

Effekte kurzer Lernvideos für den Studieneignungstest Psychologie

Zwei experimentelle Studien

Oliver Dickhäuser, Rebecca Ax, Aker Özçevik, Catharina Schlie, Anna Fouqué, Nhat Lam Nguyen, Giulia Palmer, Isabell Schlöder, Marla Sjöberg und Mirjam Rommen

Universität Mannheim, Deutschland

Zusammenfassung: Basierend auf früheren Studien und theoretischen Überlegungen zu Testcoaching und -training untersuchten wir in zwei experimentellen Studien die Wirksamkeit regelbasierter Lernvideos zur Vorbereitung auf Aufgaben des schlussfolgernden Denkens im Studieneignungstest Psychologie (BaPsy). Insgesamt 265 Schüler_innen der gymnasialen Oberstufe betrachteten zufällig entweder ein relevantes oder ein irrelevantes Lernvideo. In Studie 1 (verbal-logisches Schlussfolgern) und Studie 2 (figural-logisches Schlussfolgern) zeigten sich nachfolgend jeweils substanziell höhere Leistungen in der Experimental- gegenüber der Kontrollgruppe. Die Effektstärken waren groß ($d = 0.99$) bis sehr groß ($d = 2.19$). Die Ergebnisse legen nahe, dass bereits das einmalige Ansehen eines regelbasierten Videos korrespondierende Aufgabenleistung fördern kann. Die Befunde ergänzen bestehende Evidenz zu Testtraining und -coaching. Wir diskutieren die Bedeutung und potenzielle Herausforderungen frei zugänglicher, qualitätsgesicherter Vorbereitungsmaterialien, u. a. im Hinblick auf eine Erhöhung von Chancengleichheit beim Hochschulzugang.

Schlüsselwörter: Intelligenz, Studieneignung, Testvorbereitung, Testcoaching, Testtraining, schlussfolgerndes Denken

Effects of Instructional Videos on Test Performance: Two Experiments

Abstract: Building on previous studies and theoretical considerations regarding test coaching and training, we examine the effectiveness of rule-based videos for tasks involving reasoning skills in the BaPsy (a psychology admission test at universities). In two experimental studies, we randomly assigned 265 upper secondary school students to either watch a relevant or an irrelevant instructional video. Study 1 targeted verbal reasoning, while Study 2 addressed figural reasoning tasks. In both studies, participants in the experimental groups subsequently showed substantially higher performance than controls. The effect sizes were large ($d = 0.99$) to very large ($d = 2.19$). The results suggest that even a single exposure to a rule-based instructional video can enhance performance in corresponding test tasks. These findings add to the existing evidence on test training and coaching. We discuss the implications and potential challenges of providing free, high-quality preparation materials to promote fairness and equal opportunities in university admissions.

Keywords: intelligence, academic aptitude, test preparation, test coaching, test training, reasoning skills

Theoretischer Teil

Der BaPsy-DGPs (Studieneignungstest für das Bachelorstudium Psychologie der Deutschen Gesellschaft für Psychologie, nachfolgend kurz BaPsy) ist ein standardisierter, bundesweit eingesetzter Auswahltest, der ergänzend zur Abiturnote die Eignung von Studieninteressierten für ein Psychologiestudium auf Bachelorniveau erfassen soll. Er wurde vor dem Hintergrund anhaltend hoher Zahlen an Bewerber_innen bei begrenzten Studienplatzkapazitäten entwickelt, um die Vergabe von Studienplätzen im Rahmen der hochschuleigenen Auswahlverfahren gerechter,

rechtssicherer und stärker eignungsbasiert zu gestalten (Menz et al., 2021). Ausgangspunkt war unter anderem das sogenannte „NC-Urteil“ des Bundesverfassungsgerichts von 2017, das für den Fall des Studiengangs Medizin eine stärkere Ausrichtung an weiteren Eignungskriterien bei der Auswahl von Studienbewerber_innen forderte und die alleinige Orientierung an der Abiturnote als unzureichend bewertete.

Der BaPsy ist aus entsprechenden Vorgängerverfahren aus Baden-Württemberg (Watrin et al., 2022) und Berlin (Horstmann et al., 2023) hervorgegangen (vgl. Dickhäuser et al., 2022). Auf Initiative der Deutschen Gesellschaft

für Psychologie und des Psychologischen Fakultätentages ist das Verfahren nunmehr bundesweit etabliert: In allen deutschen Bundesländern an aktuell 40 Standorten (Stand 2025) setzen die Universitäten bei ihren hochschuleigenen Auswahlverfahren für Bachelorstudiengänge im Fach Psychologie auf den BaPsy als ein Auswahlkriterium zusätzlich zur Note der Hochschulzugangsberechtigung (HZB) und ggf. weiteren Kriterien.

Da die nahezu bundesweite Etablierung des BaPsy als Zulassungskriterium das Absolvieren des freiwilligen Tests faktisch zu einer Standardhürde bei der Bewerbung um einen Bachelorstudienplatz im Fach Psychologie werden lässt, ist die Nachfrage nach Vorbereitungsmaterialien groß. In der vorliegenden Arbeit wird in diesem Zusammenhang die Wirksamkeit zweier regelbasierter Lernvideos für Subtests des BaPsy untersucht, die frei zugänglich zu Vorbereitungszwecken bereitgestellt werden. In zwei Experimenten prüfen wir, inwieweit diese frei verfügbaren Angebote für den BaPsy ihren Zweck erfüllen und zu höheren Leistungen in relevanten Aufgaben beitragen. Zur Einordnung der Forschungsfrage stellen wir zunächst die Zielsetzung der BaPsy-Einführung sowie erste empirische Befunde vor. Anschließend gehen wir auf die besondere Situation der verfügbaren Vorbereitungsmaterialien für den BaPsy ein und leiten daraus unsere Fragestellung ab.

Ein Studieneignungstest wie der BaPsy soll ein zusätzliches, standardisiertes Auswahlkriterium bieten, um die Vorhersage von Studienerfolg zu verbessern und leistungsfähige Bewerber_innen auch über die Note der Hochschulzugangsberechtigung hinaus zu identifizieren. Gleichzeitig zielt er – zusammen mit einem idealerweise vorgelagert bearbeiteten Online-Self-Assessment als Instrument der Selbsterkundung (etwa Merkle et al., 2022) – mittelfristig auf eine Erhöhung der Passung zwischen Person und Studienfach ab. Auf dieser Grundlage soll es zu geringeren Raten von Studienabbrüchen und zu besseren Leistungen kommen. Darüber hinaus soll der Test die Vergleichbarkeit zwischen Bewerber_innen aus unterschiedlichen Bundesländern und Schulen erhöhen, da er losgelöst von schulischen Lehrplänen objektifizierbare kognitive und fachbezogene Kompetenzen prüft.

Der BaPsy besteht aus verschiedenen Testteilen. Diese umfassen Aufgaben zum schlussfolgernden Denken (verbal, numerisch, figural), zum Textverständnis psychologischer Fachtexte in deutscher und englischer Sprache sowie zu grundlegenden Kenntnissen in Mathematik. Die Subtests des BaPsy lassen sich im Rahmen des CHC-Modells (Cattell-Horn-Carroll-Theorie der Intelligenz) kognitiven Fähigkeitsdomänen zuordnen (vgl. Horstmann et al., 2023, für eine solche Zuordnung anhand des Berliner Studierfähigkeitstests Psychologie). Die Aufgaben zum schlussfolgernden Denken erfassen vor allem fluide

Denkfähigkeit (Gf), also die Fähigkeit, neue Probleme durch logisches Schlussfolgern und Regelableitung zu lösen. Die mathematischen Aufgaben lassen sich der quantitativen Fähigkeit (Gq) zuordnen. Die Testteile zum Textverständnis psychologischer Fachtexte beziehen sich dagegen primär auf Lesen und Schreiben (Grw) sowie auf fachspezifisches Wissen (Gkn), da hier sowohl sprachgebundene Verstehensprozesse als auch die Anwendung domänenspezifischer Kenntnisse erforderlich sind.

Über Studienfächer hinweg belegen bisherige Meta-Analysen sowohl international (Kuncel et al., 2001) als auch speziell für den deutschsprachigen Raum (Hell et al., 2007) beträchtliche prognostische Validität von Studieneignungstests für spätere Studienleistungen. Dies zeigte sich auch in einem meta-analytischen Update für den DACH-Raum, das Schult et al. (2019) publizierten: Auch unter Berücksichtigung neuerer Studien zeigte sich die prognostische Validität von Studieneignungstests für späteren Studienerfolg ($\rho = .43$). Die Testergebnisse erwiesen sich als vorhersagestark für Studienleistungen – auch über die Note der Hochschulzugangsberechtigung hinaus.

Speziell für Studieneignungstests für Psychologie in Deutschland liefern erste Validierungsstudien zu den Vorläuferverfahren des BaPsy Hinweise darauf, dass die Veränderung der Zulassungskriterien beträchtliche Veränderungen in den Gruppen der Zugelassenen und Nichtzugelassenen nach sich ziehen (Horstmann et al., 2024; Marquardt et al., 2024). Zwar korrelierten die Testscores mit den HZB-Noten als Indikatoren früherer Leistungsfähigkeit (.42 bei Formazin et al., 2011; .21 bei Watrin et al., 2022; und .26 bei Horstmann et al., 2024), zugleich zeigten die nur mäßigen Überlappungen zwischen HZB-Note und Testleistung, dass Abiturnote und Test beträchtliche Varianzanteile nicht teilen und damit potenziell unabhängig voneinander prognostisch valide für den Studienerfolg sein könnten. In der Arbeit von Formazin et al. (2011) sagten die Testscores bei den später zugelassenen Studierenden auch die Leistung im Psychologiestudium voraus. Der durch den Testscore (über die HZB-Note hinaus) zusätzlich vorhergesagte Varianzanteil der Studienleistung betrug dabei 6 %. Auch in einer jüngeren Studie von Watrin et al. (2022) auf Basis von 371 Psychologiestudierenden zeigten sich analoge Beziehungen zwischen Testscore und HZB-Note, späterer Studienleistung sowie ein inkrementeller Beitrag des Tests in Höhe von 8.7 % (Watrin et al., 2022, Appendix) bei der Vorhersage der Studienleistung über die HZB-Note hinaus.

Ein zentrales Element für die Akzeptanz des BaPsy – sowohl im Vorfeld seiner Implementierung im Rahmen hochschuleigener Auswahlverfahren als auch bei den Testteilnehmenden – dürfte die transparente Bereitstellung von erklärenden Informationen über den Test und

das Auswahlverfahren sein. Das legt auch eine Meta-Analyse von Truxillo et al. (2009) nahe, die zeigt, dass das Geben von Erklärungen über einen Auswahltest mit höherer wahrgenommener Fairness des Tests einhergeht.

Entsprechend hat sich die DGPs entschieden, die Etablierung des BaPsy mit der Bereitstellung von niederschwellig verfügbaren, kostenlosen Vorbereitungsmaterialien zu flankieren. Diese umfassen neben transparenten Informationen zur späteren Instruktion und Beispielaufgaben zu den Untertests auch Vorbereitungsvideos, in denen für die verschiedenen Subtests späteren Testand_innen Strategien für die Bearbeitung der Aufgaben an die Hand gegeben und anhand von Beispielaufgaben erläutert werden.

Analog zu anderen Studieneignungstests (z. B. in der Medizin oder Pharmazie) hat sich seit der Einführung des BaPsy ein umfangreicher Markt an kommerziellen Vorbereitungsangeboten etabliert. Diese Angebote umfassen den Zugang zu Übungsaufgaben, Online-Tutorials, Testsimulationen, Live-Kursen, Tutorien oder Büchern und Skripten. Die Preisgestaltung erfolgt in der Regel modular – je nach gewünschtem Leistungsumfang – und variiert aktuell in der Praxis zwischen rund 28 € (z. B. MedGurus, n.d.) und bis zu 359 € (z. B. IFS Studentenkurse, n.d.). Angesichts der Verfügbarkeit dieser kommerziellen Vorbereitungsangebote in großer Zahl und von beträchtlichen Kosten ist die Bereitstellung von Vorbereitungsmaterialien auch eine denkbare Strategie, um einem möglichen ökonomischen Bias entgegenzuwirken, der entstehen könnte, wenn zur Vorbereitung ausschließlich kostenpflichtige Vorbereitungsangebote existieren würden.

Testcoaching bezeichnet systematische Maßnahmen, die darauf abzielen, die Testleistung durch das Erlernen und Einüben teststrategischer Verhaltensweisen zu verbessern – etwa durch Hinweise zur Zeiteinteilung, Strategien zum Ausschluss falscher Antwortalternativen oder durch die Analyse typischer Aufgabenformate (vgl. Jensen, 1980). Solche Maßnahmen zielen primär auf eine Steigerung der sogenannten *testwiseness* und damit auf eine effektivere Nutzung vorhandener Fähigkeiten, ohne dass ein substanzieller Fähigkeitszuwachs stattfindet.

Davon abzugrenzen sind Trainings, die auf den Erwerb und die Anwendung von Problemlösestrategien zielen. Hier steht nicht das strategische Testverhalten im Vordergrund, sondern das Verständnis der den Aufgaben zugrunde liegenden Strukturen und Regeln. Typischerweise sehen kognitive Trainings (vgl. Klauer, 2001) neben der Vermittlung von kognitiven Lösungsparadigmen auch wiederholtes Anwenden der vermittelten Regeln im Rahmen von Übungseinheiten vor. Allerdings können zumindest die zugrunde liegenden Prinzipien deduktiv vermittelt und anhand typischer Beispiele veranschaulicht werden, ohne dass dies wiederholtes Üben notwendig macht.

Dass derartige regelbasierte Lernprozesse auch in video-basierten Formaten effektiv angeregt werden können, lassen mehrere experimentelle Studien vermuten (Schneider et al., 2020; Schneider & Sparfeldt, 2021a, 2021b). Schneider et al. (2020) konnten in zwei Experimenten mit Psychologie- bzw. Lehramtsstudierenden nachweisen, dass bereits das einmalige Ansehen eines etwa 14-minütigen Videos, in dem typische Regeln figuraler Matrizen (z. B. Addition, Rotation) systematisch erklärt werden, zu deutlich höheren Leistungen in nachfolgenden Matrizenaufgaben führte (Effektstärken $d \geq 1.19$) als bei Personen aus einer Kontrollgruppe. Auch Schneider und Sparfeldt (2021b) fanden in einem experimentellen Design mit drei Gruppen (relevantes Video, Itembeispiele, themenfremdes Video) signifikante höhere Leistungen nach einem kurzen Video zur Bearbeitung von Zahlenreihen ($d = 0.30 - 0.44$). In beiden Fällen basierten die Videos auf deduktiv entwickelten Regelmodellen, die die Bearbeitungsschritte systematisch erläutern und typische Itemstrukturen veranschaulichen.

Die Wirkung solcher Videos geht vermutlich auf den gezielten Aufbau kognitiver Schemata zurück: Testpersonen erhalten ein mentales Regelmodell, das sie bei der Bearbeitung ähnlicher Aufgaben gezielt anwenden können. Dabei kommt es – im Gegensatz zu reiner Wiederholung – nicht primär zu einem Testwiseness-Effekt (z. B. durch kluges Raten oder dem geschickten Ausschluss unplausibler Distraktoren), sondern zu einer Leistungssteigerung durch bessere Regelidentifikation und effizientere Lösestrategien. Hinweise darauf liefern die Befunde der Studie von Schneider und Sparfeldt (2021a): In dieser Studie fertigten die Versuchspersonen, die ein regelvermittelndes Video gesehen hatten, bei der anschließenden Bearbeitung von Zahlenreihenaufgaben signifikant häufiger Notizen an als die Kontrollgruppe. Die Anzahl der Items mit Notizen medierte dabei den Effekt des Videos auf die Testleistung. Diese Veränderung im Testverhalten lässt sich als ein Indikator für eine aktive Anwendung der im Video präsentierten Lösungsstrategien interpretieren.

Bemerkenswert ist, dass diese Leistungssteigerungen bereits durch einmaliges Ansehen der Videos erzielt wurden, ohne die für Trainings sonst charakteristischen Übungsphasen. In beiden Studien zeigten sich zudem substanzielle Korrelationen der Leistungsgüte in den für das Video fokalen Aufgaben mit externen Kriterien – insbesondere mit einem separaten Intelligenztest (z. B. $r \approx .56-.63$ in Schneider & Sparfeldt, 2021b; $r \approx .40-.55$ in Schneider et al., 2020) –, die in Experimental- und Kontrollgruppen (ohne bzw. mit irrelevanten Videos) vergleichbar ausfielen. In der Summe deutet dies darauf hin, dass die regelbasierten Videos zu beträchtlich höheren Testwerten führen, die Rangordnung der Testleistungen durch das Video weitgehend erhalten bleibt und die Kon-

struktvalidität nicht beeinträchtigt wird. Zusammengekommen legen diese Befunde auch für den BaPsy nahe, dass sich zentrale Aufgabenformate durch kurze, digital verbreitete regelbasierte Videos effektiv und validitätskonform vorbereiten lassen.

Die beiden vorliegenden Experimente verfolgen das Ziel, die Wirksamkeit zweier Videos zur Vorbereitung auf Aufgaben des schlussfolgernden Denkens im Rahmen des BaPsy zu überprüfen. In beiden Studien wurde die Hypothese geprüft, dass Versuchspersonen, die ein aufgabenrelevantes, regelbasiertes Vorbereitungsvideo gesehen haben, in den nachfolgenden Testaufgaben höhere Leistungen erzielen als Personen, die ein aufgabenirrelevantes Kontrollvideo gesehen haben. Dabei wurde in Studie 1 der Effekt für verbales Schlussfolgern, in Studie 2 für figural-schlussfolgerndes Denken untersucht.

Der praktische Nutzen dieser Untersuchungen liegt darin, empirisch zu überprüfen, ob die im Rahmen des BaPsy bereitgestellten Vorbereitungsmaterialien tatsächlich die Bearbeitungsleistung beeinflussen können – eine Voraussetzung, um deren faire und transparente Nutzung im diagnostischen Kontext angemessen zu bewerten.

Unter wissenschaftlicher Perspektive sollen die Studien darüber hinaus zwei zentrale Beiträge leisten: Erstens replizieren sie frühere Befunde zur Wirksamkeit videobasierter Instruktionen auf kognitive Testleistungen (Schneider et al., 2020; Schneider & Sparfeldt, 2021a, 2021b) und erweitern diese auf eine neue Population (Schüler_innen). Zweitens prüfen sie, ob sich die beobachteten Effekte auch bei unterschiedlichen Aufgabenformaten des schlussfolgernden Denkens zeigen. Damit wird die Generalisierbarkeit der Befunde über verschiedene Facetten des schlussfolgernden Denkens hinweg überprüft.

Zwei Experimente

Ziel der zwei Experimente war es, den Effekt des Instruktionsvideos zum verbalen schlussfolgernden Denken (Studie 1) bzw. zum figuralen schlussfolgernden Denken (Studie 2) der BaPsy-Vorbereitungsmaterialien auf entsprechende Aufgabenleistung zu prüfen. Vermutet wurde, dass das Ansehen eines relevanten Instruktionsvideos zu mehr korrekt gelösten Aufgaben führt als ein jeweils thematisch irrelevantes Video.

Methode

Versuchspersonen

Versuchspersonen waren 161 (Studie 1) bzw. 106 (Studie 2) Schüler_innen der gymnasialen Oberstufe. Die Akquise erfolgte durch persönliche Kontaktaufnahme eines Teils der Koautor_innen mit Schulen in Baden-Württemberg, Bayern, Nordrhein-Westfalen, Rheinland-Pfalz und Hessen. Die Schulen (vier Schulen in Studie 1, drei Schulen in Studie 2) stimmten nach Informationen über Ziel und Ablauf zu; die Teilnehmenden gaben schriftlich ihr informiertes Einverständnis (in Studie 2 gaben zwei Personen kein Einverständnis). Die Testungen fanden während der Schulzeit statt. Die Versuchspersonen waren zwischen 16 und 19 Jahren alt, die meisten waren 17 Jahre alt. Eine Vergütung erfolgte nicht.

Überblick

In beiden Studien wurden die Teilnehmenden zufällig einer von zwei Bedingungen zugewiesen. Die Experimentalgruppe (Studie 1: $n = 81$, Studie 2: $n = 54$) sah ein relevantes Instruktionsvideo zum Schlussfolgernden Denken (Studie 1: verbal, Studie 2: figural). Die Kontrollgruppe (Studie 1: $n = 80$, Studie 2: $n = 50$) sah ein in Aufbau und Länge vergleichbares, für die nachfolgenden Aufgaben jedoch thematisch irrelevantes Video. Anschließend bearbeiteten alle Versuchspersonen Aufgaben zum Schlussfolgernden Denken (Studie 1: verbal, Studie 2: figural). Beiden Studien lag also ein einfaktorielles Between-Subjects-Design zugrunde, in dem die Art der regelbasierten Videoinstruktion (relevant vs. irrelevant) variiert wurde. Die abhängige Variable war die Anzahl korrekt gelöster Aufgaben im schlussfolgernden Denken (Studie 1: verbal, Studie 2: figural).

Material und Durchführung

Die Teilnehmenden wurden in beiden Studien zufällig zwei Gruppen zugeteilt und begaben sich in separate Räume. Vor der Präsentation der Instruktionsvideos erhielten die Teilnehmenden eine kurze Einführung durch die Versuchsleitung sowie die schriftliche Einwilligungserklärung. Die Videos begannen mit allgemeinen Hinweisen.

In Studie 1 wurden im Video der Experimentalgruppe (15:39 Minuten) typische Operatoren des verbalen Schlussfolgerns wie „und“, „oder“ und „genau dann, wenn“ erklärt. Es wurde dargestellt, wie diese Operatoren in standardisierten Schlussfolgerungsaufgaben logisch interpretiert und angewendet werden. Das Video vermittelte eine strukturierte Strategie zur Bearbeitung der Aufgaben und enthielt eine Beispielaufgabe mit anschließender Lösung. Das Kontrollvideo (15:07 Minuten) behandelte

Tabelle 1. Mittelwert (M) und Standardabweichung (SD) für die schulische Leistung sowie Testleistung zum schlussfolgernden Denken getrennt nach Art des betrachteten Videos in Studie 1 (verbales Schlussfolgern) und Studie 2 (figurales Schlussfolgern)

	Studie 1 (verbales Schlussfolgern)		Studie 2 (figurales Schlussfolgern)	
	EG: Relevantes Video (n = 81)	KG: Irrelevantes Video (n = 80)	EG: Relevantes Video (n = 54)	KG: Irrelevantes Video (n = 50)
Schulische Leistung (in Punkten)	M = 9.80 SD = 2.23	M = 9.38 SD = 2.10	M = 9.70 SD = 1.96	M = 10.20 SD = 1.87
Prüfstatistik	t (159) = 1.21, p = .23		t (102) = 1.44, p = .15	
Testscore schluss- folgerndes Denken	M = 9.67 SD = 3.15	M = 6.70 SD = 2.87	M = 6.09 SD = 2.44	M = 1.60 SD = 1.51
Prüfstatistik	t (159) = 6.24, p < .001, d = 0.99		t (102) = 11.17, p < .001, d = 2.19	

mit figural-logischem Denken eine für die nachfolgenden Testaufgaben irrelevante Aufgabenart.

In Studie 2 thematisierte das Video der Experimentalgruppe (20:47 Minuten) typische Regeln zur Bearbeitung figuraler Matrizen, darunter z. B. Addition, Schnittmenge und Drehung. Die Regeln wurden im Sinne eines deduktiv entwickelten Regelmodells systematisch eingeführt: Jede Regel wurde erklärt, anhand einer Beispielaufgabe demonstriert und anschließend durch eine schrittweise Lösung verdeutlicht. Ziel war es, den Teilnehmenden ein mentales Modell zur strukturierten Bearbeitung figuraler Matrizen zu vermitteln. Das Kontrollvideo (20:46 Minuten) war thematisch irrelevant und behandelte Strategien zur Bewältigung von Prüfungsangst. Beide Videos waren in Länge und Präsentationsweise vergleichbar.

In beiden Studien handelte es sich bei den Experimentalgruppenvideos um eine minimal nachbearbeitete Version der im Jahr 2024 auf der offiziellen Website des BaPsy verfügbaren regelbasierten Instruktionsvideos. Die verwendeten Materialien (Videos und Aufgaben) sind auf Anfrage beim Erstautor erhältlich.

Nach der Betrachtung des Videos erhielten die Versuchspersonen in Studie 1 17 Aufgaben zum verbalen Schlussfolgern und 17 Minuten Zeit. In Studie 2 erhielten sie 20 Aufgaben zum figuralen Schlussfolgern und 18 Minuten Zeit, wobei eine der 20 Aufgaben exakt so im Video erläutert wurde, sodass die Lösung dieses Kontrollitems als Aufmerksamkeitscheck betrachtet werden kann. Bei den Aufgaben handelt es sich um offizielle Übungsaufgaben, die über die BaPsy-Website zur Verfügung gestellt werden. Im Anschluss wurden in beiden Studien demografische Daten (Alter, Geschlecht), die Punktzahl im letzten Abschlusszeugnis in Deutsch, Mathematik, Eng-

lisch sowie ggf. einer weiteren Fremdsprache sowie Angaben zur Vertrautheit mit dem BaPsy-Test und dem Instruktionsvideo erfasst.¹

Ergebnisse

Die mittlere schulische Leistung (durchschnittliche Punktzahl in den Fächern Deutsch, Mathematik, Englisch und ggf. einer weiteren Fremdsprache) für Personen der Experimentalgruppe wie der Kontrollgruppe ist in Tabelle 1 für beide Studien dargestellt. Die mittlere Leistung unterschied sich weder in Studie 1 noch in Studie 2 statistisch signifikant zwischen den Personen in der Experimentalgruppe und denen in der Kontrollgruppe. Dies ist ein Hinweis auf eine weitgehend gelungene Randomisierung mit Blick auf die allgemeine Leistungsfähigkeit der Teilnehmer_innen.

In Studie 2 lösten 47 Personen der Experimentalgruppe das Kontrollitem korrekt und 7 Personen falsch, in der Kontrollgruppe lösten lediglich 11 Personen das Kontrollitem richtig, 39 dagegen falsch. Diese Verteilungen unterschieden sich statistisch signifikant, $\chi^2(1) = 41.92$, $p < .001$.

Die mittlere Anzahl korrekt gelöster Aufgaben zum Schlussfolgernden Denken fiel in beiden Studien bei den Personen in der Experimentalgruppe statistisch signifikant höher aus als bei den Personen in der Kontrollgruppe. Tabelle 1 zeigt, dass nach dem Betrachten des jeweils relevanten Videos fast drei (Studie 1) bzw. über vier (Studie 2) Aufgaben zum schlussfolgernden Denken mehr korrekt gelöst wurden als nach Betrachten des jeweils irrelevanten Videos. Mit einem d von 0.99 in Studie 1 und

¹ In Studie 2 wurden als eine Nebenfragestellung noch Need for Cognition, Growth Mindset und Lernzielorientierung als mögliche Moderatoren eines Effektes erfasst. Diese Fragestellung ist im Kontext dieses Manuskriptes nicht von Bedeutung, die möglichen Moderatoren erwiesen sich auch nicht als statistisch bedeutsam.

einem d von 2.19 in Studie 2 handelte es sich jeweils um einen großen Effekt.²

Diskussion

Die beiden vorliegenden Experimente untersuchten die Wirksamkeit regelbasierter Instruktionsvideos zur Vorbereitung auf Aufgaben des schlussfolgernden Denkens im BaPsy. In beiden Studien zeigte sich ein substantieller Vorteil für diejenigen Versuchspersonen, die zuvor ein aufgabenrelevantes Instruktionsvideo gesehen hatten. Die Effektstärken waren in beiden Fällen groß bis sehr groß ($d = 0.99$ für verbales, $d = 2.19$ für figurales Schlussfolgern). Besonders das in Studie 2 eingesetzte Kontrollitem belegt, dass Inhalte des Videos vom großen Teil der Versuchspersonen verstanden und korrekt reproduziert wurden. Zusammengefasst stützen diese Befunde die Hypothese, dass relevante Videoinstruktionen die Leistung in kognitiven Testaufgaben des BaPsy unmittelbar verbessern können. Möglicherweise handelt es sich bei den verwendeten Matrizen-Aufgaben zum figuralen Schlussfolgern um Aufgaben, bei denen die vermittelten Regeln zu einem besonders großen Leistungszuwachs führen.

Die Ergebnisse reihen sich ein in eine wachsende Zahl empirischer Studien, die den Nutzen regelbasierter Instruktionen für Testleistungen belegen. Insbesondere schließen sie an die Studien von Schneider et al. (2020) sowie Schneider und Sparfeldt (2021a, 2021b) an, die zeigen konnten, dass kurze Videos auf der Basis deduktiv entwickelter Regelmodelle substanzielle Leistungssteigerungen in Aufgaben des schlussfolgernden Denkens bewirken. Die hier vorgelegten Studien erweitern diesen Forschungsstand in mehreren Punkten: Erstens wird die Wirksamkeit der regelbasierten Instruktionsvideos nicht nur bei Universitätsstudierenden, sondern auch bei Schüler_innen der gymnasialen Oberstufe nachgewiesen – also genau jener Zielgruppe, die typischerweise am BaPsy teilnimmt. Zweitens konnte gezeigt werden, dass regelbasierte Instruktionen nicht nur bei figuralen, sondern auch bei verbalen Aufgabenformaten des schlussfolgernden Denkens leistungsförderlich sind.

Die Resultate deuten wir vor dem Hintergrund theoretischer Überlegungen so, dass durch die Instruktionsvideos kognitive Schemata aufgebaut werden, die ein sys-

tematisches Vorgehen bei der Bearbeitung der Aufgaben ermöglichen. Anders als bei rein strategischen Coachings, bei denen lediglich Antwortverhalten oder Testformate vermittelt werden, scheint hier ein inhaltlich fundierter Lernerfolg vorzuliegen, der auf der Identifikation von Regeln und dem Erwerb effizienter Problemlösestrategien beruht. Allerdings sind diese Überlegungen hier spekulativ, da die Frage nach den vermittelnden kognitiven Prozessen nicht im Fokus der vorliegenden Studien stand und Prozessdaten nicht erfasst wurden.

Aus den vorliegenden Befunden ergeben sich mehrere Implikationen für die Gestaltung von Vorbereitungsmaterialien für den BaPsy. Zunächst bestätigen die Ergebnisse die Entscheidung der Deutschen Gesellschaft für Psychologie (DGPs), kostenlose videobasierte Vorbereitungsmaterialien bereitzustellen. Solche Angebote sind nicht nur niedrigschwellig und zeitlich flexibel nutzbar, sondern offenbar auch geeignet, substanzielle Leistungsgewinne zu erzielen. Angesichts der Vielzahl kommerzieller Angebote stellt die Bereitstellung frei verfügbarer, qualitätsgesicherter Materialien eine wichtige Maßnahme zur Erhöhung von Fairness und Chancengleichheit dar – insbesondere für Bewerber_innen ohne Zugang zu kostenpflichtigen Vorbereitungskursen.

Darüber hinaus lassen sich Empfehlungen für die Weiterentwicklung von Studieneignungstests und deren Vorbereitung ableiten: Vorbereitungsmaterialien sollten weiterhin auf deduktiv entwickelten Regelmodellen basieren, die den Aufbau mentaler Modelle fördern, anstatt lediglich Beispielaufgaben bereitzustellen. Die explizite Vermittlung von Bearbeitungsstrategien scheint dabei zentral zu sein. Auch sollte in der Kommunikation mit Testteilnehmenden betont werden, dass ein vertieftes Verständnis der zugrunde liegenden Aufgabentypen eine effektive Vorbereitung auf Teilbereiche des Tests ermöglicht.

Die Studien weisen mehrere methodische Stärken auf. Durch die randomisierte Zuweisung zu den Bedingungen – und die dokumentierte Vergleichbarkeit der Gruppen hinsichtlich ihrer schulischen Leistungen – konnte ein hohes Maß an interner Validität erreicht werden. Der in Studie 2 eingesetzte Manipulationscheck durch ein Kontrollitem erhärtet die Deutung, dass die Leistungssteigerung auf der Verarbeitung der Videoinhalte beruht. Auch die Verwendung von Schüler_innen der gymnasialen Oberstufe als Stichproben erhöht die externe Validität der Ergebnisse mit Blick auf die Zielgruppe des BaPsy.

² Wurden bei den Analysen diejenigen Personen ausgeschlossen, die angaben, den BaPsy oder die Testaufgaben schon zu kennen ($n = 12$ in Studie 1, $n = 21$ in Studie 2), veränderten sich die Befunde hinsichtlich der zentralen abhängigen Variable nicht wesentlich (Studie 1: $t(147) = 6.08$, $p < .001$, $d = 1.00$, $M_{EG} = 9.71$, $SD = 3.14$; $M_{KG} = 6.69$, $SD = 2.91$ / Studie 2: $t(81) = 8.81$, $p < .001$, $d = 1.94$, $M_{EG} = 5.70$, $SD = 2.41$; $M_{KG} = 1.73$, $SD = 1.58$)

Dennoch sind einige Einschränkungen zu berücksichtigen. Erstens wurden nur kurzfristige Effekte erfasst – Aussagen zur Stabilität über die Zeit sind daher nicht möglich. Gleichzeitig können die Befunde natürlich auch nicht ohne weiteres generalisiert werden, wenn sich Teile des Tests verändern. So wurde der Test zum figuralen Schlussfolgern des BaPsy nach der Durchführung von Studie 2 zum Jahr 2025 verändert (andere Regeln, generischere Instruktion, andere Figurelemente). Inwieweit die Befunde zur Validität des Vorbereitungsvideos und des Tests selbst Bestand haben, wenn am Test Veränderungen vorgenommen werden, bleibt offen.

Zweitens ist unklar, ob die außergewöhnlich hohe Effektivitätsstärke in Studie 2 zumindest teilweise darauf zurückzuführen ist, dass das Video nicht nur zu einer effizienteren Bearbeitung, sondern überhaupt erst zu einem grundlegenden Verständnis der Aufgabeninstruktion führte. Insbesondere die sehr geringe Lösungsrate in der Kontrollgruppe von Studie 2 könnte dadurch erklärt werden, dass die im Vorfeld der Aufgabenbearbeitung gegebene Instruktion bei einem Teil der Proband_innen möglicherweise nicht ausreichte, um ein gutes Aufgabenverständnis zu erreichen. Dies wurde unter Umständen in der Experimentalgruppe durch das Video kompensiert. Bei realen BaPsy-Testteilnehmenden, die sich in der Regel bereits mit den Aufgabenformaten vertraut gemacht haben, dürften dann etwaige Effekte – u.U. besonders beim neuen Testformat – geringer ausfallen.

Drittens ist die vorliegende Stichprobe – trotz ihres Bezugs zur Zielgruppe – vermutlich nicht repräsentativ für die Gesamtgruppe späterer BaPsy-Testteilnehmender, da wahrscheinlich nur ein kleiner Teil der untersuchten Personen den Test tatsächlich absolvieren würde. Die hier vorgestellten Befunde basieren daher auf einer vermutlich hinsichtlich Motivation und Leistungsstärke deutlich heterogeneren Gruppe als jener Personen, die sich später unter Realbedingungen auf den Test vorbereiten. In letzterer Gruppe könnten die Effekte möglicherweise weniger stark ausfallen. Allerdings wird man in dieser Gruppe kaum Personen finden, die das Vorbereitungsmaterial nicht ohnehin bereits kennen und damit arbeiten, was den Nachweis seiner Wirksamkeit in einem experimentellen Design wie dem vorliegenden unmöglich macht.

Zur Überprüfung der Wirksamkeit der Videos wurden in den vorliegenden Experimenten die Übungsaufgaben der jeweiligen Subtests eingesetzt, nicht jedoch die tatsächlichen Testaufgaben des BaPsy. Letztere unterliegen einem strengen Schutz und standen hier für Forschungszwecke nicht zur Verfügung. Es ist daher offen, ob die gefundenen Effekte bei den eigentlichen Testaufgaben in gleicher Stärke auftreten würden. Allerdings wurden die Übungsaufgaben auf Grundlage beispielhafter originaler Testaufgaben entwickelt und hinsichtlich Struktur, An-

forderungsniveau (soweit beurteilbar) und Lösungsformat konzeptionell eng an die echten Aufgaben angelehnt. Entsprechend ist zumindest von einer inhaltlichen Konvergenz zwischen Übungs- und Testaufgaben auszugehen, auch wenn diese bislang nicht empirisch quantifiziert wurde.

Weitere Forschung ist erforderlich, um die hier gefundenen Effekte differenzierter zu verstehen. So sollte untersucht werden, wie stabil die Effekte über längere Zeiträume hinweg sind – etwa bei mehrwöchigem Abstand zwischen Regelvermittlung und Test (so wie dies unter realen Vorbereitungsbedingungen der Fall sein dürfte). Unklar ist zudem, ob sich vergleichbare Effekte auch bei Aufgabenformaten zeigen, die eher quantitative Fähigkeiten oder kristallisiertes Wissen erfassen. Hier ist zu vermuten, dass kurze regelbasierte Instruktionen deutlich geringere Effekte erzielen können.

In der vorliegenden Studie wurde kein Training im engeren Sinne eines wiederholten Übens oder Retesting-Designs durchgeführt. Vielmehr handelte es sich um eine einmalige, regelbasierte Instruktion in Form eines Videos, das Teilnehmenden die zugrunde liegenden Lösungsprinzipien der Aufgaben erklärte. Im weiteren Sinn kann dies als regelbasierte kognitive Instruktion verstanden werden, die zum Aufbau kognitiver Schemata und der Förderung systematischer Problemlösestrategien dient (vgl. Schneider & Sparfeldt, 2021a, 2021b), nicht lediglich auf teststrategisches Verhalten etwa im Sinne von intelligentem Ausschließen unplausibler Antwortalternativen abzielt (vgl. Hausknecht et al., 2007).

Gleichwohl ist zu betonen, dass in unserem Design keine wiederholte Aufgabenbearbeitung erfolgte und daher nicht von Übungseffekten, sondern von Instruktionseffekten gesprochen werden muss. Wir halten es jedoch für plausibel, dass es die Vermittlung regelbasierter Bearbeitungsstrategien ist, deren Anwendung nachfolgend die höheren Aufgabenleistungen bedingt.

Ein offener Punkt betrifft in diesem Zusammenhang die Frage, wie sich frei zugängliche Vorbereitungsmaterialien langfristig auf die prädiktive Validität des BaPsy im Hinblick auf den Studienerfolg auswirken. Zwar ist davon auszugehen, dass Personen, die sich mithilfe der Videos vorbereiten, bessere Testergebnisse erzielen, und dass die Möglichkeit zur gleichmäßigen Nutzung solcher Materialien die wahrgenommene Fairness des Auswahlverfahrens erhöht. Empirische Befunde zu anderen Videos und anderen Testaufgaben sprechen jedoch dagegen, dass sich die Validität dadurch verringert. In der Studie von Schneider und Sparfeldt (2021b) blieben die Korrelationen zwischen Testleistungen und Intelligenz in Experimental- und Kontrollgruppen nahezu identisch, obwohl die Experimentalgruppen signifikant höhere Testwerte erzielten. Dies legt nahe, dass regelbasierte Videoinstruk-

tionen zwar das Leistungsniveau anheben, die Rangordnung der Teilnehmenden jedoch weitgehend erhalten bleibt. Dass dies auch ganz spezifisch auf den BaPsy und die frei zur Verfügung stehenden Vorbereitungsmaterialien zutrifft, mag man angesichts dieser Befunde plausibel finden, wäre jedoch zugleich eine wertvolle Fragestellung für zukünftige Studien.

Nicht zuletzt ist zu diskutieren, wie sich die zunehmende Etablierung von Vorbereitungsmaterialien auf die zukünftige Testentwicklung auswirkt. Wenn große Teile der Testpopulation gezielt auf bestimmte Subtests vorbereitet sind, stellt sich die Frage, ob die Differenzierungsleistung in höheren Leistungsbereichen aufrechterhalten werden kann und inwieweit die Entwicklung neuer Aufgaben dadurch anspruchsvoller wird. Hierbei könnte eine gewisse Spannung entstehen zwischen den Zielen des Tests (Differenzierung im oberen Leistungsbereich) und den Zielen der Vorbereitungsangebote (Transparenz und Zugänglichkeit).

Abschließend erscheint es – auch angesichts der mittlerweile bundesweiten Verbreitung des BaPsy und der dazugehörigen Vorbereitungsmaterialien – wünschenswert, weitere Studien zur Validität dieses Tests durchzuführen. Besonders relevant ist dabei die Frage, inwieweit das neue Auswahlverfahren tatsächlich zu einem faireren Zugang zum Psychologiestudium beiträgt und bislang verborgene Talente identifizieren kann. Ein Vergleich der meta-analytischen Befunde von Hell et al. (2007) und Schult et al. (2019) zeigt, dass die prognostische Validität von Eignungstests für die Vorhersage von Studienleistung insgesamt über die Jahre hinweg relativ stabil geblieben ist – trotz struktureller Veränderungen im Bildungsbereich, wie etwa der vollständigen Umsetzung der Bologna-Reform bis etwa 2010. Für den BaPsy im Speziellen sollten jedoch weitere empirische Studien klären, ob er sich unter den nun etablierten Auswahlprozeduren als ebenso valide erweist. Dabei ist auch zu untersuchen, inwieweit der BaPsy – in Kombination mit weiteren Elementen des Auswahlverfahrens wie dem Online-Self-Assessment (OSA) und den bereitgestellten Vorbereitungsmaterialien – wie erhofft zu einem chancengerechteren Zugang zum Psychologiestudium beiträgt.

Literatur

- Dickhäuser, O., Merkle, B., Becker, N. & Spinath, B. (2022). Studierendenauswahlverfahren Psychologie: Konzeption, Implementation und Implikationen für die Beratung von Studieninteressierten. *Zeitschrift für Beratung und Studium* : ZBS, 17(3), 71–76.
- Formazin, M., Schroeders, U., Köller, O., Wilhelm, O. & Westmeyer, H. (2011). Studierendenauswahl im Fach Psychologie: Testent-

- wicklung und Validitätsbefunde. *Psychologische Rundschau*, 62, 221–236. <https://econtent.hogrefe.com/doi/10.1026/0033-3042/a000093>
- 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, 373–385. <https://doi.org/10.1037/0021-9010.92.2.373>
- Hell, B., Trapmann, S. & Schuler, H. (2007). Eine Metaanalyse der Validität von fachspezifischen Studierfähigkeitstests im deutschsprachigen Raum. *Empirische Pädagogik*, 21, 251–270.
- Horstmann, K. T., Biesok, A., Witte, K., Godmann, H. R., Fliedner, K., Wilm, L., Doran, L. et al. (2023). Berliner Studierfähigkeitstest–Psychologie (BSF-P). *Psychologische Rundschau*, 74, 221–238. <https://doi.org/10.1026/0033-3042/a000628>
- IFS Studentenkurse. (n.d.). *BaPsy-DGPs Intensivkurs inkl. Testsimulation Online* [Webseite]. Verfügbar unter: <https://www.studentenkurse.de/standorte/online/bapsy-dgps-intensivkurs/bapsy-dgps-intensivkurs-inkl-testsimulation/>
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Klauer, K. J. (Hrsg.). (2001). *Handbuch Kognitives Training* (2. Aufl.). Göttingen: Hogrefe.
- Kuncel, N. R., Hezlett, S. A. & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, 127, 162–181. <https://psycnet.apa.org/doiLanding?doi=10.1037%2F0033-2909.127.1.162>
- Marquardt, C., Dickhäuser, O. & Spinath, B. (2024). Was bewirkt ein Studieneignungstest? Erfahrungen aus dem neuen Studierendenauswahlverfahren im Fach Psychologie. *Zeitschrift für Beratung und Studium*, 19 (3–4), 101–106.
- MedGurus. (n.d.). *BaPsy Basic Paket* [Produktseite]. <https://medgurus.de/produkt/bapsy-basic-paket/>
- Menz, C., Rutsch, J. & Spinath, B. (2021). Strukturelle Gerechtigkeitsaspekte bei der Einführung eines Studieneignungstests im Studiengang Bachelor Psychologie. *Das Hochschulwesen*, 69, 24–30.
- Merkle, B., Schiltenswolf, M., Kiesel, A. & Dickhäuser, O. (2022). Entwicklung und Validierung eines Erwartungs- und Interessentests (Ex I-Test) zur Erkundung studienfachspezifischer Passung in einem Online-Self-Assessment. *ZeHf–Zeitschrift für empirische Hochschulforschung*, 5(2), 162–183. <https://doi.org/10.3224/zehf.5i2.05>
- Schneider, B., Becker, N., Krieger, F., Spinath, F. M. & Sparfeldt, J. R. (2020). Teaching the underlying rules of figural matrices in a short video increases test scores. *Intelligence*, 82, 101473. <https://doi.org/10.1016/j.intell.2020.101473>
- Schneider, B. & Sparfeldt, J. R. (2021a). How to get better: Taking notes mediates the effect of a video tutorial on number series. *Journal of Intelligence*, 9, 55. <https://doi.org/10.3390/jintelli9040055>
- Schneider, B. & Sparfeldt, J. R. (2021 b). How to solve number series items: Can watching video tutorials increase test scores? *Intelligence*, 87, 10154. <https://doi.org/10.1016/j.intell.2021.101547>
- Schult, J., Hofmann, A. & Stegt, S. J. (2019). Leisten fachspezifische Studierfähigkeitstests im deutschsprachigen Raum eine valide Studienerfolgsprognose? Ein metaanalytisches update. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 51, 16–30. <https://doi.org/10.1026/0049-8637/a000204>
- Truxillo, D. M., Bodner, T. E., Bertolino, M., Bauer, T. N. & Yonce, C. A. (2009). Effects of explanations on applicant reactions: A meta-analytic review. *International Journal of Selection and*

Assessment, 17, 346–361. <https://onlinelibrary.wiley.com/doi/10.1111/j.1468-2389.2009.00478.x>
Watrin, L., Geiger, M., Levacher, J., Spinath, B. & Wilhelm, O. (2022). Development and initial validation of an admission test for bachelor psychology studies. *Frontiers in Education*, 7, 909818. <https://www.frontiersin.org/journals/education/articles/10.3389/educ.2022.909818/full>

Onlineveröffentlichung: 03.02.2026

Interessenkonflikt

Oliver Dickhäuser hat der Deutschen Gesellschaft für Psychologie (DGPs) die Nutzungsrechte an den im Manuskript beschriebenen Videos übertragen und hierfür eine Lizenzgebühr erhalten. Die Durchführung der in diesem Manuskript berichteten Studie erfolgte unabhängig von der DGPs.

Förderung

Open Access-Veröffentlichung ermöglicht durch die Universität Mannheim.

Prof. Dr. Oliver Dickhäuser

Fakultät für Sozialwissenschaften
Universität Mannheim
A 5, 6
68159 Mannheim
Deutschland
oliver.dickhaeuser@uni-mannheim.de