

Intelligenz und Subjektives Wohlbefinden -
Der Beitrag differentieller Psychologie in
gesellschaftlichen Umbruchsphasen

Dissertation

zur Erlangung des akademischen Grades
eines Doktors der Naturwissenschaften der Fakultät HW
Bereich Empirische Humanwissenschaften
der Universität des Saarlandes

vorgelegt von

Julie Levacher

aus Saarlouis

Saarbrücken, 2023

Dekan:

Prof. Dr. Peter Loos, Universität des Saarlandes

Berichterstatter:innen:

PD Dr. Nicolas Becker, Universität Greifswald

Prof. Dr. Monika Equit, Universität des Saarlandes

Tag der Disputation:

18.07.2023

Inhaltsverzeichnis

Inhaltsverzeichnis.....	I
Abbildungsverzeichnis.....	III
Tabellenverzeichnis.....	IV
Abkürzungsverzeichnis.....	V
Überblick über die relevanten Studien.....	VI
1 Einleitung.....	1
2 Teil I – Studierendenauswahlverfahren.....	2
2.1 Hinführung.....	2
2.2 Intelligenzdiagnostik im Kontext der Studierendenauswahl.....	3
2.3 Konstruktvalidität kognitiver Leistungstests zur Selektion von Medizin-Studierenden.....	10
2.4 Der Einfluss von Testvorbereitung auf die Güte eines kognitiven Leistungstests.....	13
2.5 Zusammenhang zwischen Antwortgeschwindigkeit und erfolgreicher Beantwortung kognitiver Leistungstests.....	17
3 Teil II – Die COVID-19 Pandemie.....	22
3.1 Hinführung.....	22
3.2 Der Einfluss des ersten Lockdowns auf das Wohlbefinden.....	23
3.3 Der Einfluss der Big Five auf das Wohlbefinden.....	30
3.4 Der Einfluss von Systemrelevanz, Risikogruppenzugehörigkeit und Geschlecht auf das Wohlbefinden.....	36
4 Zusammenfassende Betrachtung.....	38
4.1 Zusammenfassung der wichtigsten Ergebnisse.....	38

4.2	Praktische Implikationen für zukünftige Studierendenauswahltests im Hinblick auf die COVID-19 Pandemie	39
	Literaturverzeichnis.....	41
	Anhänge	56
	Anhang A. – Beispielaufgabe TMS: Textverständnis	56
	Anhang B. – Danksagung.....	59
	Anhang C. – Studien als Grundlage der kumulativen Promotion.....	61

Abbildungsverzeichnis

Abbildung 1. Beispielitem des Untertests relationales Schließen.....	5
Abbildung 2. Beispielitem des Untertests arithmetisches Problemlösen	6
Abbildung 3. Beispielitem des Hamburger Naturwissenschaftstests aus dem Bereich Physik.....	6
Abbildung 4. Beispielitem des konstruktionsbasierten, figuralen Matrizen- tests.....	7
Abbildung 5. Beispielitem der TMS-Subskala Diagramme und Tabellen	8
Abbildung 6. Beispielitem der TMS-Subskala medizinisch- naturwissenschaftliches Grundverständnis	9
Abbildung 7. Beispielitem der TMS-Subskala quantitative und formale Probleme....	9
Abbildung 8. Strukturgleichungsmodelle zur Schätzung allgemeiner und spezifischer Faktoren beider Zulassungstests	12
Abbildung 9. Wahrscheinlichkeit einer korrekten Antwort bei figuralen Matrizen als Funktion der Antwortzeit	19
Abbildung 10. Verteilung des Wohlbefindens über die Messzeitpunkte hinweg.....	26
Abbildung 11. Quadratisches Wachstumskurvenmodell für Lebenszufriedenheit und Einsamkeit.....	27
Abbildung 12. Latentes Wachstumskurvenmodell für Stress und psychische Belastung	27

Tabellenverzeichnis

Tabelle 1. Part-whole-korrigierte Korrelationen der Summenwerte aller Subtests.....	11
Tabelle 2. Passung der vier getesteten Modelle zur Erfassung der Konstruktvalidität	13
Tabelle 3. Ergebnisse der MGCFAs zum Vergleich der Faktorstruktur des Matrizentests zwischen beiden Gruppen	17
Tabelle 4. Deskriptive Statistik und Korrelation der ToT-Effekte	21
Tabelle 5. Auswirkungen der Persönlichkeit und des Alters auf das Wohlbefinden.....	32

Abkürzungsverzeichnis

AP	Arithmetisches Problemlösen
DT	Diagramme und Tabellen
FM	Figurale Matrizen
<i>g</i>	Generalfaktor der Intelligenz
G _c	Kristalline Intelligenz
G _f	Figurale Intelligenz
G _s	Kognitive Verarbeitungsgeschwindigkeit
G _v	Visuelle Wahrnehmung
HAM-Nat	Hamburger Naturwissenschaftstest
LGCM	Latentes Wachstumskurvenmodell
MGCFA	Konfirmatorische Multigruppen-Faktorenanalyse
MNG	Medizinisch-naturwissenschaftliches Grundverständnis
Nat	Basisskala des Hamburger Naturwissenschaftstests
QFP	Quantitative und Formale Probleme
RS	Relationales Schließen
SES	Sozioökonomischer Status
SOEP	Sozioökonomisches Panel
SWB	Subjektives Wohlbefinden
TMS	Test für medizinische Studiengänge
ToT	Time on Task
TV	Textverständnis

Überblick über die relevanten Studien

Die vorliegende Dissertation beinhaltet insgesamt vier Studien. Dabei wurden die Studien II und IV bereits in international anerkannten wissenschaftlichen Zeitschriften mit Peer-Review publiziert, Studie I zur Publikation akzeptiert und Studie III - zum Zeitpunkt der Abgabe der vorliegenden Arbeit - zur Veröffentlichung eingereicht. Studie III befindet sich aktuell entsprechend im Begutachtungsprozess.

- Studie I Levacher, J., Koch, M., Stegt, S., Hissbach, J., Spinath, F. M., Escher, M., & Becker, N. (2023). The construct validity of the main student selection tests for medical studies in Germany. *Frontiers in Education, Section Assessment, Testing and Applied Measurement*.
<https://doi.org/10.3389/feduc.2023.1120129>
- Studie II¹ Levacher, J., Koch, M., Hissbach, J., Spinath, F. M., & Becker, N. (2021). You can play the game without knowing the rules – But you're better off knowing them: The influence of rule knowledge on Figural Matrices Tests. *European Journal of Psychological Assessment, 38*, 15-23. <https://doi.org/10.1027/1015-5759/a000637>
- Studie III Krämer R., Koch M., Levacher J., & Schmitz F. (2023). Testing Replicability and Generalizability of the Time on Task Effect. *Journal of Intelligence, 11(5)*, 82.
- Studie IV² Levacher, J., Spinath, F. M., Becker, N., & Hahn, E. (2023). How did the beginnings of the global COVID-19 pandemic affect mental well-being? *PLOS ONE, 18(1)*. <https://doi.org/10.1371/journal.pone.0279753>

¹ Studie II wurde am 28. Februar 2023 publiziert.

² Studie IV wurde am 25. April 2023 akzeptiert und am 28. April 2023 publiziert.

1 Einleitung

Die vorliegende Dissertation umfasst differentialpsychologische Forschung zu zwei gesellschaftskritischen Themen der letzten fünf Jahre. Als Reaktion auf das Bundesverfassungsgerichtsurteil von Dezember 2017 zur Neugestaltung von Hochschulauswahlverfahren, setzt Kapitel 2 das Thema Intelligenzdiagnostik im Rahmen von Studierendenauswahl in den Fokus. Die vorliegende Arbeit führt zunächst den aktuellen Stand der Studierendenauswahlverfahren in Deutschland ein, definiert Anforderungen an einen Studierendenauswahltest und stellt verschiedene Skalen der beiden etablierten Auswahltests des Fachs Medizin vor (Kapitel 2.1 und Kapitel 2.2). Im Anschluss wird in Kapitel 2.3 eine Übersicht über die Konstruktvalidität der beiden etablierten Testverfahren gegeben (Studie I) und in Kapitel 2.4 der Einfluss von Testvorbereitung auf die psychometrische Güte von Auswahlverfahren am Beispiel eines figuralen Matrizen tests erläutert (Studie II). Überdies findet zusätzlich eine Betrachtung der Time on Task-Effekte bei einem der beiden etablierten Studierendeneignungstests statt (Kapitel 2.5; Studie III). Mit Hilfe von Kapitel 2 sollen folglich erste Erkenntnisse zur Neugestaltung von diagnostischen Hochschulauswahlverfahren geliefert werden. Die daraus resultierende Verbesserung der Selektion soll als Grundlage für eine langfristig hohe fachliche Fähigkeit der praktizierenden Ärzte in Deutschland dienen. Dabei wurde die Wichtigkeit von fähigem medizinischem Personal im Gesundheitssystem im Rahmen der COVID-19 Pandemie besonders salient. Diese nie dagewesenen Situation, welche weitreichende Einschränkungen im Alltag und unter anderem den Ausfall von Studierendenauswahltestungen mit sich brachte, steht im Fokus des zweiten Teils dieser Dissertation (Kapitel 3). Nach einer kurzen Hinführung (Kapitel 3.1) wird die Auswirkung des ersten COVID-19 Lockdowns in Deutschland auf das Wohlbefinden anhand von Studie IV erläutert (Kapitel 3.2). Zudem wird der Verlauf während des ersten Lockdowns in Deutschland in Abhängigkeit der Big Five unter Kontrolle des Alters (Kapitel 3.3) und weiterer individueller Unterschiede (Systemrelevanz, Zugehörigkeit zur Risikogruppe, Geschlecht; Kapitel 3.4) analysiert. Die Dissertationsarbeit schließt mit einer Zusammenfassung aller vier Studien (Kapitel 4.1) und einer Diskussion zu den praktischen Implikationen für zukünftige Studierendenauswahlverfahren (Kapitel 4.2).

2 Teil I – Studierendenauswahlverfahren

Im folgenden Kapitel wird ein kurzer Überblick über den theoretischen Rahmen des ersten Teils des Dissertationsprojektes gegeben. Hierzu wird zunächst die Intelligenzdiagnostik im Rahmen von Studierendenauswahlverfahren betrachtet und es werden bisherige Methoden zur Selektion von Studierenden in Deutschland vorgestellt. Anschließend erfolgt eine Erläuterung zur Konstruktvalidität der etablierten Verfahren. Nachdem überdies der Einfluss von Testvorbereitung auf die Güte der Testverfahren diskutiert wurde, werden empirische Zusammenhänge zwischen Antwortgeschwindigkeit und erfolgreicher Beantwortung aufgezeigt.

2.1 Hinführung

Generell kommen Auswahlverfahren immer dann zum Einsatz, wenn sich mehr Personen um einen Studienplatz bewerben als Plätze vorhanden sind. Es ist daher besonders wichtig zu beachten, dass Auswahlentscheidungen hohe Ansprüche an die Validität und Fairness stellen. Dies ist besonders relevant, da diese Entscheidung abgelehnten Bewerbern die Möglichkeit verwehrt, ein Studium in dem angestrebten Fach am bevorzugten Studienort zu beginnen. Die Verweigerung der Zulassung hat somit Auswirkungen auf die Möglichkeiten der zukünftigen Beschäftigung und die Chancen der beruflichen Selbstverwirklichung (Formazin et al., 2011). Dennoch sind in Deutschland Studierendenauswahlverfahren kaum verbreitet, obwohl etwa 43% der Studiengänge in Deutschland zulassungsbeschränkt sind (Hochschulrektorenkonferenz, 2021). Überdies sind die Hochschulen aufgrund der föderalen Struktur des Bildungssystems in Deutschland souverän in der Entscheidung über die Auswahlkriterien für ihre Studierenden.

In den meisten Fällen erfolgt die Selektion potenzieller Studierender aller zulassungsbeschränkten Studiengänge in Deutschland (z. B. Medizin, Psychologie, Pharmazie) über die Abiturdurchschnittsnoten der Bewerbenden, indem der Numerus Clausus in jedem Bewerbungsdurchgang neu festgelegt wird. Die häufige Verwendung von Schulnoten als Kriterium für die Auswahl von Studierenden lässt sich mit ihrer hohen prognostischen Validität für den späteren Studienerfolg begründen (Brandstätter & Farthofer, 2002; Gold & Souvignier, 2005; Schmidt-Atzert, 2005;

Steyer et al., 2005; Trapmann et al., 2007). Obwohl die hohe Validität der Abiturdurchschnittsnote auch metaanalytisch gestützt wird ($\rho = .53$, korrigiert für Reliabilität der Studiennoten und Varianzeinschränkung durch Selektion; Trapmann et al., 2007), zeigt die Metaanalyse auch, dass Einflüsse des Bundeslandes, der Fächer und der Zeitspanne zwischen Schul- und Studienleistung diesen Zusammenhang moderieren. Gründe für diese Unterschiede sind vielschichtig und reichen von unterschiedlichen Bewertungsmaßstäben über Unterschiede in der Fächerwahl bis hin zum Einfluss des sozioökonomischen Status (SES) (Dietze, 2011; Helbig & Nikolai, 2015; Köller & Maaz, 2017; Neumann et al., 2009; Paulus et al., 2021).

Basierend auf diesen Gegebenheiten entschied das Bundesverfassungsgericht im Dezember 2017, dass die bisherigen Selektionsarten das Recht auf freie Berufswahl einschränken und das aktuelle Hochschulauswahlverfahren im Fach Medizin daher reformiert werden muss (BVerfG, 2017). Die Vergabe von Studienplätzen sollte sich entsprechend weniger an der Durchschnittsnote orientieren als vielmehr an der Eignung der Bewerber für das jeweilige Fach und der damit einhergehenden Eignung für die berufliche Tätigkeit nach dem Studium. Da nicht-kognitive Tests (z. B. Persönlichkeits- oder Interessenstests) leicht manipuliert werden können (Alliger & Dwight, 2000; Barrick & Mount, 1996; Furnham, 1990; Griffith et al., 2007; Rentzsch & Schütz, 2009; Robie et al., 2007; Viswesvaran & Ones, 1999), hat sich der Einsatz von Tests, die kognitive Fähigkeiten erfassen, bewährt. Die derzeitige Praxis zur Selektion von Medizinstudierenden sieht vor, dass die Universitäten einen von zwei explizit für diese spezifische Auswahl entwickelten Tests zur Erfassung kognitiver Fähigkeiten verwenden (Schwibbe et al., 2018): den Hamburger Naturwissenschaftstest (HAM-Nat; Hissbach et al., 2011) oder den Test für medizinische Studiengänge (TMS; Kadmon et al., 2012).

2.2 Intelligenzdiagnostik im Kontext der Studierendenauswahl

Intelligenzdiagnostik weist in der Psychologie eine lange Forschungsgeschichte mit einem großen wissenschaftlichen Erkenntnisgewinn und hoher praktischer Relevanz auf (Sparfeldt et al., 2022). Intelligenz gilt daher als zentrales Konstrukt für differenzielle psychologische Forschung. Allgemein wird Intelligenz als eine breite kognitive Fähigkeit beschrieben, die das Verständnis von komplexen Ideen, die

Anpassungsfähigkeit an Umweltbedingungen, das Lernen aus Erfahrung und das Lösen von Problem durch Analyse beinhaltet (Neisser et al., 1996). Entsprechend gilt es als gesichert, dass sich Intelligenz aus vielen verschiedenen Fähigkeitsfacetten zusammensetzt. Heutzutage wird davon ausgegangen, dass dem Konstrukt der Intelligenz eine hierarchische Struktur zugrunde liegt (McGrew, 2009). Diese beinhaltet an oberster Stelle den von Spearman (1904) postulierten Generalfaktor der Intelligenz (g). Dieser wurde als gemeinsamer Varianzanteil mehrerer psychologischer Leistungstests identifiziert. Die darunterliegende Hierarchieebene beinhaltet mehrere breite Intelligenzfacetten wie Cattells (1963) kristalline und fluide Intelligenz. Während die fluide Intelligenz (G_f) nicht mit spezifischem Wissen, sondern eher mit der Fähigkeit, sich an neue Situationen anzupassen, in Verbindung gebracht wird, spiegelt die kristalline Intelligenz (G_c) Wissen aus früheren Lernprozessen und Erfahrungen wider. Neuere Ansätze nehmen auf dieser Hierarchieebene zusätzlich weitere Faktoren der Wahrnehmung (z. B. visuelle Wahrnehmung; G_v), der Informationsverarbeitung (z. B. kognitive Verarbeitungsgeschwindigkeit; G_s) und der Sorgfalt wie auch spezifische sprachliche und numerische Fähigkeiten auf (McGrew, 2009). Es ist daher wenig überraschend, dass Intelligenz mit vielen unterschiedlichen Facetten, wie zum Beispiel mit Kreativität (Kim, 2005) und dem Berufserfolg einer Person (Schmidt & Hunter, 1998), korreliert ist. Folglich fungiert Intelligenz als zentrale Erklärung interindividueller Leistungsunterschiede und stellt überdies einen guten Prädiktor für unterschiedliche Verhaltensweisen dar. Daher können beispielsweise unter anderem adaptives Verhalten (Alexander & Reynolds, 2020), Betrugsversuche bei Klausuren (Cuadrado et al., 2021) sowie die Schulleistungen (Roth et al., 2015) durch Intelligenz vorhergesagt werden.

Bei Testverfahren zur Erfassung kognitiver Fähigkeiten hat sich gezeigt, dass diese hoch korreliert sind, was die Extraktion des allgemeinen g -Faktors ermöglicht (Spearman, 1904). Der g -Faktor steht im Zusammenhang mit schulischen Leistungen $\rho = .54$ (Roth et al., 2015), Bildungsabschlüssen $r = .81$ (Deary et al., 2007) sowie beruflichem Erfolg $r = .51$ (Schmidt & Hunter, 1998). Daher sollten Zulassungstests den g -Faktor möglichst breit abbilden und zumindest G_f und G_c erfassen. Die beiden aktuell an deutschen Medizinfakultäten verwendeten Verfahren sind jedoch entlang

eines Kontinuums als Studierfähigkeitstests (Messung G_f ; TMS) oder als Kenntnistests (Messung G_c ; Basisskala des HAM-Nat) einzuordnen (Schwibbe et al., 2018). Im Folgenden werden zunächst die vier Skalen des HAM-Nat vorgestellt. Im Anschluss werden aus ökonomischen Gründen vier der insgesamt neun Subtests des TMS näher betrachtet. Zudem werden die in Studie I gefundenen internen Konsistenzen für jede Skala erwähnt.

HAM-Nat

Relationales Schließen (RS)

Der Untertest relationales Schließen zielt auf die Beurteilung der verbalen Fähigkeit ab und erfasst somit verbales Schlussfolgern. Ziel der Aufgabe ist es, Schlussfolgerungen anhand eines Textes abzuleiten. Dazu müssen verschiedene Prämissen verglichen und durch Beziehungen miteinander verbunden werden (Abbildung 1). Diese Skala weist eine akzeptable Reliabilität auf ($\alpha = .73$).

- Bei Dominik, Merle, Jannis und Laurin werden das Lungenvolumen und der Ruhepuls gemessen. Welche Aussage ist korrekt, wenn Folgendes gilt und kein Messwert doppelt vorkommt?
- Die Person mit dem größten Lungenvolumen hat weder den niedrigsten noch den höchsten Ruhepuls.
 - Jannis' Ruhepuls ist höher als der von Kathrin aber niedriger als der von Laurin.
 - Dominik hat den zweitniedrigsten Ruhepuls und sein Lungenvolumen ist kleinere als das der Person mit dem höchsten Ruhepuls.
 - Kathrin hat nicht das größte Lungenvolumen.
 - Merles Ruhepuls ist am höchsten.
- (A) Merle hat das kleinste Lungenvolumen und Laurin den zweithöchsten Ruhepuls.
(B) Jannis hat den höchsten Ruhepuls und Laurin hat das kleinste Lungenvolumen.
(C) Kathrin hat den niedrigsten Ruhepuls und Dominik hat das größte Lungenvolumen.
(D) Kathrin hat den niedrigsten Ruhepuls und Merle hat nicht das größte Lungenvolumen.

Abbildung 1. Beispielitem des Untertests relationales Schließen (viaMINT, 2022)

Arithmetisches Problemlösen (AP)

Der Untertest arithmetisches Problemlösen prüft die kognitive Rechenfähigkeit von Studieninteressierten und bildet somit numerisches Schlussfolgern ab. Die Aufgabe besteht darin, logische Rechenaufgaben zu lösen, die an einen Dreisatz erinnern (Abbildung 2). Die dabei präsentierten Aufgaben nehmen stetig an Schwierigkeit zu. Der Test verfügt über eine akzeptable interne Konsistenz ($\alpha = .78$).

Um eine Flüssigkeit mit niedrigem pH-Wert zu bekommen, sollen eine Base und eine Säure im Verhältnis 1 zu 25 gemischt werden. Der Behälter mit der Base hat ein Fassungsvermögen von 280 ml und ist genau zur Hälfte gefüllt.

Welchen Anteil der Base muss man nehmen, um 350 ml der Flüssigkeit mit niedrigem pH-Wert zu erhalten?

- (A) 1/12
- (B) 2/5
- (C) 1/10
- (D) 1/4
- (E) 1/3

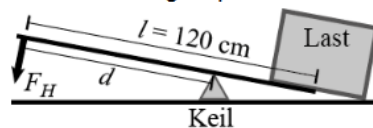
Abbildung 2. Beispielitem des Untertests arithmetisches Problemlösen (UKE, 2022a, S. 4)

Hamburger Naturwissenschaftstest (Nat)

Ziel dieses Untertests ist es, studienrelevantes Vorwissen für das Fach Medizin zu erfassen. Der Hamburger Naturwissenschaftstest umfasst hierzu Fragen zu folgenden Fächern: Biologie, Chemie, Physik und Mathematik auf dem Niveau der gymnasialen Oberstufe (Abbildung 3). Die Testautor:innen konnten in verschiedenen Studien zeigen, dass das Nat-Ergebnis den vorklinischen Studienerfolg inkrementell zur Abiturnote vorhersagt und dass dies insbesondere bei Bewerbenden mit sehr guten Abiturnoten der Fall ist (Hissbach et al., 2012; Hissbach et al., 2011). Die interne Konsistenz des Nat ist akzeptabel ($\alpha = .74$).

Eine Last, deren Gewichtskraft 1050 N beträgt, soll mit einem als Hebel darunter geschobenen Stab angehoben werden. Die Länge l des Stabs (vom Angriffspunkt der Hand bis zur Last) beträgt 120 cm. Der unter den Stab zu legende Keil (Höhe 1 cm), der den Drehpunkt bildet, soll so positioniert werden, dass die notwendige Kraft F_H gerade 210 N beträgt.

Wie groß muss der Abstand d des Keils vom Angriffspunkt der Hand am Stab sein?



- (A) 20 cm
- (B) 24 cm
- (C) 96 cm
- (D) 100 cm
- (E) 110 cm

Abbildung 3. Beispielitem des Hamburger Naturwissenschaftstests aus dem Bereich Physik (UKEb, 2022, S. 2)

Figurale Matrizen (FM)

Figurale Matrizen gehören zu jenen Intelligenztests, die besonders hoch auf fluider Intelligenz (Gignac, 2015) und damit auch auf g laden (Carpenter et al., 1990; Jensen,

1998). Im Allgemeinen sind Matrizen sprachfrei und sollten unabhängig von kulturellen Hintergründen sein. Konstruktionsbasierte Matrizen stellen eine Weiterentwicklung des klassischen Item Formats dar, indem sie die Probanden die Lösung aktiv konstruieren lassen (Becker, Schmitz, Göritz, et al., 2016, Abbildung 4). Dabei können Response-Elimination-Strategien vermieden werden, die zu einer Reduktion der Konstruktvalidität führen können (Becker, Schmitz, Falk, et al., 2016). Die gefundene interne Konsistenz ist exzellent ($\alpha = .93$).

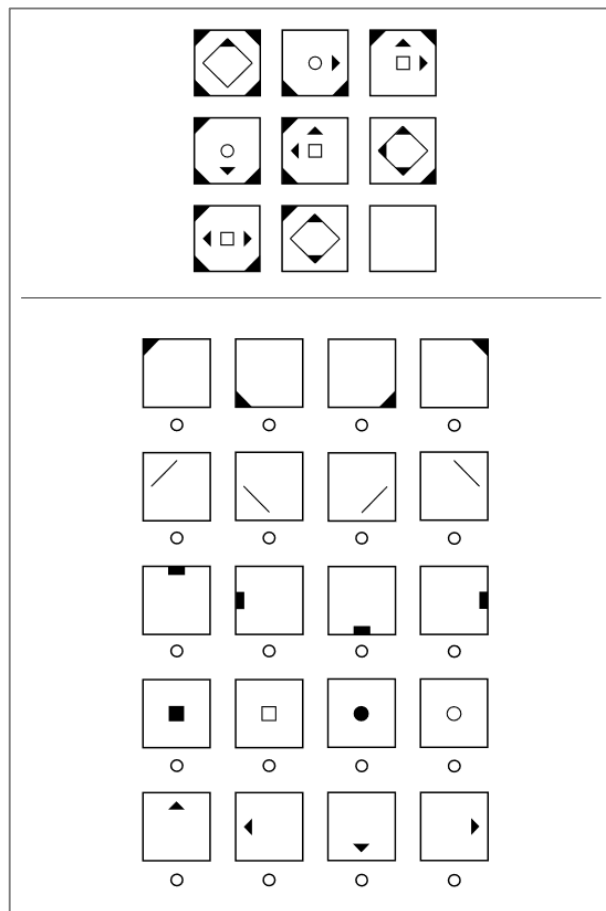


Abbildung 4. Beispielitem des konstruktionsbasierten, figuralen Matrizentests (Koch et al., 2022)

TMS

Diagramme und Tabellen (DT)

Diese Aufgabengruppe prüft die Fähigkeit, Diagramme und Tabellen richtig zu analysieren und zu interpretieren. Dazu werden komplizierte Zusammenhänge aus naturwissenschaftlichen Fachgebieten in einer Tabelle oder einem Diagramm (z. B. Kreisdiagramm und Kurvendiagramm) dargestellt. Zusätzlich wird eine kurze Beschreibung der Tabelle oder des Diagramms dargeboten. Anschließend soll aus den

Lösungsvorschlägen die richtige Antwort ausgewählt werden (Abbildung 5). Die interne Konsistenz ist akzeptabel ($\alpha = .78$).

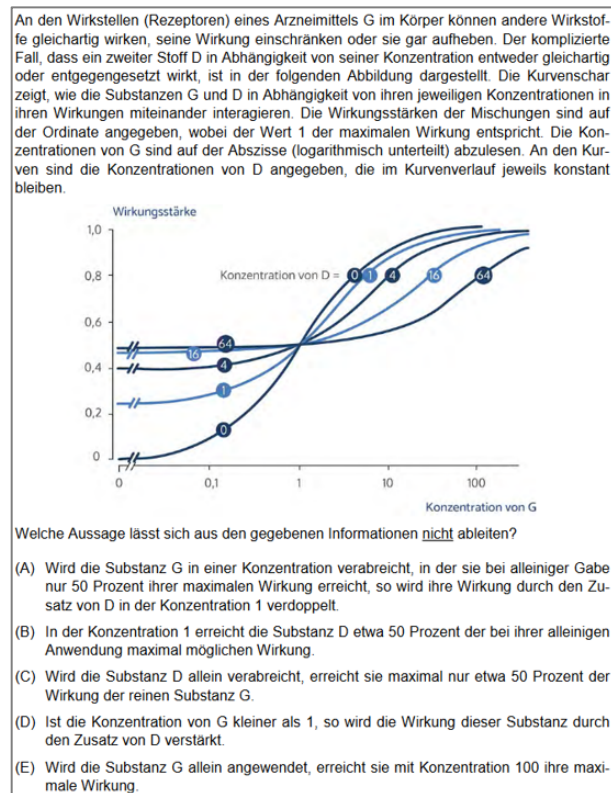


Abbildung 5. Beispielitem der TMS-Subskala Diagramme und Tabellen (ITB, 2021, S. 49)

Textverständnis (TV)

Beim Textverständnis wird die Fähigkeit getestet, umfangreiches und komplexes Textmaterial aufzunehmen und zu verarbeiten. Dazu sind sechs verschiedene Fragen zu beantworten, die sich ausschließlich auf den Inhalt des vorangegangenen, relativ langen Textes beziehen (Anhang A.). Für diesen Test des TMS wurde eine hohe interne Konsistenz gezeigt ($\alpha = .82$).

Medizinisch-naturwissenschaftliches Grundverständnis (MNG)

Zusätzlich wurde der Test zur Messung des medizinisch-naturwissenschaftlichen Grundverständnisses des TMS betrachtet. Ziel ist es, zu testen, wie gut komplexe und anspruchsvolle Informationen aus einem Text extrahiert werden können. Die Themen der Texte beziehen sich zum Beispiel auf die menschliche Anatomie oder auf Vorgänge im menschlichen Organismus. Zur Beantwortung der Fragen ist daher kaum Vorwissen erforderlich, sondern die Fähigkeit, Informationen schnell zu verarbeiten

und zu filtern (Abbildung 6). Für diese Skala wurde ein akzeptables Cronbachs Alpha nachgewiesen ($\alpha = .75$).

Röntgenstrahlen werden in umso stärkerem Maße absorbiert, je dichter oder je dicker die zu bestrahlende Substanz ist. Bei einer Röntgenaufnahme des Brustkorbs wird daher die strahlenempfindliche Schicht des Films dort stärker geschwärzt, wo die Röntgenstrahlen auf ihrem Weg durch den Körper hauptsächlich Lungengewebe getroffen haben. Das Gebiet, in dem das Herz liegt, bleibt etwas heller, und auch die Rippen zeichnen sich wegen ihrer durch den Kalkreichtum höheren Dichte als hellere Streifen ab.

Welcher der folgenden krankhaften Befunde würde sich demnach im Röntgenbild als eine etwas stärker geschwärzte Stelle von seiner Umgebung abheben?

- (A) ein alter verkalkter Tuberkuloseherd in der Lunge
- (B) ein Nierenstein mit hohem Kalziumgehalt
- (C) ein verschluckter Nagel im Magen
- (D) ein das Knochengewebe entkalkender Tumor im Beckenknochen
- (E) eine größere Eiteransammlung in der Lunge

Abbildung 6. Beispielim der TMS-Subskala medizinisch-naturwissenschaftliches Grundverständnis (ITB, 2021, S. 23)

Quantitative und formale Probleme (QFP)

Außerdem wurde der Untertest quantitative und formale Probleme betrachtet. Der Test erfasst die Fähigkeit, mit Zahlen, Mengen, Einheiten und Formeln, im Zusammenhang mit medizinischen und naturwissenschaftlichen Fragestellungen, richtig umzugehen. Dazu werden Textaufgaben präsentiert, die in der Regel keine komplizierten Berechnungen erfordern. Vielmehr kommt es darauf an, die Aufgabe in ihrer Gesamtheit zu verstehen und den richtigen Lösungsansatz entsprechend anzuwenden (Abbildung 7). Auch dieser Test weist eine hohe interne Konsistenz auf ($\alpha = .81$).

Eine bestimmte Krankheit kann durch zwei Untersuchungsmethoden diagnostiziert werden. Mit Methode X werden 85 Prozent der tatsächlichen Erkrankungen erkannt, mit Methode Y dagegen 80 Prozent.

Wie viel Prozent der tatsächlichen Erkrankungen werden mit keiner der beiden Methoden erkannt?

- (A) 0 Prozent
- (B) höchstens 5 Prozent
- (C) höchstens 15 Prozent
- (D) höchstens 20 Prozent
- (E) höchstens 35 Prozent

Abbildung 7. Beispielim der TMS-Subskala quantitative und formale Probleme (ITB, 2021, S. 29)

2.3 Konstruktvalidität kognitiver Leistungstests zur Selektion von Medizin-Studierenden

Bevor es zur Entwicklung eines einheitlichen fachspezifischen Studierendenauswahltests kommen kann, bedarf es einer detaillierten Analyse der psychometrischen Güte beider Testverfahren. Frühere Studien haben gezeigt, dass die beiden zur Auswahl von Medizinstudierenden eingesetzten Tests (HAM-Nat und TMS) geeignete psychometrische Eigenschaften in Bezug auf interne Konsistenz und prädiktive Validität besitzen (Hell et al., 2007; Hissbach et al., 2011; Kadmon et al., 2012; Werwick et al., 2015). Da sich alle diese Studien ausschließlich mit nur einem der beiden Tests befassen, gibt es jedoch derzeit keine Erkenntnisse über die Konstruktvalidität zwischen ihnen. Jedoch wäre es im Hinblick auf die Vergleichbarkeit der Auswahlverfahren wichtig, zu wissen, ob verschiedene Universitäten unterschiedliche Standards anwenden, beide Testverfahren dasselbe Konstrukt erfassen und inwieweit inkrementelle Validität vorliegt. Ein kombinierter Test könnte daher eine bessere Vorhersage des Studienerfolgs ermöglichen als die beiden Tests allein.

Studie I ist ein erster Schritt zur Schließung dieser Forschungslücke. Ziel ist es, erste Belege für die Konstruktvalidität zwischen den beiden Tests zu liefern. Hierzu wurde Medizin-Studieninteressierten ($N = 4,528$, $M_{Alter} = 20.42$ Jahre, $SD_{Alter} = 2.74$, $16 \leq \text{Alter} \leq 56$) die Option geboten, digital, in einem low-stakes Setting, Skalen beider etablierten Testverfahren (TMS und HAM-Nat) zu bearbeiten.

Durch die Betrachtung der Korrelationen der Untertests konnte gezeigt werden, dass alle betrachteten Skalen bedeutsam und signifikant untereinander korrelieren (Tabelle 1). Die höchste mittlere Korrelation findet sich zwischen den Untertests des TMS ($M_r = .54$), während etwas geringere mittlere Korrelationen zwischen den Untertests des HAM-Nat ($M_r = .46$) und zwischen dem HAM-Nat und den Untertests des TMS ($M_r = .38$) gefunden werden können.

Tabelle 1. Part-whole-korrigierte Korrelationen der Summenwerte aller Subtests

	DT	TV	QFP	MNG	Nat	FM	AP	RS
DT		3,169	3,222	3,241	2,799	1,340	816	1,063
TV	.62*		3,396	3,410	2,871	1,374	821	1,077
QFP	.63*	.57*		3,882	3,098	1,446	892	1,162
MNG	.61*	.67*	.60*		3,331	1,526	946	1,243
Nat	.40*	.41*	.49*	.37*		1,387	865	1,118
FM	.34*	.29*	.36*	.25*	.36*		435	583
AP	.51*	.43*	.53*	.40*	.55*	.50*		570
RS	.36*	.32*	.28*	.30*	.42*	.40*	.56*	

Anmerkungen. * $p < .001$; Pearson-Korrelation mit pairwise-deletion; unter der Diagonale sind die Korrelationen dargestellt; über der Diagonale sind die Stichprobengrößen angegeben; DT = Diagramme und Tabellen; TV = Textverständnis; QFP = Quantitative und Formale Probleme; MNG = Medizinisch-naturwissenschaftliches Grundverständnis; Nat = Hamburger Naturwissenschaftstest; FM = Figurale Matrizen; AP = Arithmetisches Problemlösen; RS = Relationales Schließen.

In einem nächsten Schritt wurden zur Betrachtung der Konstruktvalidität folgende Modelle aufgestellt und verglichen:

1. *G-Modell*: Es wurde das klassische g -Faktorenmodell im Sinne von Spearman (1904) analysiert. Hier laden alle Untertests auf einen einzigen allgemeinen Faktor und die übrige Varianz wird als Messfehler betrachtet.
2. *HO-Modell*: Unter Berücksichtigung von Faktorenmodellen höherer Ordnung (Jensen, 1998; McGrew, 2009) wurde ein Modell mit separaten Gruppenfaktoren untersucht, die die gemeinsame Varianz der Untertests innerhalb des HAM-Nat bzw. des TMS repräsentieren und zu einem übergeordneten g -Faktor führen.
3. *NF-Modell*: Die Idee einer verschachtelten Faktorenstruktur (Brunner et al., 2012; Gignac, 2006, 2008; Valerius & Sparfeldt, 2014) wurde anhand eines Modells überprüft, in dem die nicht durch g gebundene Varianz durch spezifische Faktoren innerhalb des HAM-Nat oder des TMS erklärt wird.
4. *TS-Modell*: In Anlehnung an Johnson und Kollegen (2004, 2008) wurde ein testspezifisches Modell mit separaten g -Faktoren für die Untertests des HAM-Nat und des TMS analysiert. Die geteilte Varianz zwischen den Tests wird durch die Korrelation zwischen den testspezifischen Faktoren dargestellt.

Die Ergebnisse der konfirmatorischen Faktorenanalysen sind in Abbildung 8 dargestellt. Alle Faktorladungen und (falls vorhanden) latenten Korrelationen sind

signifikant und substanziell. Der statistische Wert für die Anpassungsgüte χ^2 ist für alle Modelle signifikant (Tabelle 2). Die Fit-Indizes (CFI, TLI, RMSEA, SRMR) für die Modelle, die nur einen einzigen allgemeinen Faktor enthalten (*g-Modell*) oder eine Struktur höherer Ordnung, in der testspezifische Gruppenfaktoren zu einem allgemeinen Faktor führen (*HO-Modell*), deuten auf eine Fehlanpassung der Modelle hin. Die Modelle mit testspezifischen Faktoren, die unabhängig von einem allgemeinen Faktor sind (*TS-Modell*, *NF-Modell*), zeigen bessere Fit-Indizes. Das *NF-Modell* weist dabei eine exzellente Passung zu den Daten auf. Ein Vergleich der Informationskriterien (AIC, BIC) zeigt, dass sie beim *NF-Modell* am niedrigsten waren, gefolgt vom *TS-*, *HO-* und *g-Modell*. Das gleiche Muster zeigt sich bei der Betrachtung der Δ RMSEA. Dies führt dazu, dass die Modelle mit unabhängigen testspezifischen Faktoren bevorzugt werden. Dennoch scheint es, als seien beide Tests Indikatoren für einen allgemeinen Intelligenzfaktor. Intelligenz kann somit mit Hilfe beider Tests erfasst werden, so dass Konstruktvalidität besteht.

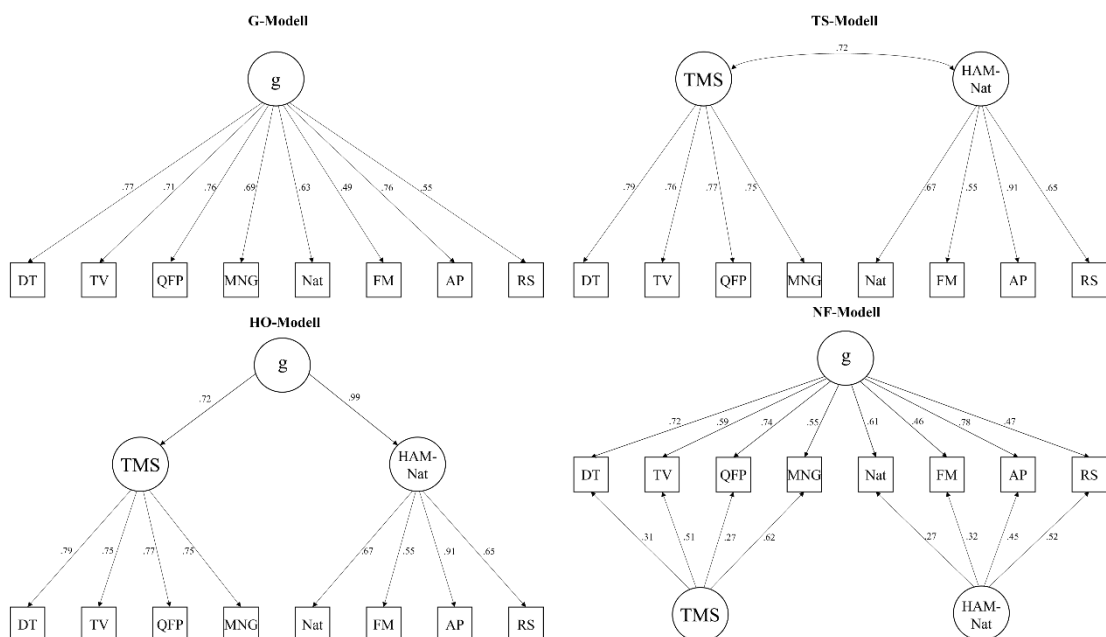


Abbildung 8. Strukturgleichungsmodelle zur Schätzung allgemeiner und spezifischer Faktoren beider Zulassungstests

Tabelle 2. Passung der vier getesteten Modelle zur Erfassung der Konstruktvalidität

	χ^2	df	p	CFI	TLI	RMSEA	SRMR	AIC	BIC	Δ RMSEA
G-Modell	2,520.84	20	< .001	.843	.781	.166	.074	198,285	198,387	
HO-Modell	745.32	17	< .001	.954	.925	.097	.037	196,516	196,638	.069
TS-Modell	745.32	19	< .001	.955	.933	.092	.037	196,511	196,620	.005
NF-Modell	324.68	12	< .001	.980	.954	.076	.022	196,104	196,259	.016

Anmerkungen. df = Freiheitsgrade; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Mean Square Error of Approximation; SRMR = Standardized Root Mean Square Residual; AIC = Akaike'sches Informationskriterium; BIC = Bayes'sches Informationskriterium; Δ = Veränderung zum jeweils vorherigen Modell.

Die Ergebnisse von Studie I stehen im Einklang mit der bisherigen Literatur zu der Konstruktvalidität von Intelligenz. Die positiven Interkorrelationen zwischen den Untertests des HAM-Nat und des TMS zeigen, dass sie einen erheblichen Anteil an Varianz teilen. Dies entspricht Spearmans (1904) Idee von *g* und Faktorenmodellen höherer Ordnung (z. B. Jensen, 1998; McGrew, 2009), die eine allgemeine intellektuelle Fähigkeit konzeptualisieren, die unabhängig von dem Test ist, mit dem sie gemessen wird. Außerdem passten die konfirmatorischen Modelle, die testspezifische, von *g* unabhängige Fähigkeitsfaktoren enthielten, besser zu den Daten als die Modelle, die sie nicht enthielten. Dies zeigt, dass es auch eine beträchtliche Menge an systematischer Varianz gibt, die nicht von den beiden Tests geteilt wird. Daher unterstützen die Ergebnisse die jüngste Literatur, die Hinweise auf testspezifische Fähigkeitsfaktoren jenseits von *g* gefunden hat (z. B. Brunner et al., 2011; Valerius & Sparfeldt, 2014). Im Hinblick auf die derzeitige Auswahlpraxis für Medizinstudierende in Deutschland ist es wichtig zu erwähnen, dass die große gemeinsame Varianz der beiden Tests zeigt, dass die Universitäten, die entweder den HAM-Nat oder den TMS verwenden, keine völlig unterschiedlichen Standards anwenden. Gleichzeitig könnte die Vernachlässigung der spezifischen Varianzen beider Tests zu einem Verlust wertvoller Informationen führen. Angesichts dieser Tatsache erscheint es sinnvoll, beide Tests oder zumindest Teile von ihnen zu kombinieren.

2.4 Der Einfluss von Testvorbereitung auf die Güte eines kognitiven Leistungstests

Im Kontext von Tests mit hohen Erfolgsanforderungen, wie Studierendenauswahlverfahren, ist die Testvorbereitung ein Thema, das berücksichtigt werden muss

(Buchmann et al., 2010). Jedoch findet durch Trainingseffekte nicht nur eine Verfälschung der Testergebnisse statt, sondern gleichzeitig auch eine Einschränkung der Validität des Auswahlverfahrens (Köbler, 2009). Mit Hilfe von Coaching und Testwiederholung kann die Leistung in einem Auswahlverfahren gesteigert werden, so dass eine vorbereitete Person im Ranking deutlich weiter oben platziert wird (Hausknecht et al., 2007). Durch Testwiederholung können auftretende Gedächtnis- sowie Regressionseffekte zur Einschränkung der Gütekriterien führen (Schuler, 2014). Alle beschriebenen Einflüsse gefährden nicht nur die Validität der Tests sondern führen ebenfalls zu einer Einschränkung der Testfairness (Höft, 2011). Die besser geeigneten können gegenüber den vorbereiteten Personen deutlich schlechtere Testergebnisse erzielen. Denn diese haben durch ein vorheriges Training gelernt, die Aufgaben schnell, effizient und richtig zu bearbeiten. Miller und Barrett (2008) schlussfolgerten daraus, dass Personen, die sich nicht vorbereiten, schlechtere Chancen haben, durch ein solches Verfahren ausgewählt zu werden als Personen mit einem vorherigen Training. Gleichzeitig steht die Möglichkeit der Nutzung von kommerziellen Vorbereitungsangeboten, aufgrund der damit verbundenen Kosten, nicht allen Studieninteressierten in gleichem Maße zur Verfügung. Die Testfairness ist jedoch nicht das einzige Nebengütekriterium welches beeinflusst wird. Vielmehr hat die Vorbereitung bzw. ein Training gleichzeitig Einfluss auf die Nützlichkeit des Testresultates, da es zu Rangreihen-Unterschieden kommen kann (Hausknecht et al., 2007).

Weithin bekannt ist, dass ein Anstieg der Punktzahl infolge der Testvorbereitung nicht mit einer Verbesserung der zugrundeliegenden Fähigkeiten einhergeht (Estrada et al., 2015; te Nijenhuis et al., 2007). Folglich sind gefundene Mittelwertsunterschiede zwischen Experimental- und Kontrollgruppen, beispielsweise nach Erklärung der Konstruktionsprinzipien von Matrizenaufgaben (Loesche et al., 2015; Schneider et al., 2020), nicht mit einer Veränderung der Intelligenz assoziiert (Haier, 2014; Hayes et al., 2015). Individuelle Unterschiede in der Testvorbereitung können daher Fähigkeitsunterschiede überdecken und die Testvalidität negativ beeinflussen (Messick, 1982).

Im Rahmen von Studie II wurde ein kurzes Testvorbereitungsmaterial entwickelt, das den Befragten ($N = 882$, $M_{Alter} = 19.67$ Jahre, $SD_{Alter} = 2.01$, $16 \leq \text{Alter} \leq 53$)

schriftliche Informationen über die im Test verwendeten Regeln gibt. Das erste Ziel der Studie war es, die Effektivität dieses Materials zu evaluieren und es mit den umfangreicheren Ansätzen von Schneider und Kollegen (2020) sowie Loesche und Kollegen (2015) zu vergleichen. Darüber hinaus wurde geprüft, ob die konvergente Validität in Bezug auf den Hamburger Naturwissenschaftstest (Nat) und die Kriteriumsvalidität in Bezug auf die schulische Leistung abnimmt, wenn Gruppen von vorbereiteten und unvorbereiteten Probanden betrachtet werden. Zudem sollten mögliche Einflüsse der Testvorbereitung auf die Itemeigenschaften untersucht werden. Dabei wurde die Vorbereitung auf den figuralen Matrizen test betrachtet.

Hierzu wurden die Teilnehmenden nach dem Zufallsprinzip zwei verschiedenen Hörsälen zugewiesen, wobei die Teilnehmenden des einen Hörsaals das Material zur Testvorbereitung erhielten, während die Teilnehmenden des anderen Hörsaals als Kontrollgruppe dienten. Das Vorbereitungsmaterial beinhaltet eine ausführliche schriftliche Information über die im Matrizen test verwendeten Regeln (d.h. Rotation, Addition, Vollständigkeit, Subtraktion, Einzelkomponentenaddition, Schnittmenge), welche zur Konstruktion der Matrizen, auch in Kombination, verwendet wurden. Beide Gruppen hatten zur Bearbeitung der 28 konstruktionsbasierten Matrizen 20 Minuten Zeit. Zusätzlich lag für alle Teilnehmenden das Ergebnis der 80 Items des Nats ($M = 29.86$; $SD = 8.75$) sowie die Abiturdurchschnittsnote ($M = 1.56$; $SD = 0.19$) vor.

Wie postuliert zeigte sich, dass die vorbereiteten Personen ($M = 17.08$; $SD = 0.19$) im Vergleich zur Kontrollgruppe ($M = 11.48$, $SD = 0.25$) ein signifikant bessere Ergebnis erzielten ($t(877.66) = -12.96$, $p < .001$, $d = .94$). Die Effektgröße liegt dabei im Bereich der in früheren Studien gefundenen Effekte (Loesche et al., 2015: $M(d) = .51$; Schneider et al., 2020: $M(d) = 1,24$). Daraus kann geschlussfolgert werden, dass selbst eine eher minimale Testvorbereitung ausreicht, um die Testergebnisse der Befragten zu verbessern.

Die Split-Half-Reliabilität war in beiden Gruppen vergleichbar hoch (Testvorbereitungsgruppe: $r_{sh} = .88$; Kontrollgruppe: $r_{sh} = .83$). Darüber hinaus waren auch die latenten Reliabilitäten ausgezeichnet (Testvorbereitungsgruppe: $\omega = .93$, 95% CI = [.91; .94]; Kontrollgruppe: $\omega = .90$, 95% CI = [.88; .92]).

Die mittlere Wahrscheinlichkeit, die Aufgaben zu lösen, war in der Testvorbereitungsgruppe höher ($M = .61$; $SD = .19$) als in der Kontrollgruppe ($M = .41$; $SD = .25$). Dieser Unterschied war signifikant und stark ($t(54) = 3,42$, $p = .001$, $d = .92$). Die Korrelation der Itemschwierigkeiten zwischen der Testvorbereitungs- und Kontrollgruppe war ebenfalls signifikant und stark ($r = .99$, $p < .001$). Vergleichbare Korrelationen wurden zwischen den Lösungswahrscheinlichkeiten in den beiden Gruppen gefunden ($r = .94$, $p < .001$).

Zur Betrachtung der Kriteriumsvalidität und der konvergenten Validität wurden Korrelationen des Matrizentests mit dem HAM-Nat-Ergebnis und der Abiturdurchschnittsnote berechnet. Die Testergebnisse in beiden Gruppen sind signifikant mit dem Testergebnis des HAM-Nat korreliert (Vorbereitungsgruppe: $r = .28$, $p < .001$, Kontrollgruppe: $r = .24$, $p < .001$). Die Ergebnisse des Matrizentests waren in keiner der beiden Gruppen signifikant mit dem Abiturschnitt korreliert (Testvorbereitungsgruppe: $r = -.08$, $p = .07$, Kontrollgruppe: $r = .01$, $p = .77$). Im Gegensatz zur Studie von Schneider und Kollegen (2020) führte die Kombination von Testvorbereitungs- und Kontrollgruppe nicht zu einer geringeren Kriteriumsvalidität oder konvergenten Validität. Da die Varianz der Abiturnote stark eingeschränkt war ($M = 1.56$; $SD = 0.19$), können die entsprechenden Korrelationen nicht plausibel interpretiert werden. Die Korrelationen zwischen dem figuralen Matrizentest und dem Nat deuten nicht darauf hin, dass die gemeinsame Testung von vorbereiteten und unvorbereiteten Probanden problematisch war. Eine mögliche Erklärung für die Unterschiede zur Studie von Schneider und Kollegen (2020) ist, dass sie einen anderen Intelligenztest verwendeten, der kognitive Fähigkeiten recht breit erfasst (G_f , G_c , G_v , G_s) und nicht wie in Studie II einen naturwissenschaftlichen Test, der nur G_c abbildet. Unterschiede aufgrund der Testvorbereitung könnten daher durch Unterschiede zwischen den Konstrukten, die bei der Analyse der konvergenten Validität berücksichtigt werden, überlagert werden.

Tabelle 3 zeigt die Ergebnisse für die konfirmatorische Multigruppen-Faktorenanalyse (MGCFA). Folglich kann eine starke Messinvarianz zwischen den beiden Gruppen angenommen werden. Der ΔCFI blieb unter dem von Chen (2007) vorgeschlagenen Schwellenwert für das schwache und das starke Modell. Beim $\Delta RMSEA$

überschritt das schwache Modell den Schwellenwert, während er beim starken Modell unter den Schwellenwert fiel. Für das strenge Modell überstiegen beide Indizes den Schwellenwert.

Tabelle 3. Ergebnisse der MGCFAs zum Vergleich der Faktorstruktur des Matrizentests zwischen beiden Gruppen

Modell	χ^2	df	$p(\chi^2)$	CFI	Δ CFI	RMSEA	Δ RMSEA	$\Delta\chi^2$	Δ df	$p(\Delta\chi^2)$
Konfigural	3.21	4	.52	1.00	-	<.001	-	-	-	-
Schwach	12.22	7	.09	.99	.001	.04	.04	9.01	3	.03
Stark	22.96	10	.01	.99	.002	.05	.01	10.75	3	.01
Streng	177.05	11	<.01	.94	.05	.19	.14	154.09	1	<.01

Anmerkungen. χ^2 = Teststatistik des Modellfits; *df* = Freiheitsgrad; $p(\chi^2)$ = Wahrscheinlichkeit der Teststatistik unter Annahme der Nullhypothese; CFI = Comparative Fit Index; RMSEA = Residual Mean Square Error of Approximation; Δ = Veränderung zum jeweils vorherigen Modell.

Was die Item-Eigenschaften betrifft, so konvergierte die MGCFa nicht auf der Ebene einzelner Items. Die Ergebnisse, die auf Item-Pacels basieren, sind zwar als Vereinfachung zu betrachten, legen aber dennoch nahe, dass sich die Item-Eigenschaften nicht verändert haben, da die Faktorladungen und Intercepts zwischen den beiden Gruppen gleich waren. Unterschiede in den Residualabweichungen können als Ausdruck individueller Unterschiede in der Testvorbereitung interpretiert werden.

Obwohl die Ergebnisse zeigen, dass schriftliche Informationen zu den Regeln ausreichen, um die Testergebnisse mit einer vergleichbaren Effektgröße zu steigern, bleibt unklar, ob eine umfangreichere Testvorbereitung die Eigenschaften der Items stärker verändern würde. Zur Beantwortung dieser Frage sind zukünftige Studien erforderlich, die Testvorbereitungsmaterialien mit unterschiedlichem Umfang vergleichen. In diesem Zusammenhang wäre es auch interessant zu wissen, ob es eine Sättigungsgrenze gibt, ab der sich zusätzliche Vorbereitungsanstrengungen nicht mehr lohnen, da dies die Mindestmenge an Vorbereitung wäre, die allen Kandidaten zur Verfügung gestellt werden sollte, um Testfairness zu gewährleisten.

2.5 Zusammenhang zwischen Antwortgeschwindigkeit und erfolgreicher Beantwortung kognitiver Leistungstests

Seit den Anfängen der Bewertung kognitiver Fähigkeiten war die Bewertung der Antwortzeit ein wichtiges Anliegen in der Psychologie der individuellen Unterschiede

(Spearman, 1904). Dies liegt daran, dass die Zeit über die Genauigkeit einer Antwort hinaus wertvolle Informationen über die kognitiven Prozesse liefert, die einer latenten Fähigkeit zugrunde liegen (De Boeck & Jeon, 2019). Time on Task (ToT) ist die Zeit vom Beginn der Aufgabe (d.h. der Präsentation des Items) bis zum Abschluss der Aufgabe (d.h. der Abgabe der Antwort) (Goldhammer et al., 2014). Es hat sich gezeigt, dass die Parameter der jeweiligen Antwortzeitverteilungen wesentlich mit der Leistung bei komplexen kognitiven Aufgaben wie dem logischen Denken zusammenhängen (Schmiedek et al., 2007; Schulz-Zhecheva et al., 2016). Basierend darauf könnte geschlossen werden, dass falsche Antworten im Allgemeinen mit längeren Antwortzeiten verbunden sind, so dass eine positive Abhängigkeit von Fähigkeit und Geschwindigkeit existiert. Dieser Zusammenhang wurde bereits sowohl für wissensunabhängige (Hornke, 2000) als auch für wissensabhängige (Lasry et al., 2013) Aufgaben gezeigt und scheint somit weitgehend unabhängig von der Art des zugrunde liegenden kognitiven Prozesses zu sein. Im Gegensatz dazu fanden Partchev und De Boeck (2012) und DiTrapani und Kollegen (2016) unterschiedliche ToT-Effekte für zwei kognitive Aufgaben mit selbstgesteuertem Tempo. Bei figuralen Matrizen lagen langsamere Probanden tendenziell richtig. Im Gegensatz dazu waren die ToT-Effekte bei verbalen Analogien gering bis nicht vorhanden. Daraus wurde gefolgert, dass schnelle und langsame Antworten auf unterschiedlichen kognitiven Prozessen beruhen. Molenaar und de Boeck (2018) erklären die bisherigen heterogenen Ergebnisse mit einer Moderation durch die Aufgabenschwierigkeit. Bei komplexeren Aufgaben hatten langsamere Antworten eine höhere Wahrscheinlichkeit, richtig zu sein.

Es wurde postuliert, dass die Richtung des ToT-Effekts davon abhängt, ob der Antwortprozess automatisch oder kontrolliert erfolgt (Goldhammer et al., 2014). Bei leichten Aufgaben, die weitgehend automatisch gelöst werden können, wie zum Beispiel Lesen, haben schnelle Antworten eine höhere Wahrscheinlichkeit, richtig zu sein. Bei schwierigen Aufgaben, die kontrollierte Prozesse erfordern, haben dagegen langsamere Antworten eine höhere Wahrscheinlichkeit, richtig zu sein.

Das Ziel von Studie III ist die Analyse der ToT-Effekte für die vier im HAM-Nat verwendeten Testteile. Einerseits ist Studie III folglich eine direkte Replikation früherer Ergebnisse und deren Konsistenz über Teilstichproben hinweg in Bezug auf

figurale Matrizen. Andererseits handelt es sich auch um eine konzeptionelle Replikation des ToT-Effekts über verschiedene Aufgaben hinweg, die Aufschluss über die Generalisierbarkeit des ToT-Effekts über verschiedene Kontexte hinweg gibt. Hierzu wurden Daten von $N = 2,640$ ($M_{\text{Alter}} = 21.34$ Jahre, $SD_{\text{Alter}} = 2.48$, $17 \leq \text{Alter} \leq 40$) Teilnehmenden der Vorbereitungsstudie analysiert. Zur Betrachtung der ToT-Effekte wurden nur die Personen betrachtet, die alle Aufgaben vollständig bearbeitet haben.

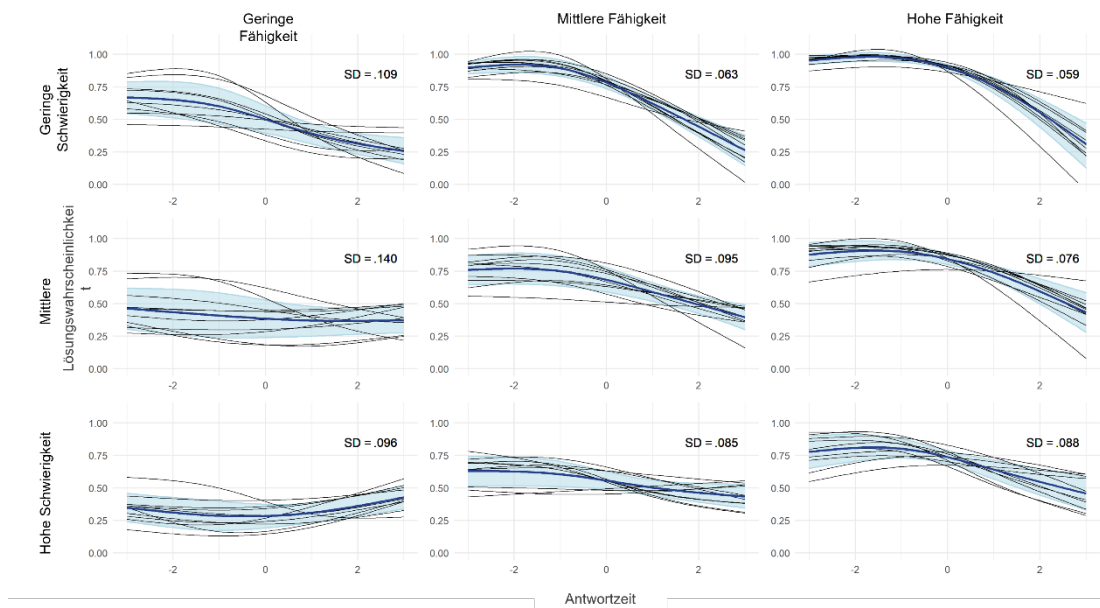


Abbildung 9. Wahrscheinlichkeit einer korrekten Antwort bei figuralen Matrizen als Funktion der Antwortzeit, aufgeteilt nach Schwierigkeitsgrad der Aufgabe, Fähigkeit der Person und Teilstichprobe.

Abbildung 9 zeigt die modellimplizierten und meta-analysierten bedingten Genauigkeitsfunktionen, aufgeteilt nach Itemschwierigkeit und Personenfähigkeit für die figuralen Matrizen. Wie aus den dunkelblauen Linien ersichtlich ist, waren schnellere Antworten im Allgemeinen eher korrekt. Dieser Effekt wurde jedoch mit zunehmender Schwierigkeit der Aufgaben und abnehmender Fähigkeit der Teilnehmenden weniger konsistent. Ein positiver ToT-Effekt wurde sogar für Teilnehmende mit geringen Fähigkeiten beim Lösen schwieriger Aufgaben beobachtet, hier stieg also die Lösungswahrscheinlichkeit mit der Reaktionszeit an. Was die Replizierbarkeit des ToT-Effekts betrifft, so waren die schwarzen Kurven, die die verschiedenen Stichproben repräsentieren, weitgehend vergleichbar. Die Heterogenität war bei Teilnehmenden

mit geringen Fähigkeiten ($0.096 \leq SD \leq 0.140$) höher als bei Teilnehmenden mit mittleren und hohen Fähigkeiten ($0.059 \leq SD \leq 0.095$).

Der ToT-Effekt für figurale Matrizen war insgesamt negativ, das heißt schnelle Antworten waren im Durchschnitt eher richtig. Negative ToT-Effekte wurden bereits in Studien mit demselben Aufgabentyp berichtet (Becker, Schmitz, Göritz, et al., 2016; Goldhammer et al., 2015). Im Sinne der dualen Verarbeitungstheorie deutet dies auf eine relativ automatische Verarbeitung hin. Die effektive mentale Belastung, die durch ein Item erzeugt wird, hängt jedoch von den Fähigkeiten der Person sowie von der Anzahl der Itemregeln und deren Interaktion ab. Es ist daher möglich, dass fähige Teilnehmende die Regeln aus den ersten Aufgaben ableiteten und dieses Wissen auf spätere Aufgaben anwenden, so dass sie diese mit geringerem Aufwand bearbeiten konnten. Dies wird dadurch gestützt, dass Regelwissen die Antwortgenauigkeit bei figuralen Matrizen-Aufgaben erhöht (Studie II).

Der Moderationseffekt von Itemschwierigkeit und Personenfähigkeit innerhalb der figuralen Matrizen replizierte das von Becker, Schmitz, Göritz und Kollegen (2016) sowie Goldhammer und Kollegen (2015) berichtete Muster. Insbesondere stieg mit zunehmender Schwierigkeit der Aufgaben die Genauigkeit der Antworten mit der Bearbeitungszeit. In Übereinstimmung mit diesen Ergebnissen wurde ein positiver ToT-Effekt für Personen mit geringen Fähigkeiten beim Lösen schwieriger Aufgaben beobachtet. Dies deutet darauf hin, dass nur in diesem Fall die kognitive Belastung durch die Aufgaben im Durchschnitt die kognitiven Ressourcen der Testteilnehmenden überstieg und eine weitgehend kontrollierte Verarbeitung erforderte, um eine korrekte Antwort zu geben. Dieses spezifische Muster konnte über verschiedene Stichproben von Items und Personen hinweg gut repliziert werden. Da alle Teilstichproben von Personen aus derselben Grundgesamtheit von Bewerbenden gezogen wurden und die Items auf analoge Weise generiert wurden, waren vergleichbare Personenfähigkeiten und Itemschwierigkeiten zu erwarten.

Deskriptive Statistiken und bivariate Korrelationen der ToT-Effekte für die vier kognitiven Leistungsmaße finden sich in Tabelle 4. Im Vergleich zu figuralen Matrizen waren die ToT-Effekte für arithmetisches Problemlösen und naturwissenschaftliches Wissen weniger negativ und für relationales Schließen sogar positiv. Die unkorrigierte

Split-Half-Reliabilität war mäßig für die figuralen Matrizen und den Hamburger Naturwissenschaftstest, aber niedrig für relationales Schließen und arithmetisches Problemlösen.

Tabelle 4. Deskriptive Statistik und Korrelation der ToT-Effekte

	Time on Task			Korrelationen			
	<i>M</i>	<i>SD</i>	ρ_{tt}	FM	AP	RS	Nat
Figurale Matrizen	- 0.60	0.66	.41	—			
Arithmetisches Problemlösen	- 0.23	0.25	.22	.16	—		
Relationales Schließen	0.34	0.15	.04	.08	.08	—	
Hamburger Naturwissenschaftstest	- 0.08	0.31	.40	.16	.22	.09	—

Anmerkungen. *M* = Mittelwert; *SD* = Standardabweichung; ρ_{tt} = Split-Half-Reliabilität der ToT-Effekte; alle Korrelationen sind signifikant bei $p < .001$

Daraus resultiert, dass die Verallgemeinerbarkeit des ToT-Effekts über verschiedene Aufgabentypen hinweg gering war. Die allgemeinen ToT-Effekte für die vier kognitiven Tests unterschieden sich sowohl in ihrer Größe als auch in ihrer Richtung. Überraschenderweise war der Effekt bei der Aufgabe zum relationalen Schließen positiv, das heißt langsamere Antworten hatten eine höhere Chance, richtig zu sein. Dies kann nicht durch einen höheren Schwierigkeitsgrad der Aufgaben erklärt werden, da die mittlere Trefferquote im Grunde die gleiche war wie bei den figuralen Matrizen. Vielmehr könnten tatsächliche Unterschiede in den geforderten kognitiven Prozessen und den geforderten Kompetenzen zum Tragen kommen. Im Falle des relationalen Schließens und der figuralen Matrizen sind diese Unterschiede vermutlich gering. Beide Fähigkeiten erfordern die Erstellung und Manipulation von mentalen Repräsentationen und sind daher grundsätzlich auf die Kapazität des Arbeitsgedächtnisses angewiesen (Wilhelm, 2005). Überdies sei erwähnt, dass die drei Aufgaben zum logischen Denken Indikatoren für G_f sind, während naturwissenschaftliches Wissen ein Indikator für G_c ist (Cattell 1963). Die unterschiedlichen kognitiven Prozesse, die hier zum Tragen kommen, sind wahrscheinlich der Grund für die geringe Korrelation der ToT-Effekte zwischen dem Nat und den figuralen Matrizen Aufgaben.

3 Teil II – Die COVID-19 Pandemie

3.1 Hinführung

In den ersten Monaten des Jahres 2020 erschütterte eine Krise die Welt: Das neuartige Coronavirus (SARS-CoV-2) verbreitete sich rasend schnell auf der ganzen Welt und führte in nur wenigen Monaten zum Ausbruch der COVID-19-Pandemie. Dies führte nicht nur für die mit dem Virus Infizierten, sondern auch für "gesunde" Individuen und die Gesellschaft als Ganzes zu einem komplexen Muster von Folgen (Bao et al., 2020; Duan & Zhu, 2020).

Am 16. März 2020 kündigte die Bundesregierung einen bundesweiten Lockdown beginnend am 22. März 2020 an (Schmid, 2021). Zu diesem Zeitpunkt war dies eine beispiellose Maßnahme zur Eindämmung der Ausbreitung von COVID-19. Die Menschen mussten mit enormen sozialen und wirtschaftlichen Herausforderungen lernen, umzugehen. Die Isolation, die sich aus der sozialen Distanz und der Schließung von Schulen und Arbeitsplätzen ergab, betraf wahrscheinlich fast alle Menschen und verstärkte die Erfahrung von Stress, Angst, Furcht und Einsamkeit. Da fast jeder die Folgen der Pandemie erfahren hat, kann die COVID-19-Pandemie als ein globales belastendes Lebensereignis angesehen werden (Brooks et al., 2020). Das globale belastende Lebensereignis wird im Allgemeinen als Ereignis definiert, das eine Veränderung der normalen Aktivitäten und Gewohnheiten der Menschen erfordert und oft zu einer Neuanpassung führt (Dohrenwend, 2006). Eine gesunde Neuanpassung erfordert jedoch auch individuelle Ressourcen (wie kognitive Fähigkeiten und psychische Gesundheit), die nicht alle Menschen in gleichem Maße besitzen. Daher ist es wahrscheinlich, dass nicht alle Menschen gleichermaßen erfolgreich in diesem Anpassungsprozess sind. Aus diesem Grund startete, kurz nach dem Beginn des Lockdowns in Deutschland, Studie IV, die vier Erhebungen in wöchentlichen Abständen und eine Abschlusserhebung vier Wochen nach der vierten Erhebung (14. Mai 2020) beinhaltet. Dieses Design ermöglichte es, einen wichtigen Zeitraum zu erfassen, der mit der Ankündigung eines landesweiten Lockdowns beginnt und nach der Aufhebung dieser ersten Abriegelung endet. Insgesamt nahmen 272 Erwachsene an der ersten Befragungswelle teil ($M_{Alter} = 36.94$ Jahre, $SD_{Alter} = 16,46$, $19 \leq \text{Alter} \leq 80$).

3.2 Der Einfluss des ersten Lockdowns auf das Wohlbefinden

Im Allgemeinen hat sich die Forschung - noch vor der COVID-19-Pandemie - eingehend mit der Bedeutung, den Determinanten und den Auswirkungen des subjektiven Wohlbefindens als zentrales Konstrukt zur Erfassung der Lebensqualität befasst. Das subjektive Wohlbefinden (SWB) kann als ein breit gefächertes Konzept angesehen werden, das als kognitive Komponenten allgemeine Lebenszufriedenheit sowie verschiedene Formen bereichsspezifischer Zufriedenheit und als emotionale Komponenten positive (z. B. Freude, Vergnügen) und negative Affekte (Angst, Stress, Einsamkeit) umfasst (Diener, 2006; Diener et al., 1999). Dabei hat sich ein großer Teil der Forschung auf das Ausmaß konzentriert, in dem wichtige Lebensereignisse das individuelle SWB beeinflussen können. Daher ist bekannt, dass beispielsweise Heirat, Scheidung und Verwitwung Ereignisse sind, die das SWB zumindest vorübergehend beeinflussen können. Frühere Forschungen in diesem Bereich haben jedoch gezeigt, dass das subjektive Wohlbefinden nach einem anfänglichen Anstieg oder Rückgang, in Abhängigkeit des Ereignisses, tendenziell wieder auf das Ausgangsniveau zurückkehrt (Campbell et al., 1976; Diener et al., 2006; Lykken & Tellegen, 1996). Diese hohe Konsistenz des SWB in empirischen Studien stützt die Set-Point-Theorie. Diese besagt, dass das Niveau des SWB von Erwachsenen um einen stabilen individuellen Set-Point herum schwankt. Daher können selbst dramatisch negative oder positive Ereignisse das SWB in der Regel nur für einen bestimmten Zeitraum beeinflussen (Diener et al., 2006). Im Gegensatz dazu haben einige Studien wie die von Hahn und Kollegen (2015) Muster des SWB entdeckt, die der Set-Point-Theorie widersprechen. Bei der Untersuchung von Arbeitslosigkeit stellten sie fest, dass selbst nach der Suche nach einem neuen Arbeitsplatz nicht alle Personen zu ihrem ursprünglichen SWB-Niveau zurückkehrten. Diese Unterschiede zwischen den einzelnen Personen waren in erster Linie auf Unterschiede in den Persönlichkeitsmerkmalen zurückzuführen. Zusammenfassend lässt sich sagen, dass die Set-Point-Theorie möglicherweise nicht für alle Personen und jedes Lebensereignis gleichermaßen gültig ist (Specht et al., 2013; Steel et al., 2008). Dies wirft die Frage auf, inwieweit eine Pandemie die Lebenszufriedenheit beeinflusst und wie unterschiedliche Persönlichkeiten darauf reagieren.

Es könnte folglich angenommen werden, dass ein dramatisches Ereignis wie die COVID-19-Pandemie die Entwicklung des SWB zunächst negativ beeinflussen sollte. Diese postulierte anfängliche Verringerung des SWB und eine gleichzeitige Zunahme psychischer Gesundheitsprobleme konnte in verschiedenen Ländern gezeigt werden (Anglim & Horwood, 2021; Daly et al., 2022; Kuper et al., 2021). Andere Studien kamen jedoch zu widersprüchlichen Schlussfolgerungen, da sie positive Veränderungen des Wohlbefindens während der COVID-19-Pandemie zeigten (O'Connor et al., 2021; Recchi et al., 2020). Dieser überraschende und etwas kontraintuitive Anstieg des Wohlbefindens deutet auf eine beträchtliche Anzahl von Personen hin, die die Situation positiv bewertet haben. Wieder andere Forschergruppen zeigten, dass das Wohlbefinden im Jahr 2020 im Vergleich zu den Vorjahren relativ stabil blieb (Kivi et al., 2020; Wang et al., 2020; Zacher & Rudolph, 2021). Zusammenfassend lässt sich sagen, dass die Ergebnisse zum SWB nicht eindeutig sind und verschiedene Studien zu widersprüchlichen Schlussfolgerungen kommen. Ein potenzieller Grund für die konträren Ergebnisse könnte darin bestehen, dass sich die Studien auf verschiedene Länder beziehen. Diese unterschieden sich hinsichtlich ihres Infektionsaufkommens, ihrer Beschränkungen sowie deren Dauer stark. Dennoch gibt es einige Gemeinsamkeiten in den Ergebnissen. So wurde zum Beispiel in Australien und Deutschland ein anfänglicher Rückgang des SWB während der Sperrzeiten festgestellt (Anglim & Horwood, 2021; Kuper et al., 2021).

Um zu erfassen, wie Individuen auf die globale Pandemie und die entsprechenden, von der Regierung diktierten Einschränkungen unmittelbar und langfristig reagieren, ist eine Längsschnittbewertung des Wohlbefindens unter Einbeziehung verschiedener Aspekte unerlässlich. Dabei ist es sinnvoll, sowohl kognitive als auch emotionale Aspekte zu erfassen, da bei den kognitiven Aspekten ein Vergleichsprozess abläuft. Die Lebenszufriedenheit spiegelt als wichtigster kognitiver Faktor des SWB den Vergleich des eigenen Ist- und Sollzustands mit anderen, früheren Zuständen und dem Idealzustand wider (Diener et al., 1985). Es hat sich gezeigt, dass die Lebenszufriedenheit relativ stabil ist, aber durch wichtige Lebensereignisse beeinflusst werden kann (Fujita & Diener, 2005; Schimmack et al., 2009). Erst im Laufe der Zeit zeigt sich in zahlreichen individuellen Vergleichen, vielleicht auch mit dem

eigenen Ideal und früheren Phasen, dass das Ereignis negativ bewertet wird und die Lebenszufriedenheit daher nicht zum Ausgangspunkt zurückkehrt. Da es sich bei der COVID-19-Pandemie um ein globales Ereignis handelt, das alle Menschen betrifft, könnte dieser Vergleichsprozess weniger negativ sein, so dass die Lebenszufriedenheit weniger stark sinkt. Bei den emotionalen Aspekten des SWB wäre dies jedoch nicht der Fall. Bei diesen sind stärkere Reaktionen zu erwarten, wenn auch nur kurzfristig. Aus diesem Grund sollte auch die affektive Komponente des SWB berücksichtigt werden. Für die Notwendigkeit, die affektive Komponente zu erfassen, spricht, dass diese nicht von logischen Überlegungen und kognitiven Bewertungen abhängt. Im Kontext der Pandemie könnte daher Einsamkeit, die unabhängig vom Alter mit der Qualität sozialer Beziehungen assoziiert ist (Luhmann & Hawkley, 2016) und höchstwahrscheinlich direkt durch Kontaktbeschränkungen beeinflusst wurde, besonders relevant sein. Außerdem könnten auch die psychische Belastung und der wahrgenommene Stress durch die Pandemie beeinflusst worden sein. Besonders das wahrgenommene Stresserleben, das eng mit der Lebensqualität zusammenhängt (Ng et al., 2009), könnte durch verschiedene Determinanten der Pandemie beeinflusst worden sein. Dabei sind der plötzliche Wechsel ins Homeoffice, die Schließung von Schulen und der daraus resultierende Heimunterricht oder auch die Tatsache, dass es sich bei der Pandemie um eine neuartige Situation handelt, die eine gewisse Anpassung erfordert, als potenzielle Einflussfaktoren auf das Stressniveau zu nennen. Diese exemplarischen Faktoren veranschaulichen, dass Lebenszufriedenheit, Einsamkeit sowie Stress und psychische Belastung wahrscheinlich unterschiedlich durch die Pandemie betroffen sind. Die Verteilung dieser vier erfassten Maße psychischen Wohlbefindens inklusive der jeweiligen Mediane ist in Abbildung 10 über alle Messwellen hinweg dargestellt.

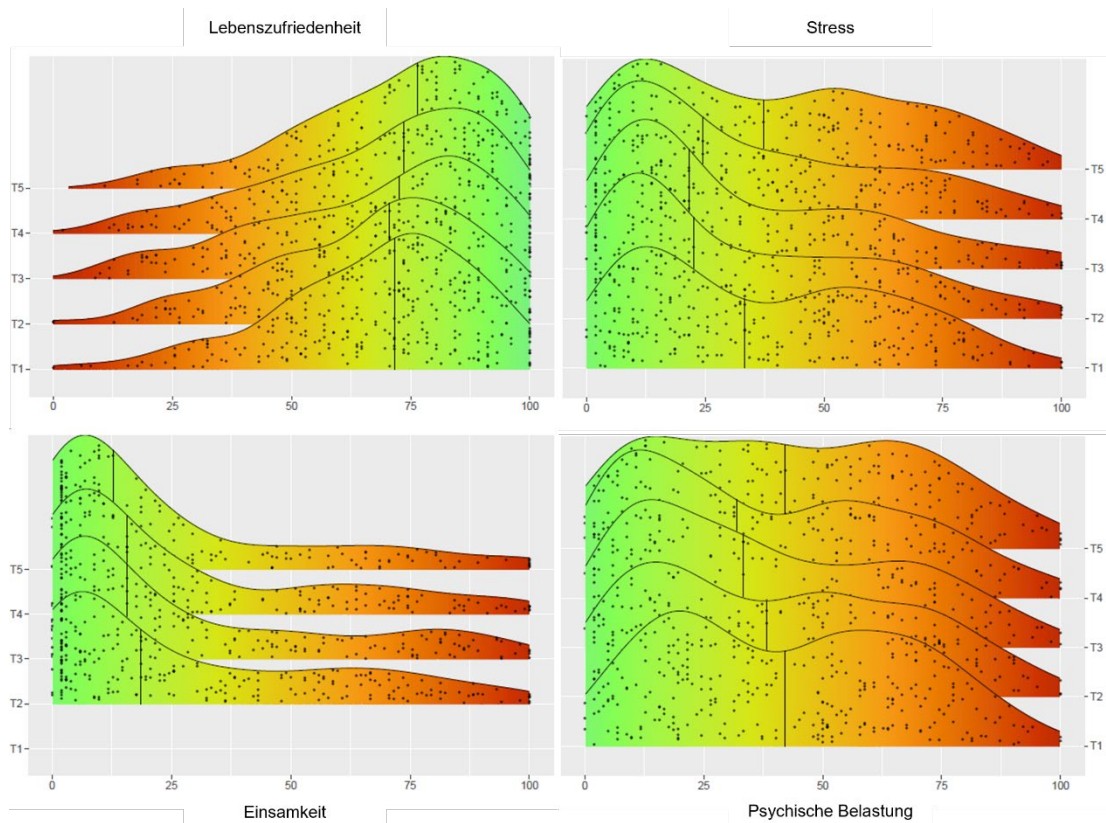


Abbildung 10. Verteilung des Wohlbefindens über die Messzeitpunkte hinweg. Der Median ist als schwarze Linie dargestellt. Hinsichtlich der Lebenszufriedenheit sind höhere Werte mit mehr Zufriedenheit assoziiert (0 – negativ vs. 100 – positiv), während bei Stress, psychischer Belastung und Einsamkeit höhere Werte negativ konnotiert sind (0 – positiv vs. 100 – negativ)

Da Unterschiede hinsichtlich des Verlaufs der jeweiligen Facette möglich sind, wurden zu dessen Analyse vier Modelle für jeden Wohlbefindensindikator aufgestellt. Entsprechend wurde für jede SWB-Facette das am besten passende Modell ermittelt. Die Ergebnisse der endgültigen bestpassenden Modelle, sind in Abbildung 11 für Lebenszufriedenheit und Einsamkeit und in Abbildung 12 für Stress und psychische Belastung dargestellt. Dabei spiegelt der Mittelwert des Intercepts (\bar{x}_i) das durchschnittliche SWB-Niveau zum Zeitpunkt 5 (nach dem Lockdown) wider. Der Mittelwert des slope-Faktors ($\bar{x}_{s1,2}$) steht für den linearen Steigungsfaktor und jener des quadratischen slope-Faktors (\bar{x}_q) für den quadratischen Steigungsfaktor. Die Varianzen s geben an, welche individuelle Varianz in diesen Komponenten zu finden ist. Die Korrelationen zwischen Intercept und Slope bedeuten, dass die Größe des Basisniveaus mit der Veränderung zusammenhängt.

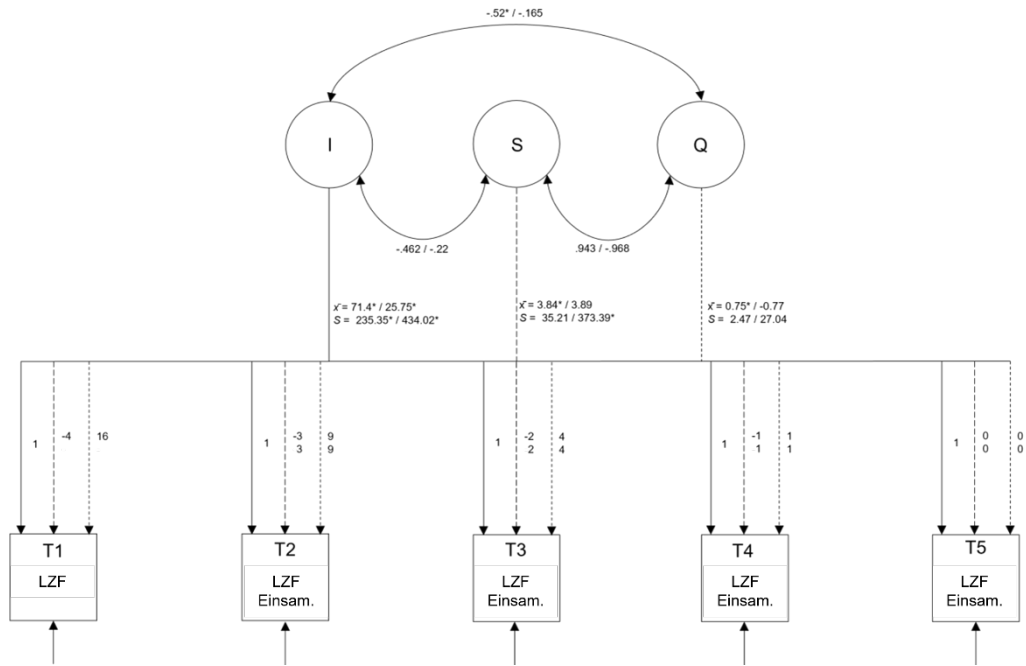


Abbildung 11. Quadratisches Wachstumskurvenmodell für Lebenszufriedenheit und Einsamkeit

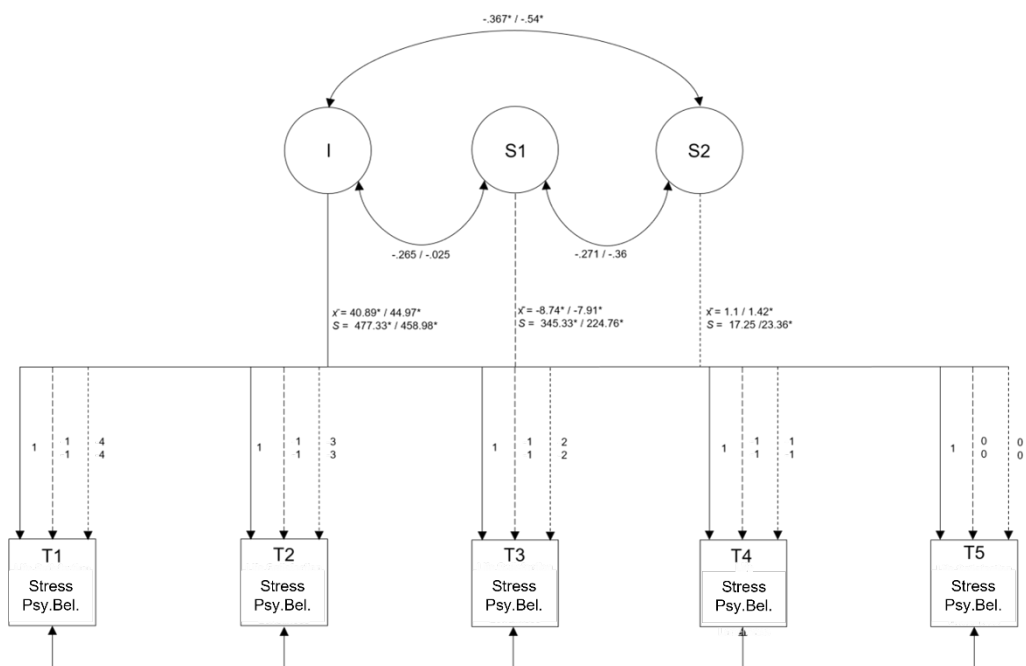


Abbildung 12. Latentes Wachstumskurvenmodell für Stress und psychische Belastung

Wie erwartet, sinkt die Lebenszufriedenheit nach Beginn des Lockdowns zunächst und erreicht nach Aufhebung der Beschränkungen wieder den ursprünglichen Wert. Dies unterstützt die Set-Point-Theorie. Wie in früheren Studien gezeigt wurde, kehrt das Wohlbefinden nach einem kurzfristigen Rückgang oder Anstieg häufig auf das Ausgangsniveau zurück (Campbell et al., 1976; Diener, 2006; Lykken & Tellegen,

1996). Insbesondere der in Studie IV gefundene Anstieg unmittelbar nach dem Ende des Lockdowns, deutet darauf hin, dass dieser nur einen kurzfristigen negativen Einfluss auf die Lebenszufriedenheit hatte. So stieg die Lebenszufriedenheit nach der Lockerung der Beschränkungen wieder an, was darauf hindeutet, dass die Personen froh waren, zum "Leben, wie es war" zurückzukehren. Dieses Ergebnis stimmt mit jenem von Foa und Kollegen (2020) überein. Die Autoren fanden einen Rückgang der Lebenszufriedenheit während des gesamten Lockdowns und eine anschließende Zunahme, wobei das Niveau vor dem Lockdown wieder erreicht (oder sogar übertroffen) wurde. In Studie IV wurden Unterschiede zwischen den Teilnehmenden nur in Bezug auf das endgültige Niveau der Lebenszufriedenheit festgestellt, nicht aber in Bezug auf den Verlauf der Lebenszufriedenheit. Offensichtlich war der anfängliche Rückgang der Lebenszufriedenheit bei allen vergleichbar, was wenig überraschend ist, da die COVID-19-Pandemie für alle eine noch nie dagewesene Situation darstellte. Die Auswirkungen waren gleich zu Beginn am stärksten und nahmen im Laufe der folgenden Wochen rasch ab. Daher können wir davon ausgehen, dass das Niveau nach der Abriegelung das übliche Lebenszufriedenheitsniveau ist. Personen, die vor dem Lockdown tendenziell zufriedener waren, zeigten auch eine flachere Kurve, das heißt sie konnten die Situation besser bewältigen oder haben den Lockdown sogar als Chance gesehen.

In Studie IV wurde außerdem ein Anstieg des Stressniveaus nach der Lockerung der Beschränkungen gefunden. Dies deckt sich mit dem Ergebnis einer anderen Studie, wonach das Stressniveau im ersten Monat des Lockdowns erheblich abnahm (Modersitzki et al., 2021). Möglicherweise wurde die Abriegelung von vielen als Gelegenheit zur Entschleunigung gesehen, da viele soziale Verpflichtungen und Termine weggefallen sind. Ein Blick auf die Mittelwerte über die Erhebungswellen hinweg zeigt, dass die Reduktion des Stresses ab T2 besonders deutlich ist, was möglicherweise widerspiegelt, dass zu diesem Zeitpunkt ein mögliches Gefühl der überwältigenden Unsicherheit und der weitreichenden Veränderungen einer eher adaptiven Reaktion gewichen war. Außerdem zeigt Studie IV, dass es individuelle Unterschiede hinsichtlich der Stressreduktion während des Lockdowns sowie hinsichtlich des Endniveaus an Stress gibt. Ein flacherer Stressverlauf während der Abriegelung war mit

höheren Stresswerten nach dem Lockdown assoziiert. Dies deutet darauf hin, dass Teilnehmende, bei denen der Lockdown keine oder nur eine schwache Reduktion des Stresses bedingte, nach der Abriegelung höhere Stresswerte aufwiesen als Teilnehmende, bei denen der Stress stark reduziert wurde.

Bezogen auf die psychische Belastung waren die in Studie IV gefundenen Ergebnisse vergleichbar mit jenen des Stresses. Entsprechend war die psychische Belastung der Teilnehmenden während des Lockdowns geringer, so dass davon ausgegangen werden kann, dass der Lockdown eine positive Wirkung hatte. Die lineare Kurve zeigt, dass die Verringerung der Belastung von Woche zu Woche zunahm. Wie der Stress kann auch die psychische Belastung als eine Form des negativen Affekts betrachtet werden; daher stimmen auch diese Ergebnisse mit denen von Modersitzki und Kollegen (2021) überein. Außerdem zeigte sich in Studie IV, dass ein flacherer Verlauf während des Lockdowns zu einem höheren Belastungsniveau danach führte. Personen, die während des Lockdowns weniger entspannten, waren danach stärker belastet. Dies verdeutlicht, dass es vermutlich wichtig war, während des Lockdowns verschiedene Bewältigungsstrategien (z. B. aufgabenorientierte Bewältigung, emotionsbasierte Bewältigung oder Ablenkung) zu verwenden, um sowohl psychische Belastung als auch Stress zu reduzieren (Polizzi et al., 2020).

Ferner analysierte Studie IV das Einsamkeitserleben der Probanden während und nach dem Lockdown. Im Allgemeinen hatten die Menschen während des Lockdowns höhere Werte für Einsamkeit als danach. Allerdings könnten die Einsamkeitswerte nach dem Lockdown immer noch höher sein als in den Vorjahren. Dies wird durch die Daten des deutschen Sozioökonomischen Panels (SOEP) gestützt, das im April 2020 einen Anstieg der Einsamkeitswerte im Vergleich zu den in den Vorjahren feststellte (Entringer et al., 2020). Im Gegensatz zu den SOEP-Daten wurden in Studie IV die Einsamkeitswerte über einen kürzeren Zeitraum, aber mit höherer Frequenz (d.h. wöchentliche Erhebungen) und somit genauer untersucht. Die Teilnehmenden der Studie IV wiesen im April 2020 ebenfalls erhöhte Einsamkeitswerte auf, die jedoch nach Aufhebung der Beschränkungen deutlich zurückgingen. Ähnliche Ergebnisse liegen für eine US-Stichprobe vor (Luchetti et al., 2020). Gleichzeitig ist es möglich, dass nicht nur das Ende des Lockdowns weniger Einsamkeitserleben bedingt,

sondern bereits während der Abriegelung alternative Formen der digitalen Kommunikation dabei halfen, mit Familie und Freunden in Kontakt zu bleiben (Nguyen et al., 2020).

3.3 Der Einfluss der Big Five auf das Wohlbefinden

Studie IV betrachtet überdies, ob Unterschiede in der Persönlichkeit die oben gezeigten Variationen im Endniveau und Verlauf des Wohlbefindens moderieren. Durch frühere Studien ist bereits bekannt, dass Persönlichkeitsmerkmale stark mit dem SWB zusammenhängen, während soziodemografische Faktoren (z. B. Alter, Geschlecht, Einkommen) nur mäßig damit verbunden zu sein scheinen (Ozer & Benet-Martínez, 2005; Schwarzer & Schulz, 2001). Zudem hat eine Studie von Steel und Kollegen (2008) gezeigt, dass die Persönlichkeit bis zu 39 % der Varianz in Lebensqualitätsmaßen erklärt. Insbesondere die Big Five Persönlichkeitsmerkmale beeinflussen das SWB auf konsistente Weise. Neurotizismus zum Beispiel steht in einem mäßig negativen Verhältnis mit SWB. Im Gegensatz dazu korrelieren Extraversion und Gewissenhaftigkeit mäßig positiv mit SWB. Verträglichkeit scheint leicht positiv mit dem Wohlbefinden zusammenzuhängen (Specht et al., 2013; Steel et al., 2008). Darüber hinaus wurde postuliert, dass Verträglichkeit und Gewissenhaftigkeit Voraussetzungen für die Schaffung wünschenswerter Lebensbedingungen sind und daher mit einem höheren SWB verbunden sind (McCrae & Costa, 1991).

Entsprechend kann angenommen werden, dass die Persönlichkeitsstruktur die Bewältigung und das Wohlbefinden während der COVID-19-Pandemie beeinflusst hat. Daher wurde ihr Einfluss in mehreren neueren Studien untersucht. Extraversion und in geringerem Maße auch Verträglichkeit schienen schützend zu wirken. Ein möglicher Grund dafür ist, dass Personen, die auf diesen beiden Dimensionen hohe Werte erzielten, in der Lage waren, den negativen Affekt während der Isolation durch Aktivierung von Bewältigungsmechanismen zu verringern (Nikčević et al., 2021). Gleichzeitig waren höhere Werte in den Persönlichkeitsdimensionen Extraversion und Neurotizismus mit einer negativeren Bewertung der Pandemie verbunden (Modersitzki et al., 2021). Insbesondere die Persönlichkeitsdimension Neurotizismus scheint einen Einfluss darauf zu haben, wie Menschen mit der Pandemie umgehen. So schienen

Personen mit hohen Neurotizismus-Werten eher besorgt und beunruhigt über COVID-19 zu sein (Aschwanden et al., 2021; Kroencke et al., 2020).

Insgesamt kann angenommen werden, dass die beiden Persönlichkeitsdimensionen Extraversion und Neurotizismus am stärksten mit dem Wohlbefinden während der COVID-19-Pandemie verbunden waren. Die Zusammenhänge zwischen SWB und Persönlichkeit wurden jedoch in erster Linie in Querschnittsstudien nachgewiesen (Modersitzki et al., 2021; Nikčević et al., 2021), die im Hinblick auf Gruppenunterschiede analysiert wurden. Die Auswirkungen könnten aber je nach Zeitpunkt und Dauer des Lockdowns unterschiedlich stark sein, weshalb sie genauer und längsschnittlich betrachtet werden sollten.

Dennoch blieben die üblichen Korrelationen (positive Korrelation mit SWB: Extraversion, Verträglichkeit, Gewissenhaftigkeit und Offenheit; negativ: Neurotizismus) während der Pandemie bestehen (Modersitzki et al., 2021). Das entgegengesetzte Muster der Korrelationen wurde bei Angst und depressiven Symptomen gezeigt (Nikčević et al., 2021). In Bezug auf die negative Bewertung wurde der erwartete Effekt gefunden, dass Personen mit einem hohen Neurotizismuswert negativer auf ein Ereignis wie die COVID-19 Pandemie reagieren. Bei der Extraversion hingegen sind die Ergebnisse unerwartet. Normalerweise reagieren extravertierte Menschen weniger negativ auf ein negatives Ereignis. Im Falle der Pandemie reagieren sie jedoch auch stark negativ, vermutlich wegen der Kontaktbeschränkungen während des Lockdowns. Dies könnte daran liegen, dass soziales Verhalten normalerweise ein Schutzfaktor ist, der während der Abriegelung nicht zum Tragen kam. Es ist also denkbar, dass während der Pandemie ansonsten schützende oder hinderliche Ressourcen und Merkmale aufgrund der zahlreichen Einschränkungen im Leben einen anderen Einfluss auf das SWB hatten und dass es zu Wechselwirkungen kam. Dies deutet darauf hin, dass die Pandemie zwar ein einschneidendes Ereignis war, der Einfluss aber nicht so stark ist, dass er die Variation überdeckt. Es sind also immer noch interindividuelle Unterschiede zu erkennen.

Aus diesem Grund sollte das Alter neben der Persönlichkeit als eine Form grundlegender soziodemographischer Information miteinbezogen werden. In einer britischen Stichprobe zeigten O'Connor und Kollegen (2021), dass insbesondere

junge Erwachsene im Alter von 18 bis 29 Jahren die größten Probleme mit der psychischen Gesundheit hatten. Diese Ergebnisse stimmen mit denen von Daly und Kollegen (2022) überein, die in ihrer britischen Stichprobe bei den 18- bis 34-jährigen Teilnehmenden die schlechtesten Ergebnisse in Bezug auf die psychische Gesundheit feststellten.

Zur Analyse des Einflusses der Persönlichkeit wurden die zuvor definierten latenten Wachstumskurvenmodelle (LGCM) um die Big-Five-Persönlichkeitsfaktoren sowie das Alter als erklärende Faktoren erweitert. Alle Modelle wiesen auch nach der Integration einen guten Model-Fit auf. Die Schätzungen der Intercept- und Slope-Faktoren sowie der Pfadkoeffizienten für die Auswirkungen der Persönlichkeit auf die Trajektorien und das Niveau des SWB nach dem Lockdown sind in Tabelle 5 dargestellt.

Tabelle 5. Auswirkungen der Persönlichkeit und des Alters auf das Wohlbefinden

Model		\bar{x}	s	E	V	G	N	O	PO	Alter	R ²
Lebenszufriedenheit	i	58.44*	160.31*	-0.08	2.83	4.35	-7.21**	3.05	-2.63	0.03	.314
	s	3.59	28.75	-2.39	-0.858	3.52	0.48	0.71	-0.93	-0.11	.258
	q	1.93	2.01	-0.65	-0.02	0.82	0.1	-0.16	-0.3	-0.03	.224
Stress	i	31.62	370.8*	4.52*	-1.23	-3.52	10.16**	-2.46	-1.67	-0.13	.192
	s1	-11.69	233.81	-8.8*	-1.92	10.68*	-1.15	-1.68	4.48	-0.1	.268
	s2	-9.12	12.81	1.49*	-0.85	-0.62	0.3	1.18	0.34	0.1*	.278
Psychische Belastung	i	-16.85	329.44*	2.46	1.8	-0.59	11.6*	1.06	5.81	0.01	.243
	s1	14.43	151.51	-5.77	-0.12	8.56*	-0.62	-4.32	-3.03	-0.27	.301
	s2	-3.56	19.32*	0.78	-1.39	-0.26	-0.51	1.09	0.92	0.09*	.173
Einsamkeit	i	15.74	281.75*	0.29	-5.07*	-5.71*	5.27*	9.24*	3.56	-0.3*	.325
	s	-10.94	332.96	-2.88	6.24	5.36	3.26	-7.93*	0.47	0.14	.137
	q	2.63	24.22	0.94	-1.29	-1.31	-0.82	1.94	-0.49	-0.05	.123

Anmerkungen. * $p < .05$; ** $p < .001$; i = intercept; s = slope; q = quadratic slope; E = Extraversion; V = Verträglichkeit; G = Gewissenhaftigkeit; N = Neurotizismus; O = Offenheit für Erfahrungen; PO = Persönlichkeitsorganisation³

³ Da Persönlichkeit auch im klinischen Sinne auf einer basaleren Ebene der psychologischen Funktionsweise verstanden werden kann, wurde neben den Big Five und dem Alter in Studie IV zusätzlich die Persönlichkeitsorganisation (PO) betrachtet. Nach Kernbergs (1984) Modell der Persönlichkeitsorganisation hängt diese mit dem Auftreten, der Schwere und dem Verlauf von psychischen Störungen zusammen. Da in Studie IV keine Zusammenhänge mit einer der betrachteten Facetten des Wohlbefindens gefunden wurden, wird die Persönlichkeitsorganisation aus ökonomischen Gründen in der vorliegenden Arbeit nicht separat eingeführt. Methodik, Ergebnisse und Diskussionspunkte zu diesem Persönlichkeitsmaß können direkt dem Originalmanuskript von Studie IV entnommen werden.

Die Ergebnisse zeigen, dass die Persönlichkeitsfaktoren in unterschiedlichem Maße mit dem Wohlbefinden zusammenhängen. Wie erwartet und aus der Literatur bekannt, war Neurotizismus mit dem durchschnittlichen Niveau aller Maße psychischen Wohlbefindens nach dem Lockdown verbunden. Ein höherer Neurotizismus war mit einer geringeren Lebenszufriedenheit nach Aufhebung der Einschränkungen verbunden. Im Einklang mit der "Set-Point"-Theorie könnte dies zum Teil eine Rückkehr zu den Unterschieden in der Lebenszufriedenheit widerspiegeln, die vor der Pandemie mit Neurotizismus verbunden waren. In ihrer Meta-Analyse berichteten Anglim und Kollegen (2020) über eine mäßige Korrelation zwischen Neurotizismus und Lebenszufriedenheit ($r = -.39$; 95% CI = $[-.41; -.38]$), die größer war als die gefundene Effektgröße nach der Aufhebung der Beschränkungen ($r = -.19$). Da in Studie IV keine Daten aus der Zeit vor der Pandemie vorliegen, kann nicht genau bestimmt werden, ob der für die vorliegende Stichprobe festgestellte Effekt gleich groß oder größer ist als jener vor dem Lockdown. Es kann jedoch geschlussfolgert werden, dass der Unterschied zwischen neurotischen und weniger neurotischen Personen bestehen bleibt. Frühere Untersuchungen zur Lebenszufriedenheit im Zusammenhang mit der COVID-19-Pandemie haben gezeigt, dass neurotische Personen den Informationen über die Pandemie mehr Aufmerksamkeit schenkten, sich mehr Sorgen machten (Aschwanden et al., 2021; Kroencke et al., 2020) und die Pandemie negativer bewerteten (Modersitzki et al., 2021). Diese mit dem Neurotizismus zusammenhängenden Unterschiede müssen jedoch nicht unbedingt die Lebenszufriedenheit nach der Pandemie beeinflussen, da die Unterschiede auch vor der Pandemie vorhanden waren.

Für die Facette Stress zeigte sich, dass durch den Einbezug der Persönlichkeit und des Alters in die Regression mehr Varianz erklärt werden konnte, was die wichtige Rolle von Persönlichkeitsunterschieden als Ursachen für individuelle Unterschiede im SWB untermauert. In Studie IV waren die drei Persönlichkeitsdimensionen Extraversion, Gewissenhaftigkeit und Neurotizismus am relevantesten für Stress, was sowohl die Ergebnisse der Metaanalyse vor der Pandemie (Anglim et al., 2020) als auch die Ergebnisse während der COVID-19-Pandemie (Liu et al., 2021) widerspiegeln. In Studie IV wiesen Teilnehmende mit höherer Extraversion eine stärkere Reduzierung des Stressses während des Lockdowns und ein höheres Stressniveau danach

auf. Dies deutet darauf hin, dass diese Personen in Bezug auf ihre Stressbelastung größere Unterschiede während des Lockdowns und dem anschließenden Zeitraum erlebten. Es scheint plausibel, dass sich die verringerten sozialen Verpflichtungen stärker auf extravertierte Personen auswirkten, die tendenziell mehr soziale Kontakte und potenziell mehr Planungsstress haben als introvertierte Personen. Aus diesem Grund ist es denkbar, dass das Stressniveau der Introvertierten recht stabil blieb, während bei den Extravertierten nach der Aufhebung der Sperre eine Rückkehr zu einem volleren und stressigeren Terminplan stattfand. Eine weitere Persönlichkeitsdimension, die mit dem Stressniveau zusammenhängt, ist die Gewissenhaftigkeit. Gewissenhafte Menschen sind im Allgemeinen gut organisiert, zuverlässig und umsichtig. Während des ersten Lockdowns wurde eine Reihe zusätzlicher Einschränkungen und Empfehlungen ausgesprochen (z. B. Kontaktreduzierung, verstärkte Reinigung und Desinfektion). Entsprechend könnte das erhöhte Stressniveau während des ersten Lockdowns darauf zurückzuführen sein, dass besonders gewissenhafte Personen die Einschränkungen und Empfehlungen ernster genommen haben; so könnte zum Beispiel der Mangel an Desinfektionsmitteln bei dieser Gruppe zu erhöhtem Stress geführt haben. Dies könnte insbesondere damit zu begründen sein, dass gewissenhafte Menschen in der Regel weniger spontan sind und sich daher langsamer an neue Situationen anpassen (Pierre, 2007). Außerdem ist bei besonders gewissenhaften Personen eine höhere Bereitschaft vorhanden, die COVID-19 Präventionsmaßnahmen zu befolgen (Liu et al., 2021). Überdies war die Persönlichkeitsfacette Neurotizismus mit dem Post-Lockdown-Niveau des empfundenen Stresses im Mai 2020 verbunden. Neurotische Personen waren nach dem Lockdown stärker gestresst. Dieser positive Zusammenhang zwischen Neurotizismus und Stress wurde auch von Liu und Kollegen (2021) festgestellt. Generell steht dieses Ergebnis im Einklang mit der Assoziation von Neurotizismus mit negativen Bewertungen und der Tendenz, mehr über die COVID-19-Pandemie zu grübeln und sich Sorgen zu machen (Aschwanden et al., 2021; Kroencke et al., 2020).

Hinsichtlich der psychischen Belastung zeigte sich, dass Gewissenhaftigkeit und Neurotizismus signifikante Prädiktoren waren. Neurotizismus war ebenfalls mit dem Ausmaß der Belastung nach dem Lockdown verbunden. Auch hier waren

Personen, die auf dieser Dimension höhere Werte erreichten, während der letzten Erhebungswelle stärker belastet. Dies stimmt mit der Literatur überein, die besagt, dass Neurotizismus und negative Beurteilungen miteinander verbunden sind (Aschwanden et al., 2021; Kroencke et al., 2020). Darüber hinaus erlebten gewissenhaftere Personen während des Lockdowns mehr psychische Belastung. Dies steht im Einklang mit Venkatesh und Kollegen (2021), die berichten, dass gewissenhaftere Arbeitnehmer während der Pandemie ebenfalls eine höhere Belastung und eine geringere Zufriedenheit angegeben haben.

Zusätzlich zeigt Studie IV, dass sich die Menschen darüber hinaus hinsichtlich des Verlaufs und des Niveaus ihrer Einsamkeit nach dem Lockdown unterscheiden. Die Einbeziehung der Big Five in die Regressionsanalysen ergab, dass verträglichere und gewissenhaftere Personen niedrigere Werte für Einsamkeit aufwiesen, während neurotische und offene Personen höhere Werte hatten. Extraversion hingegen war, im Einklang mit der Literatur (Folk et al., 2020), nicht mit Einsamkeit assoziiert. Die Unterschiede, die für Verträglichkeit, Gewissenhaftigkeit und Neurotizismus gefunden wurden, stimmen mit denen aus der Meta-Analyse von Buecker und Kollegen (2020) überein. Diese Autoren zeigten, dass Unterschiede in der Einsamkeit bereits vor der COVID-19-Pandemie mit der Verträglichkeit ($r = -.27$, 95% CI = $[-.30; -.23]$), der Gewissenhaftigkeit ($r = -.22$, 95% CI = $[-.25; -.19]$) und dem Neurotizismus ($r = .39$, 95% CI = $[.35; .43]$) in den jeweiligen Richtungen zusammenhingen. Zwar waren die in Studie IV gefundenen Effekte bei Personen mit höheren Werten von Neurotizismus und Gewissenhaftigkeit im Vergleich zu früheren Studien geringer, jedoch bildet die Datenerhebung während des ersten Lockdowns eine besondere Situation ab. Entsprechend ist es möglich, dass der Einfluss der Persönlichkeit in dieser Zeit geschwächt war, weil die Pandemie für jeden überwältigend schien. Diese These wird dadurch untermauert, dass in Studie IV eine positive Korrelation zwischen Offenheit und Einsamkeit ($r = .12$) gefunden wurde, während die Metaanalyse vor der Pandemie einen kleinen negativen Zusammenhang ergab ($r = -.12$, 95% CI = $[-.15; -.09]$). Die positive Korrelation aus Studie IV ist möglicherweise darauf zurückzuführen, dass offene Personen eher neue Erfahrungen suchen und gleichzeitig stärker von fehlender

sozialer Interaktion aufgrund von Kontaktbeschränkungen während des Lockdowns betroffen waren.

Darüber hinaus schien das Alter eine Rolle zu spielen. Wir fanden einen positiven Zusammenhang zwischen dem Alter und dem Verlauf von Stress und psychologischer Belastung. Darüber hinaus stand das Alter in einem negativen Zusammenhang mit der Einsamkeit, die nach dem Lockdown berichtet wurde. Ältere Menschen fühlten sich folglich weniger einsam. Jedoch könnte aufgrund dieser außergewöhnlichen Situation neben dem Alter auch andere situative oder individuelle Faktoren (z. B. Heimarbeit, Kinderbetreuung/Hausunterricht, Lebensbedingungen, sozioökonomischer Status) von Bedeutung gewesen sein.

3.4 Der Einfluss von Systemrelevanz, Risikogruppenzugehörigkeit und Geschlecht auf das Wohlbefinden

Die individuellen Unterschiede im Wohlbefinden während der COVID-19-Pandemie können jedoch nicht nur auf die Persönlichkeit zurückgeführt werden. Um weitere Einflussfaktoren zu identifizieren, wurden verschiedene Gruppenvergleiche durchgeführt. Demnach schienen Menschen mit gesundheitlichen Risikofaktoren (chronische Krankheit, Schwangerschaft, Alter, Behinderung) sowie Berufstätige durch die Pandemie stärker belastet zu sein (Recchi et al., 2020). Weitere Einflussfaktoren wurden von Möhring und Kollegen (2020) festgestellt. In ihrer Studie mit deutschen Teilnehmenden wurde ein Rückgang der Arbeitszufriedenheit bei Personen in Kurzarbeit und bei Müttern gefunden. Im Gegensatz dazu scheint die Zufriedenheit von Vätern im Vergleich zu Müttern, weniger negativ beeinflusst worden zu sein.

Um Gruppenunterschiede in Studie IV zu untersuchen, wurde in die Modelle der gesundheitliche Risiko-Status, die Systemrelevanz sowie das Geschlecht und das Alter integriert. Insgesamt wurde in Bezug auf das Wohlbefinden kein Unterschied zwischen systemrelevanten Arbeitnehmenden und den übrigen Teilnehmenden festgestellt.

Beim Status der Hochrisikogruppe zeigte sich ein Unterschied hinsichtlich der Einsamkeit: Personen, die dieser Gruppe zugehörig sind, unterschieden sich in der Trajektorie des Steigungsfaktors von den übrigen Teilnehmenden, indem sie

geringere Einsamkeitswerte aufwiesen ($\bar{x}_s = -18,17, p < .05$). Während die Einsamkeit in der Gruppe mit geringerem Risiko über die vier Erhebungszeitpunkte hinweg relativ stabil blieb, nahm das angegebene Einsamkeitsniveau in der Hochrisikogruppe stetig ab.

Bei der Lebenszufriedenheit wurden für Frauen und Männer unterschiedliche Verläufe gefunden ($\bar{x}_s = -5,07, p < .05$). Es konnte gezeigt werden, dass Frauen während des Lockdowns einen flacheren Verlauf aufwiesen und danach tendenziell ein niedrigeres Niveau der Lebenszufriedenheit hatten. Dieses Ergebnis steht im Einklang mit der Tatsache, dass das SWB von Müttern durch die Pandemie besonders negativ beeinflusst wurde und dass Frauen im Allgemeinen mehr psychische Probleme hatten (Möhring et al., 2020; O'Connor et al., 2020).

Außerdem gab es geschlechtsspezifische Unterschiede im Hinblick auf das Stressniveau nach Beendigung des Lockdowns ($\bar{x}_i = -10,97, p < .05$). Im Durchschnitt gaben weibliche Teilnehmende an, dass sie sowohl während des Lockdowns ein höheres Stressniveau hatten als auch nach Aufhebung der Beschränkungen. Der Einfluss des Alters auf den Stressverlauf, der bereits in Kapitel 3.3 beschrieben wurde, konnte auch in diesem Modell repliziert werden ($\bar{x}_{s2} = 0,13, p < .05$). Ältere Teilnehmende schienen weniger Stressabbau zu erleben als jüngere. Dies könnte auf die zusätzliche Belastung durch kleine Kinder zu Hause zurückzuführen sein. So haben während des ersten Lockdowns Eltern oft nicht nur selbst von zu Hause arbeiteten, sondern auch ihre Kinder betreuen und/oder zu Hause unterrichten müssen (Zinn & Bayer, 2021). Darüber hinaus deuten Studien darauf hin, dass die Kinderbetreuung häufig die Aufgabe der Frauen bleibt (Huebener et al., 2021; Möhring et al., 2020). Daher können höhere Stresswerte für Frauen und ältere Teilnehmende möglicherweise durch unterschiedliche Verantwortlichkeiten (z. B. Kinderbetreuung, Haushalt, Beruf, Einkommen) erklärt werden.

4 Zusammenfassende Betrachtung

4.1 Zusammenfassung der wichtigsten Ergebnisse

Studie I zeigt, dass beide untersuchten Auswahlverfahren (TMS und HAM-Nat) zusammengenommen, trotz unterschiedlicher theoretischer Konzepte, eine sehr ähnliche kognitive Fähigkeit zu messen scheinen. Es kann davon ausgegangen werden, dass beide Subtestgruppen sowohl mit g als auch mit einem weiteren testspezifischen Faktor zusammenhängen. Selbst wenn beide spezifischen Faktoren korreliert sind, verbleiben spezifische Varianzanteile. Daher sollte diese nicht geteilte Varianz der einzelnen Tests weiter analysiert werden, indem die Hochschulnoten in die Modelle aufgenommen werden, um ihre inkrementelle Validität zu untersuchen. Die parallele Verwendung beider Verfahren zur Auswahl von Studierenden scheint aus testtheoretischer Sicht legitim zu sein.

Die Ergebnisse der Studie II zeigen zusammenfassend, dass die Testvorbereitung ein Thema ist, das bei der Verwendung von kognitiven Tests im Auswahlkontext berücksichtigt werden sollte. Dies ist unter anderem damit zu begründen, dass durch Testvorbereitung zusätzliches Wissen vermittelt wird, was Unterschiede in der Testleistung bedingen kann, die jedoch unabhängig von tatsächlichen Unterschieden in den gemessenen Fähigkeiten sind. Dennoch zeigen die Ergebnisse, dass die Testvorbereitung in dieser Studie nicht zu Veränderungen in der Konstrukt- und Kriteriumsvalidität führt. Zur Steigerung der Fairness im Auswahlprozess sollte folglich allen Studieninteressierten ein kostenfreier Zugang zu den Vorbereitungsmaterialien ermöglicht werden.

Studie III repliziert den negativen ToT-Effekt bei figuralen Matrizen Aufgaben, das heißt, schnellere Antworten waren im Durchschnitt eher richtig. In Übereinstimmung mit den Berichten über die doppelte Verarbeitung und die kognitive Belastung wird der Effekt durch die Schwierigkeit der Aufgaben und die Fähigkeiten der Personen moderiert. Bei weniger fähigen Teilnehmenden, die schwierige Aufgaben lösten, steigt mit steigender Bearbeitungszeit auch die Lösungswahrscheinlichkeit. Im Gegensatz dazu ist die Verallgemeinerbarkeit des ToT-Effekts über verschiedene Aufgaben hinweg nur mäßig, was auf eine relevante Rolle aufgabenabhängiger

Moderatoren hinweist und darauf, dass die verwendeten Aufgaben wahrscheinlich unterschiedliche Lösungsprozesse anregen.

Studie IV untersuchte die Auswirkungen des ersten Lockdowns auf das subjektive Wohlbefinden (Lebenszufriedenheit, Stress, psychische Belastung, Einsamkeit) in Deutschland. Die Ergebnisse deuten darauf hin, dass die Lebenszufriedenheit als Reaktion auf den durch COVID-19 bedingten Lockdown kurzfristige Veränderungen aufwies. So verringerte die Abriegelung nach der zweiten Woche tatsächlich Stress und psychische Belastung. Darüber hinaus verbesserten sich sogar die Werte für Einsamkeit im Verlauf des Lockdowns. Wenn Persönlichkeitsmerkmale und andere potenzielle Einflussfaktoren hinzugefügt werden, zeigen die Ergebnisse, dass Neurotizismus und Gewissenhaftigkeit die beiden Dimensionen waren, die am stärksten mit dem Wohlbefinden während des ersten Monats der Pandemie in Verbindung standen. Neurotizismus war negativ mit dem endgültigen Niveau aller untersuchten Maße des Wohlbefindens assoziiert, während Gewissenhaftigkeit positiv mit den Verläufen von Stress und psychologischer Belastung korrelierte. Darüber hinaus schienen auch Alter und Geschlecht von Bedeutung zu sein, allerdings hauptsächlich für den Verlauf des Wohlbefindens. Die Ergebnisse legen nahe, dass Einschränkungen als Reaktion auf die Pandemie einen kurzfristigen Einfluss auf alle betrachteten Maße psychischen Wohlbefindens haben können und dass Persönlichkeitseigenschaften bei der Analyse des psychischen Wohlbefindens berücksichtigt werden sollten.

4.2 Praktische Implikationen für zukünftige Studierendenauswahltests im Hinblick auf die COVID-19 Pandemie

Die vorliegende Dissertation zeigt, dass der Beitrag differentialpsychologischer Forschung für unser alltägliches Leben ungemein wichtig ist. Dabei kann sie einen Beitrag zu den unterschiedlichsten Bereichen und Themen leisten, wie zum Beispiel zur Studierendenauswahl im Fach Medizin und den Auswirkungen der COVID-19 Pandemie.

Überdies verdeutlicht die globale COVID-19 Pandemie, dass die Auswahl des medizinischen Personals besonders wichtig ist und dringend verbessert werden sollte. Dabei ist nicht nur die Passung der Bewerbenden von Bedeutung, vielmehr muss perspektivisch die Durchführung fairer Auswahlverfahren auch unter

besonderen Bedingungen gewährleistet sein. Daher sollte ein zukünftiges Studierendenauswahlverfahren für das Fach Medizin die folgenden Aspekte berücksichtigen:

- Kombination beider bisheriger Auswahlverfahren zur Vereinheitlichung der Selektion (Studie I)
- Bereitstellung von kostenlosem Vorbereitungsmaterial (Studie II)
- Testungen in Abhängigkeit der individuellen Leistungsfähigkeit (Studie III)
- Orts- und zeitunabhängige Testungen, um bei zukünftigem Pandemiegeschehen eine valide Selektion gewährleisten zu können (Studie IV)

Eine Zukunftsvision zur Auswahl von Medizinstudierenden, bei der die oben genannten Aspekte berücksichtigt werden, ist daher die Einrichtung von Testzentren. Diese Testzentren könnten an den einzelnen Universitäten eingerichtet werden, ganzjährig geöffnet sein und daher zur Entzerrung bei Auswahltestungen beitragen. Dadurch wird nicht nur die klassische papierbasierte Testung in Hallen abgelöst, sondern es werden auch kürzere Anreisewege gewährleistet. Eine Vereinheitlichung des Auswahlverfahrens würde überdies die Ökonomie weiterhin positiv beeinflussen, da Medizinstudierende, unabhängig von der präferierten Universität, nur einen Auswahltest durchlaufen müssten. Gleichzeitig würde eine Anpassung des Testmodus den Selektionsprozess ökonomischer gestalten, indem in besagten Testzentren eine computergestützte Selektion durchgeführt wird. Dabei könnten Linking-Items verwendet werden, um die Testergebnisse der Teilnehmenden miteinander in Relation zu setzen. Zusätzlich bestünde die Option des adaptiven Testens, so dass die Auswahl weiterer Aufgaben in Abhängigkeit der Leistung der Teilnehmenden bei vorherigen Aufgaben erfolgt. Durch die Individualisierung der Testungen würden Teilnehmende zum einen in Abhängigkeit der individuellen Leistungsfähigkeit sowie zum anderen in einer kürzeren Zeit getestet werden. Zusätzlich sollte allen angemeldeten Personen kostenfreies Testmaterial zur Vorbereitung bereitgestellt werden.

Die oben beschriebene Zukunftsperspektive zeigt, dass Auswahltestungen dringend fairer, ökonomischer und digitaler werden sollten, um so perspektivisch eine valide Selektion von Medizinstudierenden zum Erhalt unseres Gesundheitssystems gewährleisten zu können.

Literaturverzeichnis

- Alexander, R. M., & Reynolds, M. R. (2020). Intelligence and adaptive behavior: A meta-analysis. *School Psychology Review*, 49(2), 85-110. <https://doi.org/10.1080/2372966X.2020.1717374>
- Alliger, G. M., & Dwight, S. A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. *Educational and Psychological Measurement*, 60(1), 59-72. <https://doi.org/10.1177/00131640021970367>
- Anglim, J., & Horwood, S. (2021). Effect of the COVID-19 pandemic and Big Five personality on subjective and psychological well-being. *Social Psychological and Personality Science*, 12(8), 1527-1537. <https://doi.org/10.1177/1948550620983047>
- Anglim, J., Horwood, S., Smillie, L. D., Marrero, R. J., & Wood, J. K. (2020). Predicting psychological and subjective well-being from personality: A meta-analysis. *Psychological bulletin*, 146(4), 279-323. <https://doi.org/10.1037/bul0000226>
- Aschwanden, D., Strickhouser, J. E., Sesker, A. A., Lee, J. H., Luchetti, M., Stephan, Y., Sutin, A. R., & Terracciano, A. (2021). Psychological and behavioural responses to coronavirus disease 2019: The role of personality. *European Journal of Personality*, 35(1), 51-66. <https://doi.org/10.1002/per.2281>
- Bao, Y., Sun, Y., Meng, S., Shi, J., & Lu, L. (2020). 2019-nCoV epidemic: Address mental health care to empower society. *The Lancet*, 395(10224), e37-e38. [https://doi.org/10.1016/S0140-6736\(20\)30309-3](https://doi.org/10.1016/S0140-6736(20)30309-3)
- Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology*, 81(3), 261-272. <https://doi.org/10.1037/0021-9010.81.3.261>
- Becker, N., Schmitz, F., Falk, A., Feldbrügge, J., Recktenwald, D., Wilhelm, O., Preckel, F., & Spinath, F. M. (2016). Preventing response elimination strategies improves the convergent validity of figural matrices. *Journal of Intelligence*, 4(1), 2. <https://doi.org/10.3390/jintelligence4010002>

- Becker, N., Schmitz, F., Göritz, A., & Spinath, F. M. (2016). Sometimes more is better, and sometimes less is better: Task complexity moderates the response time Accuracy Correlation. *Journal of Intelligence*, 4(3), 11. <https://doi.org/10.3390/jintelligence4030011>
- BVerfG, Urteil des Ersten Senats vom 19. Dezember 2017 - 1 BvL 3/14 -, Rn. 1-253, http://www.bverfg.de/e/ls20171219_1bvl000314.html
- Brandstätter, H., & Farthofer, A. (2002). Studienerfolgsprognose -- konfigurativ oder linear additiv?. *Zeitschrift Für Differentielle Und Diagnostische Psychologie*, 23(4), 381-391. <https://doi.org/10.1024/0170-1789.23.4.381>
- Brooks, S. K., Webster, R. K., Smith, L. E., Woodland, L., Wessely, S., Greenberg, N., & Rubin, G. J. (2020). The psychological impact of quarantine and how to reduce it: Rapid review of the evidence. *The Lancet*, 395(10227), 912-920. [https://doi.org/10.1016/S0140-6736\(20\)30460-8](https://doi.org/10.1016/S0140-6736(20)30460-8)
- Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, 80(4), 796-846. <https://doi.org/10.1111/j.1467-6494.2011.00749.x>
- Buchmann, C., Condrón, D. J., & Roscigno, V. J. (2010). Shadow education, american style: Test preparation, the SAT and college enrollment. *Social Forces*, 89(2), 435-461. <https://doi.org/10.1353/sof.2010.0105>
- Buecker, S., Maes, M., Denissen, J. J. A., & Luhmann, M. (2020). Loneliness and the Big Five personality traits: A meta-analysis. *European Journal of Personality*, 34(1), 8-28. <https://doi.org/10.1002/per.2229>
- Campbell, A., Converse, P. E., & Rodgers, W. L. (1976). *The quality of American life: Perceptions, evaluations, and satisfactions*. Russell Sage Foundation.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97(3), 404-431. <https://doi.org/10.1037/0033-295X.97.3.404>

- Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1-22. <https://doi.org/10.1037/h0046743>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Cuadrado, D., Salgado, J. F., & Moscoso, S. (2021). Personality, intelligence, and counterproductive academic behaviors: A meta-analysis. *Journal of Personality and Social Psychology, 120*(2), 504–537. <https://doi.org/10.1037/pspp0000285>
- Daly, M., Sutin, A., & Robinson, E. (2022). Longitudinal changes in mental health and the COVID-19 pandemic: Evidence from the UK Household Longitudinal Study. *Psychological Medicine, 52*, 2549-2558. <https://doi.org/10.1017/S0033291720004432>
- De Boeck, P., & Jeon, M. (2019). An overview of models for response times and processes in cognitive tests. *Frontiers in Psychology, 10*. <https://doi.org/10.3389/fpsyg.2019.00102>
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13-21. <https://doi.org/10.1016/j.intell.2006.02.001>
- Diener, E. (2006). Guidelines for national indicators of subjective well-being and ill-being. *Journal of Happiness Studies, 7*(4), 397-404. <https://doi.org/10.1007/s10902-006-9000-y>
- Diener, E. D., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of personality assessment, 49*(1), 71-75.
- Diener, E., Lucas, R. E., & Scollon, C. N. (2006). Beyond the hedonic treadmill: Revising the adaptation theory of well-being. *American Psychologist, 61*(4), 305-314. <https://doi.org/10.1037/0003-066X.61.4.305>

- Diener, E., Suh, E. M., Lucas, R. E., & Smith, H. L. (1999). Subjective well-being: Three decades of progress. *Psychological Bulletin*, *125*(2), 276-302. <https://doi.org/10.1037/0033-2909.125.2.276>
- Dietze, T. (2011). *Zum Übergang auf weiterführende Schulen. Auswertung schulstatistischer Daten aus 10 Bundesländern*. pedocs. <http://nbn-resolving.de/urn:nbn:de:0111-opus-41919>
- DiTrapani, J., Jeon, M., De Boeck, P., & Partchev, I. (2016). Attempting to differentiate fast and slow intelligence: Using generalized item response trees to examine the role of speed on intelligence tests. *Intelligence*, *56*, 82-92. <https://doi.org/10.1016/j.intell.2016.02.012>
- Dohrenwend, B. P. (2006). Inventorying stressful life events as risk factors for psychopathology: Toward resolution of the problem of intracategory variability. *Psychological Bulletin*, *132*(3), 477-495. <https://doi.org/10.1037/0033-2909.132.3.477>
- Duan, L., & Zhu, G. (2020). Psychological interventions for people affected by the COVID-19 epidemic. *The Lancet Psychiatry*, *7*(4), 300-302. [https://doi.org/10.1016/S2215-0366\(20\)30073-0](https://doi.org/10.1016/S2215-0366(20)30073-0)
- Entringer, T., Kröger, H., Schupp, J., Kühne, S., Liebig, S., Goebel, J., Grabka, M. M., Graeber, D., Kroh, M., Schröder, C., Seebauer, J., & Zinn, S. (2020). *Psychische Krise durch Covid-19? Sorgen sinken, Einsamkeit steigt, Lebenszufriedenheit bleibt stabil*. (SOEPpapers on Multidisciplinary Panel Data Research, 1087). Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).
- Estrada, E., Ferrer, E., Abad, F. J., Román, F. J., & Colom, R. (2015). A general factor of intelligence fails to account for changes in tests' scores after cognitive practice: A longitudinal multi-group latent-variable study. *Intelligence*, *50*, 93-99. <https://doi.org/10.1016/j.intell.2015.02.004>
- Foa, R., Gilbert, S., & Fabian, M. O. (2020). COVID-19 and subjective well-being: Separating the effects of lockdowns from the pandemic. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3674080>

- Folk, D., Okabe-Miyamoto, K., Dunn, E., & Lyubomirsky, S. (2020). *Did social connection decline during the first wave of COVID-19?: The role of extraversion. Collabra: Psychology*, 6(1). <https://doi.org/10.1525/collabra.365>
- Formazin, M., Schroeders, U., Köller, O., Wilhelm, O., & Westmeyer, H. (2011). Studierendenauswahl im Fach Psychologie: Testentwicklung und Validitätsbefunde. *Psychologische Rundschau*, 62(4), 221-236. <https://doi.org/10.1026/0033-3042/a000093>
- Fujita, F., & Diener, E. (2005). Life satisfaction set point: Stability and change. *Journal of Personality and Social Psychology*, 88(1), 158-164. <https://doi.org/10.1037/0022-3514.88.1.158>
- Furnham, A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology*, 9(1), 46-55.
- Gignac, G. E. (2006). A confirmatory examination of the factor structure of the multidimensional aptitude battery: Contrasting oblique, higher order, and nested factor models. *Educational and Psychological Measurement*, 66(1), 136-145. <https://doi.org/10.1177/0013164405278568>
- Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: G as superordinate or breadth factor? *Psychology Science Quarterly*, 50, 21-43.
- Gignac, G. E. (2015). Raven's is not a pure measure of general intelligence: Implications for g factor theory and the brief measurement of g. *Intelligence*, 52, 71-79. <https://doi.org/10.1016/j.intell.2015.07.006>
- Gold, A., & Souvignier, E. (2005). Prognose der Studierfähigkeit. *Zeitschrift Für Entwicklungspsychologie und Pädagogische Psychologie*, 37(4), 214-222. <https://doi.org/10.1026/0049-8637.37.4.214>
- Goldhammer, F., Naumann, J., & Greiff, S. (2015). More is not always better: The relation between item response and item response time in Raven's Matrices. *Journal of Intelligence*, 3(1), 21-40. <https://doi.org/10.3390/jintelligence3010021>

- Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology, 106*(3), 608-626. <https://doi.org/10.1037/a0034716>
- Griffith, R. L., Chmielowski, T., & Yoshita, Y. (2007). Do applicants fake? An examination of the frequency of applicant faking behavior. *Personnel Review, 36*(3), 341-355. <https://doi.org/10.1108/00483480710731310>
- Hahn, E., Specht, J., Gottschling, J., & Spinath, F. M. (2015). Coping with unemployment: The impact of unemployment duration and personality on trajectories of life satisfaction. *European Journal of Personality, 29*(6), 635-646. <https://doi.org/10.1002/per.2034>
- Haier, R. J. (2014). Increased intelligence is a myth (so far). *Frontiers in Systems Neuroscience, 8*, 34. <https://doi.org/10.3389/fnsys.2014.00034>
- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology, 92*(2), 373-385. <https://doi.org/10.1037/0021-9010.92.2.373>
- Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2015). Do we really become smarter when our fluid-intelligence test scores improve? *Intelligence, 48*, 1-14. <https://doi.org/10.1016/j.intell.2014.10.005>
- Helbig, M., & Nikolai, R.. (2015). *Die Unvergleichbaren. Der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949.* <https://doi.org/10.25656/01:11095>
- Hell, B., Trapmann, S., & Schuler, H. (2007). Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum. *Empirische Pädagogik, 21*(3). 251-270.
- Hissbach, J. C., Klusmann, D., & Hampe, W. (2011). Dimensionality and predictive validity of the HAM-Nat, a test of natural sciences for medical school admission. *BMC Medical Education, 11*(1), 83. <https://doi.org/10.1186/1472-6920-11-83>

- Hissbach, J., Feddersen, L., Sehner, S., & Hampe, W. (2012). Suitability of the HAM-Nat test and TMS module “basic medical-scientific understanding” for medical school selection. *GMS Zeitschrift Für Medizinische Ausbildung*, 29(5), Doc72. <https://doi.org/10.3205/zma000842>
- Hissbach, J., Klusmann, D., & Hampe, W. (2011). Reliability of a science admission test (HAM-Nat) at Hamburg medical school. *GMS Zeitschrift Für Medizinische Ausbildung*, 28(3), Doc44. <https://doi.org/10.3205/zma000756>
- Hochschulrektorenkonferenz. (2021). *Statistische Daten zu Studienangeboten an Hochschulen in Deutschland. Studiengänge, Studierende, Absolventinnen und Absolventen. Wintersemester 2021/2022*. Statistiken Zur Hochschulpolitik. https://www.hrk.de/fileadmin/redaktion/hrk/02-Dokumente/02-10-Publikationsdatenbank/Stat-2021-01_WS_2021_22.pdf
- Höft, S. (2011). Anwendungen der Psychodiagnostik in der Arbeits- und Organisationspsychologie. In L. F. Hornke, M. Amelang, N. Birbaumer, & M. Kersting (Hrsg.), *Grundfragen und Anwendungsfelder psychologischer Diagnostik* (S. 131-158). Hogrefe.
- Hornke, L. F. (2000). Item response times in computerized adaptive testing. *Psicologica*, 21(1-2), 175-189.
- Huebener, M., Waights, S., Spiess, C. K., Siegel, N. A., & Wagner, G. G. (2021). Parental well-being in times of Covid-19 in Germany. *Review of Economics of the Household*, 19(1), 91-122. <https://doi.org/10.1007/s11150-020-09529-4>
- ITB Consulting GmbH (2021). *Test für medizinische Studiengänge TMS – Informationsbroschüre 2021*. Bonn. Zuletzt geöffnet am 10. Februar 2023, https://cip.dmed.uni-heidelberg.de/tms-info/tms-info/fileadmin/pdf/informationsbroschuere_tms.pdf
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.
- Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence*, 32(1), 95-107. [https://doi.org/10.1016/S0160-2896\(03\)00062-X](https://doi.org/10.1016/S0160-2896(03)00062-X)

- Johnson, W., Nijenhuis, J. te, & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, 36(1), 81-95. <https://doi.org/10.1016/j.intell.2007.06.001>
- Kadmon, G., Kirchner, A., Duelli, R., Resch, F., & Kadmon, M. (2012). Warum der Test für Medizinische Studiengänge (TMS)? *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 106(2), 125-130. <https://doi.org/10.1016/j.zefq.2011.07.022>
- Kim, K. H. (2005). Can only intelligent people be creative? A meta-analysis. *Journal of Secondary Gifted Education*, 16(2-3), 57-66. <https://doi.org/10.4219/jsge-2005-473>
- Kivi, M., Hansson, I., & Bjälkebring, P. (2020). Up and about: Older adults' well-being during the COVID-19 pandemic in a Swedish longitudinal study. *The Journals of Gerontology: Series B*, 76(2), e4-e9. <https://doi.org/10.1093/geronb/gbaa084>
- Koch, M., Spinath, F. M., Greiff, S., & Becker, N. (2022). Development and validation of the open matrices item bank. *Journal of Intelligence*. 10(3) <https://doi.org/10.3390/jintelligence10030041>
- Köbler, R. (2009). *Neue Wege im Recruiting: Mehr Effektivität mit Gravesmodell und Metaprogrammen. Ein praxisorientiertes Handbuch*. Junfermann Verlags.
- Köller, O., & Maaz, K. (2017). Soziale Disparitäten bei Abiturleistungen. In T. Eckert & B. Gniewosz (Hrsg.), *Bildungsgerechtigkeit* (S. 153-166). Springer Fachmedien. https://doi.org/10.1007/978-3-658-15003-7_9
- Kroencke, L., Geukes, K., Utesch, T., Kuper, N., & Back, M. D. (2020). Neuroticism and emotional risk during the COVID-19 pandemic. *Journal of Research in Personality*, 89, 104038. <https://doi.org/10.1016/j.jrp.2020.104038>
- Kuper, N., Modersitzki, N., Phan, L. V., & Rauthmann, J. F. (2021). *The situation during the COVID-19 pandemic: A snapshot in Germany*. *PloS one*, 16(2). <https://doi.org/10.1371/journal.pone.0245719>

- Lasry, N., Watkins, J., Mazur, E., & Ibrahim, A. (2013). Response times to conceptual questions. *American Journal of Physics*, *81*(9), 703-706. <https://doi.org/10.1119/1.4812583>
- Liu, S., Lithopoulos, A., Zhang, C.-Q., Garcia-Barrera, M. A., & Rhodes, R. E. (2021). Personality and perceived stress during COVID-19 pandemic: Testing the mediating role of perceived threat and efficacy. *Personality and Individual Differences*, *168*, 110351. <https://doi.org/10.1016/j.paid.2020.110351>
- Loesche, P., Wiley, J., & Hasselhorn, M. (2015). How knowing the rules affects solving the Raven Advanced Progressive Matrices Test. *Intelligence*, *48*, 58-75. <https://doi.org/10.1016/j.intell.2014.10.004>
- Luchetti, M., Lee, J. H., Aschwanden, D., Sesker, A., Strickhouser, J. E., Terracciano, A., & Sutin, A. R. (2020). The trajectory of loneliness in response to COVID-19. *The American Psychologist*, *75*(7), 897-908. <https://doi.org/10.1037/amp0000690>
- Luhmann, M., & Hawkley, L. C. (2016). Age differences in loneliness from late adolescence to oldest old age. *Developmental Psychology*, *52*(6), 943-959. <https://doi.org/10.1037/dev0000117>
- Lykken, D., & Tellegen, A. (1996). Happiness is a stochastic phenomenon. *Psychological Science*, *7*(3), 186-189. <https://doi.org/10.1111/j.1467-9280.1996.tb00355.x>
- McCrae, R. R., & Costa, P. T. (1991). Adding Liebe und Arbeit: The full five-factor model and well-being. *Personality and Social Psychology Bulletin*, *17*(2), 227-232. <https://doi.org/10.1177/014616729101700217>
- McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1-10. <https://doi.org/10.1016/j.intell.2008.08.004>
- Messick, S. (1982). The values of ability testing: Implications of multiple perspectives about criteria and standards. *Educational Measurement: Issues and Practice*, *1*(3), 9-12. <https://doi.org/10.1111/j.1745-3992.1982.tb00660.x>

- Miller, C. E., & Barrett, G. V. (2008). The coachability and fakability of personality-based selection tests used for police selection. *Public Personnel Management*, 37(3), 339-351. <https://doi.org/10.1177/009102600803700306>
- Modersitzki, N., Phan, L. V., Kuper, N., & Rauthmann, J. F. (2021). Who is impacted? Personality predicts individual differences in psychological consequences of the COVID-19 pandemic in Germany. *Social Psychological and Personality Science*, 12(6), 1110-1130. <https://doi.org/10.1177/1948550620952576>
- Möhring, K., Naumann, E., Reifenscheid, M., Wenz, A., Rettig, T., Krieger, U., Friedel, S., Finkel, M., Cornesse, C., & Blom, A. G. (2020). The COVID-19 pandemic and subjective well-being: Longitudinal evidence on satisfaction with work and family. *European Societies*, 1-17. <https://doi.org/10.1080/14616696.2020.1833066>
- Molenaar, D., & de Boeck, P. (2018). Response mixture modeling: Accounting for heterogeneity in item characteristics across response times. *Psychometrika*, 83(2), 279-297. <https://doi.org/10.1007/s11336-017-9602-9>
- Neisser, U., Boodoo, G., Bouchard Jr., T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2), 77-101. <https://doi.org/10.1037/0003-066X.51.2.77>
- Neumann, M., Nagy, G., Trautwein, U., & Lüdtke, O. (2009). Vergleichbarkeit von Abiturleistungen. *Zeitschrift für Erziehungswissenschaft*, 12(4), 691-714. <https://doi.org/10.1007/s11618-009-0099-6>
- Ng, W., Diener, E., Aurora, R., & Harter, J. (2009). Affluence, feelings of stress, and well-being. *Social Indicators Research*, 94(2), 257-271. <https://doi.org/10.1007/s11205-008-9422-5>
- Nguyen, M. H., Gruber, J., Fuchs, J., Marler, W., Hunsaker, A., & Hargittai, E. (2020). Changes in digital communication during the COVID-19 global pandemic: Implications for digital inequality and future research. *Social Media + Society*, 6(3). <https://doi.org/10.1177/2056305120948255>

- Nikčević, A. V., Marino, C., Kolubinski, D. C., Leach, D., & Spada, M. M. (2021). Modelling the contribution of the Big Five personality traits, health anxiety, and COVID-19 psychological distress to generalised anxiety and depressive symptoms during the COVID-19 pandemic. *Journal of Affective Disorders, 279*, 578-584. <https://doi.org/10.1016/j.jad.2020.10.053>
- O'Connor, R. C., Wetherall, K., Cleare, S., McClelland, H., Melson, A. J., Niedzwiedz, C. L., O'Carroll, R. E., O'Connor, D. B., Platt, S., Scowcroft, E., Watson, B., Zortea, T., Ferguson, E., & Robb, K. A. (2021). Mental health and well-being during the COVID-19 pandemic: Longitudinal analyses of adults in the UK COVID-19 Mental Health & Wellbeing study. *The British Journal of Psychiatry, 218*(6), 326-333. <https://doi.org/10.1192/bjp.2020.212>
- Ozer, D. J., & Benet-Martínez, V. (2005). Personality and the prediction of consequential outcomes. *Annual Review of Psychology, 57*(1), 401-421. <https://doi.org/10.1146/annurev.psych.57.102904.190127>
- Partchev, I., & De Boeck, P. (2012). Can fast and slow intelligence be differentiated? *Intelligence, 40*(1), 23-32. <https://doi.org/10.1016/j.intell.2011.11.002>
- Paulus, L., Spinath, F. M., & Hahn, E. (2021). How do educational inequalities develop? The role of socioeconomic status, cognitive ability, home environment, and self-efficacy along the educational path. *Intelligence, 86*, 101528. <https://doi.org/10.1016/j.intell.2021.101528>
- Pierre, J. R. (2007). *Spontaneity and personality: The relationship between spontaneity, extraversion, neuroticism, openness to experience, agreeableness, and conscientiousness* (Doctoral dissertation, Roosevelt University). proquest.com. <https://www.proquest.com/open-view/b0bc5bf5b024994b0a25a505817e19d8/1?cbl=18750&pq-origsite=gscholar&parentSessionId=ERc%2FxB0w3%2F7XH801Mh3T1r7pXCibTOGW4oFQzrE4DkQ%3D>
- Polizzi, C., Lynn, S. J., & Perry, A. (2020). Stress and coping in the time of Covid-19: Pathways to resilience and recovery. *Clinical Neuropsychiatry, 17*(2), 59-62. <https://doi.org/10.36131/CN20200204>

- Recchi, E., Ferragina, E., Helmeid, E., Pauly, S., Safi, M., Sauger, N., & Schradie, J. (2020). The “eye of the hurricane” paradox: An unexpected and unequal rise of well-being during the Covid-19 lockdown in France. *Research in Social Stratification and Mobility*, 68, 100508. <https://doi.org/10.1016/j.rssm.2020.100508>
- Rentzsch, K., & Schütz, A. (2009). *Psychologische Diagnostik: Grundlagen und Anwendungsperspektiven*. W. Kohlhammer Verlag.
- Robie, C., Brown, D. J., & Beaty, J. C. (2007). Do people fake on personality inventories? A verbal protocol analysis. *Journal of Business and Psychology*, 21(4), 489-509. <https://doi.org/10.1007/s10869-007-9038-9>
- 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. <https://doi.org/10.1016/j.intell.2015.09.002>
- Schimmack, U., Diener, E., & Oishi, S. (2009). Life-Satisfaction is a momentary judgment and a stable personality characteristic: The use of chronically accessible and stable sources. In E. Diener (Hrsg.), *Assessing Well-Being: The Collected Works of Ed Diener* (S. 181-212). Springer. https://doi.org/10.1007/978-90-481-2354-4_9
- Schmid, S. (2021). *Deutscher Bundestag - Historische Debatten (19): Milliarden-Paket gegen Corona-Krise*. Deutscher Bundestag. Zuletzt geöffnet am 09. Februar 2023, <https://www.bundestag.de/dokumente/textarchiv/pandemie-856154>
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274. <https://doi.org/10.1037/0033-2909.124.2.262>
- Schmidt-Atzert, L. (2005). Prädiktion von Studienerfolg bei Psychologiestudenten. *Psychologische Rundschau*, 56(2), 131-133. <https://doi.org/10.1026/0033-3042.56.2.131>
- Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their

- relations to working memory and intelligence. *Journal of Experimental Psychology: General*, 136, 414-429. <https://doi.org/10.1037/0096-3445.136.3.414>
- 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. <https://doi.org/10.1016/j.intell.2020.101473>
- Schuler, H. (2014). *Psychologische Personalauswahl: Eignungsdiagnostik für Personalentscheidungen und Berufsberatung* (4. Auflage). Hogrefe.
- Schulz-Zhecheva, Y., Voelkle, M. C., Beauducel, A., Biscaldi, M., & Klein, C. (2016). Predicting fluid intelligence by components of reaction time distributions from simple choice reaction time tasks. *Journal of Intelligence*, 4(3). <https://doi.org/10.3390/jintelligence4030008>
- Schwarzer, R., & Schulz, U. (2001). The role of stressful life events. *Stressful Life Events*, 1-31.
- Schwibbe, A., Lackamp, J., Knorr, M., Hissbach, J., Kadmon, M., & Hampe, W. (2018). Medizinstudierendenauswahl in Deutschland. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 61(2), 178-186. <https://doi.org/10.1007/s00103-017-2670-2>
- Sparfeldt, J. R., Becker, N., Greiff, S., Kersting, M., König, C. J., Lang, J. W. B., & Beauducel, A. (2022). Intelligenz(tests) verstehen und missverstehen. *Psychologische Rundschau*, 73(3), 161-172. <https://doi.org/10.1026/0033-3042/a000597>
- Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2), 201-293. <https://doi.org/10.2307/1412107>
- Specht, J., Egloff, B., & Schmukle, S. C. (2013). Examining mechanisms of personality maturation: The impact of life satisfaction on the development of the Big Five personality traits. *Social Psychological and Personality Science*, 4(2), 181-189. <https://doi.org/10.1177/1948550612448197>

- Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin*, *134*(1), 138-161. <https://doi.org/10.1037/0033-2909.134.1.138>
- Steyer, R., Yousfi, S., & Würfel, K. (2005). Prädiktion von Studienerfolg: *Psychologische Rundschau*, *56*(2), 129-131. <https://doi.org/10.1026/0033-3042.56.2.129>
- te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on g-loaded tests: No g. *Intelligence*, *35*(3), 283-300. <https://doi.org/10.1016/j.intell.2006.07.006>
- Trapmann, S., Hell, B., Weigand, S., & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs—Eine Metaanalyse. *Zeitschrift für Pädagogische Psychologie*, *21*(1), 11-27. <https://doi.org/10.1024/1010-0652.21.1.11>
- UKE (2022a). *Probe HAM-Nat Arithmetisches Problemlösen*. Hamburg. Zuletzt geöffnet am 10. Februar 2023, https://www.auswahltestzentrale.de/medicine2023F/training_pdf/Probe_HAM_Nat_Arithmetisches_Probleml%C3%B6sen_mit_L%C3%B6sungsschl%C3%BCssel.pdf
- UKE (2022b). *Probe HAM-Nat (Naturwissenschaftstest 6)*. Hamburg. Zuletzt geöffnet am 10. Februar 2023, https://www.auswahltestzentrale.de/medicine2023F/training_pdf/Probe_HAM_Nat_Naturwissenschaftstest_6_mit_L%C3%B6sungsschl%C3%BCssel.pdf
- Valerius, S., & Sparfeldt, J. R. (2014). Consistent g- as well as consistent verbal-, numerical- and figural-factors in nested factor models? Confirmatory factor analyses using three test batteries. *Intelligence*, *44*, 120-133. <https://doi.org/10.1016/j.intell.2014.04.003>
- Venkatesh, V., Ganster, D. C., Schuetz, S. W., & Sykes, T. A. (2021). Risks and rewards of conscientiousness during the COVID-19 pandemic. *Journal of Applied Psychology*, *106*(5), 643-656. <https://doi.org/10.1037/apl0000919>
- viaMINT (2022). *Probe HAM-Nat Relationales Schließen*. Hamburg. Zuletzt geöffnet am 23. November 2023, <https://viamint.haw-hamburg.de/mod/quiz/attempt.php?attempt=530893&cmid=3171>

- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*(2), 197-210. <https://doi.org/10.1177/00131649921969802>
- Wang, C., Pan, R., Wan, X., Tan, Y., Xu, L., McIntyre, R. S., Choo, F. N., Tran, B., Ho, R., Sharma, V. K., & Ho, C. (2020). A longitudinal study on the mental health of general population during the COVID-19 epidemic in China. *Brain, Behavior, and Immunity, 87*, 40-48. <https://doi.org/10.1016/j.bbi.2020.04.028>
- Werwick, K., Winkler-Stuck, K., Hampe, W., Albrecht, P., & Robra, B.-P. (2015). Introduction of the HAM-Nat examination - applicants and students admitted to the medical faculty in 2012-2014. *GMS Zeitschrift Für Medizinische Ausbildung, 32*(5). <https://doi.org/10.3205/zma000995>
- Wilhelm, O. (2005). Measuring reasoning ability. In *Handbook of understanding and measuring intelligence* (S. 373-392). Sage Publications, Inc. <https://doi.org/10.4135/9781452233529.n21>
- Zacher, H., & Rudolph, C. W. (2021). Individual differences and changes in subjective wellbeing during the early stages of the COVID-19 pandemic. *American Psychologist, 76*(1). <https://doi.org/10.1037/amp0000702>
- Zinn, S., & Bayer, M. (2021). Subjektive Belastung der Eltern durch die Beschulung ihrer Kinder zu Hause zu Zeiten des Corona-bedingten Lockdowns im Frühjahr 2020. *Zeitschrift für Erziehungswissenschaft, 24*(2), 339-365. <https://doi.org/10.1007/s11618-021-01012-9>

Anhänge

Anhang A. – Beispielaufgabe TMS: Textverständnis (ITB, 2021, S. 35-39)

Zu den Aufgaben der Schilddrüse gehören Bildung, Speicherung und Freisetzung der jodhaltigen Hormone Trijodthyronin (T3) und Thyroxin (T4). In der Schilddrüse befinden sich zahlreiche Hohlräume, Follikel genannt, deren Wände von einer Schicht sogenannter Epithelzellen gebildet werden. Diese Follikel sind mit einer Substanz gefüllt, in der die Hormone T3 und T4 als inaktive Speicherformen enthalten sind. Beim Menschen ist in den Follikeln so viel T3 und T4 gespeichert, dass der Organismus damit für etwa 10 Monate versorgt werden kann. Das für die Hormonbildung erforderliche Jod entstammt der Nahrung und wird von den Epithelzellen als Jodid aus dem Blut aufgenommen. Die Jodid Aufnahme erfolgt an der äußeren Zellmembran der Epithelzellen durch eine sogenannte Jodpumpe. Diese wird durch ein Hormon aus der Hirnanhangsdrüse, das TSH, stimuliert und kann pharmakologisch durch die Gabe von Perchlorat gehemmt werden. Ferner gibt es erbliche Schilddrüsenerkrankungen, bei deren Vorliegen die Jodpumpe nicht funktioniert. Bei Gesunden wird das in die Epithelzellen aufgenommene Jodid im nächsten Schritt unter dem Einfluss eines Enzyms in freies Jod umgewandelt und in die Follikel abgegeben. Die Aktivität dieses Enzyms kann ebenfalls pharmakologisch gehemmt werden. Die letzten Schritte der Hormonbildung finden in den Follikeln, also außerhalb der einzelnen Epithelzellen, statt. In dort vorhandene sogenannte Tyrosin-Reste (des Thyreoglobulins) wird zunächst ein Jodatombau eingebaut. So entstehen Monojodtyrosin-Reste (MIT), von denen ein Teil durch die Bindung je eines weiteren Jodatoms in Dijodtyrosin-Reste (DIT) umgewandelt wird. Durch die Verknüpfung von je zwei DIT-Resten entsteht schließlich T4, während aus der Verbindung je eines MIT-Restes mit einem DIT-Rest T3 hervorgeht. T3 und T4 werden dann in den Follikeln gespeichert und bei Bedarf über die Epithelzellen ins Blut freigesetzt. Diese Freisetzung von T3 und T4 ins Blut (Sekretion) wird über die Hirnanhangsdrüse und den Hypothalamus, einen Teil des Zwischenhirns gesteuert: Das erwähnte Hormon TSH stimuliert außer der Bildung auch die Sekretion von T3 und T4; es ist hinsichtlich seiner eigenen Sekretionsrate jedoch abhängig von der Stimulation durch das hypothalamische Hormon TRH. Die TRH-Sekretion wiederum wird z.B. durch Kälte stimuliert, während Wärme hemmend wirken kann. Neben diesen übergeordneten Steuerungsmechanismen existiert noch ein sogenannter Rückkopplungsmechanismus: Eine hohe Konzentration von T3 und T4 im Blut hemmt die TSH- und die TRH-Sekretion, eine niedrige Konzentration stimuliert sie. Bei den an der Steuerung der Schilddrüsenhormon-Sekretion beteiligten Arealen von Hirnanhangsdrüse und Hypothalamus können krankheitsbedingte Störungen auftreten, die zu einer Über- oder Unterfunktion der Schilddrüse führen. Eine der Hauptwirkungen von T3 und T4 ist die Beeinflussung des Energieumsatzes durch eine Steigerung des Sauerstoffverbrauchs in stoffwechselaktiven Organen. Entsprechend senkt eine zu

niedrige Konzentration der beiden Hormone im Blut (Hypothyreose) den Energieumsatz bzw. die Stoffwechselaktivität unter den normalen Wert, während bei einer zu hohen Konzentration (Hyperthyreose) die Stoffwechselaktivität gesteigert wird. Die Hormone T3 und T4 können ebenso wie TSH und TRH für diagnostische und therapeutische Zwecke synthetisch hergestellt werden.

1. Welcher der folgenden Vorgänge gehört nicht zu den im Text beschriebenen Schritten, die zur Bildung von T3 führen?
 - a) Transport von Jod aus den Epithelzellen in die Follikel
 - b) Umwandlung von Jod in Jodid in den Follikeln
 - c) Transport von Jodid aus dem Blut in die Epithelzellen
 - d) Verknüpfung von MIT- und DIT-Resten in den Follikeln
 - e) Verknüpfung von Jod und Tyrosin-Resten in den Follikeln

2. Durch welche(n) der nachstehenden Sachverhalte kann dem Text zufolge eine verminderte Stoffwechselaktivität bedingt sein?
 - I. operative Entfernung der Schilddrüse
 - II. mehrjährig verminderte Jodaufnahme über die Nahrung
 - III. Funktionsstörung in der Hirnanhangsdrüse
 - a) Nur durch Sachverhalt I kann sie bedingt sein.
 - b) Nur durch die Sachverhalte I und II kann sie bedingt sein.
 - c) Nur durch die Sachverhalte I und III kann sie bedingt sein.
 - d) Nur durch die Sachverhalte II und III kann sie bedingt sein.
 - e) Durch alle drei Sachverhalte kann sie bedingt sein.

3. Welche der nachstehenden Hypothyreosen können dem Text zufolge durch TSH-Gaben positiv beeinflusst werden?
 - I. Hypothyreosen, die durch eine Funktionsstörung in der Hirnanhangsdrüse bedingt sind
 - II. Hypothyreosen, die durch eine Funktionsstörung im Hypothalamus bedingt sind
 - III. Hypothyreosen, die durch einen erbten Funktionsausfall der Jodpumpe bedingt sind
 - a) nur die unter I genannten Hypothyreosen
 - b) nur die unter II genannten Hypothyreosen
 - c) nur die unter I und II genannten Hypothyreosen
 - d) nur die unter I und III genannten Hypothyreosen
 - e) die unter I, II und III genannten Hypothyreosen

4. Die Basedowsche Krankheit beruht auf der Produktion einer Substanz, die auf die Schilddrüse die gleiche Wirkung hat wie TSH, jedoch im Gegensatz zu diesem weder einer Steuerung über das TRH noch einer Kontrolle durch Rückkoppelung unterliegt.

Welche(s) der nachstehenden Symptome können (kann) dem Text zufolge durch die Basedowsche Krankheit bedingt sein?

- I. verringerte Bildung von T3 und T4
 - II. gesteigerter Sauerstoffverbrauch in stoffwechselaktiven Organen
 - III. verringerte TRH-Sekretion
- a) Nur Symptom I kann durch sie bedingt sein.
 - b) Nur Symptom III kann durch sie bedingt sein.
 - c) Nur die Symptome I und II können durch sie bedingt sein.
 - d) Nur die Symptome II und III können durch sie bedingt sein.
 - e) Alle drei Symptome können durch sie bedingt sein.
5. Welche der folgenden Aussagen über die T3- und T4-Sekretion lässt bzw. lassen sich aus dem Text ableiten?

Eine Steigerung der T3- und T4-Sekretion kann bedingt sein durch eine...

- I. vorangegangene zu niedrige T3- und T4-Sekretion.
 - II. Verringerung des Energiebedarfs.
 - III. Überfunktion der Jodpumpe.
- a) Nur Aussage I lässt sich ableiten.
 - b) Nur Aussage II lässt sich ableiten.
 - c) Nur Aussage III lässt sich ableiten.
 - d) Nur die Aussagen I und III lassen sich ableiten.
 - e) Keine der drei Aussagen lässt sich ableiten.
6. Welche der folgenden Aussagen über die TSH-Sekretion lässt bzw. lassen sich aus dem Text ableiten?

Die TSH-Sekretion kann unter bestimmten Bedingungen verändert werden durch...

- I. Kälte.
 - II. T3- und T4-Gaben.
 - III. über Jahre hinweg erfolgende Perchlorat-Gaben.
- a) Nur Aussage I lässt sich ableiten.
 - b) Nur die Aussagen I und II lassen sich ableiten.
 - c) Nur die Aussagen I und III lassen sich ableiten.
 - d) Nur die Aussagen II und III lassen sich ableiten.
 - e) Alle drei Aussagen lassen sich ableiten.

Anhang B. – Danksagung

In den letzten fünf Jahren haben mich viele Menschen auf unterschiedlichste Art und Weise unterstützt. Im Folgenden möchte ich mich von Herzen bei einigen besonderen bedanken. Ihr alle habt dazu beigetragen, dass ich in all der Zeit das Ziel nicht aus den Augen verloren habe und diese Jahre für immer unvergessen sein werden.

Zunächst danke ich an dieser Stelle meinem Doktorvater PD Dr. Nicolas Becker für die hervorragende Betreuung meiner Dissertation. Deinen Rat, deine Rückmeldungen und deinen Rückhalt habe ich stets sehr geschätzt. All das hat dazu beigetragen, dass ich sehr viel in dieser Zeit gelernt habe.

Selbstverständlich gilt mein Dank auch Prof Dr. Frank M. Spinath für die Unterstützung der letzten Jahre auf so vielfältige Art. Vielen Dank für die ganzen Gespräche und Ratschläge, die mir nicht nur im Rahmen der Promotion halfen, sondern gleichzeitig einen großen Einfluss auf meine persönliche Entwicklung hatten.

Darüber hinaus danke ich Prof. Dr. Monika Equit für die Erstellung des Zweitgutachtens dieser Dissertation.

Ein besonderer Dank geht an meinen Qualis-Leidensgenossen Marco Koch, ohne den ich die Möglichkeit zu promovieren nicht bekommen hätte. Daher ein großes „Danke“ für das viele Korrekturlesen, die Hilfen bei diversen Journal-E-Mails, den Technik-Support und natürlich für das Aushalten von neurotischen Anfällen meinerseits – ich gelobe Besserung.

Ein herzliches Dankeschön auch an Dr. Elisabeth Hahn, die trotz Elternzeit immer ein offenes Ohr hatte, mich unterstützte, mir Mut zusprach und hilfreiches Feedback gab, so dass „CoCo laufen lernte“.

Auch allen anderen Co-Autoren danke ich für die stets produktive und unkomplizierte Zusammenarbeit.

Nun zu meinen geschätzten Kolleg: innen: ein großes Dankeschön an euch alle, denn ihr habt die Zeit am Lehrstuhl „Differentielle Psychologie und psychologische Diagnostik“ für mich zu etwas ganz Besonderem gemacht. Besonders zu nennen ist Christoph, der

mir die Zeit als fabelhafter Konditor wortwörtlich versüßt hat. Vielen Dank für viele schöne Gespräche, feine Kuchen und den ein oder anderen Crémant.

Bedanken möchte ich mich überdies bei meiner ganzen Familie. Ihr habt immer an mich geglaubt, auch wenn ich es selbst nicht mehr tat. Es ist unfassbar schön, dass ich mich immer auf euch verlassen kann, ihr mir stets den Rücken stärkt und ich über jedes Problem mit euch sprechen kann – ihr könnt euch freuen, das wiederkehrende Thema der Dissertation wäre nun nicht mehr auf der Agenda. Vielen Dank, dass ihr alle mich besonders in den letzten Monaten ertragen, abgelenkt und aufgebaut habt. Ein besonderer Dank geht an dieser Stelle an meine Mama für das Korrekturlesen der vorliegenden Arbeit und natürlich auch an meinen Papa, dessen Glaube an mich Berge versetzt. In all den Jahren habt ihr immer hinter mir gestanden und mich stets unterstützt. Ich bin dankbar, dass es euch gibt und wir als Familie immer zusammenhalten.

Zum Abschluss möchte ich mich noch bei einem besonderen Menschen bedanken: meinem Felix. Du hast versucht, jedes Tränchen zu trocknen, immer auf Leni aufzupassen und mich immer wieder motiviert, an dem Dissertationsvorhaben festzuhalten. Dir von Herzen mein persönlicher Dank, dass du mir den Freiraum gegeben hast mich weiterzuentwickeln, mich dabei immer unterstützt und an mich geglaubt hast. Ich bin froh, dass es dich in meinem Leben gibt.

Wie ich beim Schreiben dieser Worte gemerkt habe, geht eine sehr aufregende Zeit nun zu Ende. Ich bin glücklich, dass ich so viel Unterstützung von euch allen erhalten habe und ihr mich auf diesem Weg begleitet habt – DANKE ❤️

Anhang C. – Studien als Grundlage der kumulativen Promotion

- Studie I Levacher, J., Koch, M., Stegt, S., Hissbach, J., Spinath, F. M., Escher, M., & Becker, N. (2023). The construct validity of the main student selection tests for medical studies in Germany. *Frontiers in Education, Section Assessment, Testing and Applied Measurement*.
<https://doi.org/10.3389/feduc.2023.1120129>
- Studie II Levacher, J., Koch, M., Hissbach, J., Spinath, F. M., & Becker, N. (2021). You can play the game without knowing the rules – But you're better off knowing them: The influence of rule knowledge on Figural Matrices Tests. *European Journal of Psychological Assessment, 38*, 15-23. <https://doi.org/10.1027/1015-5759/a000637>
- Studie III Krämer R., Koch M., Levacher J., & Schmitz F. (2023). Testing Replicability and Generalizability of the Time on Task Effect. *Journal of Intelligence, 11(5)*, 82.
- Studie IV Levacher, J., Spinath, F. M., Becker, N., & Hahn, E. (2023). How did the beginnings of the global COVID-19 pandemic affect mental well-being? *PLOS ONE, 18(1)*. <https://doi.org/10.1371/journal.pone.0279753>