



MASTERARBEIT | MASTER'S THESIS

Titel | Title

Der Guthke-Effekt auf dem Prüfstand – Alles eine Frage des
Mindsets?

verfasst von | submitted by

Steven Lennart Hobrecht BA BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt | Degree
programme code as it appears on the
student record sheet:

UA 066 840

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student
record sheet:

Masterstudium Psychologie

Betreut von | Supervisor:

Univ.-Prof. Dr. Veronika Job Sutnar

Inhaltsverzeichnis

1	Einleitung.....	1
2	Theoretische Grundlagen	3
2.1	Dynamisches Testen	3
2.1.1	Statusdiagnostik und ihre Limitationen	3
2.1.2	Ablauf und Funktionsweise des dynamischen Testens	4
2.1.3	Dynamische Diagnostik – ein Sammelbegriff	5
2.1.4	Die Zone der aktuellen Entwicklung.....	7
2.2	Intelligenz	8
2.2.1	Intelligenz-Definitionen	8
2.2.2	Intelligenzmodelle	9
2.3	Guthke Effekt:.....	11
2.4	Konzept der impliziten Intelligenz-Theorien	12
2.4.1	Messung der impliziten Intelligenz-Theorien	15
2.4.2	Implizite Messungen	16
2.4.3	Propositional Evaluation Paradigm (PEP)	17
3	Forschungsfragen und Hypothesen.....	17
4	Methode	18
4.1	Untersuchungsdesign	18
4.2	Erhebungsinstrument und Operationalisierung	19
4.2.1	Variablen	19
4.2.2	Operationalisierung.....	19
4.3	Stichprobe	23
4.4	Untersuchungsdurchführung.....	23
5	Ergebnisse	24
5.1	Stichprobenbeschreibung	24
5.2	Deskriptive Ergebnisse: APM.....	24
5.3	Deskriptive Ergebnisse: SPM.....	25
5.4	Deskriptive Ergebnisse: WST & MWT-A	25
5.5	Deskriptive Ergebnisse: PEP	25
6	Hypothesenprüfung	25
6.1	Vorraussetzungsprüfung der geplanten Analyseverfahren	25
6.2	Guthke Effekt.....	28
6.3	Moderationseffekt: Intelligenz Mindsets	29
6.4	Weitere Analysen	30
6.4.1	Explorative Itemanalyse	30

7	Diskussion	32
	8.1 Limitationen und Ausblick.....	34
8	Literaturverzeichnis	37
9	Anhang	45
	9.1 Abstract	45
	9.2 Nutzung künstlicher Intelligenz	46

1 Einleitung

Österreich gesteht allen Kindern in der Republik im Rahmen der Unterzeichnung der VN-Kinderrechtskonvention Artikel 28 das gleiche Recht auf Bildung zu, unabhängig von ihrem Geschlecht, religiösem, kulturellem oder sozioökonomischem Hintergrund (UNICEF, 1989). Kinder unterscheiden sich allerdings stark in ihrer genetischen Veranlagung und wachsen in verschiedensten Kontexten auf (Plomin & DeFries, 2013). Aufgrund dessen ist es mit der Chancengleichheit, wie in Artikel 28 beschrieben, nicht Genüge getan. Manche Kinder brauchen in ihrer (schulischen) Entwicklung mehr Unterstützung und Förderung als andere (Guthke, 1972).

Zu Beginn einer sinnvollen Förderung steht eine Diagnostik, welche Auskunft über den aktuellen Zustand gibt und auf Grundlage derer Prognosen sowie konkrete Förderungspläne erstellt werden können (Kaufmann & Kaufmann, 2001). In diesem Kontext ist es wichtig zu ermitteln, worauf festgestellte Defizite zurückzuführen sind. Hier ist eine Diagnostik förderlich, welche wenig auf Vorwissen beruht und daher möglichst fair ist (Wygotski, 1978).

Bei einem Kind, das Defizite in schulischen Leistungen zeigt, ist im Rahmen einer Diagnostik die Frage zu beantworten, ob die unterdurchschnittliche Leistung auf eine mögliche Lernschwäche oder beispielsweise sozioökonomische Faktoren zurückzuführen ist. Durch die reine Testung der aktuellen Fähigkeiten lässt sich diese Frage nur unzureichend beantworten (Guthke & Heller, 1995). Sinnvoller wäre es stattdessen, möglichst unabhängig von Vorwissen und Umwelteinflüssen das Potenzial einer Person zu ermitteln, indem man die Lernfähigkeit „live in action“ beobachtet (Wygotski, 1978). Genau diesen Weg geht das dynamische Testen (Guthke, 1972).

Beim dynamischen Testen findet Leistungsdiagnostik über mehrere Testungen statt. Zwischen den Testungen finden Förderungsmaßnahmen statt. So kann ein Potenzial ermittelt werden, wohin sich die getestete Person bei entsprechender Förderung entwickeln könnte (Tzuriel, 2001). So wären beispielsweise bei einem begabten, aber leistungsschwachen Kind, das bis dahin nur ungenügend unterstützt wurde, große Potenziale beobachtbar. Diese Potenziale wären bei einmaliger konventioneller Testung möglicherweise unentdeckt geblieben und eine sinnvolle Förderung wäre ausgeblieben.

Anhand dieses Beispiels zeigt sich, welche Chancen in einer fairen Leistungsdiagnostik liegen. Aufgrund des vergleichsweise hohen Zeit- und Kostenaufwands, den das dynamische Testen erfordert, wird es allerdings nur sehr vereinzelt angewandt.

Eine Möglichkeit, das dynamische Testen attraktiver zu gestalten, wäre es, die Kosten und den Aufwand zu reduzieren. In diesem Kontext könnte man sich den sogenannten Guthke-Effekt zunutze machen. Dieser besagt, dass Lernprozesse bzw. Lernpotenziale auch durch die reine Testwiederholung, also ohne Förderung zwischen den Testungen beobachtbar sind (Guthke, 1972). Folglich wäre die Durchführung der Interventionen zwischen den Testungen möglicherweise vermeidbar.

Der Guthke-Effekt wurde erstmals in den 1960er Jahren festgestellt und fand seit seiner Entdeckung kaum Beachtung im wissenschaftlichen Diskurs. Diese Arbeit möchte nun nach über 60 Jahren den Versuch unternehmen, den Guthke-Effekt zu replizieren.

Darüber hinaus verfolgt die Arbeit das Ziel, zur Erklärung des Effekts beizutragen. Hierfür liegt der Fokus auf dem motivationalen Faktor der Intelligenz-Mindsets, die weitestgehend unbewusste Überzeugungen der eigenen Intelligenz sind und in Lernprozessen erwiesenermaßen eine elementare Rolle einnehmen (Dweck & Leggett, 1988). Sie beeinflussen wesentlich wie wir mit Herausforderungen umgehen und mit welcher Art von Motivation wir Problemen begegnen (Dweck & Leggett, 1988) und stellen daher einen vielversprechenden Ansatz zur Erklärung des Guthke-Effekts dar.

2 Theoretische Grundlagen

In den folgenden Abschnitten werden grundlegende Begrifflichkeiten und Theorien vorgestellt sowie relevante Studienergebnisse eingeordnet. Hierfür wird zu Beginn das Konzept des dynamischen Testens vorgestellt. Anschließend findet ein Bestimmungsversuch des Intelligenzbegriffs statt, bevor das Mindset-Framework sowie seine methodische Implikationen diskutiert werden.

2.1 Dynamisches Testen

2.1.1 Statusdiagnostik und ihre Limitationen

Das dynamische Testen wurde aufbauend auf der Kritik an der Statusdiagnostik entwickelt (Guthke, 1972). Um die Idee des dynamischen Testens zu verstehen, ist es daher wichtig, sich eingehend mit dem Konzept der Statusdiagnostik inklusive ihrer Limitationen zu beschäftigen.

Das Ziel von psychologischer Statusdiagnostik ist es, zeit- und situationsstabile psychische Merkmale einer Person zu erfassen. Die Erhebung des Ist-Zustands ist in vielen Bereichen bereits ausreichend. Beispielsweise spielt die Statusdiagnostik eine zentrale Rolle in klinischen und forensischen Frage- und Problemstellungen. Anhand einer Diagnose können fundierte Entscheidungen getroffen und Handlungsempfehlungen ausgesprochen werden (Lange & Schmidt, 2019).

Statusdiagnostik, wie die Namensgebung bereits vermuten lässt, trifft primär Aussagen bezüglich des derzeitigen Zustands und ist nur bedingt dazu geeignet, Prognosen zu machen. Besonders im Kontext von Entwicklungs- und Bildungsfragen ist dies jedoch essenziell, um Heranwachsende sowie generell lernende Personen optimal fördern zu können (Guthke, 1972).

Würde man z.B. einem leistungsschwachen Schulkind einen konventionellen Intelligenztest vorlegen, so würde das Ergebnis für die Prognose und darauf aufbauende Förderung wichtige Fragen unbeantwortet lassen. Ist das Schulkind kognitiv zu nicht mehr fähig? Gibt es ungenutzte Potenziale? Welche Form von Unterstützung benötigt das Schulkind, um ihr oder sein Potenzial auszuschöpfen? Aus der Kritik an der Statusdiagnostik heraus entstand das dynamische Testen, mit dem Ziel, sich genau diesen Fragen besser annähern zu können (Guthke, 1972). In den folgenden Abschnitten

wird das Konzept der dynamischen Diagnostik und insbesondere des dynamischen Testens näher beleuchtet.

2.1.2 Ablauf und Funktionsweise des dynamischen Testens

Das dynamische Testen setzt sich üblicherweise aus drei Phasen zusammen: dem Prätest, der Pädagogisierungs- bzw. Belehrungsphase sowie einem Posttest (Guthke, 1972). Der Prätest kann mit der Testung der Statusdiagnostik gleichgesetzt werden. Hier wird der Status Quo, also die derzeitige Leistungsfähigkeit, erhoben. In der zweiten Phase werden Interventionen mit den Proband:innen durchgeführt, um die im Prätest abgefragten Fähigkeiten zu verbessern und dadurch Lernprozesse zu simulieren (Guthke, 1972). In dieser Phase stehen also testdurchführende und teilnehmende Personen in einem aktiven Austausch (Guthke, 1972). Dies steht im starken Kontrast zur konventionellen Leistungsdiagnostik, welche üblicherweise versucht, den potenziellen Einfluss durch Testleiter:innen möglichst gering zu halten (Bosma & Resing, 2012).

Im Anschluss an die Pädagogisierungsphase findet erneut eine zum Prätest äquivalente Testung der Fähigkeiten statt. Anhand der Differenz zwischen Prä- und Posttest lassen sich nun mögliche Potenziale identifizieren. Diese Potenziale geben Hinweise darauf, wozu die getestete Person durch Förderung theoretisch imstande wäre (Guthke, 1972).

Zur Bildung eines numerischen Potenzials gibt es verschiedene Ansätze (siehe für einen ausführlichen Überblick: Guthke, 1972). Die reine Differenz zwischen den beiden Testzeitpunkten als Potenzial zu definieren, ist aus verschiedenen Gründen nicht sinnvoll. So wird die Ausgangslage (Ergebnis des Prätests) nicht berücksichtigt (Guthke, 1972). Beispielsweise hätte eine Person, die im Prätest sehr gut abschneidet (z.B. 90 % richtig gelöste Aufgaben), nur noch die Möglichkeit, sich im Posttest um weitere 10 % zu verbessern. Wohingegen eine Person, die im Prätest z.B. 20 % der Aufgaben lösen konnte, deutlich mehr Verbesserungsmöglichkeiten hat. Darüber hinaus sind bei vielen verwendeten Verfahren der Leistungsdiagnostik die Schwierigkeitsunterschiede zwischen den Items nicht homogen. So gibt es in der Regel mehr Items mit einem mittleren Schwierigkeitsniveau als Items mit einem hohen bis sehr hohen Niveau (Guthke, 1972). Dies würde im oben angeführten Beispiel dazu führen, dass es der im

Prätest fähigeren Person deutlich schwieriger fallen wird, sich numerisch in einem gleichen Maße zu verbessern wie der leistungsschwächeren Person.

Um dem entgegenzuwirken werden von Guthke (1972) verschiedene Formeln zur Berechnung der Lernfähigkeit diskutiert. Exemplarisch kann hier folgende Formel von McGuigan (1965) genannt werden:

$$L = \frac{D}{X_{max} - X_{prä}}$$

(L = Lernfähigkeit, D = Differenz der Punktwerte von Post- und Prätest, X_{max} = maximal erreichbarer Wert im Test, X_{prä} = Wert des Prätests)

Durch diese und ähnliche Formeln findet eine Gewichtung der Testaufgaben nach ansteigender Schwierigkeit der Aufgaben statt. So wird gewährleistet, dass eine Person, die den Maximalwert in einem Test erreicht, auch der Optimalwert zugesprochen wird und so leistungsstarke Personen nicht benachteiligt werden. Jedoch sorgt auch diese Formel dafür, dass z.B. eine Person, welche sich innerhalb der Testungen von 0 auf 60 Punkten verbessert hat, die gleiche Lernfähigkeit attestiert wird, wie einer Person, welche sich von 59 auf 60 Punkte verbessert (Guthke, 1972).

Ein wesentlicher Vorteil des dynamischen Testens gegenüber der Statusdiagnostik besteht darin, dass es weit weniger stark von Vorwissen und anderen Faktoren beeinflusst wird. Kumuliertes Wissen ist nur bedingt als Indikator für die Lernfähigkeit geeignet. Besser ist es hingegen, diese Fähigkeit „live in action“ zu beobachten, wie dies beim dynamischen Testen der Fall ist (Haywood & Tzuriel, 2002).

Außerdem spielen Faktoren wie z.B. der sozioökonomische Hintergrund oder die Bildungskarriere einer Person eine geringere Rolle. Diese Faktoren sind neben der genetischen Veranlagung wesentlich für die Entwicklung von Fähigkeiten. Indem im dynamischen Testen alle entsprechend ihren Fähigkeiten gefördert werden, findet hier eine potenziell fairere Testung statt (Röder et al., 2021).

2.1.3 Dynamische Diagnostik – ein Sammelbegriff

In der Literatur fallen im Kontext dynamischer Diagnostik die Begrifflichkeiten des dynamischen Testens, der Lerntests oder des dynamischen Assessments (Röder et al., 2021). Diese werden tendenziell synonym verwendet und stellen daher einen Überbegriff dar, der Verfahren beschreibt, welche dem in Abs. 3.1.2 beschriebenen Ablauf und der Funktionsweise ähneln. Allerdings weisen diese Verfahren teilweise starke Differenzen

bezüglich ihrer Durchführungszeit, der Art der Pädagogisierungsphase sowie des Merkmals, welches sie testen, auf (Börnert-Ringleb & Wilbert, 2021). Im Folgenden soll diese methodische Vielfalt in Kürze skizziert werden.

Auf der zeitlichen Dimension kann grundlegend zwischen Test-train-test-Ansätzen sowie Within-Test-Ansätzen differenziert werden (Guthke et al., 2003). In Abs. 2.1.2 wurde bisher lediglich der Test-train-Ansatz vorgestellt. Hierbei handelt es sich um ein Prä-Post-Design, bei dem die verschiedenen Phasen getrennt ablaufen und aufeinander folgen. In welchem Ausmaß diese zeitliche Trennung der Phasen erfolgt, ist weiter differenzierbar (Guthke et al., 2003). Guthke (1972) schlägt in diesem Zusammenhang eine Unterscheidung zwischen der Intervalltestmethode und der Trainingsmethode vor.

Bei der Intervalltestmethode umfasst die Pädagogisierungsphase ein sehr ausführliches und potenziell über Monate andauerndes Training. Diese Methode ist also extrem zeitintensiv und aufwändig. Hinzu kommt, dass aus ihr resultierende Testergebnisse nur schwer vergleichbar sind, da mit zunehmender Dauer auch der Einfluss von unerwünschten Faktoren ansteigt (z.B. Qualifizierung des Fachpersonals) ansteigt (Guthke, 1972).

Der deutlich üblichere Test-train-test-Ansatz ist die Trainingsmethode (Langzeit-Lerntests). Ähnlich wie bei der Intervalltestmethode sind auch hier die drei Phasen zeitlich getrennt. Jedoch beträgt der Abstand zwischen Prä- und Posttest nur mehrere Tage bis wenige Wochen. Des Weiteren ist das in der zweiten Phase absolvierte Training wesentlich kürzer und standardisierter (Guthke, 1972).

Einen anderen Weg gehen Within-Test-Ansätze oder auch Kurzzeitleerntests. Bei diesen Verfahren findet eine Verschmelzung zwischen Test- und Trainingsphasen statt. Während der Testung erhält die getestete Person beim Bearbeiten der Testaufgaben direktes Feedback (z.B. Richtig/Falsch) oder Denkhilfen beim Scheitern einer Aufgabe (Börnert-Ringleb & Wilbert, 2021).

Wie bereits angedeutet, kann der Standardisierungsgrad der Pädagogisierungsphase je nach Ansatz stark variieren. Selbst innerhalb der verschiedenen Subtypen gibt es teilweise große Unterschiede (Guthke, 1972). Es lässt sich jedoch feststellen, dass die Intervalltestmethode tendenziell einen geringeren Standardisierungsgrad aufweist als Langzeit-/Kurzzeitleerntests, da hier die Pädagogisierungsphase eher einen

Unterrichtsscharakter aufweist (Guthke, 1972). Des Weiteren lassen sich unterschiedliche Herangehensweisen methodenübergreifend identifizieren. Beim dynamischen Assessment sind die Interventionen in der Pädagogisierungsphase von einer hohen interaktiven Dichte zwischen den Testleiter:innen und den Proband:innen geprägt. Den Testleiter:innen kommt hier also eine zentrale Rolle im simulierten Lernprozess der getesteten Person zu. Insbesondere hier kommen nicht standardisierte Denkhilfen bzw. Unterstützungen zur Anwendung (Lidz, 2002).

Im Kontrast dazu stehen Verfahren des dynamischen Testens, welche auf eine stärkere Standardisierung der Hilfestellungen setzen. In der Literatur werden die Begriffe des dynamischen Assessments und des dynamischen Testens sowie die damit einhergehenden methodischen Implikationen jedoch zum Teil synonym verwendet (Röder et al., 2021).

Die dynamische Diagnostik entstand, wie bereits ausgeführt, unter anderem aus der Kritik gegenüber der konventionellen Intelligenzdiagnostik (Guthke, 1972). Aufgrund dieser Entstehungsgeschichte beschäftigt sich eventuell bis heute ein wesentlicher Teil der ausgearbeiteten Verfahren der dynamischen Diagnostik mit der Prüfung kognitiver Leistungsfähigkeit im Rahmen von Intelligenzdiagnostik. Darüber hinaus wird die Methode des dynamischen Testens zunehmend auch auf domänenspezifischere Kompetenzbereiche angewendet, wie beispielsweise der Rechen- (Fuchs et al., 2008) oder Lesefähigkeit (Dörfler et al., 2009) sowie im sprachlichen Kontext (Ehlert, 2021). Im Rahmen dieser Arbeit liegt der Fokus allerdings auf dem dynamischen Testen im Bereich der Intelligenzdiagnostik.

2.1.4 Die Zone der aktuellen Entwicklung

Das Konzept des dynamischen Testens stellt einen Versuch dar, Wygotskis (1978) theoretische Überlegungen in der Praxis Rechnung zu tragen. Nach Wygotski muss für die Erfassung der Leistungsfähigkeit neben dem aktuellen Entwicklungsstand („Zone der aktuellen Entwicklung“) auch das Entwicklungspotential („Zone der nächsten Entwicklung“) einer Person berücksichtigt werden. Die Zone der nächsten Entwicklung beschreibt hierbei die potenzielle Leistungsfähigkeit, die eine Person durch förderliche Bedingungen erreichen kann. Für die Prognose von zukünftigen Leistungsentwicklungen ist die Erfassung der Zone der nächsten Entwicklung wesentlich aufschlussreicher, da sie

die Entwicklungsmöglichkeiten der Proband:in unter gezielter pädagogischer Förderung aufzeigt. Die Zone der aktuellen Entwicklung ist hingegen als ein Ergebnis voriger Lernerfahrungen zu verstehen (Wood et al., 1976).

2.2 Intelligenz

Wie bereits erwähnt sind die Anwendungsbereiche des dynamischen Testens sehr vielfältig. Ihren Ursprung hat die dynamische Diagnostik in der Intelligenzdiagnostik. In diesem Kontext wurde auch der in Abschnitt 2.3 eingeführte Guthke-Effekt erstmals festgestellt (Guthke, 1972). Da diese Arbeit sich im Wesentlichen mit diesem Effekt beschäftigt ist auch das Konzept der Intelligenz von großer Bedeutung und wird daher im Folgenden diskutiert.

Diese Arbeit erhebt keinerlei Anspruch, das komplexe Themenfeld der Intelligenz und die damit einhergehende vielschichtige Forschung allumfassend darzustellen und zu durchleuchten. Dennoch wird aufgrund ihrer Bedeutung in dieser Arbeit ein Versuch unternommen, das Thema Intelligenz in seinen Grundzügen zu umreißen und so für die Lesenden ein grundlegendes Verständnis zu schaffen (für ein tieferes Verständnis siehe z.B. Eder, 2018 oder Schmidt-Atzert & Kuhl, 2019).

2.2.1 Intelligenz-Definitionen

Themen der Intelligenz, ihrer Konstitution und ihrer Erfassung mittels diagnostischer Verfahren nehmen im Forschungsfeld der Persönlichkeits- und Differentiellen Psychologie seit ihrer Begründung eine zentrale Rolle ein (Gittler & Pietschnig, 2019). Darüber, was unter dem Konzept der Intelligenz zu verstehen ist, gibt es im alltäglichen und wissenschaftlichen Gebrauch verschiedene Auffassungen. So wird Intelligenz im alltäglichen Gebrauch beispielsweise mit Begriffen der Begabung (Asendorpf, 1996), Klugheit (Hassenstein, 1992) und Weisheit (Baltes & Smith, 1990) assoziiert. Dieses diffuse Verständnis, welches sich im Alltagsgebrauch vorfinden lässt, deutet bereits an, dass eine eindeutige und klare Definition von Intelligenz schwierig ist. Auch analytische Intelligenzdefinitionen tun sich schwer, den Begriff der Intelligenz durch konkrete Fähigkeiten zu charakterisieren und bleiben stattdessen häufig vage (Gittler & Pietschnig, 2019). Vielfach referenzierte Definitionsversuche, die dies veranschaulichen, stammen beispielsweise von Wechsler (1956): „Intelligenz ist die Fähigkeit, zweckvoll zu handeln, vernünftig zu denken und sich mit seiner Umgebung wirkungsvoll auseinanderzusetzen“

(S. 13) oder Stern: „Intelligenz ist die allgemeine Fähigkeit des Individuums, sein Denken bewusst auf neue Forderungen einzustellen; sie ist die allgemeine geistige Anpassungsfähigkeit an neue Aufgaben u. Bedingungen des Lebens“ (Stern, 1912 zitiert nach Gittler & Pietschnig, 2019, S. 55). Wie sich exemplarisch an den Definitionsversuchen von Wechsler (1956) und Stern (1912) aufzeigen lässt, wird Intelligenz im wissenschaftlichen Kontext in der Regel mit den Fähigkeiten assoziiert, Probleme zu lösen, sich anzupassen und sich effizient mit seiner Umwelt auseinanderzusetzen. Gittler und Pietschnig (2019) sowie Guthke (1872) kritisieren diese Begriffsbestimmungen und betonen, dass diese zu global sowie unpräzise seien und über wenig Erklärungswert verfügen würden.

2.2.2 Intelligenzmodelle

Die bisher angeführten Definitionen beschreiben Intelligenz als eine einheitliche Persönlichkeitseigenschaft. Diese Auffassung wird jedoch durch faktoranalytisch gewonnene Modelle infrage gestellt, welche Intelligenz als ein aus mehreren Komponenten bestehendes Konstrukt verstehen (Lohhaus & Vierhaus, 2019). Um dies zu verdeutlichen, werden nachfolgend einige der einflussreichsten Intelligenzmodelle vorgestellt.

Spearman (1927) postuliert, dass es eine allgemeine Intelligenz (Generalfaktor der Intelligenz oder g-Faktor) gibt, die sich auf die gesamte Leistungs- und Lernfähigkeit auswirkt. Zusätzlich zum g-Faktor gibt es die s-Faktoren, die in spezifischen Bereichen ausgeprägt sind (z.B. räumliches Denken). Diese s-Faktoren werden jedoch auf Basis des g-Faktors ausgebildet, sie werden also stark von ihm beeinflusst. Empirisch lassen sich vielfach Studienergebnisse auffinden, die Spearmans These stützen und die allgemeine Intelligenz beispielsweise mit akademischen oder beruflichen Erfolgen in Zusammenhang bringen (z.B. Dalayga & Baskaran, 2019; Broady, 1997; Deary et al., 2007 & Laidra et al., 2007).

Cattell (1987) hingegen charakterisiert Intelligenz als ein aus zwei hierarchisch gleichwertigen Faktoren bestehendes Konstrukt: der fluiden und kristallinen Intelligenz. Die fluide Intelligenz entspricht im Wesentlichen Spearmans g-Faktor (1927), sie umfasst also die Fähigkeiten des schlussfolgernden Denkens und der kognitiven Anpassungsfähigkeit. Sie wird primär von der genetischen Veranlagung bestimmt,

während äußere Umweltfaktoren bei ihrer Ausbildung eine untergeordnete Rolle spielen. Die kristalline Intelligenz hingegen repräsentiert Wissen und Fähigkeiten, die durch Erfahrungen erworben wurden, wie beispielsweise Faktenwissen oder Sprachkenntnisse. Folglich spielen äußere Umweltfaktoren bei der kristallinen Intelligenz eine deutlich größere Rolle als bei der fluiden Intelligenz. Die fluide Intelligenz wird neben Umwelteinflüssen auch von der kristallinen Intelligenz beeinflusst, da beim Erwerb von fluider Intelligenz kristalline Fähigkeiten wie das logische Denken hilfreich sein können (Eder, 2018). So fällt das Erlernen einer Fremdsprache (fluide Intelligenz) möglicherweise leichter mit einer stark ausgeprägten kristallinen Intelligenz, da die Logik der fremden Grammatik besser und schneller verinnerlicht werden kann.

Thurstone (1938) beschreibt Intelligenz ebenso wie Cattell (1987) als eine aus mehreren Faktoren bestehende Eigenschaft: den Primärfaktoren. So bilden folgende sieben voneinander unabhängig bestehende Primärfaktoren zusammen die Intelligenz: 1. Räumliches Vorstellungsvermögen, 2. Schlussfolgerndes Denken, 3. Merkfähigkeit, 4. Wahrnehmungsgeschwindigkeit, 5. Rechenfähigkeit, 6. Sprachverständnis und 7. Wortflüssigkeit. Theoretisch lassen sich die Primärfaktoren 1-4 der fluiden Intelligenz und 5-7 der kristallinen Intelligenz zuordnen. Der große Unterschied zu Spearman und Cattells Modellen besteht darin, dass Thurstone die Unabhängigkeit der einzelnen Faktoren annimmt (Lohhaus & Vierhaus, 2019). Diese These konnte allerdings bisher noch nicht empirisch verifiziert werden (Myers, 2008).

Aufbauend auf den verschiedenen Überlegungen von Spearman, Cattell, Thurstone und weiteren Kolleg:innen entwickelte Carroll (1993) die *Three-Stratum-Theorie* der Intelligenz. Diese versucht, die bis dahin nebeneinander existierenden Ansätze in einem Modell zu vereinen. Nach Carroll (1993) ist Intelligenz in drei Schichten organisiert: Übergeordnet findet sich angelehnt an Spearman die allgemeine Intelligenz, darunter in der zweiten Schicht acht weitere Intelligenzfaktoren, welche Thurstones Primärfaktoren stark ähneln. In der dritten Ebene befinden sich 69 spezifische Einzelfaktoren, welche den Faktoren der zweiten Schicht zugeordnet werden.

Zusammenfassend lässt sich feststellen, dass durchaus verschiedene Ansichten darüber existieren, was Intelligenz ist, wie sie strukturiert ist und anhand welcher

Fähigkeiten sie sich operationalisieren lässt. Im Rahmen dieser Arbeit wird weiter mit dem Intelligenzverständnis nach Cattell (1987) gearbeitet.

2.3 Guthke Effekt:

Im bisherigen Verlauf der Arbeit wurde die Idee des dynamischen Testens vorgestellt sowie das Konzept der Intelligenz diskutiert. Auf diesem theoretischen Fundament wird nun der Guthke-Effekt behandelt, der im Mittelpunkt dieses Projektes steht. Um die mögliche Tragweite dieses Effektes zu begreifen, lohnt es sich nochmals eingehend mit den Nachteilen der dynamischen Diagnostik zu beschäftigen.

Verfahren der dynamischen Diagnostik finden in der Praxis weit weniger Anwendung als Verfahren der Statusdiagnostik. Dies ist vermutlich darauf zurückzuführen, dass sie wesentlich zeitintensiver für Testende und Testleitende sind. Zudem erfordert das dynamische Testen je nach Standardisierungsgrad für das leitende Personal zusätzliche Qualifikationen, um in der Pädagogisierungsphase mit den Proband:innen arbeiten zu können. Diese Umstände führen zu erhöhten Kosten und machen die dynamische Diagnostik im Vergleich zur Statusdiagnostik unattraktiv (Tzuriel, 2001). Insbesondere im Entwicklungs- und Bildungskontext können durch Verfahren der dynamischen Diagnostik wertvolle Informationen gewonnen werden, um Menschen bestmöglich zu unterstützen und zu fördern (Guthke, 1972). Ein möglicher Weg, die dynamische Diagnostik attraktiver zu machen, könnte darin bestehen, das dynamische Testen zu verschlanken und somit insbesondere kostengünstiger zu gestalten.

Spuren eines möglichen Lösungsansatzes lassen sich bei genauem Hinsehen in den 60er Jahren verorten. Zu dieser Zeit war das dynamische Testen noch in seinen Anfängen begriffen und wurde vermehrt diskutiert. Um die Wirksamkeit verschiedener Ansätze der Pädagogisierungsphasen zu untersuchen, wurden in verschiedenen Studien mit Kontrollgruppen gearbeitet. Während Versuchsgruppen nach dem Prätest verschiedenste Interventionen durchliefen, absolvierten die Kontrollgruppen nur Prä- und Posttest (z.B. Driesel & Nosko, 1970 zitiert nach Guthke, 1972; Sauer, 1968 zitiert nach Guthke, 1972; Schmidt 1969 zitiert nach Guthke, 1972; Maiwald & Stübe, 1971 zitiert nach Guthke, 1972). Die Versuchsgruppen verbesserten sich studienübergreifend in einem signifikant größeren Maße als die Kontrollgruppen. Jedoch waren auch bei den

Kontrollgruppen signifikante Lernleistungen festzustellen. Dieser von Guthke (1972) erstmals festgestellte Effekt wird als Guthke-Effekt bezeichnet.

Die festgestellten Leistungsverbesserungen durch die reine Testwiederholung ergeben im Kontext des dynamischen Testens Sinn, da Proband:innen durch das Absolvieren des Prätests bereits Lernprozesse durchschreiten, welche sich im Ergebnis des Posttests wiederfinden lassen. Somit kann der Prätest für sich als eine Art Intervention verstanden werden. Untersuchungen, bei denen der Guthke-Effekt in den 60er und 70er Jahren festgestellt werden konnte, beschäftigten sich mit der Prüfung des schlussfolgernden Denkens (fluide Intelligenz). So wurden Matrizen-, Analogien- und Zahlenfolgenaufgaben bearbeitet (Guthke, 1972). Der Guthke-Effekt fand in den folgenden Jahren im wissenschaftlichen Diskurs kaum Beachtung. Ein Ziel dieser Arbeit ist es zu überprüfen, ob der Guthke-Effekt replizierbar ist und möglicherweise dazu dienen kann, die dynamische Diagnostik für die Praxis attraktiver zu machen.

2.4 Konzept der impliziten Intelligenz-Theorien

Nun stellt sich die Frage, wodurch sich die im Kontext des Guthke-Effektes beobachtbaren Leistungsverbesserungen erklären lassen bzw. wodurch sie beeinflusst werden. Dieser Frage nachzugehen, stellt das zweite Ziel dieser Arbeit dar.

Die Erbringung von Leistung bei der Bearbeitung einer Aufgabe hängt hauptsächlich vom Schwierigkeitsgrad der Aufgabe, den Fähigkeiten sowie der Leistungsmotivation des Individuums ab (Gittler & Pietschnig, 2019).

Entspricht die Schwierigkeit einer Aufgabe den Fähigkeiten einer Person und ist diese zudem motiviert diese konzentriert zu bearbeiten, ist die Wahrscheinlichkeit hoch, dass die Person die Aufgabe erfolgreich bearbeiten wird. Fehlt der Person jedoch die Motivation, sich anzustrengen, vermindert dies die Erfolgsaussichten. An diesem simplen Beispiel lässt sich veranschaulichen, wie wichtig motivationale Faktoren für die Erbringung von Leistung sein können.

Der Guthke-Effekt beschreibt Leistungsverbesserungen bzw. Lerneffekte beim dynamischen Testen ohne externe Hilfeleistungen (Guthke, 1972). Ob und wie stark sich Personen in diesem Kontext verbessern, ist möglicherweise eine Frage der Motivation, die eigenen Fähigkeiten so zu verbessern, sodass sie dem Schwierigkeitsgrad der gestellten Aufgaben entsprechen.

Mit diesem Hintergrund scheint das motivationale Konzept der impliziten Intelligenz-Theorien ein vielversprechender Erklärungsansatz zu sein.

Im folgenden Abschnitt soll das Konzept der impliziten Intelligenz-Theorien, welches wesentlich für diese Arbeit ist, in seinen Grundzügen vorgestellt, seine Bedeutung für Lernprozesse anhand von Studienergebnissen aufgezeigt und im Kontext dieser Arbeit eingeordnet werden.

Das Konzept der impliziten Intelligenz-Theorien erklärt, warum Menschen mit potenziell vergleichbaren Fähigkeiten unterschiedlich mit Rückschlägen oder Herausforderungen umgehen (Dweck & Leggett, 1988). So gibt es grundsätzlich die Möglichkeit, auf eine Herausforderung adaptiv zu reagieren, indem beispielsweise die Anstrengung und der Fokus auf diese erhöht werden. Im Gegenzug dazu kann aber ebenso maladaptiv reagiert werden, was zur Folge haben kann, dass betroffene Personen sich weniger anstrengen und versuchen, diese Herausforderung und ähnliche Situationen zukünftig zu vermeiden (Diener & Dweck 1978). Diese unterschiedlichen Verhaltensmuster stehen nach Elliot und Dweck (1988) im Zusammenhang mit verschiedenen Zielsetzungen: den Lern- und Performanz-Zielen. Diese Zielsetzungen schaffen einen Rahmen, der ausschlaggebend ist für den Umgang mit Herausforderungen.

Personen, die Lernziele verfolgen, streben in Situationen der Herausforderung danach, in diesen Situationen zu lernen und an ihnen zu partizipieren. Bei Leistungszielen steht das Erzielen der bestmöglichen Leistung im Mittelpunkt (Elliot & Dweck, 1988).

Die adaptive oder maladaptive Reaktion auf einen Rückschlag ist also auf verschiedene Zielsetzungen zurückzuführen. Diese Zielsetzungen wiederum lassen sich durch verschiedene implizite Intelligenz-Mindsets erklären (Dweck & Leggett, 1988).

Mindsets im Allgemeinen sind weitestgehend unbewusste Annahmen und Überzeugungen bezüglich der Konstitution und Wandelbarkeit persönlicher Eigenschaften. Diese lassen sich auf einem Kontinuum zwischen den beiden Polen verorten: die eigenen Fähigkeiten und Eigenschaften als unveränderbar (Entitätstheorie; Fixed Mindset) oder wandelbar (Veränderbarkeitstheorie; Growth Mindset) anzusehen (Hertel & Kerlen, 2021). Diese Annahmen schaffen einen Bedeutungsrahmen, der wie bereits geschildert, sich auf Attributionsstile, Zielsetzungen und Verhaltensweisen

auswirkt (Hong et al., 1999). Im Kontext dieser Theorie stehen sich folglich das Fixed Mindset und Growth Mindset diametral gegenüber.

Ein Fixed Mindset postuliert die eigene Intelligenz als angeborene, unveränderbare Eigenschaft, wohingegen ein Growth Mindset auf der Annahme beruht, dass die eigene Intelligenz veränderbar ist (Dweck & Leggett, 1988). Eine Person, die mit einer Herausforderung konfrontiert ist und ein Fixed Mindset aufweist, empfindet diese Situation als Prüfung ihrer eigenen Fähigkeiten (Dweck & Leggett, 1988). Wenn in dieser Situation ein Rückschlag erfahren wird, wird dieser auf die mangelnden Fähigkeiten zurückgeführt, was wiederum eine Gefahr für das Selbstwertgefühl darstellt (Dweck & Leggett, 1988).

Aufgrund dieser Annahmen ergibt es Sinn, zur Bewältigung von Herausforderungen Performanz-Ziele verfolgt werden, um bestmögliche Leistungen zu erzielen und so nicht das eigene Selbstwertgefühl zu gefährden. Wenn im Rahmen eines Rückschlags die Zielerreichung gefährdet ist, sind maladaptive vermeidende Verhaltensweisen die Folge, um das Selbstwertgefühl zu schützen (Nussbaum, 2008).

Wenn nun eine Person mit einem Growth Mindset einen Rückschlag erfährt, dann interpretiert sie dies als Chance, ihre Fähigkeiten durch Mobilisierung zusätzlicher Ressourcen zu verbessern (Dweck & Leggett, 1988). Der Rückschlag bzw. das Scheitern ist hier also deutlich positiver konnotiert als beim Fixed Mindset, da der Status des Scheiterns als Zwischenschritt wahrgenommen wird, der überwunden werden kann und als Teil des Lernprozesses verstanden wird. So lassen sich auch die adaptiven Reaktionen auf einen Rückschlag im Rahmen von Lernzielen erklären (Dweck & Leggett, 1988).

Das Konzept inklusive der Zusammenhänge zwischen den impliziten Intelligenz-Theorien, den verschiedenen Zielsetzungen und den Verhaltensmustern wurde seit seiner Einführung durch Dweck und Kolleg:innen (z.B. Dweck & Leggett, 1988) vielfach untersucht findet seitdem großen Anklang in der Literatur (Bempechat et al., 1991; Cury et al., 2006; Elliot & Dweck, 1988; Farrell & Dweck, 1985; Leggett, 1986; Robins & Pals, 2002). Anzumerken ist, dass diese adaptiven bzw. maladaptiven Verhaltensweisen – die in Verbindung mit den impliziten Intelligenz-Theorien stehen – primär zum Tragen kommen, wenn Menschen mit Rückschlägen konfrontiert sind. Sind die Anforderungen

einer Herausforderung den Fähigkeiten einer Person entsprechend, so performen Personen unabhängig von ihrem Mindset ähnlich (Diener & Dweck, 1978; Diener & Dweck, 1980 zitiert nach Dweck & Leggett, 1988).

Untersuchungen bezüglich impliziter Intelligenz-Theorien deuten darauf hin, dass – wie sich bereits aus dem vorgestellten Konzept von Dweck und Kolleg:innen (z.B. Dweck & Leggett, 1988) ableiten lässt – Intelligenz-Mindsets von Bedeutung sind für das Lernen und den Erfolg im Schul- und Hochschulsystem in Form von Noten und Prüfungsleistungen. Konkret wirken sie sich auf die Attribution und Interpretation von Lernsituationen, die Auswahl von Zielen und Strategien, konkretes Lernverhalten sowie die Motivation der lernenden Personen aus (Blackwell et al., 2007; Burnette et al., 2013; Haimovitz & Dweck, 2017; Hertel & Karlen, 2021; Hong et al., 1999).

So ist es wahrscheinlicher, dass Lernende mit einem Growth Mindset nach einem Rückschlag Strategien anwenden, um ihre betreffenden Fähigkeiten zu verbessern und dabei mehr Aufwand betreiben als Lernende mit einem Fixed Mindset (Hong et al., 1999). Dies führen Dweck und Nussbaum (2008) darauf zurück, dass bei Personen mit Fixed- und Growth Mindset unterschiedliche Abwehrmechanismen vorliegen, um ihr Selbstwertgefühl zu bewahren. Hier wird unterschieden zwischen primären Kontrollmechanismen, bei denen an der Ursache der Bedrohung des Selbstwertgefühls angesetzt wird (→ Growth Mindset; Verbesserung der Fähigkeiten), sowie sekundären Kontrollmechanismen (→ Fixed Mindset; z.B. abwärts gerichtete Vergleiche sowie Vermeidung oder Leugnung des „Scheiterns“) (Dweck & Nussbaum, 2008).

Des Weiteren konnte Hong et al. (1999) feststellen, dass Personen mit einem Growth Mindset Situationen der Herausforderung als deutlich kontrollierbarer empfinden als Personen mit einem Fixed Mindset. Die direkten Kontrollmechanismen (Growth Mindset) wurden in der Literatur wiederholt mit akademischem Erfolg in Form von Noten oder Prüfungsleistungen in Zusammenhang gebracht (z.B. Blackwell et al., 2007; Dweck & Yeager, 2020).

2.4.1 Messung der impliziten Intelligenz-Theorien

Ein wesentliches Charakteristikum von impliziten Intelligenz-Theorien, wie bereits thematisiert, besteht darin, dass sie implizit, also in weiten Teilen unbewusst sind (Dweck & Yeager, 2019). Naheliegender wäre es daher, diese impliziten Überzeugungen mit

Erhebungsmethoden zu erforschen, die diese Überzeugungen implizit abfragen (Sik et al., 2024). Bisher wurden implizite Intelligenz-Theorien primär durch explizite Erhebungsmethoden in Form von Selbstbeurteilungen beforscht, bei denen Teilnehmer:innen darum gebeten wurden, selbst zu beurteilen, welche Überzeugungen sie bezüglich Intelligenz haben (Dweck, 2000).

2.4.2 Implizite Messungen

Zwischen der Theorie, die Intelligenz-Mindsets als implizit charakterisiert, und der Empirie, die versucht, diese mit expliziten Methoden zu messen, besteht eine Diskrepanz. Daher wäre es adäquat, implizite Theorien mit impliziten Verfahren zu messen (Sik et al., 2024). Implizite Verfahren erfassen unbewusste, automatische kognitive Prozesse. Diese Methoden ermöglichen es beispielsweise, implizite Einstellungen wie z.B. Vorurteile, Stereotype oder auch Intelligenz-Mindsets zu messen, die unbewusst sind oder aus verschiedenen Gründen (z.B. soziale Erwünschtheit) teilweise nicht aufrichtig von Proband:innen offen angegeben werden (Greenwald et al., 2009).

Diese Verfahren beruhen auf dem Modell der semantischen Netzwerke, welches die Struktur und Organisation von menschlichem Wissen beschreibt. Das Modell beruht auf der Annahme, dass Wissen in Form von miteinander verbundenen Konzepten oder Begriffen neuronal gespeichert wird. Diese Konzepte werden als Knoten dargestellt, die über Verbindungen verschiedene Arten von Beziehungen mit anderen Knoten eingehen. Hierbei können hierarchische (z.B. Apfel – Obst), assoziative (z.B. Strand – Meer) und kausale (z.B. Rauchen – Lungenkrebs) Beziehungen eingegangen werden. Wird nun ein Konzept aktiviert, so aktiviert es indirekt über seine Verbindungen umliegende Konzepte (Collins & Loftus, 1975).

Jedoch wird zuweilen kritisiert, dass die meisten Verfahren des impliziten Testens lediglich erfassen, wie stark die Verbindungen zwischen den Konzepten sind und nicht erfasst wird, welche Form von Beziehung zwischen den Konzepten vorliegt (De Houwer et al., 2021; Hughes & Barnes-Holmes, 2011). Dieser Fokus auf Assoziationen anstelle von propositionalen Überzeugungen führt dazu, dass die Ergebnisse dieser Verfahren zu unspezifisch sind. Des Weiteren ist ihr Nutzen als Prädiktor für Verhalten gering und klassische implizite Messungen weisen nur eine geringfügig höhere inkrementelle

Validität auf als vergleichbare Erhebungsinstrumente, die mit Selbstauskünften arbeiten (Meissner et al., 2019).

2.4.3 Propositional Evaluation Paradigm (PEP)

Als Reaktion auf die Limitationen, welche klassische implizite Messungen wie beispielsweise das IAT aufweisen, wurden in den letzten Jahren vermehrt neue Verfahren entwickelt, die es nun auch erlauben, nicht nur Assoziationen, sondern auch Überzeugungen abzufragen (Meissner et al., 2019). Eines dieser Verfahren stellt das sequenzielle Priming-Verfahren Propositional Evaluation Paradigm (PEP) dar, welches im Rahmen dieser Arbeit angewandt wird und im Abschnitt 4.2.2.6 genauer beschrieben wird.

3 Forschungsfragen und Hypothesen

Im Rahmen dieser Untersuchung sollen folgende Fragestellungen untersucht werden:

FF1: Sind Lerneffekte beim dynamischen Testen ohne Pädagogisierungsphase vorfindbar?

H1: Bei der Vorgabe zweier identischer Tests sind im Posttest bessere Testergebnisse beobachtbar als im Prätest.

H2: Bei der Vorgabe zweier äquivalenter Tests sind im Posttest bessere Testergebnisse beobachtbar als im Prätest.

FF2: Inwiefern beeinflussen implizite Intelligenz-Theorien den Guthke Effekt?

H3: Je stärker ein Growth Mindset ausgeprägt ist, desto größer ist die beobachtete Verbesserung von Prä- und Posttest bei der Vorgabe identischer Tests.

H4: Je stärker ein Growth Mindset ausgeprägt ist, desto größer ist die beobachtete Verbesserung von Prä- und Posttest bei der Vorgabe äquivalenter Tests.

Das Ziel dieser Arbeit besteht darin, den Guthke-Effekt, welcher zuletzt in den 1960er-Jahren festgestellt wurde, zu replizieren. Wie in Abschnitt 2.4 ausgeführt wurde, spielen Intelligenz-Mindsets eine umfassende Rolle bei Lernprozessen und akademischen Leistungen. Da beim Guthke-Effekt Leistungsänderungen und Lernprozesse im Mittelpunkt stehen, wird vermutet, dass dieser durch implizite Intelligenz-Theorien moderiert wird. Darüber hinaus wird aufgrund positiver Auswirkungen des Growth Mindsets auf verschiedene Aspekte des Lernprozesses (Verhaltensweisen,

Zielsetzungen, Attribution, Motivation) vermutet, dass ein positiver Zusammenhang zwischen dem Ausmaß, in dem ein Growth Mindset vorhanden ist und der Größe der Lerneffekte besteht.

4 Methode

4.1 Untersuchungsdesign

Im Folgenden wird das Untersuchungsdesign vorgestellt, anhand dessen die aufgestellten Hypothesen untersucht werden sollen. Die Erhebung beinhaltet zwei Messzeitpunkte, welche mit ca. 14 Tagen Abstand erfolgen. Nach der obligatorischen Einwilligung und dem Abfragen soziodemographischer Daten durchlaufen die Teilnehmer:innen das PEP-Verfahren zur Ermittlung des Intelligenz-Mindsets. Anschließend werden sie darum gebeten, Items aus verschiedenen IQ-Tests zu bearbeiten. Die Items weisen verschiedene Schwierigkeitsgrade auf. Es ist wichtig, dass auch schwer lösbare Items vorgegeben werden. Zum einen, da Verbesserungen bzw. Lerneffekte zwischen Prä- und Posttest nur potenziell vorfindbar sind, wenn nach dem Prätest noch Verbesserungsmöglichkeiten bestehen. Zum anderen lässt sich der Einfluss von impliziten Intelligenz-Theorien auf den Guthke-Effekt nur untersuchen, wenn im Rahmen des Prätests Rückschlüsse durch ungelöste Items erfahren werden, da sich Auswirkungen von Intelligenz-Mindsets auf Lernprozesse primär in Situationen des Misserfolges beobachten lassen (Diener & Dweck, 1978, 1980).

Zum zweiten Messzeitpunkt werden den Teilnehmerinnen einerseits identische Intelligenz-Items aus dem Prätest vorgelegt und andererseits dazu äquivalente Aufgaben. Dies ermöglicht es, die möglichen Lerneffekte systematisch zu untersuchen. Mittels der Vorgabe zweier identischer Tests lässt sich untersuchen, in welchem Ausmaß im Prätest nicht gelöste Items im Posttest gelöst werden konnten. Durch die Testung mit äquivalenten Items lassen sich Transfer-Lerneffekte untersuchen, also inwiefern die im Prätest abgefragten Fähigkeiten durch diesen trainiert wurden und inwieweit dieses so erworbene Wissen auf neue ähnliche Aufgaben übertragen werden kann. Die Reihenfolge, in der die identischen oder äquivalenten Aufgaben vorgegeben werden, variiert, um so Reihenfolgeeffekte zwischen diesen beiden Item-Arten über alle Teilnehmer:innen zu eliminieren.

4.2 Erhebungsinstrument und Operationalisierung

4.2.1 Variablen

Im Rahmen der Untersuchung werden folgende Variablen erhoben: 1. Lerneffekte, 2. Ergebnisse in den IQ-Testungen im Prätest, 3. Ergebnisse in den IQ-Testungen im Posttest sowie 4. Intelligenz-Mindsets.

4.2.2 Operationalisierung

4.2.2.1 Lerneffekte

Die Variable des Lerneffekts wird aus der Differenz der erbrachten Leistungen aus den beiden Erhebungszeitpunkten (Variable 2 & 3) ermittelt, welche im Rahmen der vorgelegten IQ-Test-Items erbracht wurden. In Abschnitt 2.1.2 wurde bereits erläutert, warum es problematisch ist, Lernpotentiale bzw. Lerneffekte anhand der Verbesserung zwischen Prätest und Posttest zu operationalisieren. Die Nachteile kommen jedoch primär zu tragen, wenn die Potentiale von einzelnen Personen und deren Vergleichbarkeit von Interesse ist. Da im Rahmen dieser Erhebung nur die potenziellen Leistungssteigerung der Stichprobe untersucht wird, ist dieses Vorgehen zulässig.

Zur Testung der kognitiven Leistungsfähigkeit wird auf folgende Testungen zurückgegriffen: Raven-APM-Intelligenztest (APM), Ravens Standard Progressive Matrices (SPM), Wortschatztest-WST und Mehrfachwahl-Wortschatz-Intelligenztest (MWT-A). Diese Verfahren sind langerprobte und etablierte Erhebungsinstrumente zur Erhebung von kognitiver Leistungsfähigkeit. APM und SPM prüfen mittels Matrizen die fluide Komponente von Intelligenz. WST und MWT-A prüfen hingegen durch Prüfung des Wortschatzes die kristalline Intelligenz.

4.2.2.2 APM: Raven-APM-Intelligenztest

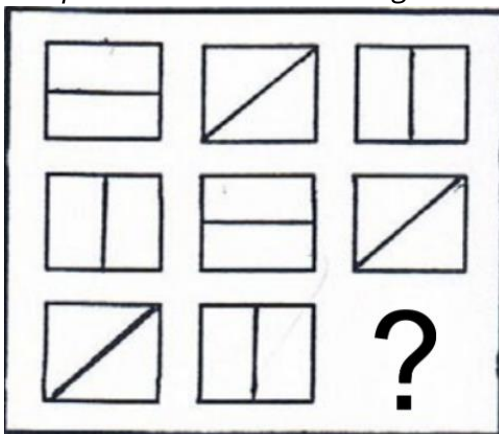
Der Raven-APM-Intelligenztest ist ein Verfahren zur Messung der allgemeinen (fluiden) Intelligenz. Er umfasst 12 (Teil 1) oder 36 (Teil 2) Matrizen, die es richtig zu ergänzen gilt (s. Abb. 1). Beide Teile können gemeinsam geprüft oder einzeln als Kurzform durchgeführt werden (Raven et al., 1998). Im Rahmen dieser Erhebung wurden hierfür insgesamt 16 Items genutzt. Acht Items bekamen alle Teilnehmenden an Testzeitpunkt 1 vorgelegt. Diese selben acht Items wurden im Posttest erneut abgefragt; zusätzlich wurden die anderen acht Items im Posttest zur Bearbeitung vorgelegt.

4.2.2.3 SPM: Raven´s Standard Progressive Matrics

Bei dem SPM handelt es sich wie beim APM ebenso um einen Matrizentest zur Erfassung der allgemeinen Intelligenz (siehe Abb. 1). Der SPM besteht aus 5 Testteilen zu jeweils 12 Items in ansteigender Schwierigkeit (Horn, 2009). Aus diesem Verfahren sollten ursprünglich 24 Items verwendet werden. Auch hier sollte die Hälfte der Items im Prätest und wiederholt im Posttest vorgegeben sowie die andere Hälfte nur im Posttest vorgelegt werden. Jedoch unterliefen Fehler beim Einarbeiten der analog vorliegenden SPM-Items in das digitale Erhebungsinstrument. Um eine Vergleichbarkeit zwischen den SPM-Itemsets hinsichtlich ihrer Anzahl und Schwierigkeit zu gewährleisten, wurden die 24 SPM-Items auf 18 gekürzt. Folglich wurden äquivalent zum Vorgehen bei den APM-Items die Hälfte der Items im Prätest sowie alle 18 Items im Posttest vorgelegt.

Abbildung 1

Beispiel für eine Matrizenaufgabe.



Anmerkung. Diese Matrizen stammt weder aus dem SPM noch aus dem APM und soll nur der Veranschaulichung einer Matrizenaufgabe dienen.

4.2.2.4 WST: Wortschatztest

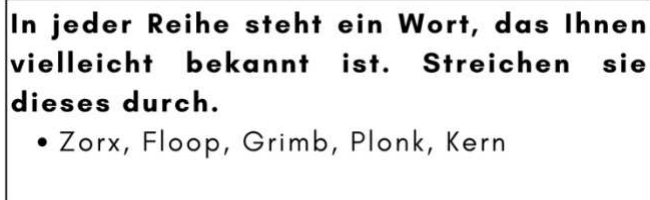
Der Wortschatztest (WST) ist ein Verfahren zur Beurteilung der verbalen Intelligenz und des Sprachverständnisses. Er beinhaltet 42 Aufgaben zur Wiedererkennung von Wörtern. Jede Aufgabe enthält hierbei Distraktoren und ein Zielwort, bei dem es sich um ein existierendes Wort handelt. Die Aufgabe besteht darin, das Zielwort zu identifizieren (Schmidt & Metzler, 1992). Ein Beispiel für eine Aufgabe eines Wortschatztests findet sich in Abb. 2.

4.2.2.5 MWT-A: Mehrfachwahl-Wortschatz-Intelligenztest

Beim Mehrfachwahl-Wortschatz-Intelligenztest (MWT-A) handelt es sich um ein Verfahren, das dem WST sehr ähnelt. Er gleicht dem WST in seiner Funktion, seiner Struktur und der Art der Aufgabenstellungen, wobei der MWT-A 37 Items beinhaltet.

Abbildung 2

Beispiel für Aufgabe eines Wortschatztests



Anmerkung. Diese Aufgabe stammt weder aus dem WST noch aus dem MWT-A und soll nur der Veranschaulichung der Aufgabenart dienen.

4.2.2.6 PEP: Propositional Evaluation Paradigm

Um die Ausprägung der impliziten Intelligenz-Theorien zu erheben, wird das von Sik et al. (2024) entwickelte Mousetracking-PEP-Verfahren verwendet. Bei dem von Sik et al. (2024) entwickelten Verfahren handelt es sich um eine adaptierte Version des von Cummins und De Houwer (2021) entwickelten PEPs. Die Änderungen, die von Sik et al. (2024) vorgenommen wurden, sind im Wesentlichen inhaltlicher Natur. So wurden andere Items genutzt, um implizite Intelligenz-Theorien zu erfassen anstelle von Einstellungen gegenüber Migrant:innen (Cummins & De Houwer, 2021). Im Folgenden wird die adaptierte Version von Sik et al. (2024) genauer erläutert und vorgestellt.

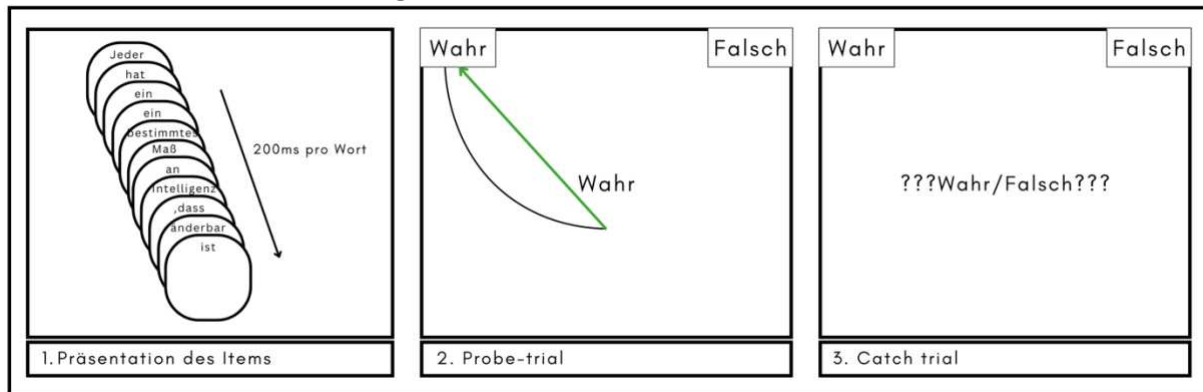
Die Grundlage für das PEP bildet die deutsche Version der Theories of Intelligence Scale (Spinath et al., 2002). Diese Skala beinhaltet drei Items mit Aussagen bezüglich der eigenen Intelligenz (z.B. Jeder hat ein bestimmtes Maß an Intelligenz, das...), die mittels fünfstufiger Likertskalen vervollständigt werden, wobei die Extreme der Skala jeweils die Intelligenz-Mindsets repräsentieren (1 = unveränderbar ist (Fixed Mindset); 5 = änderbar ist (Growth Mindset)). Um diese Items im Rahmen des PEP nutzen zu können, wurden sie von Sik et al. (2024) so adaptiert, dass aus jeder zu vervollständigen Aussage zwei Items generiert wurden, die jeweils die Intelligenz-Mindsets einzeln repräsentieren. So wurde beispielsweise aus dem zuvor angeführten Item die folgenden zwei Items generiert: „Jede:r hat ein bestimmtes Maß an Intelligenz, das änderbar ist“ (Growth Mindset) und „jede:r hat ein bestimmtes Maß an Intelligenz, das unveränderbar ist“ (Fixed Mindset). Die daraus resultierenden sechs Items werden im PEP Wort für Wort präsentiert, wobei

jedes Wort nur 200 ms angezeigt wird. Diese Items dienen als Prime Stimulus, der die impliziten Einstellungen der Teilnehmenden aktivieren soll.

Nach der Präsentation eines Items können nun zwei verschiedene Arten von Reaktionsaufgaben erfolgen, wobei den Teilnehmenden nicht bekannt ist, wann ihnen welche Aufgabe gestellt wird.

Abbildung 3

Grafische Veranschaulichung des PEP Verfahrens



Bei den sogenannten Probe Trials werden die Teilnehmenden aufgefordert, innerhalb von zwei Sekunden den Mauszeiger von der Mitte des Bildschirms in die obere linke (Wahr) oder rechte Ecke (Falsch) zu führen, abhängig davon, ob in der Mitte „Wahr“ oder „Falsch“ steht (s. Abb. 3). Die Bewegung des Zeigers wird hierbei via Mousetracking aufgezeichnet. Desto ineffizienter die Maus in das Zielfeld geführt wird, umso weniger spiegelt das zuvor präsentierte Item die impliziten Einstellungen wider. Wenn beispielsweise eine Person mit einem Growth Mindset die Aussage „Jeder hat ein bestimmtes Maß an Intelligenz, das veränderbar ist“ präsentiert bekommt und anschließend aufgefordert wird, den Mauszeiger in das „Richtig-Feld“ zu bewegen (grüne Linie; s. Abb. 3), dann wird diese Person vermutlich die Maus effizienter in das Feld bewegen als jemand mit einem Fixed Mindset (schwarze Linie; s. Abb. 3).

Neben den Probe Trials gibt es zusätzlich noch die Catch Trials. Diese sollen sicherstellen, dass die Teilnehmenden die präsentierten Aussagen genau lesen. Bei dieser Art von Reaktionsaufgabe muss die zuvor präsentierte Aussage innerhalb von 2 Sekunden als wahr oder falsch bewertet werden. Auch hier gilt es, den Mauszeiger von der Mitte des Bildschirms in das entsprechende Feld in der oberen linken oder rechten Ecke zu führen.

Insgesamt beinhaltet das PEP von Sik et al. (2024) 204 Durchgänge, in denen jeweils eine Aussage präsentiert wird und anschließend eine Reaktionsaufgabe erfolgt. Jedes der sechs verwendeten Items wird 10 Mal im Kontext eines Catch Trials und 24 Mal im Rahmen eines Probe Trials präsentiert, wobei jeweils 12 Mal aufgefordert wird, den Mauszeiger in das „Wahr-“ oder „Falsch-Feld“ zu bewegen. Die Reihenfolge der angezeigten Items erfolgt dabei zufällig.

Zur Ermittlung der Intelligenz-Mindset-Ausprägung wurde der Code, der von Cummins & De Houwer (2021) im Rahmen ihrer Untersuchungen entwickelt wurde, angewandt. Der Mindset-Score ist dabei so codiert, dass hohe Werte mit einem Growth Mindset und niedrige mit einem Fixed Mindset assoziiert sind.

Das von Sik et al. (2024) adaptierte Verfahren stellt einen innovativen Weg dar, Intelligenz-Mindsets zu erheben. Das Erhebungsinstrument erwies sich in den ersten Untersuchungen mit einer Reliabilität von .94 als äußerst zuverlässig (Sik et al., 2024).

4.3 Stichprobe

Eine Poweranalyse mittels G-Power ergab eine minimale Stichprobengröße von 258 bei folgenden Parametern: Power von .9, Alphaniveau von .05 und Effektstärke von .2. Da es derzeit nach meinem Kenntnisstand keine Studien gibt, anhand derer erwartete Effektstärken abgeleitet werden können, wurde grundsätzlich von kleinen Effekten ausgegangen. Zusätzlich wurden weitere 42 Teilnehmende eingeplant, um möglichen Dropouts entgegenzuwirken. So wurde insgesamt eine Stichprobengröße von 300 Teilnehmenden angestrebt.

Die Teilnehmenden wurden aus dem Labs-Programm der Universität Wien rekrutiert. Im Rahmen des Bachelorstudiums Psychologie sammeln Studierende Credit Points, welche sie durch das Teilnehmen an Studien wie dieser erhalten. Die Rekrutierung über dieses Programm hat vor allem praktische Gründe. Jedoch verhindert diese Form der Rekrutierung gleichzeitig auch eine repräsentative Stichprobe. Ausgeschlossen wurden Teilnehmende, die beide Teilerhebungen nicht vollständig absolviert haben.

4.4 Untersuchungsdurchführung

Die Durchführung des hier beschriebenen Studiendesigns erfolgte in Zusammenarbeit mit den Mitstudierenden Lisa Kasal und Paul Gienger, die jeweils andere inhaltliche Schwerpunkte in ihren Masterarbeiten beleuchten. Kasal beschäftigt sich in ihrem

Masterprojekt primär mit dem Phänomen des Guthke-Effekts, wohingegen Gienger dieselben Fragestellungen wie diese Arbeit untersucht, jedoch die Intelligenz-Mindsets mittels eines expliziten Verfahrens erhebt. Beide Masterarbeiten befinden sich derzeit noch in der Ausarbeitung, daher kann an dieser Stelle nicht auf sie verwiesen werden.

5 Ergebnisse

5.1 Stichprobenbeschreibung

Insgesamt wurde eine Stichprobengröße von 165 gültigen Fällen erreicht. Ausgeschlossen wurden Teilnehmende, die nicht an beiden Erhebungszeitpunkten teilnahmen oder Datensätze, die nicht eindeutig mittels der Verknüpfungsfrage zuzuordnen waren. Somit konnte die mittels Poweranalyse angestrebte Stichprobengröße von 258 nicht erreicht werden. Dies ist bei der Untersuchung des Guthke-Effekts (Unterschiedshypothesen) nicht wesentlich, jedoch potenziell für die Analyse eines Moderatoreffekts (Mindset-Hypothesen) problematisch. Die Teilnehmenden waren durchschnittlich 21.12 (*SD*: 2,96) Jahre alt und ordneten sich mehrheitlich dem weiblichen Geschlecht zu (76 % weiblich, 22 % männlich & 2 % divers/keine Angabe). 92 % der befragten Personen schätzten ihre Deutschkenntnisse auf Muttersprachenniveau ein, die restlichen 8 % als zumindest fließend. Dies ist im Kontext des PEP-Verfahrens relevant, da es nur wenig Zeit gibt, die präsentierten Aussagen zu lesen und ein Verständnis notwendig ist, um die unbewussten Überzeugungen zu aktivieren. Wie durch die Rekrutierung zu erwarten war, bestand die Stichprobe aus Bachelor-Studierenden (99 %) der Psychologie. Vier Prozent der Befragten gaben an, mit den verwendeten Verfahren der Intelligenzdiagnostik Erfahrung zu haben.

5.2 Deskriptive Ergebnisse: APM

Die Teilnehmenden erzielten in den APM-Items zu Erhebungszeitpunkt 1 von 8 möglichen Punkten durchschnittlich 6.21 (*SD*: 1.36) Punkte. Bei den identischen APM-Items an Erhebungszeitpunkt 2 wurden durchschnittlich 6.5 (*SD*: 1.42) Punkte erreicht. Bei den äquivalenten APM-Items zu Erhebungszeitpunkt 2 konnten durchschnittlich 6.06 (*SD*: 1.45) Punkte erreicht werden.

5.3 Deskriptive Ergebnisse: SPM

Bei den SPM-Items zum Erhebungszeitpunkt 1 wurden durchschnittlich 7.03 (*SD*: 1.54) von möglichen 9 Punkten erzielt. Im zweiten Erhebungszeitpunkt wurden bei den identischen Items durchschnittlich 7.13 Punkte (*SD*: 1.56) erreicht und bei den äquivalenten im Schnitt 7.21 Punkte (*SD*: 1.28).

5.4 Deskriptive Ergebnisse: WST & MWT-A

Im Rahmen der WST-Items wurde durchschnittlich 30.86 (*SD*: 3.49; 73 %) von möglichen 42 Punkten erzielt. Im zweiten Erhebungszeitpunkt zeigte sich eine durchschnittliche Leistung von 31.17 (*SD*: 3.6; 74 %) Punkten bei der wiederholten Vorgabe der Items. Beim MWT-A erzielten die Teilnehmenden durchschnittlich 24,54 (*SD*: 3.31; 66 %) von möglichen 37 Punkten.

5.5 Deskriptive Ergebnisse: PEP

Die Teilnehmenden erzielten durchschnittlich einen PEP-Wert von .12 (*SD*: .19). Der PEP-Score kann positive und negative Werte annehmen. Der PEP-Score wurde so konzipiert, dass hohe Werte für ein ausgeprägtes Growth-Mindset sprechen. Die untersuchte Stichprobe tendiert geringfügig zu einem Growth-Mindset. Neben dem impliziten PEP-Verfahren wurde im Rahmen dieses Projektes von Paul Gienger die Intelligenz-Mindsets durch Selbstauskünfte erhoben. Bei diesem expliziten Verfahren handelt es sich um eine von Sik et al. (2024) adaptierte Version von Rammstedt und Grüning (2022). Diese wiederum stellt die deutsche Version der 3 item Mindset Scale von Dweck (1999) dar. Der PEP-Score korreliert mit .41 (moderate Korrelation nach Cohen (1988)) mit dem verwendeten Verfahren.

6 Hypothesenprüfung

6.1 Vorraussetzungsprüfung der geplanten Analyseverfahren

Zur Analyse des Guthke-Effekts sollten ursprünglich t-Tests für abhängige Stichproben durchgeführt werden. Für die Anwendung eines t-Tests für abhängige Stichproben müssen abhängige Messungen vorliegen, die abhängige Variable muss mindestens

intervallskaliert sein und die Differenzen der beiden Gruppen müssen eine Normalverteilung aufweisen (Field, 2018). Die Voraussetzung der Normalverteilung ist den Daten weitestgehend nicht gegeben (s. Tab. 1). Aufgrund dessen wurde für die Analyse der äquivalenten WST-Itemsets ein t-Test für abhängige Stichproben und für die anderen Vergleiche nichtparametrische Wilcoxon-Tests durchgeführt.

Tabelle 1

Tests auf Normalverteilung

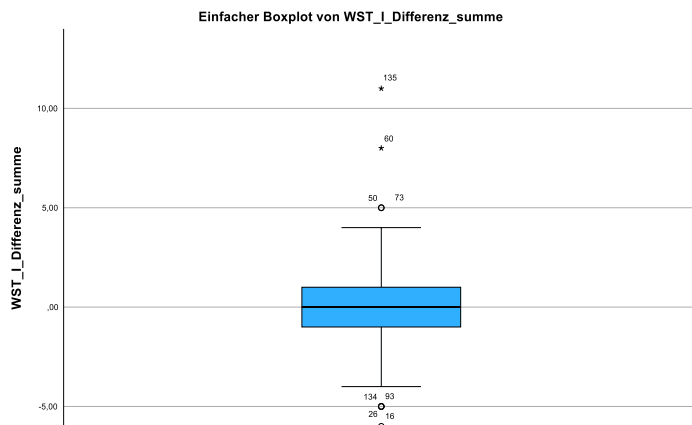
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
APM_I_Differenz_sum	.165	165	<.001	.948	165	<.001
APM_Ä_Differenz_sum	.160	165	<.001	.950	165	<.001
SPM_I_Differenz_sum	.200	165	<.001	.929	165	<.001
SPM_Ä_Differenz_sum	.140	165	<.001	.965	165	<.001
WST_I_Differenz_sum	.158	165	<.001	.905	165	<.001
WST_Ä_Differenz_mean	.039	165	.200*	.979	165	.013

Anmerkung. **APM_I_Differenz_sum** = Differenz identischer APM Items zu Erhebungszeitpunkt 1 & 2; **APM_Ä_Differenz_sum** = Differenz äquivalenter APM Items zu Erhebungszeitpunkt 1 & 2; **SPM_I_Differenz_sum** = Differenz identischer SPM Items zu Erhebungszeitpunkt 1 & 2; **SPM_Ä_Differenz_sum** = Differenz äquivalenter SPM Items zu Erhebungszeitpunkt 1 & 2; **WST_I_Differenz_sum** = Differenz identischer WST Items zu Erhebungszeitpunkt 1 & 2; **WST_Ä_Differenz_sum** = Differenz der Mittelwerte von WST- und MWT-A Items.

Zur Analyse des möglichen Moderationseffekts der Intelligenz-Mindsets auf den Guthke-Effekt wurde die Durchführung von Pearson-Produkt-Moment-Korrelationen geplant. Die Pearson-Produkt-Moment-Korrelation setzt voraus, dass metrisch skalierte Variablen, ein linearer Zusammenhang, bivariate Normalverteilung sowie keine relevanten Ausreißer vorliegen (Field, 2018). Die grafische Überprüfung der Daten ergab jedoch, dass zum Teil relevante Ausreißer auffindbar sind (s. Abb. 4) und die Voraussetzung des linearen Zusammenhangs zweifelhaft ist (s. Abb. 5 & 6). Aufgrund dessen wurden alternativ die Spearman-Korrelation zur Analyse herangezogen.

Abbildung 4

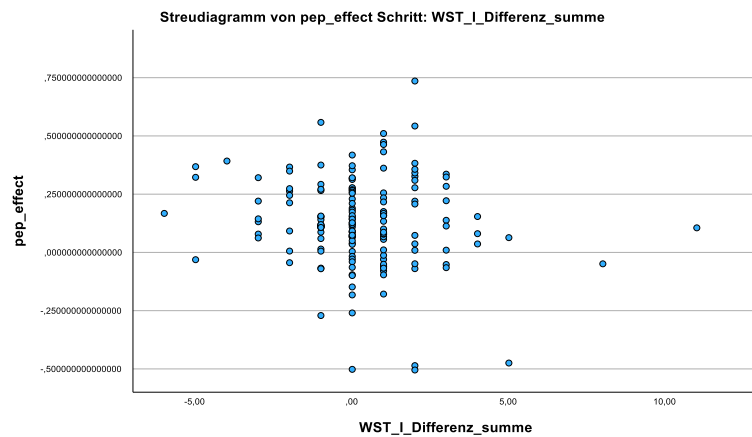
Prüfung auf relevante Ausreißer



Anmerkung. **WST_I_Differenz-summe** = Differenz der identischen Items zu Erhebungszeitpunkt 1 & 2. Ausreißer sind in Form eines Kreises markiert, extreme & relevante Ausreißer sind mit einem Stern kenntlich gemacht.

Abbildung 5

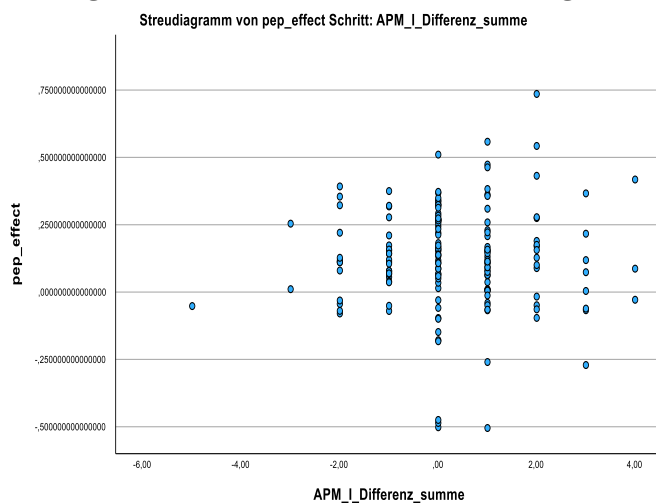
Prüfung auf Linearität des Zusammenhangs



Anmerkung. **WST_I_Differenz-summe** = Differenz der identischen WST Items zu Erhebungszeitpunkt 1 & 2.

Abbildung 6

Prüfung auf Linearität des Zusammenhangs



Anmerkung. **APM_I_Differenz-summe** = Differenz der identischen APM Items zu Erhebungszeitpunkt 1 & 2.

6.2 Guthke Effekt

Die statistische Auswertung zeigt, dass positive (s. Abschn. 6.2-6.5) und signifikante Verbesserungen bei der Vorgabe identischer Items im APM und WST (s. Tab. 2 & 3) erkennbar sind. Diese gefundenen Leistungsunterschiede weisen eine schwache Effektstärke auf (APM: .14; WST: .11).

Aufgrund dieser Ergebnisse konnte die erste Hypothese in Teilen in Bezug auf die APM- und WST-Itemsets bestätigt werden:

H1: Bei der Vorgabe zweier identischer Tests sind im Posttest bessere Testergebnisse beobachtbar als im Prätest.

Darüber hinaus waren keine signifikanten Veränderungen bei den äquivalenten Itemsets von APM und SPM auffindbar, jedoch bei den äquivalenten Itemsets des WST. Diese sind jedoch stark negativ. Die Größe des Leistungsabfalls im Vergleich zu den signifikanten Veränderungen (APM identisch und WST identisch) legt nahe, dass die WST-Items ein höheres Schwierigkeitsniveau aufweisen als die MWT-A-Items. Aufgrund dieser Ergebnisse muss folgende Hypothese vorläufig verworfen werden:

H2: Bei der Vorgabe zweier äquivalenter Tests sind im Posttest bessere Testergebnisse beobachtbar als im Prätest.

Tabelle 2

Wilcoxon-Test

	APM_I	APM_Ä	SPM_I	SPM_Ä	WST_I
Z	-2.535 ^b	-1.337 ^c	-1.073 ^b	-1.331 ^b	-2.083 ^b
Asymp. Sig. (2-seitig)	.011	.181	.283	.183	.037

Anmerkung. b. Basiert auf negativen Rängen. c. Basiert auf positiven Rängen. **APM_I** = Vergleich identischer APM Items, **APM_Ä** = Vergleich äquivalenter APM Items, **SPM_I** = Vergleich identischer SPM Items, **SPM_Ä** = Vergleich äquivalenter SPM Items & **WST_I** = Vergleich identischer WST Items.

Tabelle 3

t-Test für abhängige Stichproben

	Gepaarte Differenzen						Signifikanz		
	Mittelwert	Std.- Abweichung	Standardfehler des Mittelwertes	95% Konfidenzintervall der Differenz		T	df	Einseitiges p	Zweiseitiges p
				Unterer Wert	Oberer Wert				
WST - MWT-A	.07139	.07747	.00603	.05948	.08329	11,836	164	<.001	<.001

Anmerkung. **WST-MWT-A** = Vergleich der Mittelwerte von WST Items (Erhebungszeitpunkt 1) & MWT-A-Items

6.3 Moderationseffekt: Intelligenz Mindsets

Da ausschließlich bei der Vorgabe identischer WST- und APM-Items der Guthke-Effekt auffindbar war, konnte der mögliche Einfluss von Intelligenz-Mindsets auf den Guthke-Effekt folglich nur anhand dieser Itemsets untersucht werden. Die hierfür durchgeführten Spearman-Korrelationen konnten keine signifikanten Zusammenhänge ermitteln (s. Tab. 4). Folglich müssen folgende Hypothesen verworfen werden:

H3: Je stärker ein Growth Mindset ausgeprägt ist, desto größer ist die beobachtete Verbesserung von Prä- und Posttest bei der Vorgabe identischer Tests.

H4: Je stärker ein Growth Mindset ausgeprägt ist, desto größer ist die beobachtete Verbesserung von Prä- und Posttest bei der Vorgabe äquivalenter Tests.

Tabelle 4

Korrelationen

			APM_1	APM_I_2	APM_I_Diff	pep_effect
Spearman-Rho	APM_1	Korrelationskoeffizient	1,000	,459**	-,473**	,143
		Sig. (2-seitig)	.	<,001	<,001	,074
		N	165	165	165	158
	APM_I_2	Korrelationskoeffizient	,459**	1,000	,509**	,140
		Sig. (2-seitig)	<,001	.	<,001	,080
		N	165	165	165	158
	APM_I_Diff	Korrelationskoeffizient	-,473**	,509**	1,000	-,001
		Sig. (2-seitig)	<,001	<,001	.	,991
		N	165	165	165	158
	pep_effect	Korrelationskoeffizient	,143	,140	-,001	1,000
		Sig. (2-seitig)	,074	,080	,991	.
		N	158	158	158	158
Spearman-Rho	WST_1	Korrelationskoeffizient	1,000	,833**	-,189*	,048
		Sig. (2-seitig)	.	<,001	,015	,549
		N	165	165	165	158
	WST_I_2	Korrelationskoeffizient	,833**	1,000	,317**	-,041
		Sig. (2-seitig)	<,001	.	<,001	,606
		N	165	165	165	158
	WST_I_Differenz	Korrelationskoeffizient	-,189*	,317**	1,000	-,089
		Sig. (2-seitig)	,015	<,001	.	,268
		N	165	165	165	158
	pep_effect	Korrelationskoeffizient	,048	-,041	-,089	1,000
		Sig. (2-seitig)	,549	,606	,268	.
		N	158	158	158	158

** . Die Korrelation ist auf dem 0,01 Niveau signifikant (zweiseitig).

* . Die Korrelation ist auf dem 0,05 Niveau signifikant (zweiseitig).

Anmerkung. **APM_1** = Summenwert der APM Items zu Erhebungszeitpunkt 1, **APM_I_2** = Summenwert identischer APM Items zu Erhebungszeitpunkt 2, **APM_I_Diff** = Differenz aus APM_1 & APM_I_2, **WST_1** = Summenwert der WST Items zu Erhebungszeitpunkt 1, **WST_I_2** = Summenwert identischer WST Items zu Erhebungszeitpunkt 2, **WST_I_Diff** = Differenz aus WST_1 & WST_I_2.

6.4 Weitere Analysen

Neben der zuvor dargestellten präregistrierten Auswertung wurden weitere explorative Analysen der vorliegenden Daten vorgenommen, welche im Folgenden in Auszügen dargelegt werden.

6.4.1 Explorative Itemanalyse

Bei der Sichtung der Daten fiel auf, dass einzelne Items bereits zu Erhebungszeitpunkt 1 zu 100 % korrekt gelöst wurden. Anhand dieser Items lassen sich Lerneffekte nicht untersuchen, da sie keine Möglichkeit der Verbesserung bieten. Diese Items könnten den Guthke-Effekt hinsichtlich seiner statistischen Power geschwächt haben, was wiederum auf die Untersuchung der Mindset-Hypothesen ausgewirkt haben könnte. Vor allem beim WST (s. Tab. 5) traten diese Items vielfach auf (vs. SPM: 1 Item & APM: 0 Items).

Tabelle 5

Deskriptive Statistiken WST Scores: Erste Erhebung

	N	Minimum	Maximum	Mittelwert	Std.- Abweichung
WST_1	165	1.00	1.00	1.0000	.00000
WST_2	165	1.00	1.00	1.0000	.00000
WST_3	165	1.00	1.00	1.0000	.00000
WST_4	165	1.00	1.00	1.0000	.00000
WST_5	165	1.00	1.00	1.0000	.00000
WST_6	165	.00	1.00	.9758	.15427
WST_7	165	.00	1.00	.9273	.26048
WST_8	165	.00	1.00	.9030	.29682
WST_9	165	.00	1.00	.9818	.13402
WST_10	165	.00	1.00	.9758	.15427
WST_11	165	.00	1.00	.9636	.18776
WST_12	165	.00	1.00	.9939	.07785
WST_13	165	1.00	1.00	1.0000	.00000
WST_14	165	.00	1.00	.9576	.20217
WST_15	165	.00	1.00	.6242	.48579
WST_16	165	.00	1.00	.9758	.15427
WST_17	165	1.00	1.00	1.0000	.00000
WST_18	165	1.00	1.00	1.0000	.00000
WST_19	165	1.00	1.00	1.0000	.00000
WST_20	165	.00	1.00	.9879	.10976
WST_21	165	.00	1.00	.6061	.49011
WST_22	165	.00	1.00	.7333	.44356
WST_23	165	.00	1.00	.9939	.07785

WST_24	165	.00	1.00	.7030	.45831
WST_25	165	.00	1.00	.8000	.40122
WST_26	165	.00	1.00	.9758	.15427
WST_27	165	.00	1.00	.5697	.49663
WST_28	165	.00	1.00	.9212	.27023
WST_29	165	.00	1.00	.9697	.17194
WST_30	165	.00	1.00	.9515	.21544
WST_31	165	.00	1.00	.4667	.50041
WST_32	165	.00	1.00	.9818	.13402
WST_33	165	.00	1.00	.5455	.49945
WST_34	165	.00	1.00	.6606	.47495
WST_35	165	.00	1.00	.2182	.41427
WST_36	165	.00	1.00	.1636	.37107
WST_37	165	.00	1.00	.2364	.42614
WST_38	165	.00	.00	.0000	.00000
WST_39	165	.00	1.00	.0121	.10976
WST_40	165	.00	1.00	.0061	.07785
WST_41	165	.00	1.00	.0182	.13402
WST_42	165	.00	1.00	.0606	.23933
Gültige Werte	165				

Anmerkung. Die Zahl im Variablennamen gibt an, um welches Item es sich aus dem WST handelt.

Im Rahmen der explorativen Analyse wurden alle WST-Items, welche vollständig allen Teilnehmer:innen zu Erhebungszeitpunkt 1 gelöst werden konnten eliminiert. Dies betraf 9 Items. Anschließend wurde erneut der Zusammenhang zwischen dem Guthke-Effekt und den Intelligenz-Mindsets untersucht. Dies brachte jedoch erneut keine signifikanten Zusammenhänge zum Vorschein (s. Tab. 6)

Tabelle 6

Korrelationen

		WST_Summe_	
		Differenz_Expl	
		pep_effect	o
pep_effect	Pearson-Korrelation	1	-.068
	Sig. (2-seitig)		.399
	N	158	158
WST_Diff_Explo	Pearson-Korrelation	-.068	1
	Sig. (2-seitig)	.399	
	N	158	165

Anmerkung. **WST_Diff_Explo** = Differenz identischer Items aus Erhebungszeitpunkt 1 & 2 nach Eliminierung der Items, welche zu Zeitpunkt 1 von allen Teilnehmende gelöst werden konnten.

7 Diskussion

Diese Arbeit hat es sich zur Aufgabe gemacht, den Guthke-Effekt 60 Jahre nach seiner erstmaligen Feststellung zu replizieren und darüber hinaus zu untersuchen, ob der motivationspsychologische Faktor der Intelligenz-Mindsets zu seiner Erklärung beitragen kann.

Der Guthke-Effekt konnte in Teilen repliziert werden. Bei der Vorgabe der identischen Items von APM und WST verbesserten sich die Teilnehmenden signifikant mit niedrigen Effektstärken (APM: .14, WST: .11). Bei den identischen Items des SPM konnten keine Verbesserungen festgestellt werden. Jedoch ist hier zu betonen, dass letztendlich weniger Items zur Analyse genutzt wurden als ursprünglich präregistriert. Unter Vorgabe äquivalenter Items aus APM, SPM, WST und MWT-A konnten keine signifikanten Verbesserungen festgestellt werden.

Der Guthke-Effekt wurde bisher primär im Kontext von Aufgaben untersucht, welche fluide Fähigkeiten der Intelligenz testen (Guthke, 1972). Im Rahmen dieser Arbeit wurden mittels WST und MWT-A zusätzlich Verfahren zur Testung kristalliner Fähigkeiten vorgegeben. Auch hier zeigten sich Verbesserungen. Diese Ergebnisse legen nahe, dass der Guthke-Effekt sowohl bei der Testung von fluider als auch kristalliner Intelligenz gültig ist.

Es lässt sich diskutieren, inwiefern die beobachtbaren Verbesserungen als Lerneffekte interpretiert werden können, da unklar bleibt, worauf diese zurückzuführen sind. Dass bei den äquivalenten Items keine messbaren Verbesserungen vorzufinden sind, spricht dagegen, dass metakognitive Lernprozesse wie das Erlernen neuer Strategien im Rahmen der Testungen stattgefunden haben.

Es ließe sich argumentieren, dass die Proband:innen sich untereinander nach der ersten Testung über schwierige Items ausgetauscht haben und möglicherweise Lösungen recherchiert haben. Dies ist vor allem bei den WST-Items denkbar, da einzelne mögliche Lösungen in Form einzelner Worte leichter erinnerbar und kommunizierbar sind als Matrizenaufgaben. Allerdings sind Lösungen der Aufgaben nur durch den Erwerb der Verfahren inkl. des Manuals zugänglich. Diesem Aspekt kann daher eine untergeordnete Rolle beim Zustandekommen der Verbesserungen zugesprochen werden.

Stattdessen gibt es zahlreiche weitere mögliche Einflussfaktoren, welche eine wichtige Rolle gespielt haben können. So sind beispielsweise emotionale Faktoren zu berücksichtigen. Mehrfaches Testen führt zu Gewöhnungseffekten; die getesteten Personen sind vertrauter mit der Testsituation und es findet eine Reduktion potenzieller Testangst statt. Dies führt dazu, dass getestete Personen besser ihr Leistungspotential abrufen können (Guthke, 1972).

Neben der bewussten Beschäftigung mit den nicht gelösten Aufgaben nach der ersten Testung (z.B. das Nachschlagen von Lösungen) ist es denkbar, dass unbewusste kognitive Prozesse zu den Verbesserungen beigetragen haben. So könnten Inkubationsprozesse hier eine Rolle gespielt haben. Inkubation beschreibt den Prozess der unbewussten Beschäftigung mit einem Problem. So können alte Assoziationen und Denkmuster, die zum Scheitern geführt haben, gelöst und neue gebildet werden. Diese neuen Denkmuster können dann zu plötzlich auftretenden „Aha-Momenten“ führen und möglicherweise dazu beitragen, dass Aufgaben, welche bei der ersten Bearbeitung nicht lösbar waren, bei erneuter Bearbeitung plötzlich lösbar sind (Anderson, 1996).

Neben den bisher genannten möglichen Einflussfaktoren sind natürlich auch die motivationalen Faktoren zu diskutieren. Zentral in dieser Arbeit ist der motivationale Aspekt der Intelligenz-Mindsets. Auf Ebene des Intelligenz-Mindsets wurde ein methodisch innovativer Weg mit der Verwendung des PEP-Verfahrens beschritten. Mindsets sind weitestgehend unbewusst; naheliegend ist es daher, diese auch nicht mittels expliziter Selbstausskünfte, sondern implizit zu erheben (Sik et al., 2024). Diese Arbeit ist eine der ersten Untersuchungen, welche das von Sik et al. (2024) entwickelte PEP-Verfahren verwendet hat. Mittels dieses neuen Verfahrens konnten die Intelligenz-Mindsets als möglicher Erklärungsansatz für den Guthke-Effekt vorläufig nicht bestätigt werden. Es war kein positiver Zusammenhang zwischen Growth-Mindset und dem Grad der Verbesserung beobachtbar. Daher mussten alle in diesem Zusammenhang aufgestellten Hypothesen verworfen werden. Die Hypothesen basierten auf den Erkenntnissen zahlreicher Studien, welche nahelegen, dass Intelligenz-Mindsets einen großen Einfluss auf Lernprozesse insbesondere den Umgang mit Herausforderungen haben (s. Abschn. 2.4). Das Mindset-Framework (z.B. Dweck & Leggett, 1988) besagt, dass Personen mit einem Fixed- und Growth-Mindset unterschiedliche Ziele beim Lernen

verfolgen (Leistungs- vs. Lernziele) und aufgrund dessen entweder eher adaptiv oder maladaptiv mit Situationen des Scheiterns umgehen. Entsprechend der Theorie wurde erwartet, dass schwere Aufgaben für Personen mit einem Growth Mindset als eine Chance zum Lernen wahrgenommen werden und sich die Personen deswegen intensiver mit den Aufgaben auseinandersetzen und infolgedessen größere Lerngewinne erzielen als Personen mit einem Fixed Mindset. Diese hingegen nehmen die schweren Aufgaben als eine Gefahr für ihr Selbstbild wahr und setzen sich weniger mit diesen Aufgaben auseinander, da sie nicht ihren Fähigkeiten entsprechen. Ob der Zusammenhang zwischen Intelligenz-Mindset und Guthke-Effekt tatsächlich nicht vorliegt oder eventuell auf die Limitationen dieser Untersuchung zurückzuführen ist, kann durch zukünftige Untersuchungen geklärt werden.

Zu Beginn dieser Arbeit wurde vorgeschlagen, zukünftig dynamisches Testen ohne Pädagogisierungsphase durchzuführen und so dynamische Diagnostik in puncto Kosten und Aufwand wesentlich attraktiver für die Praxis zu machen. Diese Idee lässt sich durch die Ergebnisse dieser Arbeit nur eingeschränkt stützen, da der Guthke-Effekt nur in Teilen und nur bei den identischen Itemsets auffindbar war.

8.1 Limitationen und Ausblick

Das Studiendesign und die Durchführung dieser Arbeit brachten einige Limitationen zum Vorschein, die nun diskutiert werden sollen. Aufbauend auf diesen Erkenntnissen werden anschließend Ideen und Empfehlungen für zukünftige Forschung formuliert.

Eine wesentliche Schwäche der Erhebung liegt in ihrer Stichprobengröße. Die Poweranalyse (s. Abschn. 4.3) ergab eine Mindeststichprobengröße von 258. Die 165 gültigen Fälle, welche zur Analyse herangezogen werden konnten, waren ausreichend zur Untersuchung der Unterschiedshypothesen (Guthke-Effekt), jedoch zu gering um die Zusammenhangshypothesen (Mindsets) zu analysieren. Diese Limitation erschwert die inhaltliche Auseinandersetzung mit den Mindset-Ergebnissen erheblich. Es bleibt zu klären, ob bei ausreichender Stichprobengröße Intelligenz-Mindsets einen moderierenden Einfluss auf den Guthke-Effekt haben können oder nicht.

Hinsichtlich der SPM-Items wäre es zudem sinnvoll gewesen – wie präregistriert – mehr Items als die letzten neun pro Itemset zu erheben: Wie die explorative Analyse der MWT-A- und WST-Items ergeben hat, war das Schwierigkeitsniveau beider Verfahren nicht vergleichbar, dies machte eine Analyse des Lerneffekts bei den äquivalenten Items im Bereich kristalliner Intelligenz unmöglich. Hier wäre es sinnvoll gewesen, Verfahren zu verwenden, die vom Schwierigkeitsniveau vergleichbarer sind oder diese ebenso wie beim SPM und APM in zwei Itemsets aufzuteilen.

Der APM und SPM bestehen jeweils aus Aufgaben, die zunehmend schwerer werden. Die Erhöhung des Schwierigkeitsgrades erfolgt dabei in Etappen, sodass immer mehrere Items pro Schwierigkeitsniveau vorliegen. Bei Auswahl von Items für diese Untersuchungen wurden von jedem Niveau zwei bis drei Items ausgewählt. Die Sichtung ergab, dass einige Items im Erhebungszeitpunkt 1 bereits von (fast) allen Proband:innen gelöst werden konnten. Aufgrund dessen boten diese Items keine Chance an ihnen Lernprozesse zu beobachten. Interessanter hingegen sind Items mit mittlerem bis hohem Schwierigkeitsgrad, da sie das Potenzial bieten Lerneffekte messbar zu machen. Möglicherweise gab es ungenügend Aufgaben mit diesem Schwierigkeitsniveau, sodass die Teilnehmenden wesentlich ihre Fähigkeiten verbessern bzw. keine Lernprozesse simuliert werden konnten. Hier wäre es sicherlich interessant zu untersuchen, ob mehr mittelschwere/schwere Items dazu führen würden, dass auch bei den äquivalenten Items des Guthke-Effekts feststellbar wären. Zudem gilt es abzuklären, inwiefern sich Teilnehmende nach erster Testung bewusst mit Aufgaben beschäftigt haben oder ob diese Lerneffekte auf andere Faktoren zurückzuführen sind.

Abschließend lässt sich festhalten, im Rahmen dieses Projektes wurde ein erster Schritt unternommen, den Guthke-Effekt nach rund 60 Jahren und insbesondere sein Potential für das dynamische Testen und die Diagnostik im Allgemeinen zu untersuchen. Zwar konnte dieser nur bei der Vorgabe identischer Aufgaben ermittelt werden, jedoch sowohl für fluide sowie kristalline Fähigkeiten. Zur Erklärung des Effekts wurde der Fokus auf das Mindset-Framework von Dweck und Kolleg:innen (z.B. Dweck & Leggett, 1988) gesetzt. Der vermutete Zusammenhang zwischen dem Guthke-Effekt und dem Intelligenz-Mindset konnte jedoch nicht bestätigt werden. Aufgrund der angeführten Limitationen

und der Studienlage, welche die Bedeutung der des Intelligenz-Mindsets für Lernprozesse betont, bin ich der Meinung, dass es sich lohnt, dieser vielversprechenden Fährte zukünftig in weiteren Untersuchungen weiter nachzugehen.

8 Literaturverzeichnis

- Asendorpf, J. (1996). *Psychologie der Persönlichkeit: Grundlagen ; mit 83 Tabellen*. Springer.
- Anderson, J. R. (1996). *Kognitive Psychologie*. Spektrum.
- Baltes, P. B. & Smith, J. (1990). Weisheit und Weisheitsentwicklung: Prolegomena zu einer psychologischen Weisheitstheorie. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 22, 95–135.
- Bempechat, J., London, P. & Dweck, C. S. (1991). Children's conceptions of ability in major domains: An interview and experimental study. *Child Study Journal*, 21, 11–36.
- Blackwell, L. S., Trzesniewski, K. H. & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78(1), 246–263. <https://doi.org/10.1111/j.1467-8624.2007.00995.x>
- Börnert-Ringleb, M., & Wilbert, J. (2020). Die Vorhersage von Mathe-und Leseleistungen durch dynamisches Testen. *Lernen und Lernstörungen*. 10 (2), 102–113. <https://doi.org/10.1024/2235-0977/a000331>
- Brody, N. (1997). Intelligence, schooling, and society. *American Psychologist*, 52, 1046–1050. <https://doi.org/10.1037/0003-066X.52.9.1046>
- Burnette, J. L., O'Boyle, E. H., VanEpps, E. M., Pollack, J. M. & Finkel, E. J. (2013). Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological Bulletin*, 139(3), 655–701. <https://doi.org/10.1037/a0029531>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth and action*. North-Holland.

- Collins, A. M. & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407-428. <https://doi.org/10.1037/0033-295X.82.6.407>
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2. Aufl.). Lawrence Erlbaum Associates.
- Cummins, J. & De Houwer, J. (2021). The shape of belief: developing a mousetracking-based relational implicit measure. *Social Psychological and Personality Science*, 12(8), 1517–1526. <https://doi.org/10.1177/1948550620978019>
- Cury, F., Elliot, A. J., Da Fonseca, D. & Moller, A. C. (2006). The social-cognitive model of achievement motivation and the 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, 90(4), 666–679. <https://doi.org/10.1037/0022-3514.90.4.666>
- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13–21. <https://doi.org/10.1016/j.intell.2006.01.002>
- Dalayga, B., & Baskaran, S. (2019). Talent development practices: Does it really matter? *International Journal of Academic Research in Business and Social Sciences*, 9(6), 896–906. <https://doi.org/10.6007/IJARBSS/v9-i6/6108>
- De Houwer, J., Van Dessel, P. & Moran, T. (2021). Attitudes as propositional representations. *Trends in Cognitive Sciences*, 25(10), 870–882. <https://doi.org/10.1016/j.tics.2021.07.003>
- Diener, C. I. & Dweck, C. S. (1978). An analysis of learned helplessness: Continuous changes in performance, strategy, and achievement cognitions following failure. *Journal of Personality and Social Psychology*, 36(5), 451–462. <https://doi.org/10.1037/0022-3514.36.5.451>

- Diener, C. I. & Dweck, C. S. (1980). An analysis of learned helplessness: II. The processing of success. *Journal of Personality and Social Psychology*, 39(5), 940–952. doi: 10.1037/0022-3514.39.5.940
- Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Psychology Press.
- Dweck, C. S. & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95(2), 256–273. <https://doi.org/10.1037/0033-295X.95.2.256>
- Dweck, C. S. & Nussbaum, A. D. (2008). Defensiveness versus remediation: Self-theories and modes of self-esteem maintenance. *Personality and Social Psychology Bulletin*, 34(5), 599–612. <https://doi.org/10.1177/0146167207312960>
- Dweck, C. S. & Yeager, D. S. (2019). Mindsets: A view from two eras. *Perspectives on Psychological Science*, 14(3), 481–496. <https://doi.org/10.1177/1745691618804166>
- Eder, K.(2018).*Intelligenz: Grundlagen und Anwendungen*. Springer.
- Ehlert, H. (2021). Diagnostik: Den Einzelnen erkennen. In *Dynamic Assessment* (1–6). Springer Fachmedien Wiesbaden GmbH. https://doi.org/10.1007/978-3-658-34552-5_1
- Elliott, E. S. & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54(1), 5–12. <https://doi.org/10.1037/0022-3514.54.1.5>
- Farell, E., & Dweck C.S. (1985). *The role of motivational processes in transfer of learning*. Harvard University.
- Field, A. (2024). *Discovering statistics using IBM SPSS statistics*. Sage publications limited.
- Gittler, G. & Pitschnig, J .(2019). *Differentielle Psychologie* (5.). Facultas.

- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L. & Banaji, M. R. (2009). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17–41. <https://doi.org/10.1037/a0015575>
- Guthke, J. (1972). *Zur Diagnostik der intellektuellen Lernfähigkeit*. Deutscher Verlag der Wissenschaften.
- Guthke, J., Beckmann, J. F. & Wiedl, K. H. (2003). Dynamik im dynamischen Testen. *Psychologische Rundschau*, 54 (4), 225–232. <https://doi.org/10.1026//0033-3042.54.4.225>
- Guthke, J., & Heller, K. A. (1995). Dynamische Diagnostik: Grundlagen und Anwendungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 27(4), 241–252.
- Haimovitz, K. & Dweck, C. S. (2017). The origins of children’s growth and fixed mindsets: New research and a new proposal. *Child Development*, 88(6), 1849–1859. <https://doi.org/10.1111/cdev.12955>
- Holling, H., Preckel, F. & Vock, M. (2004). *Intelligenzdiagnostik*. Hogrefe, Verl. für Psychologie.
- Hassenstein, B. (1992). *Klugheit: Zur Natur unserer geistigen Fähigkeiten*. Piper.
- Haywood, H.C. & Tzuriel, D (2002) Applications and challenges in dynamic assessment. *Peabody Journal of Education*, 77:2, 40-63. https://doi.org/10.1207/S15327930PJE7702_5
- Hertel, S. & Karlen, Y. (2021). Machen Überzeugungen den Unterschied?: Implizite Theorien von Lernenden in Schule und Hochschule. *Unterrichtswissenschaft*, 49(4), 491–501. <https://doi.org/10.1007/s42010-021-00135-6>

- Hong, Y., Chiu, C. & Dweck, C. S. (1999). Implicit theories, attributions, and coping: A meaning system approach. *Journal of Personality and Social Psychology*, 77(3), 588–599. <https://doi.org/10.1037/0022-3514.77.3.588>
- Horn, R. (2009). *Raven's Standard Progressive Matrices*. Pearson.
- Kaufman, A. S., & Kaufman, N. L. (2001). *Intelligence testing with the WISC-V*. Wiley.
- Hughes, S., Barnes-Holmes, D. & De Houwer, J. (2011). The dominance of associative theorizing in implicit attitude research: Propositional and behavioral alternatives. *The Psychological Record*, 61(3), 465–496. <https://doi.org/10.1007/BF03395772>
- Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences*, 42, 441–451. <https://doi.org/10.1016/j.paid.2006.08.013>
- Lange, K. W., & Schmitt, M. (2019). *Psychologische Diagnostik: Ein Lehrbuch für Studium und Praxis*. Springer.
- Leggett, E. L. (1986). *Individual differences in effort/ability inference rules and goals: Implications for causal judgments (motivation, personality, helplessness)*. ProQuest Dissertations Publishing.
- Lidz, C. (2002). Mediated learning experience (MLE) as a basis for an alternative approach to assessment. *School Psychology International*, 23, 68-84. <https://doi.org/10.1177/0143034302023001731>
- Lohaus, A., Vierhaus, M., & Maass, A. (2010). *Entwicklungspsychologie des Kindes- und Jugendalters für Bachelor*. Springer.
- Meissner, F., Grigutsch, L. A., Koranyi, N., Müller, F. & Rothermund, K. (2019). Predicting behavior with implicit measures: Disillusioning findings, reasonable explanations, and

sophisticated solutions. *Frontiers in Psychology*, 10, 2483- 2483.

<https://doi.org/10.3389/fpsyg.2019.02483>

Myers, D.G.(2008). *Psychologie*. Springer.

Plomin, R., & DeFries, J. C. (2013). Genetics and intelligence differences: Five special findings. *Molecular Psychiatry*, 18(1), 1–10.

Raven, J. C., Raven, J. & Bulheller, S. (1998). *Raven's Progressive Matrices and Vocabulary Scales – Manual Teil 4: Advanced Progressive Matrices*. Swets Test Services.

Robins, R. W. & Pals, J. L. (2002). Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self and Identity*, 1(4), 313–336. <https://doi.org.1080/15298860290106805>

Röder, M., Guderjahn, L., Fingerle, M., & Büttner, G. (2021). Diagnostik im Kontext von Fluchterfahrungen: Eine kritische Betrachtung des dynamischen Testens als Möglichkeit zur Verbesserung von Bildungsteilhabechancen auf Basis eines Rapid Reviews der Literatur. *Empirische Sonderpädagogik*, 13(1), 21–33.
<https://doi.org/10.1007/s12345-021-00456-7>

Schmidt-Atzert, L., & Kuhl, J.(2019). *Intelligenz: Psychologische Grundlagen und Anwendungen*. Hogrefe.

Schmidt, K. & Metzler, P. (1992). *Wortschatztest (WST)*. Verlag Hans Huber.

Spearman, C. (1927). *The abilities of men: Their nature and measurement*. Macmillan.

Spinath, B., Stiensmeier-Pelster, J., Schöne, C., & Dickhäuser, O. (2002). *Skalen zur Erfassung der Lern- und Leistungsmotivation: SELLMO [Measurement scales for learning and performance motivation]*. Hogrefe.

- Sik, K., Cummins, J. & Job, V. (2024). An implicit measure of growth mindset uniquely predicts post-failure learning behavior. *Scientific Reports*, 14(1), 3761.
<https://doi.org/10.1038/s41598-024-52916-5>
- Spearman, C. (1928). The Abilities of Man. *Science (American Association for the Advancement of Science)*, 68(1750), 38–38.
<https://doi.org/10.1126/science.68.1750.38-b>
- Thurstone, L. L. (1938). *Primary and mental abilities*. University of Chicago Press.
- Tzuriel, D. (2001). Dynamic assessment of learning potential. In Donald, H. S., Henry, L. J. & Jac J. W. A. (2001), *Handbook of psychoeducational assessment* (451-512). Academic Press. <https://doi.org/10.1007/978-94-007-4507-0>
- UNICEF (1989). *Convention on the Rights of the Child*. <https://www.unicef.org/child-rights-convention/convention-text>
- Wechsler, D., Hardesty, A., & Riegel, K. F. (1956). *Die Messung der Intelligenz Erwachsener : Textband zum Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE)*. Huber.
- Wothly, D. (1969). *Die psychodiagnostische Relevanz von Tests mit Pädagogisierungsphase in oberen Klassen der Hilfsschule*. Unveröffentlichte Diplomarbeit.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), 89-100. <https://doi.org/10.1111/j.1469-7610.1976.tb00381.x>
- Wygotski, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard university press.

Yeager, D. S. & Dweck, C. S. (2020). What can be learned from growth mindset

controversies? *American Psychologist*, 75(9), 1269–1284.

<https://doi.org/10.1037/amp0000794>

Zaaiman, H., Van Der Flier, H. & Thijs, G. D. (2001). Dynamic testing in selection for an

educational programme: Assessing south african performance on the raven progressive matrices. *International Journal of Selection and Assessment*, 9(3), 258–269.

<https://doi.org/10.1111/1468-2389.00178>

9 Anhang

9.1 Abstract

Die vorliegende Arbeit verfolgt das Ziel den Guthke-Effekt erstmals seit seiner Entdeckung zu replizieren und zu seiner Erklärung beizutragen. Der Guthke-Effekt besagt, dass Lernfortschritte beim dynamischen Testen durch die reine Testwiederholung auch ohne gezielte Förderung beobachtbar sind. Als möglicher Erklärungsansatz wurden Intelligenz-Mindsets, weitestgehend unbewusste Überzeugungen der eigenen Intelligenz, welche einen wesentlichen Einfluss auf Lernprozesse und insbesondere auf den Umgang mit Herausforderungen haben herangezogen. Es wurde ein Studiendesign mit zwei Messzeitpunkten, mithilfe derer die Lernprozesse von 165 Studierenden untersucht wurden, durchgeführt. Die Studierenden absolvierten zu beiden Testzeitpunkten Aufgaben aus verschiedenen Intelligenztestungen. Zudem wurden die Intelligenz-Mindsets mithilfe des Propositional Evaluation Paradigm (PEP) erhoben. Der Guthke-Effekt konnte in Teilen repliziert werden. Dies betraf die Testung fluider und kristalliner Fähigkeiten und stellt eine neue Erkenntnis dar. Der vermutete Zusammenhang zwischen dem Guthke-Effekt und den Intelligenz-Mindsets war nicht vorfindbar. Ungeklärt bleibt, ob dies auf die Limitationen der Untersuchung zurückzuführen ist. So konnte beispielsweise die angestrebte Stichprobengröße nicht erreicht werden.

The objective of this study is twofold: firstly, to replicate the Guthke effect for the first time since its discovery; and secondly, to contribute to its explanation. The Guthke effect states that learning progress can be observed in dynamic testing through test repetition without targeted support. Intelligence mindsets, which are largely unconscious beliefs about one's own intelligence, were used as a possible explanation. These mindsets have a significant influence on learning processes and, in particular, on dealing with challenges. A study design incorporating two measurement points was utilised to analyse the learning processes of 165 students. The students completed tasks from various intelligence tests at both test times. Furthermore, intelligence mindsets were assessed using the Propositional Evaluation Paradigm (PEP). The Guthke effect could be replicated in parts. This study focused on the assessment of fluid and crystallised intelligence,

thereby providing novel insights into the field. The hypothesis that there is a connection between the Guthke effect and intelligence mindsets was not supported by the evidence. The question of whether these results are connected to the limitations of the study remains unresolved. For instance, the desired sample size could not be attained.

9.2 Nutzung künstlicher Intelligenz

In der vorliegenden Masterarbeit wurden verschiedene KI-Tools unterstützend bei der Recherche und Textgestaltung eingesetzt. Im Sinne der Transparenz wird im Folgenden ausgeführt, welches Tool zu welchem Zweck eingesetzt wurde. Für die Recherche relevanter Literatur wurde ergänzend das Tool Elicit genutzt. Zur Überprüfung und Korrektur von Grammatik- und Rechtschreibfehlern kamen die Plattform [rechtschreibpruefung24.de](https://www.rechtschreibpruefung24.de) sowie ChatGPT zum Einsatz. Diese Tools wurden genutzt, um potentielle Fehler zu identifizieren und Formulierungsvorschläge zu generieren, um einzelne Abschnitte noch kompakter und verständlicher zu gestalten. Für die Übersetzung des Abstracts ins Englische wurde das KI-basierte Übersetzungstool DeepL verwendet.