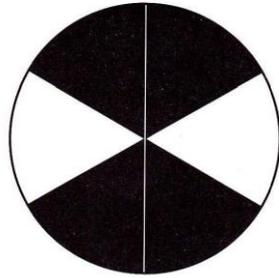
Martin hat die Fahrzeuge vor seinem Haus gezählt und eine Tabelle gemacht.

Schau Dir die Tabelle genau an!

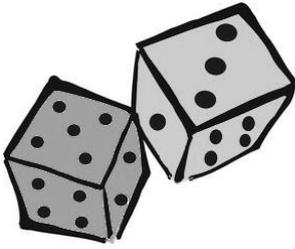
Fahrzeug	Wie viele?
Auto	
Rad	
LKW	
Motorrad	
Bus	
andere	

- Welches Fahrzeug hat er am öftesten gesehen?.....
- Welches am wenigsten oft?
- Wie viele Radfahrer hat er gesehen?

Thomas spielt an einem Glücksrad, kreuze an, was öfter gedreht wird!



Schwarz kommt öfter			
Weiß kommt öfter			
Schwarz und weiß kommen gleich oft			



Lisa und Laura würfeln mit zwei Würfeln. Sie würfeln 100-mal.

Lisa wettet, dass mit **beiden Würfeln** die Augensumme **ZEHN** am öftesten gewürfelt wird.

Laura wettet, dass die Zahl **ACHT** am öftesten gewürfelt wird.

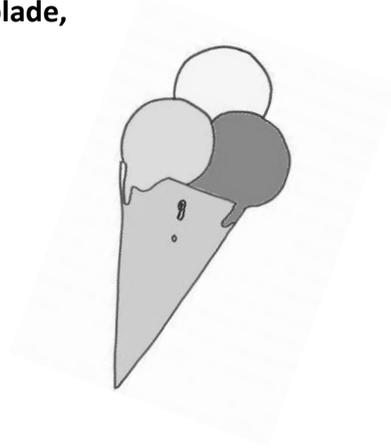
Wer wird wahrscheinlich gewinnen?

- Lisa
- Laura
- kann man nicht sagen

Warum?

Bei einem Eisverkäufer gibt es die Sorten **Vanille, Schokolade, Erdbeere** und **Nuss**.

Du darfst Dir **2 Kugeln** aussuchen, möchtest aber **keine Sorte doppelt** kaufen.



Wie viele Möglichkeiten hast Du?

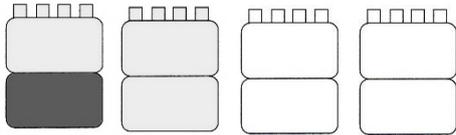
.....
.....

Wie viele Möglichkeiten hättest Du, wenn Du Dir **drei Kugeln** aussuchen dürftest und keine Sorte doppelt kaufen möchtest?

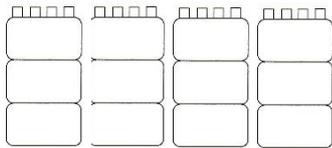
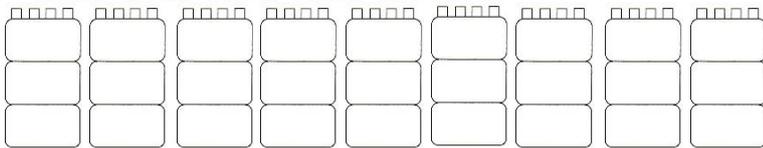
.....
.....
.....

Sabine hat gelbe und rote Legosteine und baut Türme aus 2 Steinen.

Kannst Du ihr alle möglichen Türme aufmalen?



Jetzt baut sie **Türme mit 3 Steinen**, welche Möglichkeiten hat sie? Male sie auf!

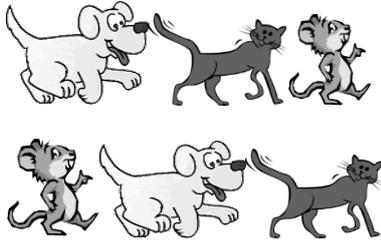


Wie viele Möglichkeiten hätte sie, wenn sie **mit 2 Farben Türme aus 4 Steinen** bauen würde?

Marco feiert seinen Geburtstag mit 2 Freunden, jedes der 3 Kinder hat ein Glas.

Wie oft macht es „kling“, wenn die Kinder miteinander anstoßen?

3 Tiere (Hund, Katze und Maus) gehen in einer Reihe, wie viele Möglichkeiten haben sie, sich in einer Reihe aufzustellen?



Die 3 Tiere haben ____ Möglichkeiten, sich in einer Reihe aufzustellen.

Wie viele Möglichkeiten hätten 4 Tiere, sich in einer Reihe aufzustellen?

4 Tiere haben ____ Möglichkeiten, sich in einer Reihe aufzustellen.

Beurteilung Mathematische Fähigkeiten (Lehrer/in)

1 – 2 – 3 – 4 – 5

[1 = sehr gut → 5 = ungenügend]

Beurteilung Mathematisches Interesse (Lehrer/in)

1 – 2 – 3 – 4 – 5

[1 = sehr gut → 5 = ungenügend]

Beurteilung Leseverständnis (Lehrer/in)

1 – 2 – 3 – 4 – 5

[1 = sehr gut → 5 = ungenügend]

3. Testentwicklung

Im Frühling 2012 wurde ein Pilot an 400 Schülerinnen und Schülern zwischen 7 und 14 Jahren in 6 Schulen aus Österreich, Deutschland und der Schweiz durchgeführt. Vorgegeben wurden jedem Schüler 40 Items aus einem Itempool mit insgesamt 250 Items zu den Bereichen Begriffe, Tabellen, Grafiken, Kombinatorik und Zufallsexperimente. Die Ideen für die 250 Aufgaben entstammen verschiedenen Lehrbüchern (Lehner & Mehlretter, 2009; Bettner & Dinges, 2010; Bettner & Dinges, 2011; Behring, 2011). Auf Basis der Pilotierung wurde die Itemselektion durchgeführt. Kriterien für die Itemauswahl war eine Schwierigkeit zwischen .05 und .95 und darauf folgend eine möglichst gute Trennschärfe von $> .30$. Zusätzlich wurden konfirmatorische Faktorenanalysen (CFA) berechnet um die Zuordenbarkeit der Items zu den 5 Konstrukten zu sichern. Ende 2012 und im ersten Halbjahr 2013 bearbeiteten 974 Schülerinnen und Schüler (49.7 % weiblich) an 51 Schulen (Volksschulen, Sonderschulen, neue Mittelschulen und Gymnasien in Österreich, Deutschland, der Schweiz und Ungarn) im Rahmen eines Projekts der Pädagogischen Hochschule Salzburg die Endversion des Tests.

Um die Retestrelabilität messen zu können wurde ein Teil der Kinder der Schulstufen 1 bis 8 aus 16 Schulen nach 3 Monaten ohne dazwischenliegende Intervention erneut mit demselben Test getestet ($N = 155$, Mittleres Alter = 9.81 Jahre, $SD = 2.44$ Jahre). Ein anderer Teil der Kinder ($N = 195$) löste beide Formen (Form A und Form B) des Tests um die Paralleltestrelabilität feststellen zu können.

4. Ermittlung der Testwerte

Zur Ermittlung der Testrohwerte wurde eine Syntax entwickelt, die die Ratewahrscheinlichkeit in allen Bereichen minimiert. Es gibt bei den Multiple Choice Aufgaben nur Punkte, wenn alle Aufgaben vom selben Typ richtig gelöst wurden. Wenn ein Kind alle vier Aufgaben zum Begriff „möglich“ korrekt bearbeitet ist die Wertung für den Begriffsbereich „möglich“ positiv (damit verringert sich die Ratewahrscheinlichkeit von $\frac{1}{2}$ auf $(\frac{1}{2})^4 = 1/16$). Bei den Balkendiagrammen aus dem Bereich Grafiken bekommen die Kinder nur Punkte, wenn beide Aufgaben dazu gelöst wurden. Die Ratewahrscheinlichkeit reduziert sich damit auf $(\frac{1}{3})^2 = 1/9$. Die offenen Fragen wurden mit Punkten bewertet, wenn sie richtig gelöst wurden. Insgesamt konnten die Kinder 60 Punkte erreichen, jeweils 12 für jeden Subtest.

5. Analysen

Um die Zuverlässigkeit des Tests zu prüfen wurden Item- und Skalenanalysen durchgeführt und Reliabilitätskoeffizienten berechnet. Einerseits wurde ein Reliabilitätskoeffizient für den gesamten Test produziert, andererseits wurde die Reliabilität der 5 Subskalen Begriffe, Tabellen, Grafiken, Kombinatorik und Zufallsexperimente geprüft. Berechnet wurden einerseits Cronbachs Alpha (α) als gebräuchlichster Reliabilitätskoeffizient (Janssen & Laatz, 2010; Cortina, 1993; Cronbach, 1951; Hogan et al., 2000) andererseits die Split-Half Reliabilität (rtt), bei welcher die Skala in 2 Hälften geteilt wird und diese miteinander korreliert werden. Dieser Koeffizient wurde aufgrund der naturgemäßen Unterschätzung der Zuverlässigkeit dadurch, dass jeweils nur eine Testhälfte in die Berechnungen miteingeht Spearman-Brown korrigiert (Janssen & Laatz, 2010).

Die Beurteilung der faktoriellen Validität des KIDS Quiz erfolgte mittels der sich ergebenden Kennwerte aus der explorativen Faktorenanalyse (EFA) und der konfirmatorischen Faktorenanalyse (CFA). Berechnet wurde explorativ eine Hauptkomponentenanalyse, konfirmatorisch wurde das klassische Modell zugrundegelegt (Cattell & Tsujioka, 1964; Grice, 2001; Raykov & ShROUT, 2002). Mithilfe der EFA wurde der Kaiser-Meyer-Olkin-Koeffizient, der Bartlett-Test auf Spharizität und die Faktorladungen berechnet. Mithilfe der konfirmatorischen Faktorenanalyse wurden unter anderem die robusten und stichprobenunabhängigen Fit-Indizes CFI (Comparative-Fit-Index), RMSEA (Root-Mean-Square-Error of Approximation) und SRMR (Standardized-Root-Mean-Residual) berechnet.

6. Ergebnisse

Es gibt in der Testleistung keinen signifikanten Unterschied zwischen Schülerinnen und Schülern ($t(714) = .39, p = .70$), weshalb die Ergebnisse nicht nach Geschlecht differenziert dargestellt werden. Weil die Testleistung signifikant mit der Schulstufe korreliert ($r = .45, p < .001$), werden Gütekriterien und Normen getrennt für die Schulstufen (es wurden aufgrund der Datenlage jeweils zwei Klassenstufen zusammengefasst) dargestellt.

Testgütekriterien

Objektivität: Die Objektivität ist durch die Standardisierung des Verfahrens, d.h. durch die strukturierte Vorgangsweise in der Anwendung des Tests gegeben. Die Durchführungsobjektivität kann durch genaue Angaben zur Vorbereitung, zum Wortlaut der

Anweisungen im Testmanual und zu den Durchführungszeiten als gesichert angesehen werden, die Auswertungsobjektivität lässt sich durch die Einschätzung der Testleistung in den Normtabellen gewährleisten.

Reliabilität:

Interne Konsistenz und Split-Half-Reliabilität des KIDS-Quiz:

Die Reliabilität des Tests ist gegeben: Cronbachs α ist hinreichend hoch und beträgt .88, die Split-Half-Reliabilität beträgt $r_{tt} = .88$. Sieht man sich die Zuverlässigkeit des Tests getrennt für die Schulstufen an, so ergeben sich auch hier hinreichend hohe α -Koeffizienten zwischen $\alpha = .83$ und $\alpha = .84$ (siehe Tabelle 2). Sieht man sich die Reliabilitäten für die 5 Testteile getrennt für die Gesamtstichprobe an (Tabelle 3), so ergeben sich Reliabilitäten zwischen .56 (Grafiken) und .92 (Tabellen). Die Testgüte ist daher für den Bereich Grafiken eher schwach, für die anderen Bereiche ist sie hinreichend (Kombinatorik, Zufallsexperimente, Begriffe) bzw. sehr gut (Tabellen).

Tabelle 1. Interne Konsistenz α und Split-Half-Reliabilität r_{tt} des KIDS-Quiz

Schulstufe	# Items	α	r_{tt}
SS 1–2	46	.837	.654
SS 3–4	46	.823	.689
SS 5–6	46	.838	.730
SS 7–9	46	.829	.752
Gesamt	46	.876	.774

Tabelle 2. Interne Konsistenz Cronbachs Alpha (α) des KIDS-Quiz getrennt für die 5 Testteile

Begriffe			Grafiken			Tabellen		
Schulstufe	# Items	α	Schulstufe	# Items	α	Schulstufe	# Items	α
SS 1–2	16	.533	SS 1–2	6	.530	SS 1–2	13	.948
SS 3–4	16	.696	SS 3–4	6	.375	SS 3–4	13	.846
SS 5–6	16	.671	SS 5–6	6	.606	SS 5–6	13	.811
SS 7–9	16	.737	SS 7–9	6	.484	SS 7–9	13	.924
Gesamt	16	.702	Gesamt	6	.557	Gesamt	13	.923

Zufall			Kombinatorik		
Schulstufe	# Items	α	Schulstufe	# Items	α
SS 1–2	4	.692	SS 1–2	7	.381
SS 3–4	4	.699	SS 3–4	7	.579
SS 5–6	4	.771	SS 5–6	7	.617
SS 7–9	4	.569	SS 7–9	7	.679
Gesamt	4	.715	Gesamt	7	.614

Schwierigkeit der KIDS-Items:

Die Schwierigkeit (Lösungshäufigkeit) der Testitems bewegt sich für die Gesamtstichprobe zwischen .08 („Wie viele Möglichkeiten gibt es, 4 Tiere in einer Reihe aufzustellen“) und .95 („Glücksrad 2“) (siehe Tabelle 3). Die mittlere Schwierigkeit beträgt .71. Tabellarisch dargestellt sind die Schwierigkeiten der Items für die Gesamtstichprobe und für die einzelnen Altersgruppen in Tabelle 3. Es zeigt sich, dass die Kombinatorik (Items 40–46; Ausnahme: Item 45) für jüngere Kinder noch sehr schwer zu lösen sind, was gut zu Befunden passt, wonach sich bestimmte Problemlösefähigkeiten erst im Alter von 9 bis 11 Jahren entwickeln (Rasch, 2001).

Trennschärfen der KIDS-Items:

Die Trennschärfen der Testitems aus der Endversion bewegen sich für die Gesamtstichprobe zwischen .10 (Item 9) und .52 (Item 27; Tabelle 4). Sie differenzieren demnach schwach bis sehr gut zwischen Schülerinnen mit hoher und niedriger Merkmalsausprägung (hier Fähigkeit auf dem Gebiet der Stochastik). Dass die Trennschärfen insgesamt eher im niedrigen Bereich angesiedelt sind ist zum Teil auch dem Faktum geschuldet, dass sehr schwere oder sehr leichte Items naturgemäß geringere Trennschärfen haben als Items im mittleren Schwierigkeitsbereich (Lienert, 1989).

Tabelle 3. Trennschärfe und Schwierigkeit der Aufgaben des KIDS Quiz

Item	Trennschärfe	Schwierigkeit				
	Gesamt	Gesamt	SS 1–2	SS 3–4	SS 5–6	SS 7–9
1	.35	.73	.68	.70	.76	.87
2	.39	.74	.65	.74	.68	.91
3	.31	.72	.65	.67	.87	.85
4	.27	.84	.75	.84	.89	.96
5	.31	.79	.62	.82	.79	.99
6	.30	.66	.58	.64	.58	.84
7	.50	.59	.40	.64	.45	.85
8	.25	.63	.64	.61	.34	.77
9	.10	.91	.90	.90	.95	.95
10	.31	.87	.75	.89	.95	.96
11	.26	.65	.59	.59	.74	.84
12	.38	.75	.66	.76	.61	.94
13	.12	.94	.94	.93	.95	.97
14	.19	.95	.91	.97	.95	.99
15	.25	.90	.83	.92	.89	.99
16	.25	.88	.87	.87	.74	.99
17	.22	.90	.82	.90	.95	.99
18	.43	.78	.59	.84	.68	.94
19	.33	.91	.83	.91	1.00	1.00
20	.38	.77	.69	.71	.82	1.00
21	.32	.47	.14	.59	.50	.67
22	.34	.16	.04	.15	.13	.38
23	.46	.90	.71	.96	1.00	1.00
24	.46	.91	.75	.96	.97	1.00
25	.48	.77	.69	.74	.87	.90
26	.47	.71	.62	.67	.79	.91

27	.52	.85	.69	.87	.95	.97
28	.43	.79	.69	.78	.84	.96
29	.48	.86	.76	.86	.95	.97
30	.44	.83	.75	.80	.92	.99
31	.49	.60	.46	.57	.61	.89
32	.46	.71	.61	.68	.74	.92
33	.33	.85	.72	.86	.87	1.00
34	.42	.84	.68	.84	.95	1.00
35	.33	.78	.64	.83	.82	.90
36	.32	.85	.79	.85	.82	.96
37	.15	.95	.94	.95	.89	.99
38	.29	.88	.83	.89	.82	.97
39	.17	.85	.85	.80	.95	.91
40	.47	.37	.14	.39	.21	.77
41	.41	.35	.27	.29	.39	.57
42	.34	.27	.15	.28	.26	.46
43	.14	.30	.31	.39	.21	.15
44	.31	.08	0.00	.05	.13	.23
45	.36	.56	.47	.52	.53	.82
46	.35	.08	0.00	.06	.05	.25

Retestreliaibilität/Paralleltestreliaibilität:

Die Retestreliaibilität liegt bei einer Testung nach 3 Monaten bei $r = .739$ (Gesamtscore). Die Paralleltestreliaibilität beträgt $.870$.

Validität:

Inhaltliche Validität:

Die Stochastik im Elementarbereich gliedert sich – wie in der Einleitung erwähnt - in die Teilbereiche (1) Daten und Häufigkeit, (2) Wahrscheinlichkeit und (3) Kombinatorik. Die Zusammenstellung der Themenbereiche im Quiz orientiert sich an den Bildungsstandards der Kultusministerkonferenz für das Fach Mathematik, die den Umgang mit Daten, Häufigkeit und Wahrscheinlichkeit festlegt. Die Aufgaben können demnach als unmittelbarer Indikator für die stochastischen Fähigkeiten angesehen werden. Die Inhaltsvalidität gilt damit als gesichert.

Konstruktvalidität/Kriteriumsvalidität:

Zu Konstruktvalidität und Kriteriumsvalidität können noch keine Angaben gemacht werden da es bislang kein Instrument gibt, das die kindlichen Fähigkeiten im Bereich Stochastik testet. Hohe Zusammenhänge ($r > .56$) ergaben sich zwischen den Testleistungen im Bereich „Kombinatorik“ und den Leistungen im HAWIK IV. Der Zusammenhang mit der Mathematiknote beträgt $r = -.15$ ($p < .001$) und ist in den höheren Schulstufen deutlich höher als in den Schulstufen 1 bis 4. Interessant ist der verhältnismäßig hohe Zusammenhang zwischen den Leistungen in Kombinatorik und dem wahrnehmungsgebundenen logischen Denken aus dem HAWIK IV ($r = .79$).

Faktorielle Validität:

Gesamtstichprobe unter Annahme eines Ein-Faktorenmodells betrug 20 %, unter Annahme von 5 Faktoren (Begriffe, Kombinatorik, Zufallsexperimente, Tabellen und Grafiken) betrug sie 36 %. Die Faktorladungen für drei Aufgaben aus dem Bereich Begriffe und für eine Aufgabe aus dem Kombinatorik-Teil (Ziehen mit Zurücklegen) waren kleiner .20 und damit zu gering. Für die anderen Items waren die Ladungen hinreichend hoch. Sieht man sich die Faktorladungen nach Bildung der Summen für die einzelnen Subskalen an, so ergeben sich für alle 5 Subskalen gute bis sehr gute Faktorladungen zwischen .48 und .73 (Tabelle 5).

Tabelle 4. Faktorladungen (EFA) nach Itemgruppen

Itemgruppe	Faktorladung
Begriffe	.699
Grafiken	.731
Tabellen	.649
Zufallsexperimente	.484
Kombinatorik	.674

Die CFA wurde sowohl für eine Ein-Faktoren-Lösung als auch für die 5-Faktor-Lösung berechnet. In Tabelle 6 finden sich der Comparative Fit Index (CFI), der Normed Fit Index (NFI), der Root Mean Square Error of Approximation (RMSEA) sowie der Standardized Root Mean Square Residual (SRMR) für das Ein-Faktor-Modell (alle Subtests laden auf einen Faktor „Fähigkeiten auf dem Gebiet der Stochastik“) und für ein 5-Faktoren-Modell mit den 5 Faktoren Begriffe, Grafiken, Tabellen, Zufallsexperimente und Kombinatorik. In der 1-

Faktoren-Lösung zeigt sich mit .96 für die Goodness-of-Fit-Indizes CFI und NFI eine ausgezeichnete Passung des Modells (Hu & Bentler, 1999) und auch der Standardized Root Mean Square Residual (SRMR) als Badness-of-Fit Index ist mit .039 ausreichend klein (der Schwellenwert beträgt hier .11; Hu & Bentler, 1999). Der Root Mean Square Error of Approximation (RMSEA) übersteigt mit .12 (Konfidenzintervall 0.096 bis 0.14) den Schwellenwert von .06 ist aber trotzdem im annehmbaren Bereich (Hu & Bentler, 1999). Bei der 5-Faktoren-Lösung ergibt sich mit CFI = .85 ein kleinerer Fit als bei der Ein-Faktoren-Lösung, der SRMR liegt auch hier mit .077 unter dem Schwellenwert und gilt damit als sehr gut, der RMSEA als zweiter Badness-of-Fit-Index ist mit .09 zwar nahe an Null, untersteigt aber den Schwellenwert (Hu & Bentler, 1999) nicht.

Tabelle 5. Fit-Indizes aus der CFA

	1 Faktor	5 Faktor	Differenz
CFI	.96	.85	-0.11
NFI	.96	.83	-0.13
RMSEA	.12	.09	-0.03
SRMR	.039	.077	+0.04

Normierung

Von der Normstichprobe, die die Endversion des Tests bearbeitete (N = 974), liegen die Ergebnisse des Tests in Form von Summenwerten vor. Diese Summenwerte („Rohwert-Summen“) wurden auf Normalverteilung geprüft und in z-Werte und T-Werte umgerechnet und sind untenstehend tabelliert.

Tabelle 6. Normen KIDS Quiz (Form A): z-Werte und T-Werte

Schulstufe	Punkte	z-Score	T-Wert
SS 1-2	2	-1.72	33
SS 1-2	4	-1.5	35
SS 1-2	5	-1.39	36
SS 1-2	6	-1.28	37
SS 1-2	7	-1.17	38
SS 1-2	8	-1.06	39
SS 1-2	9	-0.94	41
SS 1-2	10	-0.83	42
SS 1-2	11	-0.72	43
SS 1-2	12	-0.61	44

SS 1-2	13	-0.5	45
SS 1-2	14	-0.39	46
SS 1-2	15	-0.28	47
SS 1-2	16	-0.17	48
SS 1-2	17	-0.06	49
SS 1-2	18	0.06	51
SS 1-2	19	0.17	52
SS 1-2	20	0.28	53
SS 1-2	21	0.39	54
SS 1-2	22	0.5	55
SS 1-2	23	0.61	56
SS 1-2	24	0.72	57
SS 1-2	25	0.83	58
SS 1-2	26	0.95	59
SS 1-2	27	1.06	61
SS 1-2	28	1.17	62
SS 1-2	29	1.28	63
SS 1-2	30	1.39	64
SS 1-2	31	1.5	65
SS 1-2	32	1.61	66
SS 1-2	34	1.83	68
SS 1-2	36	2.06	71
SS 1-2	37	2.17	72
SS 1-2	39	2.39	74
SS 3-4	4	-1.99	30
SS 3-4	5	-1.9	31
SS 3-4	6	-1.8	32
SS 3-4	7	-1.71	33
SS 3-4	8	-1.61	34
SS 3-4	9	-1.52	35
SS 3-4	11	-1.33	37
SS 3-4	12	-1.24	38
SS 3-4	13	-1.14	39
SS 3-4	14	-1.05	40
SS 3-4	15	-0.96	40
SS 3-4	16	-0.86	41
SS 3-4	17	-0.77	42
SS 3-4	18	-0.67	43
SS 3-4	19	-0.58	44
SS 3-4	20	-0.48	45
SS 3-4	21	-0.39	46
SS 3-4	22	-0.3	47
SS 3-4	23	-0.2	48
SS 3-4	24	-0.11	49
SS 3-4	25	-0.01	50
SS 3-4	26	0.08	51
SS 3-4	27	0.18	52
SS 3-4	28	0.27	53
SS 3-4	29	0.36	54
SS 3-4	30	0.46	55
SS 3-4	31	0.55	56
SS 3-4	32	0.65	56
SS 3-4	33	0.74	57
SS 3-4	34	0.83	58
SS 3-4	35	0.93	59

SS 3-4	36	1.02	60
SS 3-4	37	1.12	61
SS 3-4	38	1.21	62
SS 3-4	39	1.31	63
SS 3-4	40	1.4	64
SS 3-4	41	1.49	65
SS 3-4	43	1.68	67
SS 3-4	44	1.78	68
SS 3-4	45	1.87	69
SS 3-4	46	1.96	70
SS 3-4	50	2.34	73
SS 3-4	52	2.53	75
SS 3-4	54	2.72	77
SS 3-4	57	3	80
SS 5-6	0	-2.11	29
SS 5-6	4	-1.76	32
SS 5-6	5	-1.67	33
SS 5-6	13	-0.97	40
SS 5-6	14	-0.88	41
SS 5-6	15	-0.8	42
SS 5-6	16	-0.71	43
SS 5-6	17	-0.62	44
SS 5-6	18	-0.53	45
SS 5-6	19	-0.44	46
SS 5-6	21	-0.27	47
SS 5-6	22	-0.18	48
SS 5-6	23	-0.09	49
SS 5-6	25	0.08	51
SS 5-6	27	0.26	53
SS 5-6	28	0.35	53
SS 5-6	34	0.88	59
SS 5-6	36	1.05	61
SS 5-6	38	1.23	62
SS 5-6	39	1.32	63
SS 5-6	40	1.4	64
SS 5-6	46	1.93	69
SS 5-6	51	2.37	74
SS 5-6	52	2.46	75

SS 7-9	0	-2.93	21
SS 7-9	8	-2.29	27
SS 7-9	10	-2.13	29
SS 7-9	11	-2.05	30
SS 7-9	12	-1.97	30
SS 7-9	13	-1.89	31
SS 7-9	16	-1.65	34
SS 7-9	17	-1.57	34
SS 7-9	18	-1.49	35
SS 7-9	24	-1.01	40
SS 7-9	25	-0.93	41
SS 7-9	26	-0.85	41
SS 7-9	28	-0.69	43
SS 7-9	29	-0.61	44
SS 7-9	30	-0.53	45
SS 7-9	31	-0.45	45
SS 7-9	32	-0.37	46
SS 7-9	34	-0.21	48
SS 7-9	35	-0.13	49
SS 7-9	36	-0.05	49
SS 7-9	37	0.03	50
SS 7-9	38	0.11	51
SS 7-9	39	0.19	52
SS 7-9	40	0.27	53
SS 7-9	41	0.35	53
SS 7-9	42	0.43	54
SS 7-9	43	0.51	55
SS 7-9	44	0.59	56
SS 7-9	45	0.67	57
SS 7-9	46	0.75	57
SS 7-9	47	0.83	58
SS 7-9	48	0.91	59
SS 7-9	49	0.99	60
SS 7-9	50	1.07	61
SS 7-9	52	1.23	62
SS 7-9	54	1.39	64
SS 7-9	56	1.55	65

7. Diskussion

Wie genau können mit diesem Verfahren Fähigkeiten im Bereich der Stochastik gemessen werden? Die wichtigsten empirischen Antworten aus dieser Studie sind: (1) Das KIDS-Quiz wies (für unsere Stichprobe) in allen Altersgruppen zufrieden stellende Gesamtreliabilitäten von $> .80$ auf. (2) Die Konstruktreliabilitäten waren, mit Ausnahme der Skala Grafiken hinreichend (Kombinatorik, Zufallsexperimente, Begriffe) bzw. sehr gut (Tabellen). (3) Die Validität des Tests kann als gesichert angesehen werden. Ausnahme sind einige wenige Items aus dem Themenbereich Begriffe, bei welchen die Faktorladungen sehr gering sind. (4) Die

Testitems sind – bis auf einige wenige Ausnahmen bei den sehr schweren und sehr leichten Items – trennscharf.

Die verhältnismäßig schwache Reliabilität für den Bereich Grafiken lässt sich durch die Art der zweiten Fragestellung um Themenbereich „Balkendiagramme“ erklären. Da viele der getesteten Kinder im Zahlenraum bis 100 noch nicht ausreichend geschult waren (Grundschulalter) und daher zwar die erste Aufgabe zu den Balkendiagrammen und das Kuchendiagramm problemlos lösen konnten, die 2. Aufgabe dazu aber – aufgrund des noch nicht ausreichend geschulten Zahlenraums – nicht, ergab sich hier eine schwache Reliabilität.

Zu niedrige Faktorladungen ergaben sich bei drei Aufgaben aus dem Bereich Begriffe und bei einer Aufgabe aus dem Kombinatorik-Teil (Ziehen mit Zurücklegen). Dies lässt sich damit erklären, dass die Kenntnis der Begriffe nicht mit der eigentlichen Fähigkeit im Bereich der Stochastik gleichgesetzt werden kann (obwohl selbstverständlich Teil des Bereichs Wahrscheinlichkeit) sondern vielmehr mit dem Sprachverständnis zusammenhängt. Hier ergeben sich naturgemäß auch keine signifikanten Korrelationen mit Reasoning-Fähigkeiten. Die Aufgabe zum Ziehen mit Zurücklegen war für alle getesteten Kinder schwer und wurde nur von einem kleinen Prozentsatz richtig gelöst, weshalb hier möglicherweise die geringere Faktorladung zu verzeichnen war. Die Analysen legen nahe, dass der Test die Fähigkeiten im Bereich der Stochastik verlässlich messen kann und dass die einzelnen Subskalen durchaus geeignet sind, Vorhersagen über die kognitiven Fähigkeiten zu machen.