

Intelligenzmessung in drei Minuten – Evaluation des *mini-q* und Konstruktion einer Parallelversion

Julius J. Weise^{1,2} , Nicolas Becker³ und Jörn R. Sparfeldt¹ 

¹Bildungswissenschaften, Universität des Saarlandes, Saarbrücken, Deutschland

²Institut für Medizinische Biometrie, Epidemiologie und Medizinische Informatik, Universität des Saarlandes, Homburg, Deutschland

³Differentielle Psychologie und Psychodiagnostik, Universität des Saarlandes, Saarbrücken, Deutschland

Zusammenfassung: Zur ökonomischen Erfassung eines Intelligenzindikators in nur drei Minuten wurde der *mini-q* (Baudson & Preckel, 2016), als deutschsprachige Adaptation und Weiterentwicklung von Baddeleys (1968) *Reasoning Test*, mit vielversprechenden Gütekennwerten vorgestellt. Im Beitrag wurde die psychometrische Güte des *mini-q* an einer unabhängigen Studierendenstichprobe überprüft und um eine neue Parallelversion erweitert (*mini-q B*). In balancierter Reihenfolge bearbeiteten Lehramtsstudierende ($N = 163$) beide Versionen. Die Summenwertverteilungen beider Versionen wiesen auf eine angemessene Varianz und keine bedeutsame Abweichung von einer Normalverteilung hin. Beide Versionen (*mini-q/mini-q B*) korrelierten bedeutsam untereinander (Paralleltestreliabilität: $r = .76$ bzw. $.80$) sowie mit dem IST-Screening ($r = .25/.47$) und teils bedeutsam mit der Abiturnote ($r = -.07/- .25$). Implikationen zur Eignung des *mini-q* und der neuen Parallelversion *mini-q B* als zeitökonomisch einsetzbare Intelligenzscreenings werden diskutiert.

Schlüsselwörter: Intelligenz, Intelligenzmessung, Intelligenz-Screening, *mini-q*

Assessment of intelligence in three minutes – An evaluation of the *mini-q* and the construction of a parallel form

Abstract: To assess intelligence with a screening test within just three minutes, Baudson and Preckel (2016) introduced the *mini-q*, an adaptation of Baddeley's (1968) verbal *Reasoning Test*. In order to examine the promising test characteristics reported by the test authors and to introduce a new parallel form (*mini-q B*), a sample of $N = 163$ students of a teacher-education program answered both *mini-q* versions in balanced order. The distributions of the sum scores of both versions showed substantial variances and no substantial deviations from a normal distribution. Both versions correlated substantially with each other (parallel test reliability: $r = .76, r = .80$). Regarding validity, both test versions (*mini-q/mini-q B*) correlated significantly with another intelligence test ($r = .25/.47$) and partially significantly with the GPA of the high school graduation certificate (Abitur; $r = -.07/- .25$). Finally, we discuss implications regarding the usefulness of the *mini-q* and the newly developed *mini-q B* as suitable and time-economic intelligence screenings.

Keywords: intelligence, assessment of intelligence, intelligence-screening, *mini-q*

Einleitung

Die besondere Relevanz der Intelligenz in der (pädagogisch-)psychologischen Forschung ergibt sich u. a. aus bedeutsamen Korrelationen mit dem (künftigen) Schul-, Ausbildungs-, Studien- und Berufserfolg (vgl. z. B. Gottfredson, 1997; Rost, 2013; Roth, Becker, Romeyke, Schäfer, Dominick & Spinath, 2015). Zur Intelligenzmessung liegen bewährte Verfahren mit guten bis sehr guten psychometrischen Kennwerten vor (vgl. z. B. Preckel & Brüll, 2008; Sparfeldt et al., 2022). Doch eine umfassende Intelligenzdiagnostik ist für umfängliche Forschungsvorhaben häufig zu aufwändig. Somit sind ökonomische und evaluierte Intelligenzscreenings mit guten psychometrischen Eigenschaften wünschenswert. Der *mini-q* liefert bei einer Be-

arbeitungszeit von nur drei Minuten einen Intelligenzindikator mit ersten, vielversprechenden Befunden zur psychometrischen Qualität, hoher Testökonomie und freier Verfügbarkeit (Baudson & Preckel, 2016). Die verbreitete Nutzung kennzeichnet den *mini-q* als attraktives Intelligenzscreening (z. B. Leichner, Prestele, Matheis, Weis, Schmitt & Lischetzke, 2022; Schlegel & Mortillaro, 2019). Daran anknüpfend sollten in dieser Arbeit zentrale Testgütekriterien des *mini-q* überprüft und unabhängig evaluiert werden. Zudem sollte der *mini-q* um eine neu konstruierte Parallelversion ergänzt werden, was die Anwendungsgebiete substantiell erweitert.

Ausgangspunkt des *mini-q* (Baudson & Preckel, 2016) war der pragmatisch entwickelte (verbale) *Reasoning Test* von Baddeley (1968), in dem Aussage (z. B. „A is not pre-

ceded by B“) in Bezug auf eine nachfolgende Buchstabenfolge (z. B. „BA“) zu beurteilen sind. In den Aussagen werden sechs Merkmale variiert, was $2^6 = 64$ Items ergibt. Mit verbalem Material wird Baddeley zufolge schlussfolgerndes Denken ([verbales] *Reasoning*; vgl. Baudson & Preckel, 2016, zu Alternativinterpretationen) ökonomisch erfasst. Im *mini-q* nutzten Baudson und Preckel (2016) das Kontrastpaar „vorziehen“ und „ablehnen“ sowie eine Figurensequenz (Viereck, Dreieck, Kreis; Abbildung E1 im Elektronischen Supplement [ESM] 1). Die ebenfalls 64 Aussagen beziehen sich auf das stets mittig abgebildete Dreieck und dessen Abstand (dichter [„vorziehen“] oder entfernter [„ablehnen“]) zu den Nachbarfiguren. Theoretisch erfasst der *mini-q* Baudson und Preckel (2016) zufolge Reasoning sensu Carroll (1993) bzw. Kombinationen aus den Operationen Verarbeitungsgeschwindigkeit und Verarbeitungskapazität sowie figuralen und verbalen Inhalten sensu Jäger (1984).

Den *mini-q* setzten Baudson und Preckel (2016) in sieben Teilstudien ein ($N = 478$, zumeist Studierende). Für eine dreiminütige Bearbeitungszeit ergab sich ein annähernd normalverteilter Summen-Rohwert mit $M = 35.05$ ($SD = 9.70$, Range = 11–63). Als Reliabilitätsmaß wurde insbesondere eine Testhalbierungsreliabilität (*odd-even*) von $r = .98$ berichtet, also eine hohe Zuverlässigkeit. Konvergente Validitätshinweise umfassten eine signifikante Abiturnotenkorrelation ($r = -.28$) sowie mindestens (mittel-)hohe Korrelationen mit in Teilstichproben bearbeiteten Intelligenzindikatoren: figurales Reasoning (*CFT 20-R*; Weiß, 2008; $r = .51$), schlussfolgerndes Denken/allgemeine Intelligenz (*IST-Screening*; Liepmann, Beauducel, Brocke & Nettelstroth, 2012; $r = .67$), kristalline Intelligenz (*Mehrfachwahl-Wortschatz-Test*; Lehrl, 1999; $r = .32$) und Verarbeitungsgeschwindigkeit (*Ulmer-Speed-Batterie, USB*; Schmitz & Wilhelm, 2016, 2019; $r = .73$). In schrittweisen Regressionen mit umgekehrter Sequenz der Prädiktoren IST-Screening und USB sowie dem *mini-q* als Kriterium ergaben sich für beide Prädiktoren ähnliche inkrementelle Varianzaufklärungen (15% Inkrement; Varianzaufklärung des ersten Prädiktors jeweils ca. 50%). Den vergleichbaren *mini-q*-Varianzaufklärungen aus Reasoning und Verarbeitungsgeschwindigkeit entsprechend, erfasst der Test nach Baudson und Preckel (2016, S. 194) „speeded reasoning“. Weiterhin erreichten Hochbegabte eine bessere *mini-q*-Leistung als durchschnittlich Begabte. Diese Befunde sprechen für die jeweiligen Validitätsaspekte des mit dem *mini-q* als Screening-Instrument erfassten Intelligenzindikators in Forschungsprojekten mit Studierenden.

Der *mini-q* wurde inzwischen in über 20 Studien eingesetzt (z. B. Lechner et al., 2022; Nalis, Schütz & Pastukhov, 2018; Schlegel & Mortillaro, 2019). Als Reliabilitätshinweise berichteten z. B. zwei Publikationen eine *odd-even* Testhalbierungsreliabilität ($r = .96$; Basch et al., 2021; Lenhart,

Dangel & Richter, 2020) sowie weitere Cronbachs Alpha (z. B. Bouley, 2017: $\alpha = .99$; Nalis et al., 2018: $\alpha = .97$), was für eine hohe Genauigkeit der *mini-q*-Testwerte spricht. Als Validitätshinweise fanden sich bedeutsame Korrelationen mit pädagogischem Wissen ($r = .21$, Lechner et al., 2022) und einem Lerntest ($r = .34$, Wenzel & Reinhard, 2021) sowie erwartete Gruppenunterschiede im *mini-q* ($\eta^2 = .18$; Hochbegabte MENSAMitglieder > Spitzenleistende > Kontrollgruppe; Pollet & Schnell, 2017). In den gesichteten weiteren Publikationen wurden weder über die erwähnten *odd-even*-Korrelationen und Cronbachs Alpha hinausgehende andere Reliabilitätsmaße noch weitere und an unabhängigen Stichproben erhobene Konstrukt- bzw. Kriteriumsvaliditätsbefunde berichtet. Kennwerte neuer Tests sollten unabhängig überprüft und ergänzt werden; dies leisten die oben vorgestellten Befunde seit der Testvorstellung zumeist nicht. Es besteht also Forschungsbedarf. Außerdem würde eine Parallelversion eine wertvolle Ergänzung für z. B. Gruppentestungen und Testwiederholungen darstellen. Allgemein wurden für Intelligenztestwiederholungen mit einer identischen/alternativen Testversion metaanalytisch beträchtliche Zugewinne berichtet (Bias Corrected Standardized Mean Change with Raw Score Standardization = 0.37/0.23; Scharfen, Peters & Holling, 2018), die mit zunehmendem zeitlichen Abstand schrumpften.

In dieser Studie sollte der *mini-q* somit an einer unabhängigen Studierendenstichprobe überprüft sowie eine neu erstellte Parallelversion evaluiert werden. Für beide Versionen sollten (1) zentrale psychometrische Kennwerte betrachtet, (2) Hinweise zur Paralleltestreliabilität, zur konvergenten Validität mit einem weiteren Intelligenztest und zur Kriteriumsvalidität mit der Abiturnote ermittelt sowie (3) Mittelwertunterschiede zwischen beiden Testversionen und Übungeffekte analysiert werden.

Methoden

Stichprobe und Ablauf

mini-q und *mini-q B* (s. u.) wurden im Rahmen einer bildungswissenschaftlichen Vorlesung $N = 163$ Lehramtsstudierenden unterschiedlicher Schulformen und studierter Fächer in einer Gruppentestung administriert. Dazu wurden zwei Papier-Testheftversionen mit getauschter Reihenfolge beider *mini-q*-Versionen den Studierenden zufällig zugeteilt. Um beide *mini-q*-Versionen nicht direkt nacheinander zu bearbeiten, wurden dazwischen 22 Persönlichkeitsitems ohne Bedeutung für unsere Fragestellungen platziert. In einer vorhergehenden Vorlesungssitzung wurden Intelligenz (s. u.), Alter, Geschlecht und die Abiturnote erhoben ($n = 136$; Alter: $M = 23.12$; $SD = 5.09$;

77.2% weiblich, 21.3% männlich, 1.5% ohne Geschlechtsangabe). Die Abiturnote lag von $n = 135$ Studierenden vor ($M = 2.01$; $SD = 0.62$; Range 1.0–3.7).

Instrumente

In Ergänzung zum *mini-q* erstellten wir eine Parallelversion (*mini-q B*), indem wir die grafischen Elemente (Kreis, Viereck, Dreieck) durch neue Symbole (Pfeil, Raute, Stern) und die Verben „ablehnen“ und „vorziehen“ durch „abstoßen“ und „anziehen“ ersetzten (Abbildung E1 im ESM 1). Die Parallelversion ähnelt also der Ursprungsversion hinsichtlich Aufgabenform und -darstellung weitgehend. Die Instruktion der Parallelversion wurde angepasst; unverändert blieb die Aufgabenreihenfolge. Somit bestanden *mini-q* und *mini-q B* jeweils aus einer Instruktion, sechs Beispielitems und 64 Aufgaben (Bearbeitungszeit: drei Minuten). Als weiteres Intelligenzmaß setzten wir das IST-Screening (Version A; Liepmann et al., 2012) mit dem Gesamtwert als Indikator der allgemeinen Intelligenz ein.

Auswertung

Zur Auswertung des *mini-q* und des *mini-q B* wurden wie bei Baudson und Preckel (2016) jede korrekt gelöste Aufgabe mit einem Punkt und jede nicht korrekt gelöste oder unbearbeitete (ausgelassene) Aufgabe mit null Punkten bewertet und zum Summenwert addiert. Neben deskriptiven Kennwerten betrachteten wir die durchschnittliche Aufgabeninangriffnahme. Bei der Bestimmung der Itemschwierigkeiten bezogen wir nur bearbeitete Aufgaben ein. Reliabilitätshinweise lieferten Cronbachs Alpha, die odd-even- und die Paralleltestreliabilität beider *mini-q*-Versionen pro Testheft. Als Validitätshinweise berechneten wir Korrelationen und 90%-Konfidenzintervalle (KI) mit dem Summenwert des IST-Screening und der Abiturnote. Reliabilitäts- und Validitätshinweise ermittelten wir ergänzend über die Testhefte hinweg (Gesamtstichprobe) für beide Testversionen. Die Summenwerte der jeweils als ers-

tes bearbeiteten *mini-q* Version wurden mit einem *t*-Test für unabhängige Stichproben ($\alpha < .05$) verglichen. Übungseffekte, also Mittelwertunterschiede zwischen *mini-q* und *mini-q B* innerhalb von Personen, analysierten wir mit einem *t*-Test für abhängige Stichproben pro Testheft.

Ergebnisse

Im *mini-q* resultierte für die als erste/zweite bearbeitete *mini-q*-Version (Testheft 1: *mini-q*, *mini-q B*; Testheft 2: *mini-q B*, *mini-q*; $n = 81/82$) ein mittlerer Summenwert von $M = 32.79/43.54$ ($SD = 7.88/11.39$; Tabelle 1). Schiefe und Kurtosis sprachen nicht für eine substantielle Abweichung der Summenwertverteilung von einer Normalverteilung; außerdem zeigten sich keine Hinweise auf Boden- oder Deckeneffekte. Im Schnitt wurden $M = 33.75/44.63$ ($SD = 7.70/10.93$) Aufgaben bearbeitet. Die Lösungswahrscheinlichkeiten der bearbeiteten Aufgaben lagen zwischen .92–1.00/.88–1.00 ($M = .98/.97$; $SD = .02/.03$). Odd-even-Reliabilitätshinweise erreichten $r = .96/.99$ (Gesamtstichprobe: $r = .98$), Cronbachs Alpha lag bei $\alpha = .94/.96$ (.96). Der *mini-q*-Summenwert der zuerst bearbeiteten *mini-q*-Version korrelierte statistisch bedeutsam mit dem IST-Screening Summenwert ($r = .25$, $p = .042$, KI: .05, .43, $n = 69$; Gesamtstichprobe: $r = .37$, $p < .001$, KI: .24, .49, $n = 136$), nicht jedoch der Abiturnote ($r = -.07$, $p = .581$, KI: $-.26$, .13, $n = 69$; Gesamtstichprobe: $r = -.24$, $p = .006$, KI: $-.37$, $-.10$, $n = 135$).

Für den *mini-q B* ergab sich für die Bearbeitung als erste/zweite *mini-q*-Version (Testheft 2/1; $n = 82/81$) ein mittlerer Summenwert von $M = 38.24/44.00$ ($SD = 9.50/9.27$; Tabelle 1). Schiefe und Kurtosis deuteten ebenfalls keine Abweichung von einer Normalverteilung an; auch zeigten sich keine Boden- oder Deckeneffekte. Die Studierenden bearbeiteten durchschnittlich $M = 38.94/45.42$ ($SD = 9.18/9.15$) Aufgaben (Lösungswahrscheinlichkeiten bearbeiteter Items: .94–1.00/.83–1.00; $M = .99/.96$; $SD = .02/.04$). Odd-even-Reliabilitätshinweise erreichten $r = .98/.96$ (Gesamtstichprobe: $r = .97$), Cronbachs Alpha lag

Tabelle 1. Deskriptive Ergebnisse der *mini-q* Versionen für Testheft 1 (Reihenfolge: *mini-q*, *mini-q B*) und Testheft 2 (Reihenfolge: *mini-q B*, *mini-q*)

| Version | Testheft 1 ($n = 81$) | | | | Testheft 2 ($n = 82$) | | | |
|-----------------|---------------------------|-------|--------------------------|---------------------------|---------------------------|-------|--------------------------|---------------------------|
| | <i>M</i> (<i>SD</i>) | Range | Schiefe (<i>SE</i>) | Kurtosis (<i>SE</i>) | <i>M</i> (<i>SD</i>) | Range | Schiefe (<i>SE</i>) | Kurtosis (<i>SE</i>) |
| <i>mini-q</i> | 32.79 (7.88) | 12–55 | .32 (.27) | .55 (.53) | 43.54 (11.39) | 19–64 | .13 (.27) | –.66 (.53) |
| <i>mini-q B</i> | 44.00 (9.27) | 19–63 | –.05 (.27) | .22 (.53) | 38.24 (9.50) | 19–58 | .11 (.27) | –.69 (.53) |

¹ Auf Anfrage beim Erstautor erhältlich.

bei $\alpha = .95/.95$ (.95). Der *mini-q B*-Summenwert der zuerst bearbeiteten *mini-q*-Version korrelierte statistisch signifikant mit dem IST-Screening ($r = .47, p < .001, KI: .29, .61, n = 67$; Gesamtstichprobe: $r = .32, p < .001, KI: .19, .44, n = 136$) und der Abiturnote ($r = -.25, p = .046, KI: -.43, -.05, n = 66$; Gesamtstichprobe: $r = -.14, p = .118, KI: -.27, .01, n = 135$).

Hinweise auf die Paralleltestrelia­bilität lieferten Korrelationen (beide $p < .001$) der *mini-q*-Versionen in Testheft 1 ($r = .76, KI: .67, .83$) bzw. 2 ($r = .80, KI: .73, .86$). Die Summenwerte der als zweites bearbeiteten Testversion lagen jeweils statistisch bedeutsam über denen der als erstes bearbeiteten ($t(80/81) = 16.50/7.07$, beide $p < .001, d = 1.84/0.78$). Ein Vergleich der jeweils als erstes bearbeiteten Testversion (Testheft 1: *mini-q*, Testheft 2: *mini-q B*) zeigte einen statistisch signifikant höheren mittleren Summenwert im *mini-q B* als im *mini-q* (Test auf Varianzhomogenität: $F(1,161) = 5.60, p = .019$; t -Test, unabhängige Stichproben: $t(156.45) = -4.00, p < .001, d = 0.63$).

Diskussion

Ausgangspunkt waren die von Baudson und Preckel (2016) berichteten vielversprechenden Ergebnisse zur Erfassung eines Intelligenzindikators in nur drei Minuten. Neben einer Prüfung zentraler Ergebnisse konnten wir die *mini-q*-Befunde in verschiedene Richtungen bedeutsam erweitern – insbesondere um eine neue Parallelversion (*mini-q B*). Bei Lehramtsstudierenden dokumentierten wir für beide Testversionen mindestens zufriedenstellende Verteilungskennwerte und Reliabilitätshinweise (inklusive den neuen Paralleltestrelia­bilitätshinweisen) sowie teils erwartungstreue konvergente Korrelationen mit einem weiteren Intelligenzindikator und der Abiturnote. Dies geht über die eingangs erwähnten weiteren gesichteten Studien hinaus², in denen der *mini-q* häufig lediglich zur Erfassung einer Kovariate oder Kontrollvariable genutzt wurde. Mittelwertunterschiede verwiesen auf die erstmalig betrachtete Bedeutung der Testversion und einer vorherigen Testbearbeitung.

Die Verteilungseigenschaften des *mini-q* sind mindestens gut ausgefallen (ähnlich Baudson & Preckel, 2016): Der Mittelwert liegt bei uns ($M = 32.79$, erst-bearbeitete Testversion) nah am theoretischen Mittelwert bei korrekter Bearbeitung der Hälfte der 64 Aufgaben; gemeinsam mit der Verteilung verweist dies auf eine Diskrimination unterschiedlich leistungsstarker Personen. Deskriptiv löste unsere Stichprobe im Mittel weniger Aufgaben als die Gesamtstichprobe der Originalpublikation bei dreiminütiger

Bearbeitungszeit ($d = 0.24$). Ähnlich gute Verteilungscharakteristika ermittelten wir für den *mini-q B*.

Ebenso fielen die Hinweise zur Messgenauigkeit beider *mini-q* Versionen mindestens gut aus: Baudson und Preckel (2016) berichteten für den *mini-q* eine *odd-even* Testhalbierungsreliabilität von $r_{tt} = .98$; unsere Werte lagen für beide *mini-q* Versionen sowie für die als erstes und zweites bearbeitete Version in gleicher Größenordnung ($.96 \leq r_{tt} \leq .99$). Üblicherweise fallen *odd-even*-Reliabilitätswerte für gespedete Leistungstests mit hoher Item-Lösungswahrscheinlichkeit hoch aus. Die Paralleltestrelia­bilität beider *mini-q* Versionen ($r_{tt} = .76/.80$) erweitert die überzeugenden Messgenauigkeitsbefunde bedeutsam.

Als konvergente Validitätshinweise ermittelten wir Korrelationen mit einem Intelligenztest und der Abiturnote. Die Korrelation des *mini-q* mit dem IST-Screening differierte zwar in erwarteter Richtung bedeutsam von null ($r = .25$), fiel allerdings numerisch niedriger als die Erwartung der Korrelation zweier Intelligenztests aus (z. B. Baudson & Preckel, 2016: $r = .67$). Interessanterweise liegt der überzeugendere konvergente Validitätshinweis für den *mini-q B* ($r = .47$) numerisch zwischen beiden Werten. Mit der Abiturnote hing der *mini-q* bei uns erstaunlicherweise nicht bedeutsam zusammen ($r = -.07$); der *mini-q B* korrelierte hingegen bedeutsam mit der Abiturnote ($r = -.25$), und zwar ähnlich wie von den Testautorinnen für die Ursprungsform berichtet ($r = -.28$). Insofern stehen diese Validitätshinweise nur zum Teil im Einklang mit den Erwartungen sowie den Befunden zum Originalverfahren; zum Teil sind sie nicht erwartungskonform und stützen mithin die Validitätsannahme der *mini-q*-Testwerte nur partiell. Dies gilt ebenfalls für den *mini-q B* trotz etwas erwartungskonformerer und damit eine Validitätsvermutung teils stärker stützender Ergebnisse. Dies überrascht, da die Reliabilitätshinweise sowie die erwartungskonformen Validitätshinweise gegen prinzipielle Einwände wie eine zu homogene Stichprobe sprechen. Unsere Stichprobe umfasste zwar „nur“ Lehramtsstudierende (was die Generalisierbarkeit einschränkt), jedoch verschiedener Schulformen und -fächer, war also diesbezüglich eine recht heterogene Studierendenstichprobe, wenn auch eine größere und noch heterogenere Stichprobe wünschenswert ist. Prinzipiell sollten Testverfahren auch in einer solchen Stichprobe überzeugen. Wir gehen von einer gewissen Generalisierbarkeit unserer Befunde aus, auch wenn zur weiteren Einordnung weiterer Forschungsbedarf besteht.

Beachtenswert sind der Versions- sowie der Übungs- bzw. Testwiederholungseffekt. Beim Vergleich der jeweils als erstes bearbeiteten *mini-q*-Versionen überrascht der substantiell höhere Mittelwert im *mini-q B* ($d = 0.63$), was beim Vergleich beider Varianten zu beachten ist. Die randomisierte Zuteilung der Testhefte und die Berücksich-

² Beachtenswert ist auch die kürzlich erschienene *mini-q*-Studie von Schubert, Löffler, Wiebel, Kaulhausen & Baudson (2024).

tigung der jeweils als erstes bearbeiteten Version beim Versionsvergleich sprechen hier für eine hohe interne Validität. Eine Klärung der Versionsunterschiede – z. B. leichtere Verständlichkeit der Verben und/oder der graphischen Elemente im *mini-q B* als im *mini-q* – erfordert weitere Forschung. Der Mittelwertunterschied zwischen beiden Versionen verweist auf die Auswirkungen selbst geringer Modifikationen; ein direkter Vergleich der Messwerte ohne Berücksichtigung des Versioneffekts wäre mit Fehlinterpretationen verbunden. Insofern finden sich zwischen beiden Parallelversionen Messäquivalenzeinschränkungen. In Bezug auf Übungs- bzw. Testwiederholungseffekte konnten wir jeweils Testwertsteigerungen von der ersten zur zweiten Bearbeitung zeigen (vgl. Schneider, Becker, Krieger, Spinath & Sparfeldt, 2020, zur Interaktion zwischen Testversion und Bearbeitungsreihenfolge bei figuralen Matrizen), was die Frage der Testautorinnen nach der Stabilität und Effekten einer Testwiederholung aufgreift. Diese Ergebnisse übersteigen den eingangs berichteten metaanalytischen mittleren Testwiederholungseffekt in Intelligenztests mit alternativen Testversionen (Scharfen et al., 2018). Interessanterweise differieren die Mittelwerte der jeweils als zweites bearbeiteten *mini-q*-Versionen kaum (Tabelle 1; $d = 0.04$). Künftigen Studien bleibt eine Klärung dieser Befunde vorbehalten.

Auch wenn die Ergebnisse zu *mini-q* und *mini-q B* teils vielversprechend sind und teils für die Eignung als Intelligenz-Screening in drei Minuten sprechen, ersetzt dies keine umfängliche Intelligenzdiagnostik (vgl. Baudson & Preckel, 2016). Möchte man die Intelligenzstruktur oder das -profil erfassen, ist eine breitere Intelligenzdiagnostik erforderlich. Abschließend weisen wir trotz der sich aus den Limitationen ergebenden Forschungsbedarfe auf Vorteile des *mini-q* und des neuen *mini-q B* hin. Diese erlauben eine sehr zeitökonomische Erfassung eines Intelligenzindikators und somit einer zentralen Variable vieler (pädagogisch-)psychologischer Forschungen mit einem Screening-Instrument.

Elektronisches Supplement (ESM)

Das elektronische Supplement ist mit der Online-Version dieses Artikels verfügbar unter <https://doi.org/10.1024/1010-0652/a000384>

ESM 1. Diese Datei enthält Beispielitems aus *mini-q* und *mini-q B*. (Abbildung E1).

Literatur

Baddeley, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science*, 10, 341 – 342.

- Basch, J. M., Brenner, F., Melchers, K. G., Krumm, S., Dräger, L., Herzer, H. & Schuwerk, E. (2021). A good thing takes time: The role of preparation time in asynchronous video interviews. *International Journal of Selection and Assessment*, 29, 378 – 392.
- Baudson, T. G. & Preckel, F. (2016). *mini-q*: Intelligenzscreening in drei Minuten. *Diagnostica*, 62, 182 – 197.
- Bouley, F. (2017). *Kompetenzerwerb im Rechnungswesenunterricht*. Berlin: Springer Fachmedien.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- Gottfredson, L. S. (1997). Mainstream science on intelligence. An editorial with 52 signatories, history, and bibliography. *Intelligence*, 24, 13 – 23.
- Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35, 21 – 35.
- Lehrl, S. (1999). *Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B)* (5. Aufl.). Balingen: Spitta.
- Lechner, N., Prestele, E., Matheis, S., Weis, S., Schmitt, M. & Lischetzke, T. (2022). Lehramt-Studienwahlmotivation sagt Zielorientierungen vorher, pädagogisches Wissen und selbst eingeschätzte Kompetenz aber nur teilweise. *Zeitschrift für Pädagogische Psychologie* [Vorab-Onlinepublikation]. <https://doi.org/10.1024/1010-0652/a000348>
- Lenhart, J., Dangel, J. & Richter, T. (2020). The relationship between lifetime book reading and empathy in adolescents: Examining transportability as a moderator. *Psychology of Aesthetics, Creativity, and the Arts*, 16, 679 – 693.
- Liepmann, D., Beauducel, A., Brocke, B. & Nettelstroth, W. (2012). *IST-Screening*. Göttingen: Hogrefe.
- Nalis, D., Schütz, A. & Pastukhov, A. (2018). The Bamberg Trucking Game: A paradigm for assessing the detection of win-win solutions in a potential conflict scenario. *Frontiers in Psychology*, 9, 138.
- Pollet, E. & Schnell, T. (2017). Brilliant: But what for? Meaning and subjective well-being in the lives of intellectually gifted and academically high-achieving adults. *Journal of Happiness Studies*, 18, 1459 – 1484.
- Preckel, F. & Brüll, M. (2008). *Intelligenztests*. München: Reinhardt.
- Rost, D. H. (2013). *Handbuch Intelligenz*. Weinheim: Beltz.
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F. & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118 – 137.
- Scharfen, J., Peters, J. M. & Holling, H. (2018). Retest effects in cognitive ability tests: A meta-analysis. *Intelligence*, 67, 44 – 66.
- Schlegel, K. & Mortillaro, M. (2019). The Geneva Emotional Competence Test (GECO): An ability measure of workplace emotional intelligence. *Journal of Applied Psychology*, 104, 559 – 580.
- Schmitz, F. & Wilhelm, O. (2016). Modeling mental speed: Decomposing response time distributions in elementary cognitive tasks and relations with working memory capacity and fluid intelligence. *Journal of Intelligence*, 4, 1 – 23.
- Schmitz, F. & Wilhelm, O. (2019). Mene mene tekkel upharsin: Clerical speed and elementary cognitive speed are different by virtue of test mode only. *Journal of Intelligence*, 7, 16.
- Schneider, B., Becker, N., Krieger, F., Spinath, F. M. & Sparfeldt, J. R. (2020). Teaching the underlying rules of figural matrices in a short video increases test scores. *Intelligence*, 82, 101473.
- Schubert, A.-L., Löffler, C., Wiebel, C., Kaulhausen, F. & Baudson, T. G. (2024). Don't waste your time measuring intelligence: Further evidence for the validity of a three-minute speeded reasoning test. *Intelligence*, 102, 101804.
- Sparfeldt, J. R., Becker, N., Greiff, S., Kersting, M., König, C. J., Lang, J. W. & Beauducel, A. (2022). Intelligenz(tests) verstehen und missverstehen. *Psychologische Rundschau*, 73, 161 – 172.
- Weiß, R. H. (2008). *Grundintelligenztest Skala 2 – Revision (CFT 20-R) mit Wortschatztest und Zahlenfolgetest – Revision (WS/ZF-R)*. Göttingen: Hogrefe.

Wenzel, K. & Reinhard, M. A. (2021). Learning with a double-edged sword? beneficial and detrimental effects of learning tests – taking a first look at linkages among tests, later learning outcomes, stress perceptions, and intelligence. *Frontiers in Psychology*, 12, 693585.

Historie

Manuskript eingereicht: 14.03.2023

Manuskript nach Revision angenommen: 03.04.2024

Onlineveröffentlichung: 22.04.2024

Förderung

Open-Access-Veröffentlichung ermöglicht durch die Saarländische Universitäts- und Landesbibliothek.

ORCID

Julius Johannes Weise

 <https://orcid.org/0000-0002-0908-5579>

Jörn Sparfeldt

 <https://orcid.org/0000-0002-7715-0651>

Julius J. Weise

Universität des Saarlandes

Bildungswissenschaften

Campus A5 4

66123 Saarbrücken

Deutschland

julius.weise@uni-saarland.de