



universität
wien

DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Über den diagnostischen Mehrwert von Antwortlatenzen
bei kognitiven Leistungstests“

verfasst von / submitted by

Joseph-Sebastian Steinlechner

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, 2016 / Vienna, 2016

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

A 298

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Psychologie

Betreut von / Supervisor:

Univ.-Prof. i.R. Mag. Dr. Klaus Kubinger

Danksagung

An erster Stelle möchte ich Prof. Klaus D. Kubinger und Mag. Jan Steinfeld für ihre Betreuung im Rahmen dieser Diplomarbeit meinen Dank aussprechen. Ohne die Unterstützung und angeregten Diskussionen wäre es mir wohl nicht möglich gewesen, diese Arbeit fertig zu stellen.

Ebenso möchte ich mich natürlich bei den zahlreichen Studenten bedanken, die ihre Zeit während den Testungen zur Verfügung gestellt haben. Eine Diplomarbeit kostet oftmals viel Zeit und Nerven, weshalb sich in dieser Zeit Freunde und Familie als besonders wertvolle Ressource herausgestellt haben, sei es beim Korrekturlesen oder beim Lindern der Verzweiflung – danke Patrica, Linda, Robert, Fabian, Gerhard, Dani, Sebi und Paul.

Abstract

Besonders in den letzten Jahren erfreuen sich computerisierte Leistungsverfahren zunehmender Beliebtheit, wofür es eine Vielzahl an Gründen gibt; abgesehen von der Einfachheit, mit der diese Verfahren unter standardisierten Bedingungen vorgegeben werden können, gibt es Variablen, die durch diese Vorgabemodalitäten einfach miterfasst werden können. Diese waren Testleiter_innen zuvor nur schwer oder sehr umständlich zugänglich.

Die vorliegende Arbeit widmet sich der Bearbeitungszeit einzelner Items eines kognitiven Leistungsverfahrens und der Frage, ob ein diagnostischer Mehrwert erzielt werden kann, wenn dieser besondere Aufmerksamkeit geschenkt wird.

Um dies zu überprüfen, wurden 211 Student_innen der Universität Wien untersucht. Es wurden logistische Regressionen durchgeführt, in der Hoffnung, den Erfolg von Untersuchungsteilnehmer_innen bei einzelnen Items durch Inklusion der Zeitkomponente (zusätzlich zum Fähigkeitsparameter der Teilnehmer_in) befriedigend vorhersagen zu können. Während gewisse Tendenzen, die im Rahmen dieser Arbeit auftraten (wie etwa jene, dass leistungsfähige Personen sich mehr Zeit für einzelne Items nehmen als weniger leistungsstarke), durchaus von Interesse für den Fachbereich der Psychologischen Diagnostik sein könnten, ergaben sich nicht die erhofften Ergebnisse – die Analyse der Bearbeitungszeit einzelner Items barg keinen diagnostischen Mehrwert hinsichtlich des Erfolgs bei der Bearbeitung.

Der Autor kommt in dieser Arbeit zu dem Schluss, dass die Interaktionen von Leistungsfähigkeit, Schwierigkeitsparameter und Bearbeitungszeit des Items nicht linearer Natur sein dürften, weshalb es komplexerer mathematischer Modelle bedarf, um diese in Bezug auf den Bearbeitungserfolg in Einklang zu bringen. Weitere Untersuchungen sind notwendig, um in dieser Hinsicht Klarheit zu schaffen.

Abstract

In psychological diagnostics, computerized testing, whether of personality or of aptitude, has significantly gained in popularity in recent times. There are several reasons for this, such as the simplicity with which standardized testing surroundings can be generated; also, certain variables that were previously difficult to assess or measure are now just a mouse click away for the test administrator.

This thesis focuses on the response latency of single items in a test of cognitive aptitude and aims to answer the question, whether additional diagnostic value can be achieved, by paying closer attention to the time individuals spend on single items.

In order to achieve this, 211 students of the University of Vienna were tested and logistic regressions were performed to ascertain whether the inclusion of a time component (in addition to a subject's ability and the difficulty of the item) can lead to satisfactory predictions in terms of a subject's success on single items.

While certain tendencies could be observed — for example, subjects with higher ability parameters were more likely to spend more time on single items than less skilled subjects — there seemed to be no added value in the inclusion of a time component in order to predict a subjects' success on single items.

There might be several reasons for this, the author concludes that the interactions between the subjects' ability, the difficulty of the item and time spent while working on the item are not linear in nature, which is why more complex statistical models are needed in order to clarify the relations between the aforementioned variables.

Inhalt

I.	Einleitung	6
II.	Theoretischer Teil	9
1.	Latenzzeit bei Leistungstests.....	11
1.1.	Das Hicksche Gesetz.....	12
1.2.	Das Falsch > Richtig-Phänomen.....	15
1.3.	Modelle der Bearbeitungszeit: Linear Ballistic Accumulator.....	16
III.	Empirischer Teil	18
2.	Ziel der Untersuchung (Hypothesen)	18
3.	Methode.....	19
3.1.	Erhebungsinstrumente.....	19
3.1.1.	Intelligenz-Struktur-Batterie.....	19
3.1.2.	Subtest Numerisch-induktives Denken.....	22
3.1.3.	Subtest Allgemeinwissen.....	23
3.1.4.	Subtest Figural-induktives Denken.....	24
3.1.5.	Subtest Verbale Flüssigkeit.....	25
3.2.	Stichprobe	26
4.	Ergebnisse	27
4.1.	Hypothese 1	27
4.2.	Hypothese 2	29
5.	Diskussion	33
6.	Zusammenfassung.....	40
7.	Literaturverzeichnis.....	42
8.	Anhang	46
	Lebenslauf und Abstract.....	46

I. Einleitung

Die Vorgabe von Leistungsverfahren am Computer gehört in Zeiten wie diesen ohne Frage zur gängigen Praxis in der Psychologie. Es ist auch nicht schwer zu verstehen, warum dies der Fall ist; das vormals oft mühsame Auswerten jeder einzelnen Testperson geschieht mittlerweile komplett automatisch, was der Testleiter_in eine Menge an Zeit und Aufwand ersparen kann. Für die Testteilnehmer_in ist es natürlich auch mit Vorteilen verbunden, immerhin kann das Ergebnis einer Untersuchung hinsichtlich der Leistungsfähigkeit einer Person unmittelbar im Anschluss an diese mitgeteilt werden. Ganz abgesehen von der Effizienz im zeitlichen Sinne gibt es natürlich auch Verfahren, die ohne computergestützter Vorgabe nicht realisierbar wären, wie zum Beispiel der objektive Persönlichkeitstest Arbeitshaltungen (Kubinger & Ebenhö, 2007), um nur ein Beispiel zu nennen.

Wenn, wie in dieser Arbeit, Verfahren zur Messung der Intelligenz eingesetzt werden, fällt es leicht, seinen Weg zu einem meiner persönlichen Lieblingszitate zu finden. Boring (1923) stellte folgende weitläufig bekannte und kritische Definition von Intelligenz auf: „Intelligenz ist das, was ein Intelligenztest misst.“

In dieser Diplomarbeit werde ich mich nicht mit dem ersten Teil des Zitats beschäftigen, vielmehr wird der Frage nachgegangen, ob ein Intelligenztest außer Intelligenz auch in Bezug auf die Bearbeitungszeit der Items etwas messen kann. Man stelle sich zum Beispiel Folgendes vor:

zwei Personen bearbeiten einen identischen Intelligenztest und der Einfachheit halber sagen wir, dass beide jeweils dieselben Aufgaben lösen konnten, was zwangsläufig dazu führt, dass beiden Teilnehmern dieselben Fähigkeitsparameter zugeschrieben werden, was im Endeffekt nichts anderes bedeutet, als dass beiden Personen derselbe Fähigkeitsparameter attestiert wird. Da die konventionelle Auswertung von Verfahren zur Messung der Intelligenz sich oftmals darauf beschränkt, die Fähigkeitsparameter der einzelnen Teilnehmer auszurechnen und mit einer Referenzstichprobe zu vergleichen, um das Intelligenzniveau in Relation zu einer gewählten Testpopulation zu bestimmen, wäre es bei konventioneller Auswertung nicht möglich, die beiden Testpersonen aus dem Gedankenexperiment zu unterscheiden.

Jetzt stelle man sich aber vor, dass eine der beiden Personen trotz derselben Intelligenzwerte für jede Aufgabe mehr als dreimal so lang braucht, wie die andere Person. Denkbar wäre auch, dass eine Testperson sich für eine einzelne Aufgabe nie länger als 20 Sekunden Zeit nimmt,

egal ob sie es schafft, die Aufgabe zu lösen oder nicht, während die andere Person mehrmals mehrere Minuten für eine Aufgabe benötigt.

Es drängt sich die Frage auf, ob Fälle wie der oben beschriebene in der Praxis auftreten, und ferner, ob sich Muster zeigen. Zum Beispiel wäre es von Interesse, ob besonders leistungsfähige Personen einen höheren Anspruch an die eigene Leistung haben und dadurch länger an Aufgaben verharren, deren Lösung sie nicht kennen, da für sie das Scheitern bei Aufgaben nur schwer mit dem Selbstwert zu vereinbaren ist.

In dieser Arbeit wird also der Frage nachgegangen, ob das Berücksichtigen der Zeitkomponente bei der Beantwortung einzelner Items einen diagnostischen Mehrwert liefert, oder ob der Erkenntnisgewinn vernachlässigbar ist und den zusätzlichen Aufwand nicht rechtfertigt.

Somit sind wir auch schon bei der Relevanz dieser Arbeit. Sollte sich herausstellen, dass durch zusätzliche Beachtung des Zeitverhaltens von Testpersonen ein diagnostischer Mehrwert erreicht werden kann, wäre es durchaus angebracht, diesem bei der Erstellung von Verfahren ebenfalls gezielt Beachtung zu schenken.

Wie sehr sich eine Testperson während der Testung anstrengt, ist für die Testleiter_in kaum ersichtlich. Wenn sich eine Testperson aber freiwillig dazu bereit erklärt, überdurchschnittlich viel von einer Ressource (in diesem Fall Zeit) in einzelne Aufgaben zu investieren, könnte man zu dem Schluss kommen, dass diese Teilnehmer_in überdurchschnittlich motiviert ist, in diesem Test eine gute Leistung zu erzielen. Wo diese Motivation herrührt, mag sich von Fall zu Fall unterscheiden; Need for Cognition und Persistenz sind nur zwei von vielen psychologischen Konstrukten, die hier eine Rolle spielen könnten. Genauso kann es aber auch daran liegen, dass die Testperson dem Resultat der Testung einen subjektiv hohen Wert zuspricht, wie zum Beispiel bei leistungsdiagnostischen Auswahlverfahren.

Es wird wohl jedem im Personalmanagement schon einmal das gängige Mantra „hire for attitude – train for skill“ untergekommen sein. Wenn man nun die motiviertesten Mitarbeiter sucht, die vor Anstrengung und Problemen nicht zurückscheuen, könnte die Bearbeitungszeit einzelner, besonders schwieriger Items durchaus dabei helfen, diese Auswahl zu treffen.

Die zugrundeliegende Überlegung lautet also, dass während eines psychologischen Leistungsverfahrens nicht nur Fähigkeitsparameter erfasst werden können, sondern auch Persönlichkeitsdimensionen über den Umweg der Bearbeitungszeit.

Wie die Leser_innen aufgrund der Zielsetzung dieser Arbeit wohl vermuten, handelt es sich um eine explorative Forschungsarbeit. Ich werde mich in erster Linie der grundlegenden Frage widmen, ob das Berücksichtigen der Zeitkomponente bei einem psychologischen Test dazu beitragen kann, dessen Qualität zu steigern. Genauer gesagt wird in dieser Arbeit der Frage nachgegangen, ob man durch Berücksichtigung der Zeitkomponente von einzelnen Testitems, in Kombination mit dem Fähigkeitsparameter der untersuchten Person, bzw. dem entsprechenden Schwierigkeitsparameter des Items, die Leistungsgüte befriedigend vorhergesagt kann.

Die genaue Vorgehensweise wird im Methodik-Teil dieser Arbeit erläutert, vorher möchte ich mich aber noch der Geschichte der Bearbeitungszeit in der Psychologie und auch den bisherigen Untersuchungen, die diesbezüglich unternommen wurden, widmen.

II. Theoretischer Teil

Das Untersuchen der Bearbeitungszeit unterschiedlicher Items bei Leistungstests ist schon seit dem Ende des 19. Jahrhunderts von Interesse für die psychologische Fachwelt. Galton (1883) vermutete, dass es Zusammenhänge zwischen der Geschwindigkeit der Informationsverarbeitung und der Intelligenz gäbe, konnte aber trotz seiner Bemühungen keinen befriedigenden Erkenntnisgewinn auf diesem Gebiet erzielen.

Trotzdem waren seine Untersuchungen wesentliche Grundbausteine für die Entstehung des sogenannten mental-speed-Ansatzes, der auch in der gegenwärtigen Forschung noch eine Rolle spielt (z.B. Preckel, Wener & Spinath, 2011).

Etliche Jahre nach Galtons ersten Ansätzen waren es Peak und Boring (1926), die aufgrund ihrer Untersuchungen zum Zusammenhang zwischen kognitiver Leistungsfähigkeit und Reaktionszeit zu folgendem Schluss kamen: "speed of reaction is an important, and probably the most important, factor in individual differences in the intelligent act." (Peak & Boring, 1926, p. 92).

Bei einer Untersuchung an fünf Versuchspersonen zeigte sich Peak und Boring (1926) eine hohe Korrelation zwischen Reaktionszeit und Intelligenz ($r = 0,7$ bis $1,0$). Hier ist die sehr kleine Stichprobe natürlich als Kritikpunkt hervorzuheben.

Untersuchungen von Lemon (1928) bargen bei einem größeren Stichprobenumfang ($n = 100$) wesentlich geringere Zusammenhänge zwischen Intelligenzmaßen und der Reaktionszeit ($r = 0,14$ bis $0,53$) bei elementarkognitiven Aufgaben.

Knapp 10 Jahre später leistete Thurstone (1937) mit seinem Modell der „ability surface“ einen wesentlichen Beitrag zum Verständnis der Relation zwischen „response time“ (für die Lösung der Aufgabe zur Verfügung stehende Zeit), „task difficulty“ (Schwierigkeitsparameter eines Items) und der „probability of success“ (Erfolgswahrscheinlichkeit).

Wie sich in *Abbildung 1* erkennen lässt und bereits vorhin angedeutet wurde, ist die Erfolgswahrscheinlichkeit in diesem Modell von der Fähigkeit der getesteten Person, der Schwierigkeit des zu bearbeitenden Items und der zur Verfügung stehenden Zeit zur Aufgabenbewältigung abhängig. Ebenso lässt sich erkennen, dass Items, die für eine Testperson besonders schwierig sind, selbst dann nicht gelöst werden können, wenn komplett von einem Zeitlimit abgesehen wird.

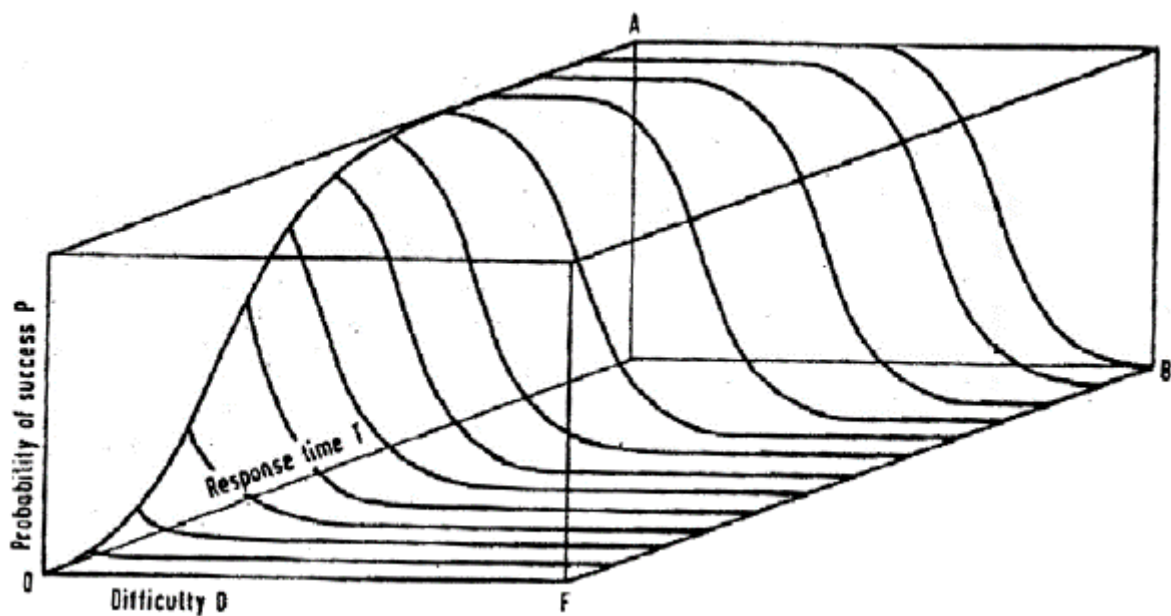


Abbildung 1: Thurstones Modell der "ability surface" (Thurstone, 1937, p. 250)

Insbesondere die letzte Erkenntnis ist für die vorliegende Arbeit von besonderer Bedeutung, da ich mich in weiterer Folge ausschließlich mit Untertests von leistungsdagnostischen Verfahren auseinandersetzen werde, die ohne zeitliche Begrenzung vorgegeben wurde. Ohne an dieser Stelle ins Detail zu gehen, gibt es bei dieser Art von Verfahren kein Abbruchkriterium seitens des Verfahrens, die Bearbeitungszeit jedes einzelnen Items ist also einzig und allein vom Verhalten der getesteten Person abhängig.

Weil die Testteilnehmer_in bei jedem Item selbst entscheiden muss, wann der Zeitpunkt erreicht ist, wo ein weiteres Investieren von Ressourcen (in diesem Fall kognitive und zeitliche) keine Erfolgsaussicht mit sich bringt, gibt es keine offensichtlichen externen Faktoren, die zu Deckeneffekten beitragen. Dementsprechend wäre es wenig verwunderlich, wenn gewisse Persönlichkeitsdimensionen, wie etwa Persistenz, einen Einfluss auf das Zeitverhalten von Testteilnehmer_innen ausüben – siehe dazu *Abschnitt 1.2*.

1. Latenzzeit bei Leistungstests

Die Latenzzeit ist bei kognitiven Leistungstests jene Zeit, die eine Testperson mit einem einzelnen Item verbringt (Hornke, 1997; Rammsayer, 1999; Beckmann, 2000). Die Latenzzeit beginnt bei Computerdiagnostischen Verfahren mit dem Aufscheinen der Aufgabenstellung am Bildschirm und endet damit, dass sich die Testperson für eine Antwortmöglichkeit entscheidet, und diese bei offenem Antwortformat eingibt bzw. bei Multiple-Choice-Antwortformat auswählt und bestätigt.

Aus dieser Definition lässt sich schlussfolgern, dass die Latenzzeit für Falschantworten sich von jener für gelösten Items unterscheiden müsste.

Bei gelösten Items ist davon auszugehen, dass die Testperson nach dem Finden einer passenden Antwortmöglichkeit nicht nach weiteren Lösungen suchen wird (s. *Abschnitt 1.3*) das Abbruchkriterium ist das Finden der Lösung. Bei ungelösten Items verhält es sich zwar eine gewisse Zeit lang ähnlich, da ebenfalls zuerst die Aufgabenstellung verstanden und Lösungsalgorithmen generiert werden müssen, allerdings dient das Finden einer geeigneten Lösung nicht als Abbruchkriterium. Dementsprechend ist es vorstellbar, dass die Person, welche die Items bearbeitet hat, sich entweder davon überzeugt hat, dass eine ungeeignete Antwort die Richtige war, oder dass die Person aufgegeben hat. Wenn eine Person aufgibt, hat sie bei den klassischen Auswertungsmodalitäten keinen Incentive, nicht zumindest einen "educated guess" abzugeben – immerhin hat die Person bei einem Multiple-Choice-Test durch Raten bei Nichtwissen nichts „zu verlieren“. Falls die geratene Antwort nicht richtig ist, hat es keine Auswirkungen auf den eigenen Score und im Falle richtigen Tippens verbessert sich der Zahlenwert am Ende des Tests.

Da die Latenzzeit für diese Arbeit von größter Bedeutung ist, werde ich im Folgenden näher auf bisherige Erkenntnisse auf diesem Gebiet eingehen. Zunächst werde ich das Hicksche Gesetz erläutern, welches sich mit der elementaren Informationsverarbeitung beschäftigt.

1.1. Das Hicksche Gesetz

Hick (1952) leistete mit seinen Untersuchungen, aus denen das Hicksche Gesetz hervorging, einen wesentlichen Beitrag zum Verständnis von Bearbeitungszeiten.

Das Hicksche Gesetz besagt, dass die Reaktionszeit bei Wahlentscheidungen von der Anzahl der Wahlalternativen abhängt – mehr Antwortmöglichkeiten bedeuten, vereinfacht gesagt, eine langsamere Reaktionszeit. Wie Neubauer (1995) in weiterer Folge zeigen konnte, hängt das Ausmaß, in welchem sich die Anzahl an Wahlalternativen auf die Reaktionszeit auswirkt, (wenig überraschend) wiederum von der Informationsverarbeitungsgeschwindigkeit der Proband_in ab.

Bei der damaligen Versuchsanordnung, welche Hick zur Postulation „seines“ Gesetzes führte, sah sich die Proband_in mit einer Apparatur konfrontiert, auf welcher sich acht im Halbkreis angeordnete Leuchten befanden. Neben jeder dieser Leuchten befand sich ein korrespondierender Knopf, der gedrückt werden musste, sobald das Aufleuchten einer dieser Leuchten von der Teilnehmer_in registriert wurde.

Der Informationsgehalt der Versuchsanordnung konnte mit einfachen Mitteln, realisiert durch das Abdecken von Leuchten mit Papier, variiert werden. Je weniger Leuchten es gibt, die möglicherweise aufleuchten können, desto geringer ist der Informationsgehalt.

Im einfachsten Fall konnten alle bis auf eine Lampe abgedeckt werden, was einer 0-Bit-Bedingung entspricht. Im Gegensatz dazu entspricht es einer achtfachen Wahlreaktions-Aufgabe bzw. 3-Bit-Bedingung, wenn keine Leuchte abgedeckt ist und alle zur Auswahl stehen. Hick (1952) konnte zeigen, dass die Reaktionszeit mit der Anzahl Wahlalternativen logarithmisch ansteigt (s. Abbildung 2).

$$R = \log n_e$$

Abbildung 2 - Das Hicksche Gesetz. (Hick, 1952, S. 26).

Wenn man zur Basis 2 logarithmiert, erhält man eine lineare Beziehung, weshalb Hick (1952) zu dem Schluss kommt, dass es nicht die Anzahl der Wahlalternativen ist, welche die Reaktionszeit determiniert, sondern der Informationsgehalt der Aufgabenstellung. Diese Überlegung wurde von Neubauer (1995) aufgegriffen und führte zu folgender Umformulierung:

$$RT = a + b \cdot \log_2 n$$

Abbildung 3 – Das Hicksche Gesetz (nach Neubauer, 1995, S. 8): RT steht für die Reaktionszeit, während a die Reaktionszeit der Proband_in in einer 0-bit-Bedingung ist. n stellt die Anzahl der Wahlalternativen dar, b steht für den Anstieg der Geraden.

Jensen (1987) stellte bei einer Metaanalyse von 32 Studien (N = 2317) zum Hick-Paradigma fest, dass die Korrelation zwischen Reaktionszeiten und Intelligenz mit steigendem Informationsgehalt der Aufgaben zunimmt ($r = -0,18$ bis $r = -0,4$). Laut Jensen (1987) spricht das dafür, dass die allgemeine Intelligenz besser durch komplexe Aufgaben erfasst werden kann, als etwa durch elementarkognitive Aufgaben.

Interessanterweise zeigen sich bezüglich der Intelligenz und Latenzzeit zwei gegenläufige Trends. Während es bei elementarkognitiven Aufgaben, wie etwa jenen Reaktionsaufgaben bei Hick (1952) in der einfachsten Bedingung, einen negativen Zusammenhang zwischen Bearbeitungszeit und Intelligenz zu geben scheint (Grudnik & Kranzler, 2001), zeigt sich bei komplexeren Aufgaben, wie etwa jenen, die im Rahmen eines kognitiven Leistungstests zur Bearbeitung vorgesetzt werden, ein komplett gegensätzlicher Trend – hier konnte von Beckmann (2000) eine positive Korrelation auf signifikantem Niveau zwischen Bearbeitungszeit und Leistungsgüte festgestellt werden. Fähigere Testteilnehmer neigten also eher dazu, mehr Zeit in einzelne Aufgaben zu investieren, als Personen mit einem geringeren Fähigkeitsparameter – zumindest, wenn es sich nicht um elementarkognitive Aufgaben handelt.

Diese Beobachtungen legen nahe, dass man in Bezug auf die Untersuchung von Bearbeitungszeiten klar zwischen einfachen Reaktionsaufgaben bzw. elementarkognitiven Aufgaben und jenen Aufgaben, bei denen für die Beantwortung der Fragestellung höhere kognitive Funktionen verlangt werden (wie etwa bei kognitiven Leistungstests), unterscheiden muss.

In dieser Arbeit widme ich mich der Latenzzeit bei kognitiven Leistungstests, weshalb auf die gegenläufigen Trends bei elementarkognitiven Aufgaben hier nicht näher eingegangen wird. Interessierte können sich unter anderem bei Sommerfeld (1994) näher zum Hickschen Gesetz und Reaktionszeiten bei elementarkognitiven Aufgaben informieren.

1.2. Das Falsch > Richtig-Phänomen

Im Lichte dieser Überlegungen wird es wohl wenig überraschend scheinen, dass ungelöste Items durchschnittlich länger als gelöste Items bearbeitet werden (Tate, 1948; Ebel, 1953; Beckmann, 2000; Hornke, 2000; Hornke, 2005; Dörfler, 2010).

Dörfler (2010) kommt bei einer Metaanalyse zu dem Schluss, dass die Latenzzeit bei ungelösten Aufgaben durchschnittlich 25% länger ist als bei Aufgaben, die gelöst werden konnten. Dörfler (2010) konnte des Weiteren zeigen, dass Personen mit einem höherem Fähigkeitsparameter eine höhere Variabilität der Latenzzeiten zeigten und sich auch im Durchschnitt signifikant länger mit den einzelnen Items auseinandersetzten als leistungsschwächere Teilnehmer_innen. Da es sich hierbei um Raven's Advanced Matrices (Raven, Raven & Court 2003) handelte, einen nonverbalen Intelligenztest, bei welchem die Items konservativ vorgegeben werden (also nicht etwa im Sinne der Item-Response-Theory aufgrund des Fähigkeitsparameters der Teilnehmer_in aus einem Itempool ausgewählt werden), kann dieser Unterschied nicht daher rühren, dass von den leistungstärkeren Testteilnehmer_innen schwierigere Aufgaben bearbeitet wurden, als von den Leistungsschwächeren.

Es scheint, als würde die Bereitschaft, Ressourcen (in diesem Fall Zeit und kognitive Arbeit) in Aufgaben zu investieren, bei intelligenteren Testteilnehmer_innen tendenziell eher gegeben sein, als bei weniger intelligenten Testteilnehmer_innen – Hornke (1997) stellt hier einen Bezug zum Persönlichkeitsmerkmal Persistenz her, welcher von der Definition her (Persistenz – „bei der Stange bleiben, sich in eine Aufgabe verbeißen“) augenscheinlich Sinn zu ergeben scheint. Einen Bezug zu Persönlichkeitsdimensionen stellte auch Dörfler (2010) her. Bei einer Untersuchung an 101 Student_innen zeigte sich eine negative Korrelation ($r = -0,26$, $p < 0,05$) zwischen durchschnittlicher Bearbeitungszeit von Items des vorhin erwähnten Leistungstests und dem Persönlichkeitsmerkmal Extraversion.

Personen, welche eine höhere Extraversion aufwiesen, neigten dazu, Items schneller zu bearbeiten, zeigten aber auch eine geringere Leistungsgüte, was sich mit den Erkenntnissen von Eysenck (1947) deckt. Dieser kam zu dem Schluss, dass Testteilnehmer_innen mit stark ausgeprägter Introversion im Vergleich zu Testteilnehmer_innen mit stark ausgeprägter Extraversion tendenziell eine langsamere Bearbeitungszeit, aber höhere Leistungsgüte aufwiesen.

1.3. Modelle der Bearbeitungszeit: Linear Ballistic Accumulator

Zum Prozess der Entscheidungsfindung gibt es eine Vielzahl an Theorien, wobei aktuell besonders den *evidence accumulation models* (Ratcliff und Smith, 2004) in der Forschung Aufmerksamkeit geschenkt wird.

In diesen Modellen wird, vereinfacht gesagt, davon ausgegangen, dass die Auswahl zwischen mehreren Lösungsalternativen getroffen wird, indem Beweise für jede der möglichen Antworten gesammelt werden, bis ein Schwellwert überschritten wird und dann die entsprechende Antwort gegeben wird. Die meisten Modelle beschränken sich in ihrer Anwendbarkeit dabei auf binäre Fragestellungen, wie zum Beispiel das *Drift Diffusion Model* (Ratcliff, 1978) oder das *EZ-diffusion Model* von Wagenmakers, van der Maas und Grasman (2007).

Es gibt aber auch vielversprechende Modelle, die für mehr als zwei Antwortalternativen eingesetzt werden können, wie zum Beispiel der *Linear Ballistic Accumulator* (Brown & Heathcote, 2008). Im Sinne der Flüssigkeit des Textes wird im Weiteren die Abkürzung LBA verwendet, gemeint ist der *Linear Ballistic Accumulator*.

Bei diesem Modell wird davon ausgegangen, dass es für jede Antwortmöglichkeit eine eigene unabhängige Beweis-Akkumulations-Funktion gibt, wobei es für jede Antwortmöglichkeit zusätzlich noch von Beginn an eine gewisse Anzahl an Indizien gibt. Die Anzahl der gesammelten Indizien für die jeweiligen Entscheidungsalternativen nimmt inkrementell mit einer Geschwindigkeit zu, die durch die sogenannte „Drift rate“ gegeben ist, bis eine Antwortschwelle erreicht ist. Die Testteilnehmer_in entscheidet sich dann für jene Antwortmöglichkeit, welche als erstes den Schwellwert erreicht hat.

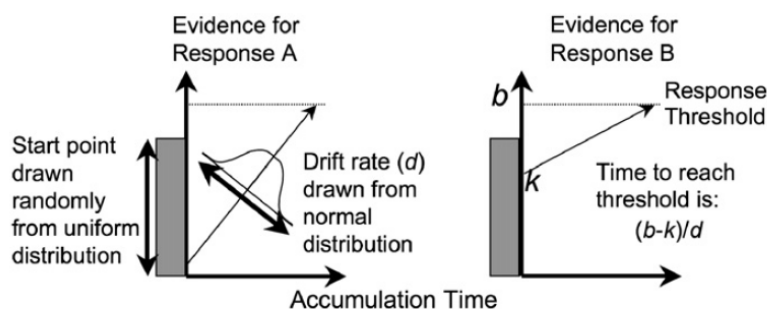


Fig. 3. A two-choice version of the LBA. Evidence for Response A is gathered in the left hand accumulator, and for Response B in the right hand accumulator. Starting values for the evidence accumulation processes (e.g., k) are drawn randomly and independently from identical uniform distributions on the interval $[0, A]$. The drift rate (the speed of evidence accumulation, e.g., d) is drawn independently for each response from normal distributions with standard deviation s . A response is triggered when the first accumulator reaches the threshold b .

Abbildung 3: Veranschaulichung des LBA. (Brown & Heathcote, 2008)

Das Modell besticht vor allem, weil es mit seinen vier Komponenten verhältnismäßig simpel gehalten ist und dennoch in der Lage ist, mehrere Phänomene und Widersprüchlichkeiten aufzuklären.

Wie bereits in Kapitel 1.1. erwähnt, werden Falschantworten bei Reaktionsaufgaben durchschnittlich schneller gegeben, während es sich bei kognitiven Leistungsaufgaben gegenteilig verhält, bei denen Falschantworten durchschnittlich mehr Zeit in Anspruch nehmen als korrekte Antworten. Im LBA wird dies durch das Zwischenspiel des Ausgangspunkts der *Beweisakkumulationsfunktion* und der Variabilität der *Drift Rate* erklärt. Wenn von den Testteilnehmer_innen eine möglichst schnelle Antwort verlangt wird (wie etwa bei Reaktionstests), werden von der Testteilnehmer_in *Antwortschwelle* und *Startpunkt* für die jeweiligen Antwortalternativen von Beginn an sehr nahe aneinander angesiedelt, dass schon ein einzelner Sinneseindruck ausreichen kann, um sich für eine Antwort zu entscheiden – sei sie korrekt oder inkorrekt.

Im Gegensatz dazu wird bei kognitiv anspruchsvollen Aufgaben – bei denen das Hauptaugenmerk auf dem Geben einer korrekten Antwort liegt und nicht auf der Zeit, welche verstreicht, bis diese gegeben wird – der Startpunkt wesentlich niedriger angesetzt. Da die korrekte Antwort eine höhere *Drift-Rate* hat als die Inkorrekte, ist die Zeit, bis die Antwortschwelle erreicht wird, hier für korrekte Antworten geringer als für inkorrekte.

Für detaillierte Informationen zu LBA wird die Lektüre des Artikels von Brown und Heathcote (2008) nahe gelegt, da in dieser Arbeit nicht weiter darauf eingegangen wird.

Um ein weiteres Modell, das von Interesse für das Verständnis von Bearbeitungszeiten bei Leistungstestungen ist, handelt es sich beim *Speed-Power-Zweistufenmodell* von Fischer. Hier werden die gegebenen Antworten bei der Auswertung nicht binär in „richtig“ und „falsch“ kodiert. Stattdessen wird bei korrekten Antworten zusätzlich noch zwischen „richtig und schnell“ und „richtig, aber langsam“ unterschieden. Die zugrundeliegende Überlegung lautet nämlich, dass es sich bei „speed“ und „power“ um nachweislich unterschiedliche Eigenschaften handelt, weshalb für zwei Fähigkeitsparameter pro Testperson argumentiert wird. Mehr hierzu findet sich in *Moderne Testtheorie* (Kubinger 1989: 71ff.).

Nachdem nun ein kurzer Überblick über Latenzzeiten in der Psychologie und Modelle zum Verständnis geliefert wurde, wendet sich diese Arbeit nun dem empirischen Teil zu.

III. Empirischer Teil

2. Ziel der Untersuchung (Hypothesen)

In dieser Arbeit wird der Einfluss des Itemschwierigkeitsparameters bzw. Fähigkeitsparameters und der Bearbeitungszeit auf die Lösungsqualität untersucht.

Die 1. Hypothese dieser Arbeit besagt, dass das *Falsch > Richtig-Phänomen* (s. Kapitel 1.2) bei allen vorgegebenen Subtests gefunden werden kann.

Wie Brown und Heathcote (2008) im LBA postulieren, werden bei anspruchsvollen Aufgaben falsche Antworten langsamer gegeben als richtige Antworten. Es handelt sich bei allen Aufgaben, die vorgegeben wurden, um solche anspruchsvollen Aufgaben. Auch wenn der Anteil der Leistung, der durch fluide bzw. kristalline Intelligenz erklärt werden kann, sich von Aufgabengruppe zu Aufgabengruppe unterscheiden mag, haben alle Aufgaben gemein, dass sie eingehender Analyse bedürfen, um gelöst zu werden. Dementsprechend wird erwartet, dass dieser Trend sich in allen Subtests zeigt.

Die 2. Hypothese besagt, dass durch Einbezug von Bearbeitungszeit, Schwierigkeitsparameter des Items und Fähigkeitsparameter der Testteilnehmer_in in befriedigendem Maß vorhergesagt werden kann, ob ein Beispiel gelöst werden konnte. Diese Hypothese baut auf der vorigen auf – es wird vermutet, dass das Falsch > Richtig-Phänomen stark genug ausgeprägt ist, um einen sinnvollen Beitrag zur Klassifikation von Antworten beitragen zu können. Natürlich darf man dabei auch den Fähigkeitsparameter der Versuchsperson bzw. den Schwierigkeitsparameter des Items nicht außer Acht lassen, da beide einen Einfluss auf die Lösungswahrscheinlichkeit haben, was augenscheinlich auch einen Einfluss auf den Erfolg beim Beantworten des Items hat. Sollten Persönlichkeitsdimensionen keinen Einfluss auf das Zeitverhalten der Teistteilnehmer_innen ausüben, müsste es möglich sein hier eine entsprechende Modellgleichung zu finden.

Das Ziel der Untersuchung ist es, eine Modellgleichung zu finden, welche unter Einbezug der vorhin genannten Faktoren in befriedigendem Ausmaß die vorliegenden Daten erklären kann.

3. Methode

3.1. Erhebungsinstrumente

3.1.1. Intelligenz-Struktur-Batterie

In dieser Arbeit wurden, wie bereits im vorherigen Kapitel erwähnt, Aufgaben einer Intelligenztestbatterie ohne zeitliche Begrenzung verwendet. Beim eingesetzten Verfahren handelt es sich um die Intelligenz-Struktur-Batterie (INSBAT) in der Version 26 – Revision 1 (Arendasy et al., 2010).

Es handelt sich hierbei um ein computergestütztes Verfahren mit adaptiver Itemvorgabe, welches modular aufgebaut ist und der „Erfassung von Intelligenzniveau und Intelligenzstruktur“ (S.5) dient. Es basiert auf dem hierarchischen Intelligenzmodell von Cattell-Horn-Carroll (Carroll, 1993; Horn, 1989; Horn & Noll, 1997).

Bezüglich der Gütekriterien sei angemerkt, dass sowohl die *Durchführungs-* als auch die *Auswertungsobjektivität* als gegeben angesehen werden können, da es sich um ein computergestütztes Verfahren handelt, bei dem Instruktionen und Vorgaben standardisiert sind, und die Testwerte automatisch berechnet werden.

Mit der *Interpretationsobjektivität* verhält es sich ähnlich, da die Testperson über den Normwert automatisch in Bezug auf die gemessenen Eigenschaften mit der gewünschten Referenzpopulation in Relation gestellt wird.

Die Innere Konsistenz der Subtests liegt zwischen 0,73 und 0,96, während die *Stabilität* der Testkennwerte der einzelnen Subtests zwischen 0,52 und 0,87 liegt, was die Annahme stützt, dass die einzelnen Aufgabengruppen zeitstabile Fähigkeitsdimensionen erfassen.

Bezüglich der *Konstruktvalidität* des Verfahrens wurden mehrere Studien durchgeführt, in denen versucht wurde, anhand des jeweiligen Konstruktionsrational die Itemschwierigkeit vorherzusagen. Die Korrelation zwischen den Itemparametern und dem Konstruktionsrational (R) bewegt sich je nach Subtest zwischen 0,72 und 0,97, woraus sich ein durch das Konstruktionsrational erklärter Varianzanteil (R^2) ergibt, der je nach Untertest zwischen 0,52 und 0,94 liegt.

Kriteriumsvalidität liegt vor, wenn ein Test mit einem für die diagnostische Fragestellung relevanten Außenkriterium korreliert (Kubinger, 2003). Dies wurde von Sommer, Häusler, Koning und Arendasy (2006) für mehrere Subtests (unter anderem *Figural-induktives Denken* und *Numerisch-induktives Denken*) im Bereich der Flugpsychologischen Diagnostik untersucht. Hierbei wurden eine Klassifikationsrate von 92,9% und ein Validitätskoeffizient von 0,83 erzielt – im Bereich der Pilotenauswahl sprechen die untersuchten Subtests also für

die Kriteriumsvalidität des INSBAT. Weitere Befunde hierzu finden sich bei Sommer und Häusler (2008).

Skalierung im Sinne des Gütekriteriums kann dann als gegeben verstanden werden, wenn Testwerte die empirischen Verhaltensrelationen exakt wiedergeben können (Kubinger, 2003). Für die Subtests des INSBAT bedeutet dies, dass der Rohwert bzw. Fähigkeitsparameter alle relevanten Informationen über die Testbearbeitung fair abbildet, was als gegeben betrachtet werden kann, da die einzelnen Items der Subtests des INSBAT das Raschmodell erfüllen.

Da es sich um ein computergestütztes Verfahren mit adaptiver Itemvorgabe handelt, das modular gegliedert ist, ist eine ökonomische Durchführung und Auswertung gewährleistet. Dies geht auch Hand in Hand mit der *Zumutbarkeit* des Verfahrens, da man aufgrund des modularen Aufbaus die Möglichkeit hat, ausschließlich jene Subtests vorzugeben, die für die zu untersuchende Fragestellung von Relevanz sind

Zur *Nützlichkeit* des Verfahrens sei gesagt, dass eine detaillierte Erfassung von unterschiedlichen Fähigkeitskonstrukten für eine Vielzahl von Fragestellungen von großem Interesse ist.

Wie bei Kubinger (2009, p. 120) festgestellt wird, ist es bei Leistungstests „möglich, aber kaum wahrscheinlich, dass eine Tp absichtlich schlechte Leistungen erbringt“. Der umgekehrte Fall, also dass eine Testperson bewusst bessere Leistungen erzielt als es die eigene Fähigkeit erlaubt, ist auf überzufälligem Niveau auszuschließen. Um weitere *Unverfälschbarkeit* zu gewährleisten, wurde Item- und Content-Exposure-Control eingesetzt, um Coached-Faking (Hiscock, Branham & Hiscock, 1994) zu unterbinden – allerdings wird hierauf im Manual des Verfahrens nicht näher eingegangen.

Das Gütekriterium *Fairness* ist damit gleichzusetzen, dass bestimmte Gruppen von Testpersonen aufgrund ihres soziokulturellen Hintergrunds nicht systematisch diskriminiert werden. Hiervon kann ausgegangen werden, da die Geltung des Raschmodells der einzelnen Subtests an psychologisch relevanten Untergruppen nachgewiesen werden konnte.

Für diese Arbeit sind vor allem die Eigenschaften des Itemmaterials besonders von Interesse; sämtliche Items aller Subtests sind nach dem 1PL-Rasch Modell skalierbar (Rasch, 1960), wodurch zu jedem Item ein zugehöriger Schwierigkeitsparameter bekannt ist.

Zuvor sei kurz erklärt, weshalb genau dieses Verfahren eingesetzt wurde. Für diese Arbeit war allen voran die Bearbeitungszeit der jeweiligen Items von Interesse. Diese wird bei der automatischen Auswertung für jedes Item miterhoben und steht somit leicht zugänglich für

weitere Berechnungen zur Verfügung. Besonders wichtig war es auch, dass es Items ohne zeitliche Begrenzung in der Testbatterie gibt, da hier Deckeneffekte, die die Bearbeitungszeit betreffen, vermieden werden können. Von den 16 Subtests, welche das Verfahren umfasst, wurden vier ohne zeitliche Begrenzung pro Item vorgegeben, welche für die weiteren Analysen in dieser Arbeit verwendet wurden.

Im Einzelnen wurden die Subtests *Numerisch-induktives Denken*, *Allgemeinwissen*, *Figural-induktives Denken* und *Verbale Flüssigkeit* eingesetzt. Zur Veranschaulichung folgt eine kurze Erläuterung inklusive Beispielaufgabe zu jedem der Subtests.

3.1.2. Subtest Numerisch-induktives Denken

Bei dieser Aufgabengruppe geht es darum, eine Zahlenfolge, die nach bestimmten Regeln aufgebaut ist, zu ergänzen. Dementsprechend muss die zugrundeliegende Regel erkannt werden und im Anschluss die Lösungszahl in der Software eingegeben werden. Die Anzahl der bearbeiteten Items variiert aufgrund der adaptiven Vorgabe, maximal werden 30 Items vorgegeben.

Anleitung...

Natürlich können die Aufgaben auch etwas schwieriger werden. Die Regel kann aus Addieren, Subtrahieren, Multiplizieren oder Dividieren bestehen. Bei manchen Zahlenfolgen können auch mehrere Rechenregeln zum Einsatz kommen.

Versuchen Sie nun die folgende Aufgabe selbst zu lösen:

4 8 16 32 64 128 ?

Löschen

1 2 3 4 5 6 7 8 9 0

Abbildung 4: Beispielim Numerisch-induktives Denken (entnommen aus Arendasy et al, 2010)

3.1.3. Figural-induktives Denken

Beim Subtest *Figural-induktives Denken* wird auf dem Bildschirm eine 3x3-Matrix präsentiert, in welcher sich acht Symbole und ein leeres Feld befinden. Die acht Figuren sind nach einer oder mehreren bestimmten zugrundeliegenden Regeln angeordnet, wobei das Ziel der getesteten Person ist, diese Regeln aufzufinden, um das leerstehende Feld sinnvoll zu ergänzen. Zur Beantwortung stehen acht verschiedene Antwortmöglichkeiten zur Verfügung – es muss die Richtige ausgewählt werden. Die Anzahl der bearbeiteten Items variiert aufgrund der adaptiven Vorgabe, maximal werden 30 Items vorgegeben.

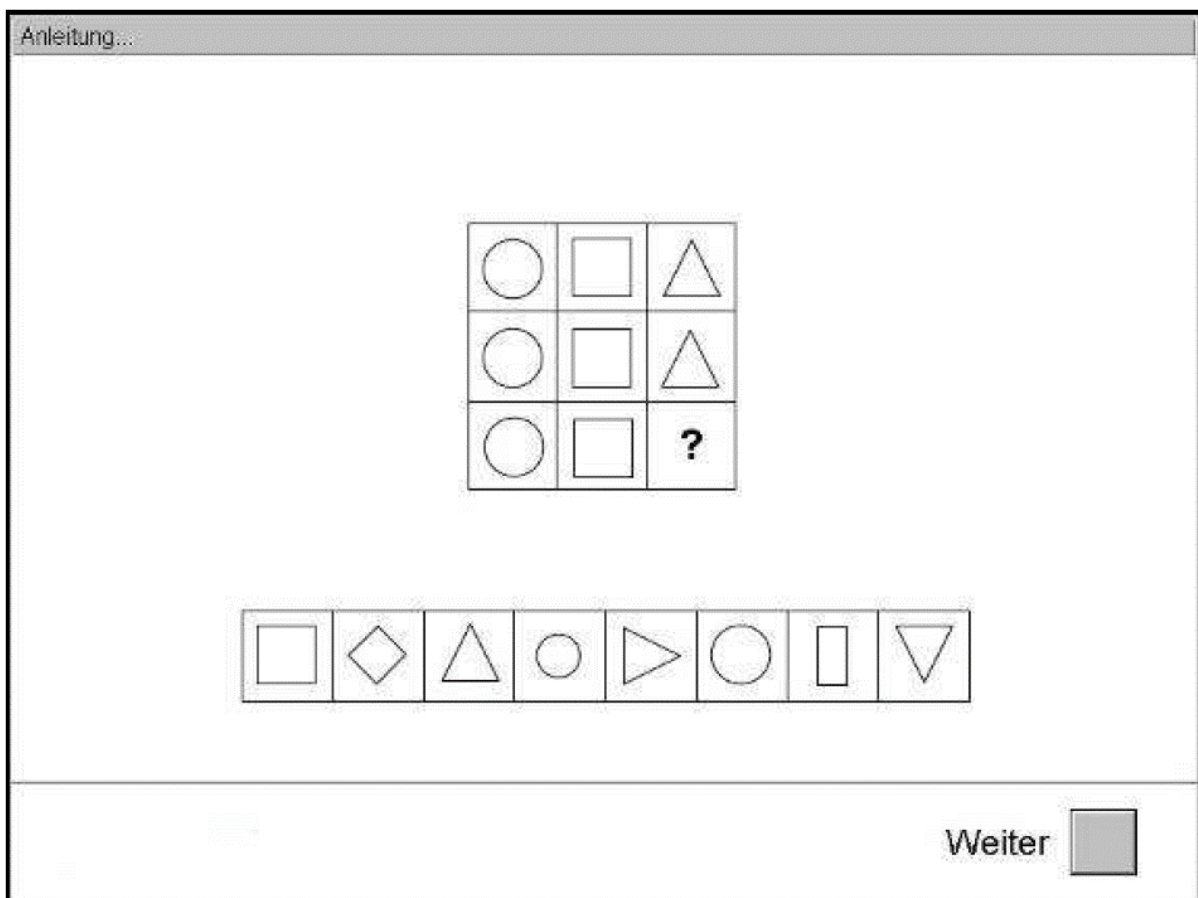


Abbildung 5: Beispielitem Figural-induktives Denken (entnommen aus Arendasy et al, 2010)

3.1.4. Allgemeinwissen

Beim Subtest *Allgemeinwissen* werden Definitionen als Lückentexte vorgegeben, wobei die Aufgabe der untersuchten Person ist, diese sinnvoll zu ergänzen. In jedem Satz sind zwei Lücken enthalten, die mit vorgegebenen Satzteilen – es stehen pro Lücke vier zur Auswahl – gefüllt werden müssen, um zu einer kohärenten und adäquaten Definition des gesuchten Begriffs zu gelangen. Die Anzahl der bearbeiteten Items variiert aufgrund der adaptiven Vorgabe, maximal werden 30 Items vorgegeben.

Anleitung...

Wenn Sie das vorhin gezeigte Wort und seine Bedeutung nicht gekannt haben, versuchen Sie bitte trotzdem den Satz zu vervollständigen.

Die MUTATION ist eine plötzlich auftretende **Veränderung der Erbsubstanz**, die dann _____ weiter gegeben wird.

☒ Veränderung der Erbsubstanz	☐ konstant
☐ Taubheit	☐ verändert
☐ Sprechunfähigkeit	☐ wechselhaft
☐ Hauterkrankung	☐ sprunghaft

☐ Zurück

Weiter ☐

Abbildung 6: Beispielitem Allgemeinwissen (entnommen aus Arendasy et al, 2010)

3.1.5. Verbale Flüssigkeit

Im Subtest *Verbale Flüssigkeit* verlangt die Aufgabenstellung, dass eine ungeordnete Buchstabenfolge durch Umreihung der Buchstaben in ein sinnvolles Hauptwort transformiert wird. Vereinfacht gesagt werden Buchstaben vorgegeben, die ein Anagramm eines deutschsprachigen Hauptwortes sind, das von der getesteten Person gefunden werden muss. Hierzu müssen die Buchstaben in der richtigen Reihenfolge angeklickt werden. Da die Items adaptiv vorgegeben werden, kann die Anzahl der Buchstaben der vorgegebenen Aufgaben von Teilnehmer zu Teilnehmer variieren, wobei fähigere Testpersonen sich mit längeren Buchstabenfolgen konfrontiert sehen werden. Die Anzahl der vorgegeben Items variiert aufgrund der adaptiven Vorgabe, maximal werden 30 Items vorgegeben.

Anleitung...

Das war richtig. Versuchen Sie nun diese Aufgabe selbst zu lösen. Falls Sie kein Wort in der Buchstabenfolge erkennen, betätigen Sie zweimal hintereinander das „Weiter“-Kästchen um die Aufgabe auszulassen.

M U S _ _ _

E ~~M~~ ~~S~~ ~~U~~ R T

Löschen

Weiter

Abbildung 7: Beispielitem Verbale Flüssigkeit (entnommen aus Arendasy et al, 2010)

3.2. Stichprobe

Bei der Stichprobe handelt es sich ausschließlich um Student_innen der Fakultät für Psychologie der Universität Wien. Im Rahmen des Studiums ist es vorgesehen, dass diese im Rahmen einer sogenannten „Testerfahrung“ gängige psychologische Verfahren am eigenen Leib ausprobieren, um die Verfahren besser kennenzulernen und auch ein realistisches Konzept der Zumutbarkeit solcher Verfahren zu entwickeln.

Es wurden nur jene Student_innen in die Stichprobe aufgenommen, die alle vier Subtests, die für die Untersuchung von Interesse sind (zur Erinnerung: *Numerisch-induktives Denken*, *Allgemeinwissen*, *Figural-induktives Denken* und *Verbale Flüssigkeit*), durchgeführt haben, ohne die Testung vorzeitig zu beenden zu haben.

Das Verfahren wurde mit den Teilnehmer_innen in den Räumlichkeiten der Fakultät für Psychologie der Universität Wien jeweils in Einzelbearbeitung am Computer durchgeführt. Als Eingabegeräte wurden Mouse und Tastatur verwendet.

Genaueres zur Stichprobe findet sich in *Abbildung 8*.

Stichprobenumfang

Männlich	66	31,30%
Weiblich	145	68,70%
Gesamt (N)	211	100%

Schulabschluss

Matura	183	86,70%
Hochschulabschluss	28	13,30%

Alter (in Jahren)

Minimum	17,5
Maximum	55,75
Mittelwert	25,37

Abbildung 8: Relevante Kennzahlen zur Stichprobe

4. Ergebnisse

4.1. Hypothese 1

Wie bereits in Kapitel 2 angedeutet, lautet das 1. Hypothesenpaar dieser Arbeit folgendermaßen:

H0: Es gibt keinen Unterschied zwischen der durchschnittlichen Bearbeitungszeit für gelöste und ungelöste Items

H1: Die durchschnittliche Bearbeitungszeit ist für ungelöste Aufgaben höher als jene für gelöste Aufgaben

Um dies zu überprüfen, wurde für jede Teilnehmer_in in jedem der Subtests jeweils die durchschnittliche Bearbeitungszeit für die gelösten und ungelösten Aufgaben berechnet. Die durchschnittliche Bearbeitungszeit für ungelöste Items wurde bestimmt, indem für jede Teilnehmer_in alle Bearbeitungszeiten eines Subtests von ungelösten Items summiert wurden, und anschliessend durch die Anzahl an ungelösten Items dividiert wurden – für gelöste Items wurde analog vorgegangen. Dies wurde für jeden Subtest durchgeführt. So ergab sich für jede Teilnehmer_in eine durchschnittliche Bearbeitungszeit für gelöste und ungelöste Items.

Es wurde ein *Wilcoxon signed-rank test* (Wilcoxon, 1945) durchgeführt, da für diesen keine Voraussetzungen verletzt wurden. Zur Auswertung wurde die Software SPSS in der Version 22.0 verwendet.

Das Ergebnis der Analyse findet sich in Abbildung 9.

	Asymp. Sig. (2-Seitig)	z-Wert
Numerisch-induktives Denken	0,000	-12,587
Allgemeinwissen	0,000	-11,169
Figural-induktives Denken	0,000	-9,506
Verbale Flüssigkeit	0,000	-12,413

Abbildung 9: Ergebnis Wilcoxon signed-rank test.

Wie sich anhand von *Abbildung 9* erkennen lässt, ist für jeden der verschiedenen Subtests die H_0 (kein Unterschied zwischen dem Median der Bearbeitungszeit für gelöste und ungelöste Items) zu verwerfen.

Beim Falsch > Richtig-Phänomen dürfte es sich also im Sinne von Ernst von Glasersfeld (1997) um ein Konstrukt handeln, das dem Anspruch der Viabilität gerecht wird.

4.2. Hypothese 2

Bei der 2. Hypothese dieser Arbeit wird der Frage nachgegangen, ob durch Einbezug von Bearbeitungszeit, Schwierigkeitsparameter des Items und Fähigkeitsparameter der getesteten Person befriedigend vorhergesagt werden kann, ob ein Beispiel gelöst werden kann.

Hierfür war ursprünglich geplant, eine Diskriminanzanalyse zu rechnen, allerdings waren die Voraussetzungen zum Einsatz des Verfahrens verletzt. Bei allen Subtests fiel ein Box's-M-Test (bei allen Subtests zeigte sich ein $p = 0,000$) signifikant aus, weshalb auf ein anderes Verfahren zurückgegriffen werden musste.

Die Wahl fiel auf die logistische Regression. In diesem Fall war die Prädiktorvariable „Bewertung Item“ (eine binäre Variable mit den Ausprägungen gelöst/nicht gelöst), während als Kriteriumsvariablen der Fähigkeitsparameter der Testperson, der Schwierigkeitsparameter des Items und die Bearbeitungszeit gewählt wurden.

Der Schwierigkeitsparameter des Items und der Fähigkeitsparameter der Testperson sind intervallskalierte Variablen, ebenso wie die Bearbeitungszeit des Items in Sekunden.

Zur Überprüfung der Hypothese wurde die Stichprobe in eine Lern- und Prüfstichprobe geteilt. Hierfür wurden die Teilnehmer_innen zufällig einer der beiden Gruppen zugewiesen, wobei 70% der Testteilnehmer der ersteren Gruppe zugewiesen wurden und 30% der Prüfstichprobe. Bei einem $N = 211$ hat dies zur Bedeutung, dass der Umfang der Lernstichprobe 63 Personen beinhaltet, während es bei der Prüfstichprobe 148 sind.

Hier zeigte sich unerfreulicherweise gleich, dass sich bei keinem der Subtests ein Modell ergab, welches signifikant bessere Vorhersagen leisten konnte

als das Baseline-Modell – in allen Subtests fiel ein χ^2 -Anpassungstest signifikant aus.

Die Ergebnisse zu den jeweiligen Subtests finden sich in *Abbildung 10* bis *Abbildung 13*.

Numerisch-induktives Denken

		<u>korrekte Vorhersagen</u>			
<u>Baseline Modell</u>		65,40%			
<u>Postuliertes Modell</u>		74,50%			
		<u>Modellübersicht</u>			
<u>Chi-Quadrat-Anpassungstest</u>					
		<u>-2Log Likelihood</u>		<u>R-Quadrat nach Nagelkerke</u>	
Chi-Quadrat	1543,741				
df	3	5827,646		0,326	
Sig.	0,000				
<u>Variablen in Gleichung</u>					
	B	Standardfehler	Wald	df	Sig
Bearbeitungszeit Item	-0,01	0,001	288,427	1	0,00
Fähigkeitsparameter	0,879	0,042	435,986	1	0,00
Schwierigkeitsparameter	-0,68	0,048	203,669	1	0,00
Konstante	0,734	0,070	108,953	1	0,00

Abbildung 10: Logistische Regression der Lernstichprobe für den Subtest Numerisch-induktives Denken

Figural-induktives Denken

		<u>korrekte Vorhersagen</u>			
<u>Baseline Modell</u>		52,20%			
<u>Postuliertes Modell</u>		70,00%			
		<u>Modellübersicht</u>			
<u>Chi-Quadrat -Anpassungstest</u>					
		<u>-2Log Likelihood</u>		<u>R-Quadrat nach Nagelkerke</u>	
Chi-Quadrat	1046,421				
df	3	6676,131		0,228	
Sig.	0,00				
<u>Variablen in Gleichung</u>					
	<u>B</u> <u>l</u> <u>a</u> <u>r</u> <u>d</u> <u>f</u> <u>e</u> <u>h</u> <u>l</u> <u>e</u> <u>r</u>		<u>Wald</u>	<u>df</u>	<u>Sig</u>
Bearbeitungszeit Item	-0,003	0,001	27,728	1	0,00
Fähigkeitsparameter	1,292	0,069	346,264	1	0,00
Schwierigkeitsparameter	-1,224	0,061	408,431	1	0,00
Konstante	0,246	0,061	16,151	1	0,00

Abbildung 11: Logistische Regression der Lernstichprobe für den Subtest figural-induktives Denken

Allgemeinwissen

		<u>korrekte Vorhersagen</u>			
<u>Baseline Modell</u>		50,50%			
<u>Postuliertes Modell</u>		68,10%			
		<u>Modellübersicht</u>			
<u>Chi-Quadrat-Anpassungstest</u>					
		-2Log Likelihood	R-Quadrat nach Nagelkerke		
Chi-Quadrat	1045,411				
df	3	7404,541	0,21		
Sig.	0,00				
<u>Variablen in Gleichung</u>					
	B	Standardfehler	Wald	df	Sig
Bearbeitungszeit Item	-0,017	0,002	71,286	1	0,00
Fähigkeitsparameter	1,281	0,072	315,082	1	0,00
Schwierigkeitsparameter	-1,243	0,060	428,675	1	0,00
Konstante	0,417	0,067	39,221	1	0,00

Abbildung 12: Logistische Regression der Lernstichprobe für den Subtest Allgemeinwissen

Verbale Flüssigkeit

		<u>korrekte Vorhersagen</u>			
<u>Baseline Modell</u>		50,40%			
<u>Postuliertes Modell</u>		74,50%			
		<u>Modellübersicht</u>			
<u>Chi-Quadrat-Anpassungstest</u>		-2Log Likelihood	R-Quadrat nach Nagelkerke		
Chi-Quadrat	1659,530				
df	3	5127,501	0,383		
Sig.	0				
<u>Variablen in Gleichung</u>					
	B	Standardfehler	Wald	df	Sig
Bearbeitungszeit Item	-0,017	0,001	307,864	1	0,00
Fähigkeitsparameter	2,074	0,102	235,385	1	0,00
Schwierigkeitsparameter	-1,569	0,108	372,048	1	0,00
Konstante	0,942	0,063	222,442	1	0,00

Abbildung 13: Logistische Regression der Lernstichprobe für den Subtest verbale Flüssigkeit

Da in diesem Fall keine der vier Lernstichproben Ergebnisse lieferte, die grundsätzlich für das postulierte Modell sprechen, muss *Hypothese 2* schon vor tiefgehender Analyse verworfen werden.

Die Inklusion von Schwierigkeitsparameter des Items, Fähigkeitsparameter der Teilnehmer_in und Bearbeitungszeit ist bei den vorliegenden Daten nicht von ausreichender diagnostischer Qualität um einen diagnostischen Mehrwert zu erzielen.

In der nun folgenden Diskussion wird reflektiert, worin die Wurzel dieser Ergebnisse liegen könnte und darauf eingegangen, worauf man bei zukünftigen Untersuchungen hinsichtlich der Antwortlatenzen das experimentelle Augenmerk legen könnte.

5. Diskussion

Wie ursprünglich erwartet und auch schon in zahlreichen Untersuchungen mit zeitlicher Beschränkung bei den einzelnen Items festgestellt wurde, konnte das Falsch > Richtig-Phänomen in allen Subtests vorgefunden werden – was nicht besonders verwunderlich ist. Im Gegenteil, da das Phänomen bereits bei Tests mit zeitlicher Begrenzung pro Item mehrfach beobachtet wurde, wäre es verwunderlich gewesen, wenn sich das Antwortverhalten der Testteilnehmer_innen durch das Weglassen einer zeitlichen Beschränkung in hohem Ausmaß geändert hätte.

Bezüglich der zweiten Hypothese, welche postulierte, dass man durch Inklusion der Itemschwierigkeit, Bearbeitungszeit und Fähigkeit der Testperson den Erfolg auf ein jeweiliges Item vorhersagen kann, hat sich gezeigt, dass das Modell hierzu wohl nicht ausgefeilt genug war und es keine allgemeine Formel gibt, die auf alle Testteilnehmer_innen gleichermaßen anzuwenden ist.

Dies kann mehrere Gründe haben, wobei meine persönliche Meinung dahingeht, dass es sich um Interaktionen handeln dürfte, die höherer Ordnung sind, weshalb es nicht möglich ist, sie mittels einer logistischen Regressionsgleichung befriedigend widerzuspiegeln. Ebenso klingt es meines Erachtens nach plausibel, dass Persönlichkeitsdimensionen einen Einfluss auf das Zeitverhalten der Testteilnehmer_innen ausüben.

Um der Thematik näher auf den Grund zu gehen, wurden zur Veranschaulichung die Teilnehmer_innen nach Leistungsfähigkeit gruppiert. Für jede Teilnehmer_in wurde zudem die durchschnittliche Bearbeitungszeit pro Item, je nachdem, ob es gelöst werden konnte oder nicht, unter Berücksichtigung des Schwierigkeitsparameters bestimmt.

Die Ergebnisse für die jeweiligen Subtests finden sich in den *Abbildungen 14 bis 21*.

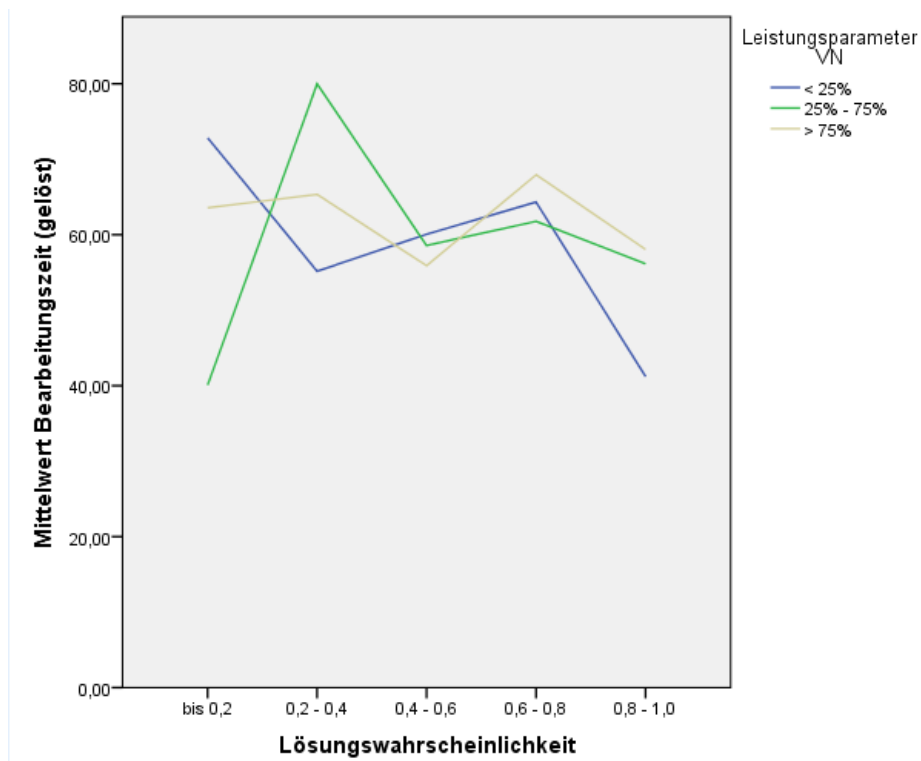


Abbildung 14: Numerisch induktives Denken – Mittelwert Bearbeitungszeit für gelöste Items

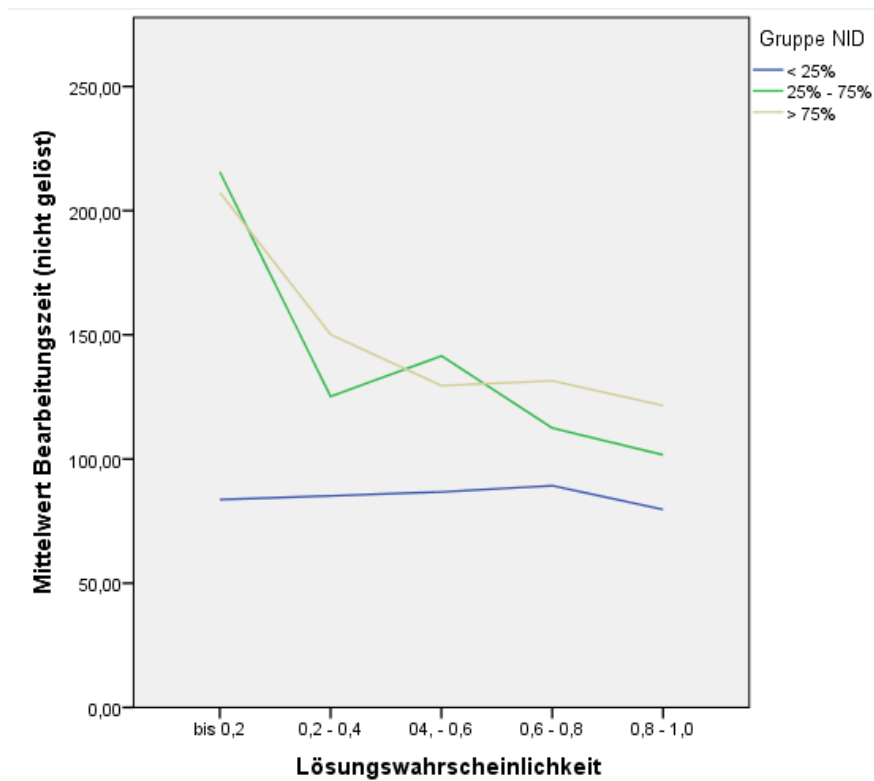


Abbildung 15: Numerisch-induktives Denken - Mittelwert Bearbeitungszeit für ungelöste Items

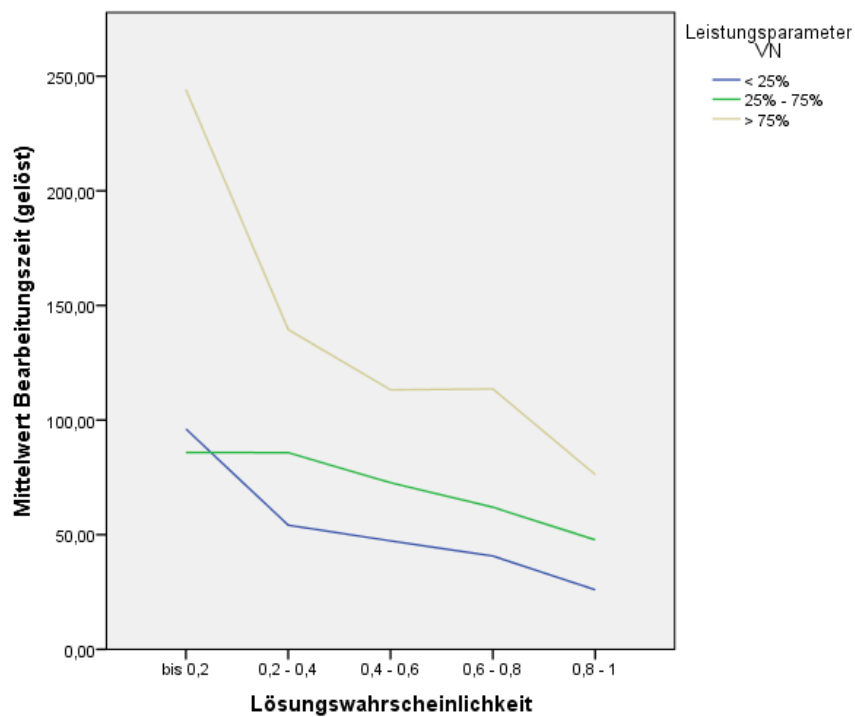


Abbildung 16: Figural-induktives Denken: Mittelwert Bearbeitungszeit für gelöste Items

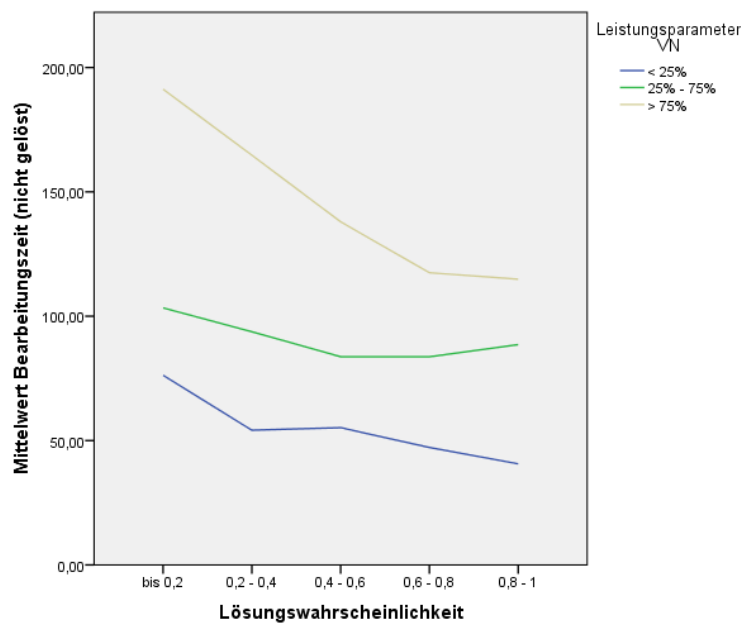


Abbildung 17: Figural-induktives Denken: Mittelwert Bearbeitungszeit für ungelöste Items

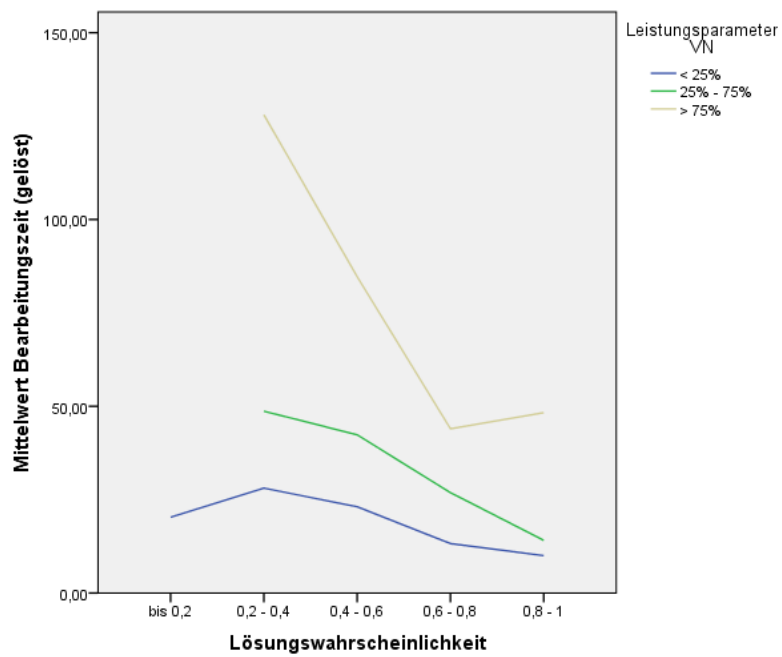


Abbildung 18: Verbale Flüssigkeit: Mittelwert Bearbeitungszeit für gelöste Items

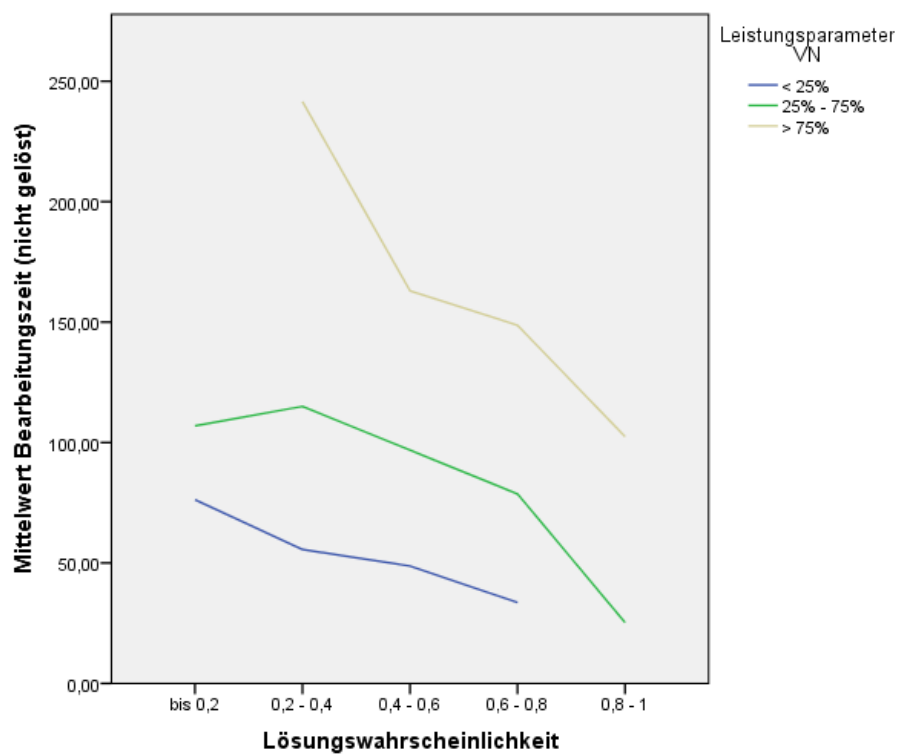


Abbildung 19: Verbale Flüssigkeit: Mittelwert Bearbeitungszeit für ungelöste Items

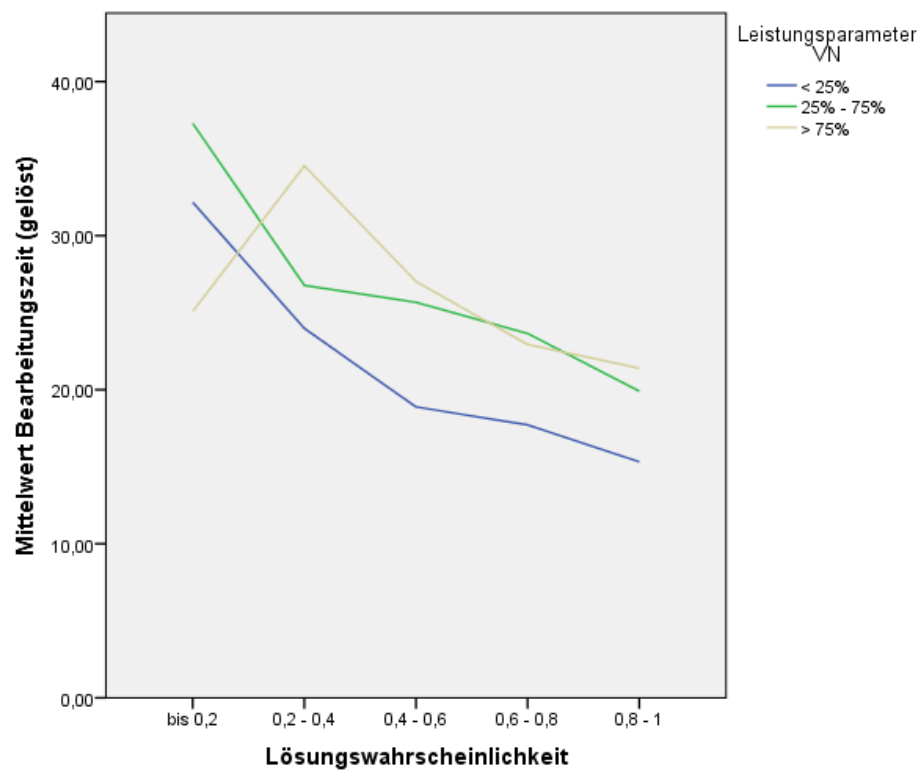


Abbildung 20: Allgemeinwissen: Mittelwert Bearbeitungszeit für gelöste Items

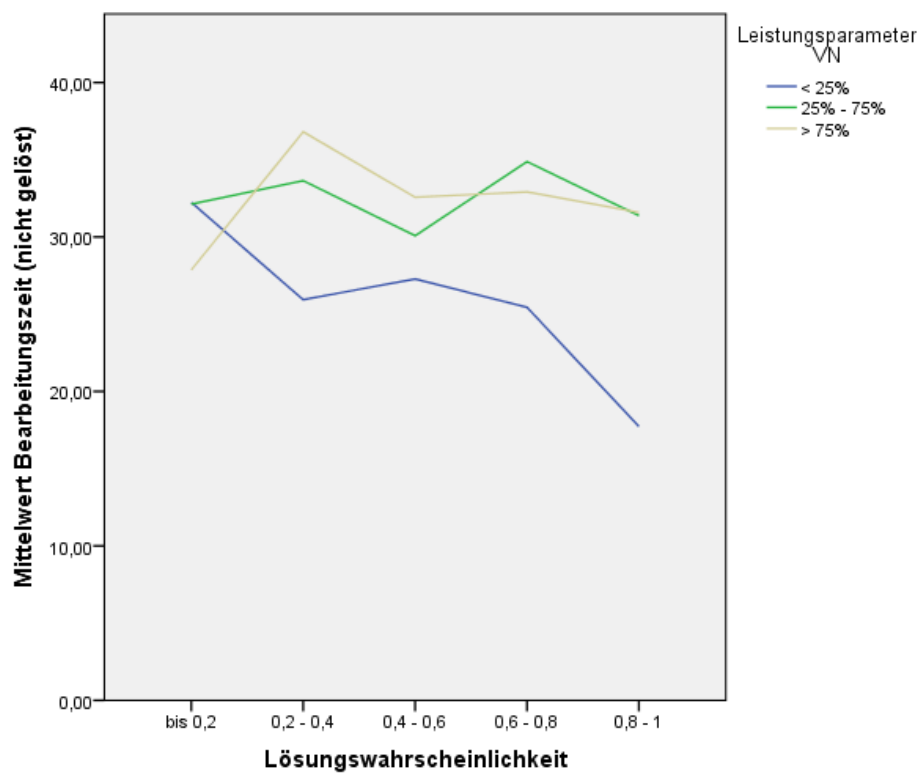


Abbildung 21: Allgemeinwissen: Mittelwert Bearbeitungszeit für ungelöste Items

Auffällig erscheint, dass es, unabhängig von Subtest, die Tendenz gibt, dass Testteilnehmer_innen, die in der Untersuchung einen gering ausgeprägten Fähigkeitssparameter zeigten, Falschantworten schneller gaben als leistungsfähigere Testteilnehmer_innen. Auch die relative Schwierigkeit der Items scheint an diesem Phänomen nichts zu ändern.

Die größenordnungsmäßig auffallendsten Ergebnisse gab es beim Subtest *Verbale Flüssigkeit*, wobei es sich hierbei auch um ein Testartefakt handeln kann. Da das Testmaterial adaptiv vorgegeben wurde, haben fähigere Teilnehmer_innen logischerweise Aufgaben mit höherem Schwierigkeitsparameter erhalten. Im Subtest *Verbale Flüssigkeit* wird die Schwierigkeit unter anderem durch das Hinzufügen weiterer Buchstaben erhöht, weshalb es bei diesem Subtest vorkommt, dass der Informationsgehalt der vorgegebenen Aufgaben sich von Teilnehmer_in zu Teilnehmer_in unterscheidet. Zur Veranschaulichung: Während es bei einem „Buchstabensalat“ aus sechs Buchstaben 720 Möglichkeiten gibt, diese anzuordnen, steigt die Anzahl der möglichen Kombinationen durch das Hinzufügen eines Buchstabens auf 5040 – ein dramatischer Unterschied, der eine mögliche Erklärung für dieses Phänomen liefert.

Über alle Subtests hinweg zeigt sich allerdings die Tendenz, dass besonders fähige Teilnehmer_innen sich von den drei Gruppen am meisten Zeit nehmen, bis eine Antwort gegeben wird.

Generell erscheint es plausibel, dass es durchaus eine Möglichkeit gibt, anhand der Bearbeitungszeit die Qualität einer Antwort zu einem gewissen Grad vorhersagen zu können. Allerdings scheint es, als wäre keine allgemeine lineare Gleichung möglich, sondern als müsse man diese, je nach Fähigkeit der jeweiligen Person, unterschiedlich gewichten.

Grundsätzlich stellt sich die Frage, ob das Verharren an einem Item, auch wenn man die Lösung nicht kennt, aufgrund der Augenscheinvalidität als Kriterium für Persistenz gesehen werden kann. Wenn man die beiden Personen aus dem Veranschaulichungsbeispiel der Einleitung heranziehen möchte, wäre es vermutlich nicht unbegründet, die Vermutung aufzustellen, dass eine Person, die tendenziell bei Aufgaben später aufgibt, eine höher ausgeprägte Persistenz aufweist. Hier kann natürlich die Verarbeitungsgeschwindigkeit mediierend einfließen, da jemand mit höherer „mental speed“ wohl auch weniger Zeit benötigt, um alle erdenklich und realistisch scheinenden Lösungsalgorithmen auf das Beispiel anzuwenden. Hinzu kommt, dass ein flexiblerer Denker wiederum in der Lage sein sollte, eine größere Anzahl an Lösungsalgorithmen zu generieren, was mit einer längeren Bearbeitungszeit bei ungelösten Items einhergehen würde.

Kurz gesagt würde ich die Fragestellung nicht aufgrund der unerwarteten Ergebnisse verwerfen, vielmehr benötigt es meiner Meinung nach ein komplexeres theoretisches Konstrukt, um mehr Klarheit in dieser Angelegenheit zu schaffen.

Für zukünftige Arbeiten wäre es meiner Meinung nach sinnvoll, zuerst mit einigen Teilnehmer_innen die Aufgaben im Form von „lautem Denken“ durchzuarbeiten, um zu sehen, ob sich tatsächlich Unterschiede im Verhalten oder den eingesetzten Lösungsstrategien der Testpersonen bemerkbar machen, die mit dem Fähigkeitsparameter der jeweiligen Person (und auch zuvor erhobenen Persönlichkeitsdimensionen) korrelieren.

6. Zusammenfassung

In dieser Arbeit wurde im Theoretischen Teil (Kapitel 1) auf die Geschichte der Analyse der Bearbeitungszeit bei Leistungstests und auf hierfür eingesetzte Modelle eingegangen. In der Vergangenheit hat sich vor allem das *Falsch > Richtig-Phänomen* als hartnäckig erwiesen und konnte in mehreren Untersuchungen repliziert werden.

Diese Untersuchungen wurden allerdings bislang mit Items mit zeitlicher Begrenzung durchgeführt, weshalb beschlossen wurde, im Rahmen dieser Arbeit Items vorzugeben, die keine zeitliche Begrenzung haben, um mögliche Deckeneffekte zu vermeiden.

Es wurden 211 Student_innen der Universität Wien mit vier Subtests der Intelligenz-Struktur-Batterie (Version 26) untersucht.

Nach ausgiebiger Umstrukturierung der Daten zeigten sich in den Untersuchungen in Bezug auf die Hypothesen nicht die erwarteten Ergebnisse.

Während das Falsch>Richtig-Phänomen, wie in Hypothese 1 (siehe Kapitel 4.1.) postuliert, bei allen Subtests vorzufinden war, ließ sich kein Modell mit zufriedenstellender diagnostischer Vorhersagefähigkeit finden (s. Kapitel 4.2.).

Basierend auf dem in Kapitel 4.2. eingesetzten Modell war bei keinem der vier Subtests eine signifikant bessere Vorhersage möglich als mit einem Baseline-Modell.

Die zweite Hypothese dieser Arbeit, die lautete, dass bei jedem der Subtests durch Einbezug der Bearbeitungszeit befriedigende Prognosen über die Bewertung (richtig/falsch) eines Items gemacht werden können, ist daher zu verwerfen.

Hierfür sind viele Gründe denkbar. Plausibel erscheint, dass es eine wesentlich komplexere Interaktion zwischen den verschiedenen Variablen, möglicherweise auch über Moderatorvariablen, gibt, die in dieser Arbeit nicht berücksichtigt wurden.

Auch wenn in dieser Arbeit kein geeignet spezifiziertes Modell gefunden werden konnte, ist das Themengebiet dennoch in der aktuellen Forschung von Interesse. Vor allem in der neuropsychologischen Forschung gibt es zahlreiche aktuelle Arbeiten zum Entscheidungsprozess bei kognitiv anspruchsvollen Aufgaben (z.B. Heathcote, Wagenmaker & Brown, 2014 oder Voss, Voss & Lerche 2015).

Möglicherweise wäre es für weiterführende Arbeiten sinnvoll, Persönlichkeitsdimensionen mit zu erheben. Insbesondere Persistenz und Need for Cognition bieten sich hier an. Bei Persistenz stellt sich allerdings die Frage, ob die Latenzzeit bei Falschantworten nicht aufgrund der

Augenscheinvalidität als Indikator für die Ausprägung des Merkmals bei einer Person interpretiert werden kann.

7. Literaturverzeichnis

Arendasy, M., Hornke, L. F., Sommer, M., Wagner, Menghin, M., Gittler, G., Häusler, J., Bognar, B. & Wenzl, M. (2010). *Intelligenz-Struktur-Batterie – INSBAT*. Mödling: Schuhfried GmbH.

Beckmann, J.F. (2000). Differentielle Latenzzeiteffekte bei der Bearbeitung von Reasoning Items. *Diagnostica*, 46, 124 - 129.

Boring, E.G. (1923) Intelligence as the tests test it. *New Republic*, 35, 35-37

S.D. Brown, A.J. Heathcote (2008). The simplest complete model of choice reaction time: Linear ballistic accumulation. *Cognitive Psychology*, 57, 153–178

Carroll, J.B. (1993). *Human cognitive abilities - A survey of factoranalytic studies*. New York: Cambridge University Press.

Doerfler, T., Hornke, L.H. (2010). Working style and extraversion: New insights into the nature of item response latencies in computer-based intelligence testing. *Journal of Research in Personality*, 44, 159–162

Ebel, R.L. (1953). The use of item response time measures in the construction of educational achievement tests. *Educational and Psychological Measurement*, 12, 391 - 401.

Eysenck, H. J. (1947). Student selection by means of psychological tests—a critical survey. *British Journal of Educational Psychology*, 17(1), 20-39.

Grudnik, J.L. & Kranzler, J.H. (2001). Meta - analysis of the relationship between intelligence and inspection time. *Intelligence*, 29, 523 - 535.

Heathcote, A., Wagenmakers, E.-J., Brown, S.D. (2014). The falsifiability of actual decision-making models. *Psychological Review*, 121, 676-678

Hick, W.E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, 4, 11 - 26.

Hiscock, C.K., Branham, J.D., Hiscock, M. (1994). Detection of feigned cognitive impairment: the two-alternative forced choice method compared with selected conventional tests. *Journal of Psychopathological Behaviour Assessment*, 16, 95 – 110.

Horn, J. L. (1989). Models for intelligence. In R. Linn (Ed.), *Intelligence: Measurement, theory and public policy* (pp. 29-73). Urbana: University of Illinois Press.

Horn, J.L., Noll, J. (1997). Human cognitive capabilities: Gf-Gc Theory. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theory, tests and issues* (pp. 49-91). New York: The Guilford Press

Hornke, L.F. (1997). Untersuchung von Itembearbeitungszeiten beim computergestützten adaptiven Testen. *Diagnostica*, 43, 27 - 39.

Hornke, L.F. (2000). Item Response Times in Computerized Adaptive Testing. *Psicológica*, 21, 175 - 189.

Hornke, L.F. (2005). Response time in computer - aided testing: a “Verbal Memory” test for routes and maps. *Psychology Science*, 47, 280 - 293. Jensen (1987)

Hosmer, D. W., & Lemeshow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics-Theory and Methods*, 9(10), 1043-1069.

Kubinger, K.D. (1989) *Moderne Testtheorie*. Weinheim: Beltz PVU

Kubinger, K. D. (2003). *Schlüsselbegriffe der psychologischen Diagnostik*. Weinheim: Beltz PVU.

Kubinger, K. D., J. Ebenhöf. "Arbeitshaltungen (AHA)." *Version 27* (2007).

Kubinger, K. D. (2009). *Psychologische Diagnostik: Theorie und Praxis psychologischen Diagnostizierens*. Hogrefe Verlag.

Lemmon, V.W. (1927). The relation of reaction time to measures of intelligence, memory and learning. *Archives of Psychology*, 15, 5 - 38.

Neubauer, A., Freudenthaler, H. H., & Pfurtscheller, G. (1995). Intelligence and spatiotemporal patterns of event-related desynchronization (ERD). *Intelligence*, 20(3), 249-266.

Peak, H. & Boring, E.G. (1926). The factor of speed in intelligence. *Journal of Experimental Psychology*, 9, 71 - 94.

Preckel, F., Wermer, C., & Spinath, F. M. (2011). The interrelationship between speeded and unspeeded divergent thinking and reasoning, and the role of mental speed. *Intelligence*, 39(5), 378-388.

Rammsayer, T. (1999). Zum Zeitverhalten beim computergestützten adaptiven Testen: Antwortlatenzen bei richtigen und falschen Lösungen. *Diagnostica*, 45, 178 - 183.

Rasch G. (1960) *Probabilistic Models for Some Intelligence and Achievement Tests*. Copenhagen: Danish Institute for Educational Research; p. 1960.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, 85(2), 59.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological review*, 111(2), 333.

Sommer, M., Häusler, J., Koning, A. J., & Arendasy, M. (2006). VALIDATION OF THE DUTCH AIRFORCE TEST BATTERY USING ARTIFICIAL NEURAL NETWORKS. *Proceedings from 48th Annual International Military Testing Association*, 1-16.

Sommer, M., Herle, M., Häusler, J., Risser, R., Schützhofer, B., & Chaloupka, C. (2008). Cognitive and personality determinants of fitness to drive. *Transportation research part F: traffic psychology and behaviour*, 11(5), 362-375.

Sommerfeld, E. (1994). *Kognitive Strukturen: Mathematisch-psychologische Elementaranalysen der Wissensstrukturierung und Informationsverarbeitung*. Waxmann.

Tate, M.W. (1948). Individual differences in speed of response in mental test materials of varying degrees of difficulty. *Educational and Psychological Measurement*, 8, 353 - 374.

Thurstone, L.L. (1937). Ability, motivation, and speed. *Psychometrika*, 2, 249 - 254.

Wagenmakers, E. J., Van Der Maas, H. L., & Grasman, R. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic bulletin & review*, 14(1), 3-22.

Von Glasersfeld, E. (1997). Radikaler Konstruktivismus. *Ideen, Ergebnisse, Probleme*, 1.

Voss, A., Voss, J., & Lerche, V. (2015). Assessing cognitive processes with diffusion model analyses: a tutorial based on fast-dm-30. *Frontiers in psychology*, 6.

8. Anhang

Lebenslauf und Abstract

Joseph-Sebastian Steinlechner

Name: Joseph-Sebastian Steinlechner
Geburtsdatum/Ort: 23.01.1987, London
Staatsbürgerschaften: Österreich, Kanada und Vereinigtes Königreich
Adresse: Türkenschanzstrasse 2/15
1180 Wien
Telefon: +43 650 9337339
E-Mail: joe@nc1.eu



BILDUNGSWEG

02/2012 - 06/2012 **Pflichtpraktikum** am Arbeitsbereich Sportpsychologie an der Universität Wien unter Dr. Björn Krenn
09/2009 - 10/2010 **Medizinstudium** an der Medizinischen Universität Wien
Seit 09/2007 **Diplomstudium der Psychologie** an der Universität Wien
08/2006 - 02/2007 **Absolvierung des Grundwehrdienstes**, Österreichisches Bundesheer
06 / 2006 **AHS-Matura**, Sir-Karl-Popper-Schule, Wien

BERUFSERFAHRUNG

02/2013 – aktuell **Leitung der Einsatzzentrale**, Call Us Assistance, Wien
04/2011 - 07/2012 **Kundenservice und Systembetreuung**, Apartmentzwerk, Wien
05/2008 - 07/2010 **Risikomanagement, Supervision und Projektleitung**, Wettbüro GmbH, tätig in Wien und Skopje
05/2007 - 09/2007 **Mitgliedswerbung**, Firma Wesser, tätig in München und Gießen
07/2006 – 09/2006 **Verkauf und Kundenführungen**, Swarovski Kristallwelten, Wattens
11/2005 - 01/2006 **Verkauf und Kundenbetreuung**, Swarovski Store Kärntnerstrasse, Wien
07/2005 – 09/2005 **Verkauf und Kundenführungen**, Swarovski Kristallwelten, Wattens

RELEVANTE FÄHIGKEITEN

Sprachen	Englisch und Deutsch perfekt in Wort und Schrift Italienisch fortgeschritten, Grundkenntnisse Französisch
Programme	Fortgeschrittene SPSS- und Office-Kenntnisse Fortgeschrittene Adobe Premiere, After Effects und Photoshop-
Kenntnisse	Basiskenntnisse MatLab und R Basiskenntnisse C#

ANDERE INTERESSEN

Sport	Hobby-Fußballspieler in der DSG-Meisterschaft Landhockeyspieler beim Post Sportverein Wien
Sonstiges	Sieger beim Bundeswettbewerb des philosoph. Essayschreibens 2006 6. Platz bei der Internationalen Philosophieolympiade 2006 NLP-Practitioner-Zertifikat von Trinergy International Absolvierung des Lehrgangs "Kommunikation und soziale Kompetenz" im Ausmaß von 220 Stunden

Abstract

Besonders in den letzten Jahren erfreuen sich computerisierte Leistungsverfahren zunehmender Beliebtheit, wofür es eine Vielzahl an Gründen gibt; abgesehen von der Einfachheit, mit der diese Verfahren unter standardisierten Bedingungen vorgegeben werden können, gibt es Variablen, die durch diese Vorgabemodalitäten einfach miterfasst werden können. Diese waren Testleiter_innen zuvor nur schwer oder sehr umständlich zugänglich.

Die vorliegende Arbeit widmet sich der Bearbeitungszeit einzelner Items eines kognitiven Leistungsverfahrens und der Frage, ob ein diagnostischer Mehrwert erzielt werden kann, wenn dieser besondere Aufmerksamkeit geschenkt wird.

Um dies zu überprüfen, wurden 211 Student_innen der Universität Wien untersucht. Es wurden logistische Regressionen durchgeführt, in der Hoffnung, den Erfolg von Untersuchungsteilnehmer_innen bei einzelnen Items durch Inklusion der Zeitkomponente (zusätzlich zum Fähigkeitsparameter der Teilnehmer_in) befriedigend vorhersagen zu können. Während gewisse Tendenzen, die im Rahmen dieser Arbeit auftraten (wie etwa jene, dass leistungsfähige Personen sich mehr Zeit für einzelne Items nehmen als weniger leistungsstarke), durchaus von Interesse für den Fachbereich der Psychologischen Diagnostik sein könnten, ergaben sich nicht die erhofften Ergebnisse – die Analyse der Bearbeitungszeit einzelner Items barg keinen diagnostischen Mehrwert hinsichtlich des Erfolgs bei der Bearbeitung.

Der Autor kommt in dieser Arbeit zu dem Schluss, dass die Interaktionen von Leistungsfähigkeit, Schwierigkeitsparameter und Bearbeitungszeit des Items nicht linearer Natur sein dürften, weshalb es komplexerer mathematischer Modelle bedarf, um diese in Bezug auf den Bearbeitungserfolg in Einklang zu bringen. Weitere Untersuchungen sind notwendig, um in dieser Hinsicht Klarheit zu schaffen.

Abstract

In psychological diagnostics, computerized testing, whether of personality or of aptitude, has significantly gained in popularity in recent times. There are several reasons for this, such as the simplicity with which standardized testing surroundings can be generated; also, certain variables that were previously difficult to assess or measure are now just a mouse click away for the test administrator.

This thesis focuses on the response latency of single items in a test of cognitive aptitude and aims to answer the question, whether additional diagnostic value can be achieved, by paying closer attention to the time individuals spend on single items.

In order to achieve this, 211 students of the University of Vienna were tested and logistic regressions were performed to ascertain whether the inclusion of a time component (in addition to a subject's ability and the difficulty of the item) can lead to satisfactory predictions in terms of a subject's success on single items.

While certain tendencies could be observed — for example, subjects with higher ability parameters were more likely to spend more time on single items than less skilled subjects — there seemed to be no added value in the inclusion of a time component in order to predict a subjects' success on single items.

There might be several reasons for this, the author concludes that the interactions between the subjects' ability, the difficulty of the item and time spent while working on the item are not linear in nature, which is why more complex statistical models are needed in order to clarify the relations between the aforementioned variables.