

**Entwicklung eines digitalen Rechtschreibtests für die erste Klasse Grundschule:
Dimensionalität und Reliabilität des SCHNAPP-Rechtschreibtests**

**Development of a digital spelling test for first grade primary school: dimensionality and
reliability of the SCHNAPP spelling test**

M. Schöfl^{1,2}, G. Steinmair¹, S. Zepnik¹, T. Gnambs³

1 Pädagogische Hochschule Oberösterreich, Linz, Österreich

*2 Forschungsinstitut für Entwicklungsmedizin (RID), Johannes Kepler Universität, Linz,
Österreich*

3 Leibniz Institut für Bildungsverläufe, Bamberg, Deutschland

Martin Schöfl
Kaplanhofstraße 40
4020 Linz, Österreich
0680/1440524

Martin.Schoefl@ph-ooe.at

[Martin Schöfl](https://orcid.org/0000-0002-7254-2621)  <https://orcid.org/0000-0002-7254-2621>

Dieses Manuskript ist ein Postprint des hier zur Verfügung stehenden Artikels:

<https://doi.org/10.1024/2235-0977/a000404>

Zusammenfassung

Hintergrund: Die Rechtschreibleistung am Ende der ersten Klasse sagt die weitere Rechtschreibentwicklung voraus. Um Schwierigkeiten im Rechtschreiben früh zu erkennen und Interventionen ansetzen zu können, braucht es ein Messinstrument, das schriftsystematische Prinzipien abbildet. Der neue digitale „SCHNAPP-Rechtschreibtest“ basiert auf Wortmaterial, das vom prototypischen trochäischen Zweisilber ausgeht und eine Hierarchie an Itemschwierigkeiten aufzeigt.

Methoden: Es wurde die Rasch-Konformität und Reliabilität des aus der Literatur deduzierten Schreibwortschatzes anhand zweier Stichproben ($N = 249$ und 148) von Grundschulkindern der ersten Klasse in Österreich überprüft.

Ergebnisse: Der SCHNAPP-Rechtschreibtest erfasst einen latenten Faktor der Rechtschreibfähigkeit mit Rasch-konformen Items unterschiedlicher Schwierigkeit aber identer Trennschärfen. Die Reliabilität konnte durch Itemselektion und Ergänzungen ausgehend von einer Erstversion hin zu einer revidierten Form verbessert werden. Die Items differenzieren insbesondere auch im niedrigen Fähigkeitsbereich gut und eignen sich damit als Screening zur Identifizierung von Kindern mit Schwierigkeiten im Rechtschreiben im ersten Grundschuljahr.

Schlüsselwörter: Rechtschreiben, SCHNAPP-Rechtschreibtest, erste Grundschulklasse, Rasch-Modellierung

Abstract

Background: Spelling performance at the end of first grade predicts further spelling development. To identify children with difficulties in spelling at an early stage and to be able to initiate interventions, an analysis procedure is needed that maps spelling-systematic principles. The new digital "SCHNAPP Spelling Test" is based on word material that starts from the prototypical trochaic two-syllable and thus shows a hierarchy of item difficulties.

Methods: We examined the Rasch model fit and reliability of the writing vocabulary by testing two samples ($n = 249$ and 148) of first graders in Austria.

Results: The SCHNAPP spelling test captures a single latent dimension of spelling abilities with Rasch-conforming items of different difficulties but identical discriminations.

Throughout test development, reliability was improved by item selection and amendments.

Importantly, the test items differentiate well in the low ability range and thus, are suitable for screening children with spelling difficulties in their first year of primary school.

Keywords: Spelling, SCHNAPP-spelling test, first elementary school grade, Rasch modeling

Einleitung

Ein Kind im Laufe der ersten Klasse Grundschule, 7 Jahre, versucht sich gegen Schuljahresende an der Schreibung eines Diktatwortes: „Kater“. Zur Lösung spricht es sich das Wort besonders deutlich vor und kommt zu folgenden Schreibweisen: „Kata“, „Katar“. Dieses Beispiel verdeutlicht, dass die reine „Hinhörschreibung“ im Deutschen fehleranfällig und daher unzureichend ist. Sie trägt nicht zum Aufbau eines systematischen Strukturverständnisses der deutschen Sprache bei. Um dieses Problem zu lösen, sind zwei Strategien denkbar. Zum einen könnte das Kind Hinweise erhalten, um das Wort mit seinen orthographischen Besonderheiten (hier: vokalisiertes <r> und getilgtes <e> in der Reduktionssilbe) auswendig zu lernen. Wird diese Vorgehensweise häufig angewendet, würde das Kind bald an die Grenzen seiner Gedächtniskapazität stoßen. Zum anderen könnte es die Systematik, die nicht nur für dieses Wort, sondern für eine Vielzahl von nativen Schreibungen im Deutschen gilt, mit der Zeit erkennen lernen. Dies nimmt jedoch – wie beim Lesen – ohne Unterstützung viel Zeit in Anspruch (Bredel & Pieper, 2015). Um Kinder im Schriftspracherwerb ausreichend zu unterstützen, wird die Kombination zweier Prozesse als optimale Vorgangsweise beschrieben: Ein gesteuerter Lernprozess, der den Wissensaufbau von außen lenkt und ein ungesteuerter Erwerbsprozess über eine eigenaktive Hypothesen- und Regelbildung (Bredel, Fuhrhop & Noack, 2017). Allerdings setzt auch die innere Regelbildung eine entsprechende Anregung voraus: Lernende müssen *„mit verständlichen schriftlichen Daten versorgt werden, die Strukturen aufzeigen, aus denen sie Hypothesen bilden und Regeln ableiten können“* (ebd. S. 72). Auf diese Anregungen sind besonders Kinder mit weniger Schrifterfahrung angewiesen, denn sie laufen Gefahr, ohne explizite didaktische Unterstützung das Schriftsystem langsamer zu erfassen und dadurch möglicherweise den Anschluss in der Schule zu verlieren (Naegele, 2011).

Um Defizite bei orthographischen Fähigkeiten möglichst frühzeitig erkennen zu können, werden daher sensitive Messinstrumente benötigt, die für Kinder in Frühstadien des Schrifterwerbs geeignet sind und insbesondere auch auf schriftsystematischen Prinzipien beruhendes Wortmaterial enthalten. Dem SCHNAPP-Rechtschreibtest liegt ein hierarchisch aufgebautes Wortmaterial zugrunde, das auf dem trochäischen Zweisilber beruht (siehe nächstes Kapitel für eine ausführliche Erläuterung).

Neben den theoretischen Überlegungen, welche die Entwicklung des Verfahrens geleitet haben, werden insbesondere erste psychometrische Ergebnisse zur Dimensionalität des neuen digitalen SCHNAPP-Rechtschreibtests vorgestellt, welche die Auswahl des verwendeten Wortschatzes empirisch bestätigen. Weiters wird die Reliabilität des SCHNAPP-

Rechtschreibtests an einer Stichprobe von Kindern am Ende der ersten Grundschulklasse beschrieben.

Prinzipien der deutschen Rechtschreibung

Alphabetschriften beruhen grundlegend auf einem phonographischen Prinzip, d.h. ganz allgemein stehen lautliche Einheiten in einer Beziehung mit den schriftlichen (in der Regel spricht man von Phonem-Graphem-Korrespondenzen, z.B. Dürscheid, 2016, S. 134ff.). Im deutschen Schriftsystem kommen aber nur relativ wenige Schreibungen auf diese Weise korrekt zustande. Weitere Prinzipien – wie das silbische, morphologische und syntaktische – überlagern das phonographische Prinzip. Annahmen über die Schreibsilbe machen vorwiegend Schreibungen in Bezug auf die Vokalquantität nachvollziehbar (ursprünglich Eisenberg, 1989). Morphologische Schreibungen beziehen sich darüber hinaus u.a. auf Morphemkonstanzschreibungen (Hand/Hände); die syntaktischen Schreibungen beziehen sich über die Wortebene hinaus auf die Groß-/Kleinschreibung (Bredel, 2010a) sowie auf die Getrennt-/Zusammenschreibung (Fuhrhop, 2010).

Zwischen der Ebene der Silbe und der Ebene des Morphems wird noch eine weitere Ebene – der Fuß – angesiedelt (z.B. Fuhrhop & Peters, 2013). Silben sowie Füße sind rhythmisch/prosodische Einheiten, die der Lautebene zugeordnet werden. Genauso wie bei der Silbe zeigt sich aber auch für den Fuß, insbesondere für den Trochäus, strukturelle Relevanz für die geschriebene Sprache. Die jüngere Graphematikforschung hat für die Wortschreibung den Trochäus als Basismuster herausgearbeitet (z.B. Röber, 2009; Bredel, 2010b; Primus, 2010). Es befinden sich somit monomorphemische zweisilbige Wörter, die aus einer betonten und einer unbetonten bzw. einer reduziert artikulierten Silbe bestehen (z.B. <Ro-se>, <ba-den>, <Rin-de>, <Win-del>, <Wol-le>), im Zentrum. Wörter, die diesem Basismuster entsprechen, werden hier als Basisschreibungen bezeichnet, deren Schwierigkeit in eine Hierarchie entsprechend der graphematischen Komplexität gebracht werden kann. Sehr viele Wörter, z.B. Infinitive bei Verben; häufig die Pluralform bei Substantiven, im Deutschen können auf dieses Basismuster zurückgeführt werden (Müller, 2010). Damit können strukturelle Aspekte wie z.B. die Schärfungsschreibung bei kurzem, ungespannten Vokal in der ersten Silbe, aber auch die <e>-Schreibung in der zweiten Silbe (Reduktionssilbe) leicht nachvollziehbar gemacht werden. Zudem bildet dieses Muster auch die Basis für die morphologischen Schreibungen. Sehr häufig umfasst das Stammmorphem genau die erste Silbe und den Anfang der zweiten Silbe (Reduktionssilbe). Als Beispiele können hier angeführt werden: <Kind> mit <d>, weil der trochäische Zweisilber auch mit

<d> verschriftet wird; <Mann> mit <nn> und <a>, weil es im trochäischen Zweisilber <Männer> heißt (z.B. Bredel, 2010b; Fuhrhop & Peters, 2013).

Wortmaterial im SCHNAPP-Rechtschreibtest

Üblicherweise wird in Rechtschreibtests sowie in traditionellen Schriftspracherwerbskonzeptionen Wortmaterial verwendet, das zwar unterschiedlichen Prinzipien des deutschen Schriftsystems folgt, aber nicht aufeinander systematisch diesbezüglich aufbauend ausgewählt ist. Oftmals wird dabei viel Zeit und Mühe auf das erste Prinzip, das phonographische Prinzip, also die grundlegende Beziehung zwischen Phonemen-Graphemen, verwendet. Über lauttreue Schreibungen wird das phonographische Prinzip überprüft und bei Bedarf gefördert. Bei der Arbeit mit lauttreuen Schreibungen werden die Kinder aufgefordert, die gesprochene Form in segmentale lautliche Bestandteile zerlegen, eine für kompetente Rechtschreiber einfache Aufgabe, die aber für Kinder zu Beginn des Schriftspracherwerbs abstrakt und herausfordernd sein kann. Im Anschluss werden diese herausgelösten lautlichen Einheiten in einem 1:1-Verfahren in Buchstaben übertragen und notiert. Wir folgen in der Konzeption des Schnapp-Rechtschreibtests *nicht* dieser Herangehensweise, sondern entwerfen das Messinstrument aufgrund der oben skizzierten neueren Theorie. Die Konzentration auf lauttreue Schreibungen berücksichtigt nämlich nicht, dass Buchstaben isoliert keinen absoluten Lautwert besitzen, sondern erst relational sowie abhängig von ihrer jeweiligen Position und der damit verbundenen Umgebung innerhalb der Silbe/des Wortes suprasegmental eine lautliche Realisierung erfahren (Bredel, 2009, S. 137; eine kritische Auseinandersetzung mit der Lauttreue gibt Röber, 2001). Zugänglich wird der konkrete Lautwert eines Buchstabens damit nur im Zusammenhang mit anderen Buchstaben innerhalb einer Silbe. Anhand von zweisilbigen Trochäen kann für den Kernbereich der deutschen Wortschreibung eine stabile Ausgangsbasis angenommen werden, die dem silbischen Prinzip zuzuordnen und sehr gut anschlussfähig für das morphologische Prinzip ist (zusammenfassend Bredel, 2009).

Die Realität in Klassenzimmern entspricht aber in den meisten Fällen nicht dieser Herangehensweise. Kinder werden immer noch selten gezielt systematisch silbenbasiert unterrichtet. Die Entscheidung, dennoch den Schwerpunkt rein auf die strukturelle Komponente und damit auf die Analyse von trochäischen Zweisilbern zu legen, ist damit zu begründen, dass Kinder später anhand ihrer reinen Rechtschreibleistung bewertet werden; für diese ist strukturelle Kenntnis maßgebend, auch wenn sie nicht systematisch im Unterricht vermittelt wird. Förderempfehlungen würden sich dementsprechend auch nicht auf das isolierte Training von Phonem-Graphem-Korrespondenzen bzw. auf Übungen zur

Einzellautsegmentierung der gesprochenen Sprache beschränken, sondern auf silbische Übungen und Arbeit mit systematischem Wortmaterial zunächst anhand des trochäischen Zweisilbers. Darauf aufbauend folgt die Auseinandersetzung mit morphologischen Schreibungen. Konzeptionen und erste Lernbeobachtungen in diesem Bereich erscheinen vielversprechend (vgl. Röber, Häusle & Berchtold, 2019). Ausreichend empirische Daten zu diesen Annahmen fehlen jedoch bislang.

Als Grundlage des SCHNAPP-Rechtschreibtests wurden trochäische Schreibungen herangezogen (Röber, 2009; Bredel, 2009; Bredel, 2010b; Krauß, 2010): Trochäen mit offenen und mit geschlossenen Vollsilben, mit Schärfungsschreibung und mit h-Schreibung (silbeninitiales h und Dehnungs-h). Trochäische Schreibungen mit offenen (<ba-den>) und geschlossenen Vollsilben (<Win-del>) können als unmarkiert bezeichnet werden und stehen damit auch am Beginn der Erwerbsreihenfolge (Bredel, 2009). Den trochäischen Schreibungen mit offenen Vollsilben wird auch die ie-Schreibung und die Schreibung mit Diphthong und Umlaut zugerechnet. Diese Schreibungen mit komplexem Nukleus werden gegenüber Schreibungen mit einfachem Nukleus generell als schwieriger eingestuft.

Der SCHNAPP-Rechtschreibtest verzichtet auf Wortmaterial mit Schärfungsschreibung, silbeninitialem h und Dehnungs-h und konzentriert sich stattdessen auf den trochäischen Zweisilber mit offener oder geschlossener Vollsilbe. Diese beiden Typen werden in Folge in weitere Schwierigkeitsgrade aufgefächert. Hervorzuheben ist, dass die weitere Untergliederung theoretisch ableitbar ist (Röber, 2009; Bredel, 2010b; Krauß, 2010), jedoch bislang keine empirische Basis aufweist und damit als explorativ zu sehen ist.

Eine hierarchische Gliederung der trochäischen Schreibungen baut auf Wörtern mit offenen Vollsilben (<Ro-se>, <ba-den>) über geschlossene Vollsilben (<Rin-de>, <Win-del>) hin zu Vollsilben mit Konsonantenhäufungen im Onset (<schwin-den>) auf. Innerhalb dieser groben Hierarchie kommen für die Testung weitere Schwierigkeitsgrade in Bezug auf Komplexität der graphematischen Silbe zum Einsatz, um durch genügend Aufgaben im unteren Leistungsbereich vor allem auch Kinder mit Schwierigkeiten zu identifizieren. Nicht nur die Markierung der Vokalquantität bei kurzen, ungespannten Vokalen über eine geschlossene Vollsilbe (z.B. <Kin-der>) stellt eine angenommene Schwierigkeitsstufe dar, sondern auch die zunehmende Komplexität von Silbenkonstituenten. Zum Beispiel wurde für unmarkierte trochäische Zweisilber mit offener Vollsilbe angenommen, dass komplexe Nuklei wie <ie>-Schreibung (z.B. <Lie – be>) sowie die Umlautschreibung (z.B. <Jä-ger>) höhere Anforderungen an die Kinder stellen als die Verschriftung einfacher Nuklei (z.B. <Ro-se>). Die <ie>-Schreibung wird zwar regelhaft im Deutschen für ein langes, gespanntes [i:]

verwendet, aber im Anfangsunterricht werden Kinder in der Regel zunächst mit lauttreuen Wörtern konfrontiert, die ihnen auch im Falle der <i>-Schreibung vermitteln, dass das „lange i“ mit <i> wiedergegeben wird (vgl. Wörter wie z.B. Timo, Limo). Der Übertrag im Anschluss von diesen einzelnen Fällen auf die regelhafte Verschriftung von <ie> macht vielen Kindern Schwierigkeiten (Röber, 2006). Diphthonge werden theoretisch ebenfalls als schwieriger eingestuft, da es sich einmal um einen komplexen Nukleus handelt und zusätzlich unterliegt die geschriebene Form nicht direkt einfachen Phonem-Graphem-Beziehungen. Im SCHNAPP-Rechtschreibtest werden lediglich Items mit <ei> und <au> verwendet, von denen ausgegangen werden kann, dass sie in Klasse 1 bereits behandelt wurden. In Bezug auf die Umlautschreibung lässt sich generell feststellen, dass Buchstaben mit Trema als komplex einzustufen sind und damit graphematisch eher schwer sind (Fuhrhop & Peters, 2013, S. 195, S. 231).

Unterschiedliche Schwierigkeiten müssen auch für die Verschriftung des Onsets der ersten Silbe (Vollsilbe) angenommen werden: Die Verschriftung von Konsonantenclustern (z.B. <schw> oder <spr>) sollte sich als schwieriger erweisen als die Verschriftung einfacher Onsets (z.B. <s> oder <p>) (z.B. Thomé, 2019). Als weitere Hürde wird <r> in der Koda der Vollsilbe (z.B. <war-ten>) oder in der Reduktionssilbe (z.B. <Fe-der>) eingestuft, weil in dieser Position im Gegensatz zur Position Onset vorwiegend vokalisches Potential lautlich assoziiert wird. In Folge treten noch Kombinationen von komplexen und weniger komplexen Voll- und Reduktionssilben auf.

Basierend auf diesen Grundlagen wird eine theoretische Hierarchie von Wortschwierigkeiten für den Erwerbszeitraum Ende erster Grundschulstufe angenommen, wie sie in Tabelle 1 dargestellt ist. Im SCHNAPP-Rechtschreibtest wurden für Version 1 und Version 2 leicht unterschiedliche Items verwendet, weil Schwächen in der Differenzierung an den Randbereichen auftraten. Neue Items wurden vor allem in Schwierigkeitsniveau 1 und 2 eingefügt, um im Besonderen die Differenzierung im unteren Leistungsbereich zu verbessern. 12 Items sind in beiden Versionen ident.

Inwiefern sich die theoretischen Annahmen dieser sechsstufigen Wortschatzstruktur auch empirisch bestätigen lassen, soll nachfolgend näher untersucht werden.

Tabelle 1 hier einfügen

Forschungsfragen

Der SCHNAPP-Rechtschreibtest ist ein neu entwickeltes digitales Testverfahren zur Prüfung orthographischer Kompetenzen von Kindern in ersten Grundschulklassen. Im Gegensatz zu etablierten Papier-Bleistift Verfahren basiert er auf Wortmaterial, das nach schriftsystematischen Prinzipien ausgewählt wurde, um eine Hierarchie an Schwierigkeiten abzubilden (siehe Tabelle 1). Gegenstand der vorliegenden Studie ist daher eine erste Prüfung der psychometrischen Qualität des Tests. Insbesondere gehen wir drei zentralen Frage nach:

- (a) Lässt sich die schriftsystematisch orientierte Auswahl des verwendeten Wortschatzes zur Erfassung eines gemeinsamen latenten Faktors bestätigen?
- (b) Lässt die Messgenauigkeit des Tests eine präzise Erfassung der Rechtschreibfähigkeit von Kindern am Ende der ersten Grundschulklasse zu, um insbesondere im unteren Fähigkeitsbereich Kinder mit Schwierigkeiten im Rechtschreiben zu identifizieren?
- (c) Ermöglichen die empirisch ermittelten Itemschwierigkeiten eine Bestätigung der aus der Literatur ableitbaren Hierarchie des Wortschatzes?

Diese Fragestellungen werden anhand von zwei Stichproben, die in zwei Jahrgängen in ersten Klassen an verschiedenen österreichischen Grundschulen erhoben wurden, untersucht.

Methode

Stichproben und Vorgehen

Das Testverfahren wurde in zwei Stichproben entwickelt, die in den Jahren 2020 (Stichprobe 1) und 2021 (Stichprobe 2) an oberösterreichischen Grundschulen erhoben wurden. Die erste Stichprobe umfasste $N = 299$ Kinder (53% Mädchen) aus 23 ersten Schulklassen von 8 Schulen. Ungefähr 68% der Kinder hatten Deutsch als Erstsprache (DaE), während 26% mit Deutsch als Zweitsprache aufwuchsen (DaZ; für 17 Kinder lag keine Information zur Sprache vor). Für 17% der Kinder lag eine Empfehlung für eine Sprachförderung vor, während 76% der Kinder keine entsprechende Empfehlung erhalten haben (für 21 Kinder lagen diesbezüglich keine Angaben vor).

Die zweite Stichprobe umfasste $N = 148$ Kinder (45% Mädchen) aus 13 ersten Klassen von 7 Grundschulen (es handelte sich um die gleichen Schulen wie in Stichprobe 1, die geringere Zahl an Probanden ist durch die Einschränkungen im Rahmen der Corona-Pandemie zu erklären). Die meisten Kinder (70%) wuchsen mit Deutsch als Erstsprache und 27% mit Deutsch als Zweitsprache auf (für 4 Kinder lag keine Information zur Muttersprache vor).

Ungefähr 17% der Kinder hatten in der Vergangenheit bereits eine Empfehlung für eine Sprachförderung erhalten, während für 77% der Kinder keine entsprechende Empfehlung vorlag (für 9 Kinder lagen keine Angaben darüber vor).

Ein Großteil der Kinder aus beiden Stichproben stammt aus einer Region, in der alle Schulen am Projekt teilnahmen. Im Österreichvergleich (Statistik Austria, 2021) ist die zweite Stichprobe hinsichtlich des Geschlechts (45% vs 48,2%), der Erstsprache Deutsch (70% vs 72%) und dem höchsten Bildungsabschluss der Eltern (Universität: 27% vs 21%; Pflichtschulabschluss: 6% vs 3%) repräsentativ, wenngleich die Bildungsabschlüsse (entweder Mutter oder Vater) leicht höher ausfielen.

Instrument

Die Erstversion des SCHNAPP-Rechtschreibtests setzte sich aus 17 Wörtern zusammen, während die revidierte Testversion 22 Wörter beinhaltet. Berücksichtigt wurden zwei Testkonstruktionsprinzipien: Bezüglich *Wortschatzfrequenz* erfolgte eine Orientierung am Corpus childlex, einer Datenbank zur Schriftsprache für Kinder (Schroeder, Würzner, Heister, Geyken & Kliegl, 2015). Die Wörter wurden entsprechend Schroeder et al. (2015) in drei Gruppen von frequenten Wörtern unterteilt: >100 hochfrequent, 10-100 mittelfrequent und 1-10 niedrigfrequent. Um zu zeigen, dass die Struktur der Wörter für eine entsprechende Realisierung verantwortlich ist und nicht die Frequenz, wurden für die Testbatterie Wörter aus den drei Frequenzbereichen herangezogen und innerhalb der Hierarchiestufen gemischt verwendet (siehe dazu Tabelle 2 für die revidierte Testversion und Supplement Tabelle S1 für die Erstversion).

Bezogen auf die *Schriftsystematik* wurde eine aus der Literatur abgeleitete Hierarchie des Wortschatzes auf Basis des trochäischen Silbenrhythmus verfolgt. Die Schwierigkeit erstreckt sich von offenen über geschlossene Vollsilben, von einfacher bis zur komplexen Besetzung im Onset oder der Koda (siehe Tabelle 1). Dem Schwierigkeitsniveau 1 wurden 6 Items zugeordnet, Schwierigkeitsniveau 2 dann 5 Items, Schwierigkeitsniveau 3 und 5 je 4 Items, Schwierigkeitsniveau 4 nur 2 und Schwierigkeitsniveau 6 schließlich 1 Item. Die Itemverteilung erfolgte aus zweierlei Gründen nicht gleichwertig: Zuerst wurde aufgrund des Nebengütekriteriums Testzumutbarkeit darauf geachtet, die Anzahl der Items einzugrenzen. Weiters steht die Differenzierung in den untersten Leistungsbereichen im Fokus, woraufhin im unteren Leistungsbereich mehr Items angeboten wurden als im oberen Spektrum.

Durchführung

Angewandt wurde der SCHNAPP-Rechtschreibtest nur bei jenen Schülerinnen und Schülern, deren Erziehungsberechtigte der Teilnahme zuvor zugestimmt hatten. Alle Untersuchungen fanden an Schulvormittagen statt. Die Erhebungen wurden von ausgebildeten Lehramtsstudierenden und Stammmitgliedern des Teams des Forschungsprojektes instruiert und begleitet. Der SCHNAPP-Rechtschreibtest wurde in der Klasse als Gruppentest durchgeführt, wobei jedes Kind ein Tablet (iPad, 17. Generation, 10,2 Zoll) mit digitalem Stift und Kopfhörern erhielt. Die Testbatterie wurde in eine motivierende Rahmengeschichte eingebettet, die Instruktion (sowohl Eingangsinstruktion als auch jedes Item) startete automatisch, konnte jedoch beliebig wiederholt werden. Da die Kinder Kopfhörer trugen, konnten sie die Übungen im eigenen Schreibtempo bewältigen. Zur Gewöhnung an den digitalen Stift und das Tablet als Schreibunterlage wurden zuvor zwei unbewertete Aufgaben („Schreibe deinen eigenen Namen!“ und ein Probe-Item) vorgegeben. Abbildung 1 illustriert die Tablet-Situation.

Abbildung 1 hier einfügen

Statistische Analysen

Das eindimensionale Messmodell des SCHNAPP-Rechtschreibtests wurde in jeder Stichprobe mit einem einparametrischen logistischen Testmodell (1PL; Rasch, 1960) überprüft. Dieses unterstellt einen logistischen Zusammenhang zwischen der zu messenden Personenfähigkeit und der Wahrscheinlichkeit, ein Item richtig zu lösen. Dabei dürfen sich die Items hinsichtlich ihrer Schwierigkeit unterscheiden, während von gleichen Trennschärfen ausgegangen wird. Bei Gültigkeit dieses Testmodells kann ein Test durch einfaches Auszählen der richtigen Antworten (z.B. als Summe oder Mittelwert) gebildet werden (McNeish & Wolf, 2020). Um die Passung der beobachteten Itemantworten zu den theoretischen Annahmen des 1PL empirisch zu belegen, wurden verschiedene psychometrische Analysen durchgeführt.

Zunächst wurde die globale Modellpassung des 1PL mit der Modellpassung eines zweiparametrischen logistischen Testmodells (2PL; Birnbaum, 1968), das auch unterschiedliche Trennschärfen berücksichtigt, anhand von Likelihood-Ratio Tests (LRT) und Bayesianischer Informationskriterien (BIC; Schwarz, 1978) verglichen. Da der LRT stark von der Stichprobengröße abhängig ist, legen wir einen größeren Wert auf das BIC, das eine bessere Modellpassung bei kleineren Werten anzeigt. Als Effektgröße für diesen Vergleich ziehen wir den Bayesfaktor (BF; Wagenmakers, 2007) heran, der nach Raftery (1995) bei

Werten zwischen 20 und 150 starke Evidenz für das Modell mit dem kleineren BIC anzeigt sowie sehr starke Evidenz bei Werten über 150.

Neben diesem globalen Modelltest wurde zusätzlich auch für jedes Item einzeln geprüft, ob die beobachteten Antworten mit den Annahmen des 1PL vereinbar sind. Dazu wurden vier Kriterien herangezogen: (a) Der Infit gibt für jedes Item an, wie gut die beobachteten Antworten durch das theoretische 1PL vorhergesagt werden können (Wright & Masters, 1982). Werte von 1 weisen eine perfekte Modellpassung aus, während in der einschlägigen Literatur von einer zufriedenstellenden Modellpassung für praktische Anwendungen ausgegangen wird, wenn der Infit kleiner als 1.15 ausfällt (Pohl & Carstensen, 2013). (b) Zusätzlich berichten wir auch die *Root Mean Squared Deviation* (RMSD), welches in internationalen Bildungsstudien kürzlich als alternatives Gütekriterium eingeführt wurde (Köhler, Robitzsch & Hartig, 2020). Hierfür werden Werte kleiner als 0.08 als Indikator für eine gute Modellpassung angesehen (Robitzsch & Lüdtke, 2020). (c) Als Signifikanztest verwenden wir den χ^2 -Test von Orlando und Thissen (2000), welcher prüft, ob sich die beobachteten Antworten überzufällig von den aufgrund des Testmodells vorhergesagten Antworten unterscheiden. Zur Kontrolle des Fehlers 1. Art wird hierfür eine Korrektur der p -Werte nach Benjamini und Hochberg (1995) angewandt. Ein nicht-signifikanter Test weist auf eine zufriedenstellende Modellpassung hin. (d) Die Annahme statistischer lokaler Unabhängigkeit wird anhand der adjustierten Teststatistik $Q3a$ von Yen (1984) geprüft. Diese gibt die durchschnittliche Residualkorrelation zwischen den Items wieder. Gängigen Empfehlungen zufolge kann von einem essentiell eindimensionalen Testmodell und damit einer guten Modellpassung ausgegangen werden, wenn das durchschnittliche absolute $Q3a$ für ein Item .20 nicht überschreitet (Yen, 1993).

Ergebnisse

Nachfolgend werden die psychometrischen Informationen im Rahmen der Entwicklung des SCHNAPP-Rechtschreibtests getrennt für die beiden Stichproben zusammengefasst. Die Ergebnisse aus der ersten Stichprobe wurden herangezogen, um den Test zu überarbeiten und eine revidierte Testversion in der zweiten Stichprobe vorzugeben.

Erstversion des SCHNAPP-Rechtschreibtests

Die Items der Erstversion des SCHNAPP-Rechtschreibtests, die in Stichprobe 1 eingesetzt wurden, zeigten sehr heterogene Schwierigkeiten. Der Prozentsatz der Kinder, die ein Item richtig gelöst hatten, variierte zwischen 8% und 86% ($Mdn = 57%$) für die 17 administrierten Items. Damit umfasste der Test sowohl einfache als auch schwierige Aufgaben und konnte

dadurch ein breites Fähigkeitsspektrum abdecken. Dies wird auch durch die geschätzten Itemschwierigkeiten des 1PL verdeutlicht, welche zwischen -2.35 und 3.06 ($Mdn = -0.37$) variierten. Ein Vergleich der Itemschwierigkeiten mit den Personenfähigkeiten erlaubt es abzuschätzen, wie gut die Testitems in der Lage sind, die Rechtschreibkompetenz der Stichprobe zu schätzen. Im 1PL wurde der Mittelwert der latenten Personenfähigkeitenverteilung zur Modellidentifizierung auf 0 fixiert. Zudem wurde eine Standardabweichung von 1.31 geschätzt; d.h. der Test war in der Lage, gut zwischen den Kindern zu differenzieren. Die Verteilung der geschätzten Personenfähigkeiten (d.h. der Rechtschreibkompetenz der Kinder) ist als Balkendiagramm in Abbildung 2 (linke Hälfte) dargestellt. Zudem werden darin auch die Itemschwierigkeiten abgebildet. Da die Personenfähigkeiten und Itemschwierigkeiten auf derselben Skala verortet sind, erlaubt ein Vergleich der beiden Verteilungen die Passung des Tests für die Schätzung der Rechtschreibkompetenz der vorliegenden Stichprobe zu beurteilen. Abbildung 2 zeigt, dass die Items im Durchschnitt etwas zu leicht waren, d.h. mehr einfache als schwierige Items vorlagen.

Abbildung 2 hier einfügen

Wie Abbildung 2 zeigt, inkludierte der Test trotz (im Durchschnitt) etwas geringerer Itemschwierigkeiten keine Items mit Itemschwierigkeiten unter -2, weshalb sehr geringe Fähigkeiten nur ungenau gemessen werden konnten. Dies wird auch durch eine Betrachtung der marginalen Reliabilität auf Basis des Item Response Modells bestätigt (vgl. Adams, 2005). Obwohl die durchschnittliche Reliabilität für die Gesamtstichprobe mit .81 zufriedenstellend ausfiel, zeigte eine differenzierte Betrachtung in Abhängigkeit des Fähigkeitsniveaus der Kinder, dass die Messgenauigkeit des Tests insbesondere in den Randbereichen (d.h. sehr niedrigen und sehr hohen Fähigkeiten) etwas eingeschränkt war (siehe dünne Linie in Abbildung 3).

Abbildung 3 hier einfügen

Daher wurde auf Basis dieser Ergebnisse eine Überarbeitung des Tests vorgenommen und auch neue Items wurden entwickelt, um insbesondere den unteren Fähigkeitsbereich messen zu können. Auf eine ausführliche Darstellung der Modellgüte, also der Passung des 1PL zu den beobachteten Antworten, wird an dieser Stelle daher verzichtet. Diese sind im Elektronischen Supplement zusammengefasst.

Revidierter SCHNAPP-Rechtschreibtest

In der zweiten Stichprobe zeigte das 1PL für die revidierte Testversion insgesamt eine zufriedenstellende Passung zu den beobachteten Daten. Obwohl der LRT mit einem $\text{Chi}^2 (df = 21) = 50.05, p < .001$, eine empirische Präferenz für das 2PL nahe zu legen schien, ergab ein Vergleich der Bayesianischen Informationskriterien eine bessere Modellpassung für das 1PL (BIC = 3226 versus 3281). Der Bayesfaktor deutete zudem auf eine sehr starke Evidenz für das 1PL hin ($\text{BF} > 150$). Auch eine Prüfung der Modellgüte auf Itemebene ergab eine zufriedenstellende Passung der beobachteten Antworten zum 1PL. Die vier Indikatoren zur Prüfung der Itemqualität, die in Tabelle 1 zusammengefasst sind, ergaben lediglich für Item 11 einen etwas zu großen RMSD von 0.10. Da die übrigen Indikatoren jedoch keine Auffälligkeiten anzeigten, wurde dies nicht als schwerwiegend betrachtet. Damit scheint die durch das 1PL implizierte Annahme gleicher Trennschärfen vertretbar und eine Interpretation der Summenscores (d.h. Anzahl richtig gelöster Aufgaben) als Indikator für die Rechtschreibkompetenz der Kinder gerechtfertigt.

Tabelle 2 hier einfügen

Die 22 Items des revidierten SCHNAPP-Rechtschreibtests deckten ein breites Fähigkeitsspektrum ab. Der Prozentsatz der Kinder, die ein Item richtig gelöst hatten, variierte zwischen 14% und 89% ($Mdn = 55\%$). Die entsprechenden Itemschwierigkeiten des 1PL reichten von -2.79 und 2.46 ($Mdn = -0.36$) und deckten damit einen größeren Bereich der niedrigen Fähigkeiten ab als die ursprüngliche Testversion. Dies wird auch in Abbildung 2 (rechte Hälfte) verdeutlicht. Der Test umfasste mehr Items mit einer Schwierigkeit kleiner als 0 (= Mittelwert der Personenfähigkeit). Insbesondere lagen nun auch Items mit Schwierigkeiten unter -2 vor, die eine präzise Fähigkeitsmessung auch bei Kindern mit sehr geringer Rechtschreibkompetenz erlauben sollten. Für die Personenfähigkeit ergab sich in der revidierten Testform eine Standardabweichung von 1.41, was eine gute Differenzierung zwischen den Kindern mithilfe des Testverfahrens nahelegt.

Die höhere Messpräzision der revidierten Testform wird auch anhand der Reliabilität bestätigt. Die durchschnittliche Reliabilität für die Gesamtstichprobe lag mit .86 leicht höher als jene der Erstversion. Besonders hervorzuheben ist jedoch, dass die Messgenauigkeit des Tests insbesondere im unteren Randbereich (d.h. für sehr niedrige Fähigkeiten) besser ausfiel (siehe fette Linie in Abbildung 3). Damit ist der SCHNAPP-Rechtschreibtest in der Lage, auch Fähigkeiten reliabel (mit $r_{xx} > .80$) zu erfassen, die rund 2 Standardabweichungen unterhalb des mittleren Fähigkeitsbereichs liegen.

Für eine Prüfung der theoretisch angenommenen Wortschatzhierarchie, die der Konstruktion des SCHNAPP-Rechtschreibtests als Grundlage diente, wurden die empirischen Itemschwierigkeiten mit den theoretischen Schwierigkeitsniveaus der Items (siehe Tabelle 2) korreliert. Die hohe Spearman Rangkorrelation von .88 legt eine sehr gute Stützung der theoretischen Annahmen über eine sechsstufige Wortschatzhierarchie zum Testzeitpunkt nahe.

Diskussion

Erhebungsformat des SCHNAPP-Rechtschreibtests

Eine Erhebung der Rechtschreibleistung in der Grundschule erfolgt meist durch Einschätzungen der Lehrkraft oder mittels Diktat von Wörtern, Sätzen oder Texten (Uppstad & Solheim, 2007). Die Güte der Einschätzung von etwaigen Rechtschreibschwierigkeiten der Schülerinnen und Schüler durch Lehrkräfte sind dabei abhängig vom Schweregrad: Während starke Symptomaten erkannt werden, sind Übereinstimmungen zwischen Lehrer- und Lehrerinnenurteil und Performanzdaten der Kinder bei geringerer Symptomausprägung maximal mittelgradig (Schöfl, Wiltsche, Holzer & Steinmair, 2019). Die Überprüfung der Rechtschreibung in Form von Diktaten wird im Schulalltag entweder mit standardisierten Verfahren oder qualitativ mit Diktaten der Klassenlehrkraft durchgeführt. Erstere zeichnen sich durch Objektivität und die Möglichkeit eines Normvergleichs aus, sind aber bezüglich Vorbereitung, Durchführung und Kosten aufwändiger als selbst erstellte Diktate. Diktate an sich stehen in der Kritik, für manche Schülerinnen und Schüler stressauslösend zu sein, insbesondere auch deshalb, weil das Tempo der Ansage im typischen Gruppensetting für alle gleich und manchen Kindern zu schnell ist, wodurch die Leistungsfähigkeit geschmälert werden kann. Digitale, Tablet-gestützte Erhebungen, wie etwa der neu entwickelte SCHNAPP-Rechtschreibtest, können diese Nachteile ausgleichen, indem sie die individuelle Bearbeitung der Aufgaben ermöglichen und weitere Nebengütekriterien, wie Fairness und Ökonomie im Vergleich zu analogen Testformen positiv bedienen (vgl. für ältere Schülerinnen und Schüler: Endlich, Lenhard, Marx & Richter, 2022).

Durch die Verwendung von Kopfhörern kann der SCHNAPP-Rechtschreibtest einerseits mit einer gesamten Klasse durchgeführt werden, andererseits kann jedes Kind die Aufgaben im individuellen Arbeitstempo erledigen. Zur zusätzlichen Motivationssteigerung wurden die Testaufgaben in eine abenteuerliche Rahmengeschichte rund um den Drachen „Schwupp“

gebettet. Die Verwendung des digitalen Stiftes sowie die Menüführung in der Softwareanwendung wurde als unkompliziert beobachtet. Vereinzelt benötigten die Kinder Hilfe, wenn der digitale Stift unabsichtlich ausgeschaltet worden war. Positiv wurde von Lehrkräften und Kindern angemerkt, dass die Kinder bei den diktierten Wörtern im eigenen Tempo schreiben konnten.

Wortschatz im SCHNAP-Rechtschreibtest

Als Besonderheit des SCHNAPP-Rechtschreibtests ist neben der Durchführung am Tablet die Konzeption zu nennen. Es wird ein Kernwortschatz abgebildet, der sich maximal an den grundlegenden Schriftsprachmustern orientiert. Dabei bilden unmarkierte, prototypische Erscheinungen die Basis, die im Regelfall früher erworben und verwendet werden als weniger prototypische (Primus, 2010). Diese Hierarchie wird bei Erhebung dieses Kernwortschatzes sichtbar und lässt erste und darauf aufbauende Schriftstrukturmuster bei Kindern erkennen. Der neu entwickelte SCHNAPP-Rechtschreibtest versucht anhand eines hierarchisch strukturierten Wortschatzes diese grundlegenden und aufbauenden Muster auf Basis des silbischen Prinzips zu erheben. Obwohl auch bisherige, weit verbreitete Rechtschreibtests Items beinhalten, die auf dem silbischen Prinzip beruhen, hebt sich der SCHNAPP-Rechtschreibwortschatz dennoch davon ab. Etablierte Tests enthalten neben wenigen rein silbischen Schreibungen vorrangig zusätzliche orthographische Markierungen. Bei Fehlschreibungen ist demnach nicht ableitbar, ob das basale Prinzip nicht beherrscht wird, oder der orthographische Marker missachtet wurde. Dieser Logik folgend wurde der SCHNAPP-Rechtschreibwortschatz ausschließlich auf Basis des silbischen Prinzips entwickelt und enthält unmarkiertes Wortmaterial, von dem auszugehen ist, dass es den geringsten Schwierigkeitsgrad aufweist.

Prüfung der Rasch-Konformität und Schwierigkeitshierarchie

Eine erste Prüfung erbrachte den Nachweis, dass der Wortschatz Rasch-konform ist, was bedeutet, dass alle Items mit unterschiedlichen Itemschwierigkeiten bei identer Trennschärfe den Faktor Rechtschreiben erfassen. Die Reliabilität konnte durch Itemselektion und Ergänzungen ausgehend von einer Erstversion für den revidierten SCHNAPP-Rechtschreibtest deutlich verbessert werden. Eine ansteigende Schwierigkeit, die die Grundlage des Wortschatzauswahlverfahren darstellt, entspricht den empirischen Ergebnissen: Der Theorie folgend, dass unmarkiertes Wortmaterial als am einfachsten eingestuft werden kann, konnten jene mit offenen Vollsilben häufiger richtig geschrieben werden als Wörter mit geschlossenen Vollsilben, gefolgt von Wörtern mit vokalisiertem r, ie-Schreibung bis hin zu komplexen Onsets und Umlauten. Drei Ausnahmen wurden für den revidierten SCHNAPP-Rechtschreibtest gefunden: Bei einem Wort (Item 14) mit

vokalisiertem r in der Vollsilbe (wie etwa beim Wort „warten“) wurde eben aufgrund der r-Besetzung der Koda der ersten Silbe eine hohe Schwierigkeit angenommen, für die meisten Kinder (80%) war jedoch eine richtige Schreibweise möglich. Diese richtigen Schreibungen lassen sich nicht durch die Wortstruktur erklären, denn ein Wort (Item 15) derselben Struktur wurde von deutlich weniger Kindern (38%) richtig realisiert. Weiters liegt für dieses Wort (Item 14) nach childlex (Schröder et al., 2015) lediglich eine Häufigkeit im niedrigfrequenten Bereich vor, man kann also die richtigen Schreibungen nicht aufgrund der Häufigkeit erklären. Eine Möglichkeit wäre die Bildung von Reimwörtern als Stützmechanismus: Das geforderte Wort lässt sich als Reimwort zu <lernen> bilden, das als hochfrequent einzustufen ist und zusätzlich auch im Übungswortschatz einer ersten Schulstufe liegt. Ähnliche Effekte wurden bei zwei Wörtern (Item 21, 18) gefunden, die aufgrund der Komplexität der Konstituenten Onset und Nukleus (wie etwa bei <Preise>) sowie der Besetzung der Koda der ersten Silbe mit einem <r> (wie etwa bei <starten>) als schwierig in der Hierarchie erwartet wurden, für die revidierte Testform aber lediglich mittelschwer abschnitten. Obwohl Item 18 als hochfrequent einzustufen ist, kann die Frequenz nicht ausschlaggebend für die vermehrten richtigen Schreibungen (50%) sein, denn Wörter mit einer vergleichbaren Frequenz bleiben innerhalb der Erwartungen. Item 21 liegt im mittleren Frequenzbereich (30). Dass bei Wörtern wie z.B. <starten> das <r> an der Position Koda der ersten Silbe verschriftet wird, ist erstaunlich, zumal auch in Verbindung mit vorangehendem <a> im Nukleus kaum konsonantisches Lautpotenzial wahrnehmbar ist. Damit kann auch eine phonographisch orientierte Strategie bei der Verschriftung durch die Kinder in diesem Fall weitgehend ausgeschlossen werden. Möglicherweise spielten diese Wörter in einigen Klassen zum Zeitpunkt der Erhebung eine Rolle im Unterricht.

Bei den restlichen Wörtern zeigt sich – den schriftstrukturellen Überlegungen entsprechend – eine ansteigende Schwierigkeit. Ergänzend zu aktuell eingesetzten und weit verbreiteten Verfahren zur Rechtschreiberfassung, wie etwa SLRT-II (Moll & Landerl, 2014) oder HSP 1-10 (May, 2018) sind Basisschreibungen ohne weitere orthographische Markierungen enthalten, Wörter mit <ie> und <ä> werden zum Zwecke der Differenzierung im höheren Fähigkeitsbereich eingesetzt. Wenn bereits Basisschreibungen zur Detektion von geringem Rechtschreibfertigkeiten dienen, sind sie auch zur Differenzierung im niedrigen Leistungsbereich geeignet, da diese Fertigkeiten aufbauend zu betrachten sind.

Zusammenfassend weist das Rasch-Modell, das für die statistischen Analysen herangezogen wurde und die Itemschwierigkeit in der Testung schätzt, darauf hin, dass eine aufsteigende Schwierigkeitshierarchie in den dargebotenen Items sichtbar wird. Kinder mit

geringem orthographischem Wissen realisieren demzufolge weniger Schreibungen korrekt, die im oberen Hierarchiebereich eingeordnet wurden.

Fibel-Unabhängigkeit

Der Einsatz der Testung ist von der gewählten Unterrichtsform und dem Fibelwerk unabhängig. Die Systematik der Wortschreibungen wird an den Schreibleistungen der Kinder auch dann sichtbar, wenn kein silbenbasierter Ansatz im Unterricht zum Einsatz kam. Nachdem der ausgewählte Rechtschreibwortschatz ein Basiswortschatz ist, baut die weitere Rechtschreibentwicklung darauf auf. Falls Kinder im Lauf ihrer orthographischen Entwicklung explizite Unterstützung bei der systematischen Erarbeitung der Basisschreibungen erhalten, kann diese Entwicklung schneller voranschreiten und Schwierigkeiten abfedern, während Kinder ohne explizite Förderung sich diese Inhalte implizit und vielleicht langsamer oder lückenhaft aneignen. Dennoch gründet die weitere orthographische Entwicklung auf dieser Basis und diese muss von allen Kindern beherrscht werden. Wenn Testergebnisse darauf hinweisen, dass diese Basisschreibungen noch nicht im erwünschten Ausmaß realisiert werden können, dient das im Besonderen jenen Kindern, die noch zusätzliche Förderung an der Basis brauchen, damit ihre orthographische Entwicklung abgesichert wird. Aus diesem Grund wird zur Förderung der orthographischen Entwicklung von Schülerinnen und Schülern angeraten, jene Hierarchien bei der Planung von Unterricht zu berücksichtigen. Vor allem für Kinder mit geringem orthographischem Wissen im ersten Lernjahr braucht es geeignete didaktische Umsetzungen, die nicht auf die Lauttreue des Deutschen verweisen, sondern die Systematik der Basisschreibungen ins Zentrum stellen. Durch frühes Detektieren von fehlendem Rechtschreibstrukturwissen kann in weiterer Folge mit Hilfe der digitalen SCHNAPP-Rechtschreibüberprüfung in der Grundstufe gegengesteuert werden, ehe sich dieses verfestigen. Voraussetzung dafür, dass orthographische Strukturen im Rechtschreibunterricht Einzug halten, ist allerdings eine dementsprechende Vorbildung der Lehrerinnen und Lehrer, damit die Strukturen erkannt werden, die der Schriftstruktur zugrunde liegen, und deren Geschick, „*diese Strukturen logisch und konsequent in didaktische Modellierungen umzusetzen*“ (Krauß, 2010, S. 134).

Wer profitiert?

Die Ergebnisse zeigen auch, dass systematische Schreibungen, die in Bezug auf die hier angenommene Schwierigkeitshierarchie als einfach gesehen werden und an der Basis angesiedelt sind, von Schülerinnen und Schülern der überprüften Klassenstufen weitgehend beherrscht werden – es zeigen sich 74-89% korrekte Schreibungen der ersten 7 Wörter für den revidierten SCHNAPP-Rechtschreibtest. Darin wird ein Vorteil gesehen, weil Kinder bei der Realisierung der ersten Schreibungen zu Beginn der Überprüfung eine positive

Einstellung entwickeln und nicht von Anfang an auf ihre Unzulänglichkeit verwiesen werden. Jene Kinder, die zu diesem Zeitpunkt diese hierarchisch niedrig angesiedelten Schreibungen nicht beherrschen, sind auf eine systematische Förderung angewiesen (Bredel et al., 2017).

Limitationen

Die Wortschatzauswahl erfolgte auf Basis von schriftstrukturellen Aspekten. Obwohl der Wortschatz aufgrund der Häufigkeit im Schriftgebrauch aus unterschiedlichen Frequenzbereichen ausgewählt wurde, kann aufgrund des umschriebenen Itempools nicht ausgeschlossen werden, dass Wörter im schulischen Kontext bereits als Merk- oder Lernwörter behandelt wurden, - was übrigens für alle standardisierten Testungen zutreffen kann - und aus diesem Grund einzelne Wörter besser geschrieben werden, als die Schriftstruktur annehmen ließe. Dieser Umstand kann die angenommene Hierarchie beeinflussen und muss im Folgenden anhand von weiteren Stichproben und Paralleltests kontrolliert werden. Weiterentwicklungen des SCHNAPP-Rechtschreibtests müssen auch verschiedene Testzeitpunkte im Schuljahr sowie Validitätsprüfungen berücksichtigen. Die vorliegenden Ergebnisse sind als vorläufig anzusehen und bedürfen weiterer Analysen.

Relevanz für die Praxis

Der revidierte SCHNAPP-Rechtschreibtest unterscheidet sich von bereits vorliegenden Tests aufgrund der Konzeption: Wortschreibungen, die die Systematik des Deutschen widerspiegeln, kommen zum Einsatz. Im Vordergrund steht das Wissen über trochäische Schreibungen, die in der deutschen Orthographie häufig vorkommen und die Basis für morphologische Ableitungen darstellen. Der Wortschatz bietet nicht nur eine gute Differenzierung im unteren Leistungsbereich der ersten Klasse, sondern zusätzlich förderdiagnostische Ableitungen durch die Hierarchie und Systematik. Als Vorbedingung dafür wurde bereits ein entsprechendes Wissen der Lehrpersonen bezüglich der Sprachstruktur genannt, ebenso braucht es didaktische Konzepte, damit die Struktur der Schreibsilben visuell unterstützt und den Kindern sichtbar gemacht werden kann. Dazu schlagen z.B. Röber (2009) und Bredel et al. (2017) den Einsatz von Silbenhäusern vor. Zusätzlich ermöglicht die digitale Anwendung neben der Gewährleistung hoher Durchführungsobjektivität aufgrund des standardisierten eingesprochenen Textes auch eine erhöhte Testökonomie für Lehrpersonen und Kinder, denn lange Wartezeiten können verkürzt werden.

Literatur

- Adams, R. J. (2005). Reliability as a measurement design effect. *Studies In Educational Evaluation*, 31(2–3), 162–172. <https://doi.org/10.1016/j.stueduc.2005.05.00>
- Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B*, 57, 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Hrsg.), *Statistical Theories of Mental Test Scores*, (S. 397–472). MA: Addison-Wesley Publishing.
- Bredel, U. (2009). Orthographie als System – Orthographieerwerb als Systemerwerb. *Zeitschrift für Literaturwissenschaft und Linguistik*, 153, 135-154
- Bredel, U. (2010a). Die satzinterne Großschreibung – System und Erwerb. In U. Bredel, A. Müller & G. Hinney (Hrsg.), *Schriftsystem und Schriffterwerb: linguistisch–didaktisch–empirisch* (S. 217–234). München: Walter de Gruyter. <https://doi.org/10.1515/9783110232257.217>
- Bredel, U. (2010b). Der Schrift vertrauen. *Praxis Deutsch*, 221, 14–21.
- Bredel, U. & Pieper, I. (2015). *Integrative Deutschdidaktik*. Paderborn: Schöningh.
- Bredel, U., Fuhrhop, N. & Noack, C. (2017). *Wie Kinder lesen und schreiben lernen*. Tübingen: Narr Francke Attempo Verlag.
- Dürscheid, C. (2016). *Einführung in die Schriftinguistik*. Göttingen: Vandenhoeck & Ruprecht.
- Endlich, D., Lenhard, W., Marx, P. & Richter, T. (2021). Tablet-basierter Fehleridentifikationstest zur ökonomischen und validen Erfassung von Rechtschreibfähigkeiten in der Grundschule. *Lernen und Lernstörungen*, 10, 29–42. <https://doi.org/10.1024/2235-0977/a000324>
- Eisenberg, P. (1989). Die Schreibsilbe im Deutschen. In P. Eisenberg & H. Günther (Hrsg.), *Schriftsystem und Orthographie* (S. 57–84). Tübingen: Max Niemayer Verlag.
- Fuhrhop, N. (2010). Getrennt- und Zusammenschreibung: Kern und Peripherie: Rechtschreibdidaktische Konsequenzen aus dieser Unterscheidung. In U. Bredel, A. Müller. & G. Hinney (Hrsg.), *Schriftsystem und Schriffterwerb: linguistisch–didaktisch–empirisch* (S. 235–258). München: Walter de Gruyter. <https://doi.org/10.1515/9783110232257.235>
- Fuhrhop, N. & Peters, J. (2013). *Einführung in die Phonologie und Graphematik*. Heidelberg: Verlag Metzler.

- Köhler, C., Robitzsch, A., & Hartig, J. (2020). A bias-corrected RMSD item fit statistic: An evaluation and comparison to alternatives. *Journal of Educational and Behavioral Statistics*, 45(3), 251-273. <https://doi.org/10.3102/1076998619890566>
- Krauß, A. (2010). Orthographieerwerb von Beginn an. In U. Bredel, A. Müller. & G. Hinney (Hrsg.), *Schriftsystem und Schrifterwerb: linguistisch–didaktisch–empirisch* (S. 133–150). München: Walter de Gruyter. <https://doi.org/10.1515/9783110232257.235>
- May, P. (2013). *HSP 1–10. Hamburger Schreib-Probe zur Erfassung der grundlegenden Rechtschreibkompetenzen. Manual/Handbuch: Diagnose orthografischer Kompetenz*. Stuttgart: Klett-Verlag.
- Naegele, I. (2011). *Jedes Kind kann lesen und schreiben lernen. LRS, Legasthenie, Rechtschreibschwäche – Wie Eltern helfen können*. Weinheim: Beltz.
- McNeish, D. & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, 52(6), 2287–2305. <https://doi.org/10.3758/s13428-020-01398-0>
- Moll, K. & Landerl, K. (2014). *SLRT-II: Lese-und Rechtschreibtest* (2. korrigierte Auflage mit erweiterten Normen). Bern: Huber.
- Müller, A. (2010). *Rechtschreiben lernen. Die Schriftstruktur entdecken – Grundlagen und Übungsvorschläge*. Seelze: Friedrich Verlag.
- Orlando, M. & Thissen, D. (2000). Likelihood-based item fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, 24, 50–64. <https://doi.org/10.1177/01466216000241003>
- Pohl, S. & Carstensen, C. H. (2013). Scaling the competence tests in the national educational panel study. *Journal of Educational Research Online*, 5, 189–216. <https://doi.org/10.1177/0013164414561785>
- Primus, B. (2010). Strukturelle Grundlagen des deutschen Schriftsystems. In U. Bredel, A. Müller & G. Hinney (Hrsg.), *Schriftsystem und Schrifterwerb: linguistisch–didaktisch–empirisch*. (S. 9–45). München: Walter de Gruyter. <https://doi.org/10.1515/9783110232257.9>
- Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Hrsg.), *Sociological Methodology* (S. 111–196). Oxford: Blackwell.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Robitzsch A. & Lüdtke O. (2020). A review of different scaling approaches under full invariance, partial invariance, and noninvariance for cross-sectional country comparisons in large-scale assessments. *Psychological Test and Assessment Modeling*, 62(2), 233–279.

- Röber, C. (2001). Der Mythos der Lauttreue. Für eine andere Präsentation der Schrift. *Grundschule*, 33, 40-42.
- Röber, C. (2006). Die Systematik der Orthographie als Basis von Analysen von Kinderschreibungen. Eine empirische Untersuchung zur Schreibung der *i*-Laute. In Bredel, U. & Günther, H. (Hrsg.), *Orthographietheorie und Rechtschreibunterricht* (S. 71-102). Berlin, Boston: Max Niemeyer Verlag.
- Röber, C. (2009). *Die Leistungen der Kinder beim Lesen- und Schreibenlernen. Grundlagen der Silbenanalytischen Methode. Ein Arbeitsbuch mit Übungsaufgaben*. Berlin: Schneider Verlag.
- Röber, C., Häusle, R. & Berchtold, M. (2019). *Erstklässler entdecken die Orthographie für das Rechtlesen und Rechtschreiben. Vorstellung des schriftsprachstrukturierenden Konzeptes PALOPE*. F&E Edition, 25, 47-60.
- Schöfl, M., Wiltsche, R., Holzer, J. & Steinmair, G. (2019). Schriftsprachrisiko im ersten Schuljahr erkennen – Beobachtung oder Testung? In A. Holzinger, S. Kopp-Sixt, S. Luttenberger, & D. Wohlhart (Hrsg.), *Fokus Grundschule Band 1: Forschungsperspektiven und Entwicklungslinien* (S. 199–209). Münster: Waxmann.
- Schroeder, S., Würzner, K. M., Heister, J., Geyken, A. & Kliegl, R. (2015). childLex–eine lexikalische Datenbank zur Schriftsprache für Kinder im Deutschen. *Psychologische Rundschau*, 66(3), 155-165. <https://doi.org/10.1026/0033-3042/a000275>
- Schwarz, G. E. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461–464. <https://doi.org/10.1214/aos/1176344136>
- Statistik Austria (2021). *Bildung in Zahlen. 2019/20. Schlüsselindikatoren und Analysen*. Abgerufen 12.10.2021 von https://www.statistik.at/web_de/services/publikationen/5/index.html?includePage=detailedView§ionName=Bildung%2C+Kultur&pubId=508
- Thomé, G. (2019). *Deutsche Orthographie. Historisch, systematisch, didaktisch. Grundlagen der Wortschreibung*. Oldenburg: ibs-Fachverlag.
- Uppstad, P. H. & Solheim, O. J. (2007). Aspects of fluency in writing. *Journal of Psycholinguistic Research*, 36(2), 79–87. <https://doi.org/10.1007/s10936-006-9034-7>.
- Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin Review*, 14, 779–804. <https://doi.org/10.3758/BF03194105>
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. San Diego: MESA Press.
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8, 125–145. <https://doi.org/10.1177/014662168400800201>

Yen, W. M. (1993). Scaling performance assessments: strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187–213.
<https://doi.org/10.1111/j.1745-3984.1993.tb00423.x>

Tabelle 1

Darstellung der hierarchischen Wortschatzstruktur des SCHNAPP-Rechtschreibtests mit analogen Beispielwörtern.

Theoretisches Schwierigkeitsniveau	Beschreibung der Wortschatzstruktur	Beispielwort
1	Offene Vollsilbe, Reduktionssilbe (besetzt oder unbesetzt)	z.B. Rose, baden
2	Geschlossene Vollsilbe, Reduktionssilbe (besetzt oder unbesetzt)	z.B. Windel, Rinde
3	Vokalisiertes <r> in der Vollsilbe oder der Reduktionssilbe	z.B. Feder, warten
4	Vollsilbe mit komplexem Nukleus <ie>	z.B. Liebe
5	Vollsilbe mit komplexem Onset	z.B. Schwester
6	Vollsilbe mit Umlaut im Nukleus	z.B. Jäger

Tabelle 2

Deskriptive Statistiken und Rasch Modellparameter des revidierten SCHNAPP-Rechtschreibtests

Item	PR	β (SE_{β})	Infit	O-S χ^2	$Q3_a$	RMSD	<i>theoretisches . Schwierigkeitsniveau</i>
1	80%	-1.90 (0.27)	0.94	.64	0.10	0.03	1
2	89%	-2.78 (0.32)	0.94	.53	0.07	0.05	1
3	85%	-2.33 (0.29)	1.08	.64	0.07	0.04	1
4	86%	-2.47 (0.30)	1.11	.64	0.09	0.04	1
5	74%	-1.44 (0.25)	0.87	.69	0.08	0.03	1
6	82%	-2.07 (0.28)	1.00	.49	0.06	0.04	1
7	80%	-1.90 (0.27)	0.77	.53	0.06	0.08	2
8	64%	-0.79 (0.23)	1.07	.13	0.06	0.06	2
9	66%	-0.91 (0.24)	0.99	.94	0.07	0.03	2
10	57%	-0.40 (0.23)	0.84	.92	0.09	0.05	2
11	47%	0.16 (0.23)	0.68	.53	0.07	0.10	2
12	46%	0.24 (0.23)	1.05	.64	0.07	0.06	3
13	52%	-0.10 (0.23)	0.93	.64	0.07	0.03	3
14	80%	-1.84 (0.26)	0.80	.64	0.08	0.07	3
15	38%	0.69 (0.23)	0.80	.92	0.07	0.05	3
16	34%	0.89 (0.23)	0.87	.97	0.10	0.03	4
17	19%	1.96 (0.27)	0.83	.94	0.08	0.04	4
18	50%	0.01 (0.23)	0.91	.68	0.08	0.03	5
19	22%	1.70 (0.26)	1.05	.64	0.07	0.06	5
20	14%	2.46 (0.30)	1.08	.53	0.07	0.05	5
21	41%	0.54 (0.23)	1.00	.64	0.08	0.05	5
22	14%	2.46 (0.30)	0.94	.64	0.06	0.04	6

Anmerkung. $N = 148$. PR = Prozentsatz richtiger Antworten; β (SE_{β}) = Itemschwierigkeit des Rasch-Modells mit Standardfehler in Klammern; Infit = Weighted mean squared Statistik (Wright & Masters, 1982); O-S $\chi^2 = p$ -Wert von Orlando und Thissen's (2000) Chi Quadrat Test auf Itemfit mit Freiheitsgraden mit Bonferroni-Holm Korrektur; $Q3_a$ = Durchschnittliche absolute Residualkorrelation (= adjustierte $Q3$ Statistik nach Yen); $RMSD$ = Root mean squared deviation (Köhler et al., 2020).

* $p < .05$

Abbildung 1.

Screenshots der Rahmengeschichte (links) und beim Schreiben mit dem digitalen Stift (rechts). Copyright PHOÖ. Mit Genehmigung reproduziert.

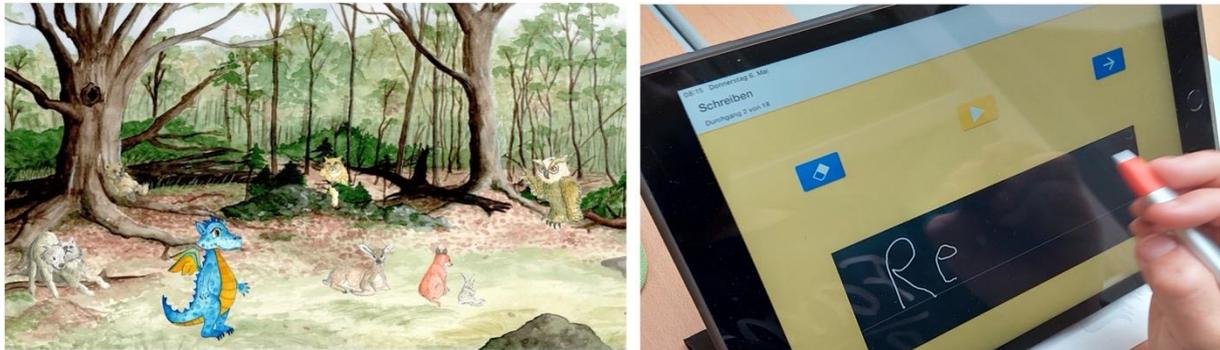
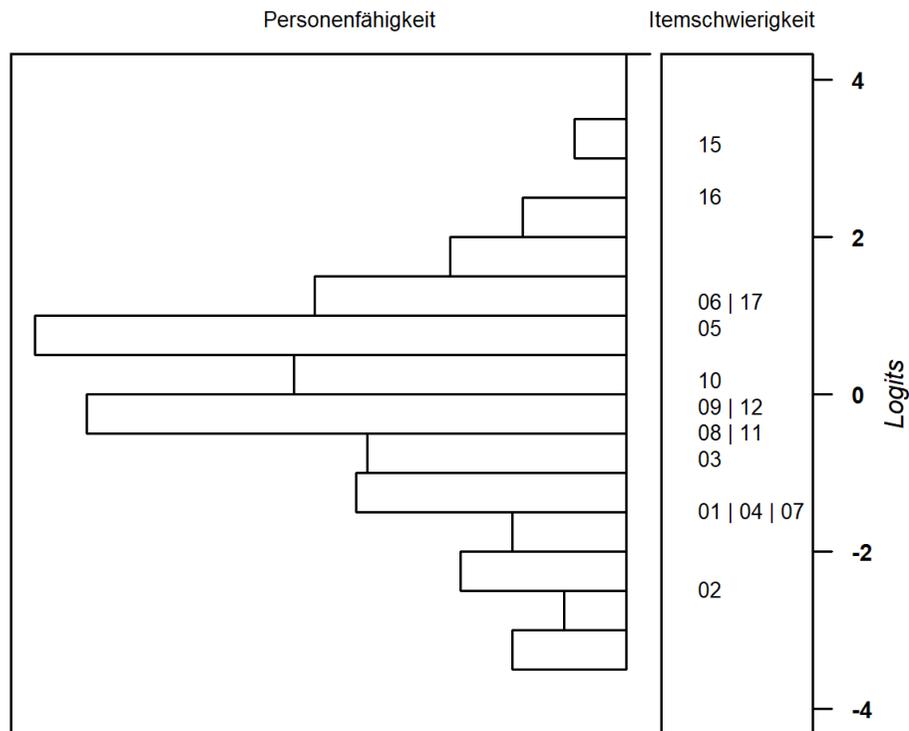


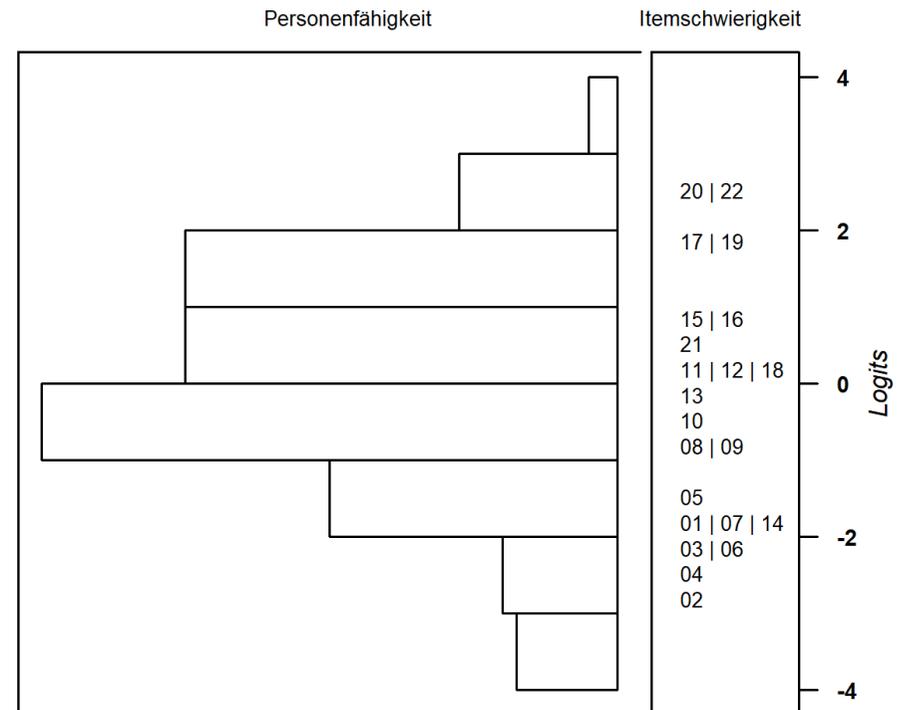
Abbildung 2.

Verteilung der Personenfähigkeiten und Itemschwierigkeiten

Erstversion des SCHNAPP-Rechtschreibtest



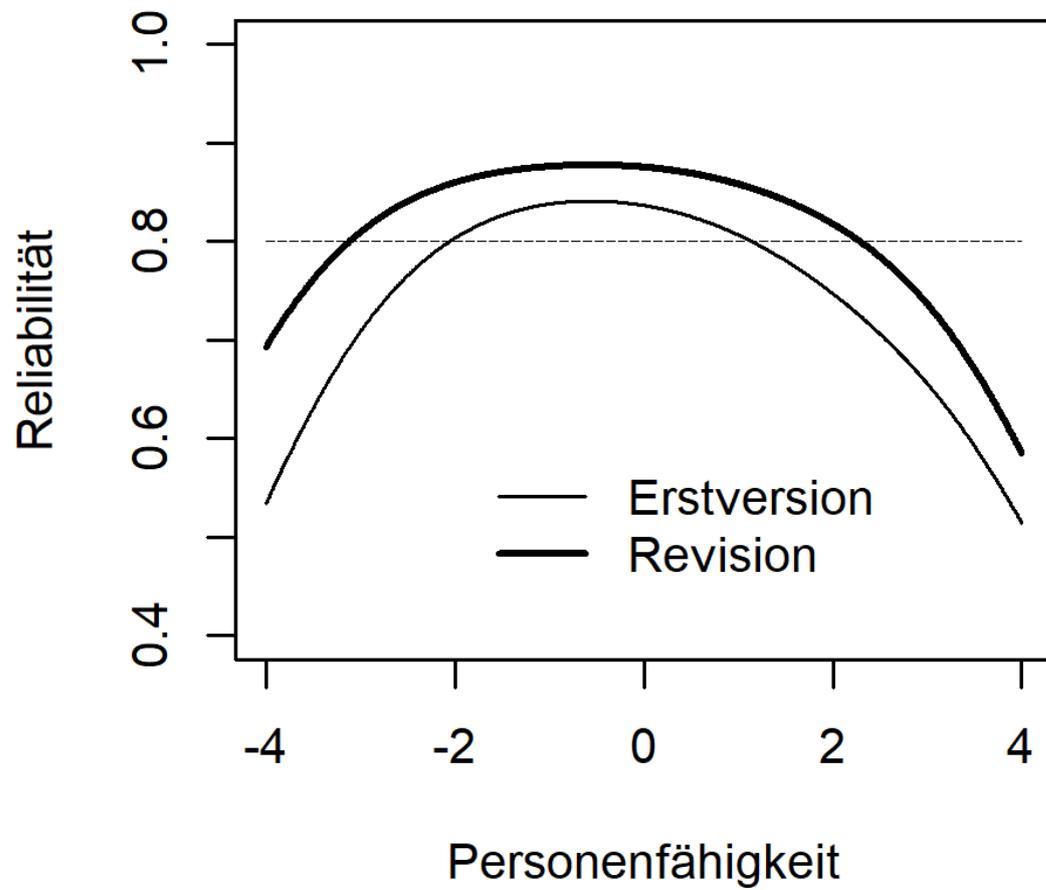
Revidierter SCHNAPP-Rechtschreibtest



Anmerkung. Die Itemnummern in der jeweils rechten Spalte entsprechen den Nummern in Tabellen 1 und S1 (im Elektronischen Supplement).

Abbildung 3.

Reliabilität des SCHNAPP-Rechtschreibtests in Abhängigkeit von der Personenfähigkeit



Anmerkung. Gestrichelte Linie stellt eine Reliabilität von .80 dar.

Elektronisches Supplement

Tabelle S1.

Deskriptive Statistiken und Rasch Modellparameter der Erstversion des SCHNAPP-Rechtschreibtests

Item	PR	β (SE_{β})	Infit	O-S χ^2	$Q3_a$	RMSD
1 (4)	78%	-1.66 (0.18)	1.02	.48	0.04	0.03
2 (5)	86%	-2.35 (0.20)	0.87	.59	0.04	0.03
3 (-) [†]	68%	-0.98 (0.16)	0.95	.48	0.07	0.03
4 (-) [†]	78%	-1.63 (0.18)	0.88	.64	0.09	0.03
5 (16)	36%	0.76 (0.16)	0.94	.70	0.09	0.03
6 (7)	29%	1.17 (0.16)	0.99	.48	0.10	0.04
7 (9)	74%	-1.34 (0.17)	0.84	.48	0.07	0.04
8 (10)	58%	-0.42 (0.16)	1.00	.48	0.06	0.04
9 (12)	51%	-0.03 (0.15)	0.80	.48	0.07	0.04
10 (-) [†]	46%	0.23 (0.16)	1.06	< .001	0.07	0.07
11 (11)	57%	-0.37 (0.16)	0.91	.87	0.06	0.03
12 (-) [†]	51%	-0.04(0.15)	0.80	.42	0.07	0.05
13 (6)	75%	-1.44 (0.17)	0.89	.48	0.04	0.02
14 (22)	75%	-1.46 (0.17)	0.85	.64	0.09	0.04
15 (20)	8%	3.06 (0.24)	0.85	.48	0.05	0.02
16 (18)	12%	2.58 (0.21)	1.07	.48	0.05	0.04
17 (-) [†]	30%	1.13 (0.16)	0.87	.48	0.06	0.03

Anmerkung. $N = 299$. Item = Itemnummer der Erstversion des SCHNAPP mit Itemnummer des revidierten SCHNAPP-Rechtschreibtests in Klammern. PR = Prozentsatz richtiger Antworten; β (SE_{β}) = Itemschwierigkeit des Rasch-Modells mit Standardfehler in Klammern; Infit = Weighted mean squared Statistik (); O-S $\chi^2 = p$ -Wert von Orlando und Thissen's (2000) Chi Quadrat Test auf Itemfit mit Bonferroni-Holm Korrektur; $Q3_a$ = Durchschnittliche absolute Residualkorrelation (= adjustierte $Q3$ Statistik nach Yen); RMSD = Root mean squared deviation (Köhler et al., 2020).

† In der revidierten Testform nicht inkludiert.