

Varianten von Reihenfolgetests und ihre Beziehung zu einigen Konzentrationstests

Abstract

Nach dem Konstruktionsprinzip des Konzentrationstests ZRF_20_5 wurde der Trail Making Test als Internetfassung umgesetzt sowie weitere Varianten konzipiert, welche statt arabischer Zahlen Zahlworte oder Mengen von Punkten darboten. Als Kriteriumsvariablen für diese Reihenfolgetests dienten Internetvarianten des TEMEKKO und d2 sowie der selbst konstruierte Konzentrationstest BSZ. Mit Ausnahme der Zahlwort- und Punktevarianten, die offenbar zu komplex erscheinen, korrelieren alle Tests hochsignifikant miteinander und laden auf einem Faktor, der am ehesten als Konzentrationsfähigkeit zu bezeichnen ist. Den Befunden nach kann der ZRF_20_5 als eine Form des Trail Making Test Part A angesehen werden. Der bereits nachgewiesene hohe Zusammenhang zwischen ZRF_20_5 und d2 in Höhe von $r=.62$ konnte erfolgreich repliziert werden. Insgesamt bestätigen die Daten dieser Studie, der ZRF_20_5 sei hinreichend geeignet, Konzentrationsfähigkeit oder Informationsverarbeitungsgeschwindigkeit zu messen.

Schlagworte

Konzentration, Konzentrationstests, Reihenfolgetests, ZRF_20_5, Trail Making Test, Temekko, d2, BSZ, psychologische Testdiagnostik

Einleitung, Zielsetzung und Vorgehen

Ziel der Studie war es, einige Varianten von Reihenfolgetests, darunter einen Trail Making Test in der von Jacobs konstruierten Testkonzeption empirisch zu überprüfen und die Beziehungen zu bekannten sowie einem neu konstruierten Konzentrationstests zu ermitteln. In der Literatur existieren etliche Varianten von Reihenfolgetests. Im deutschsprachigen Raum wäre an erster Stelle der Zahlen-Verbindungs-Test (ZVT) von Oswald und Roth (1987) zu nennen, ansonsten der wohl am meisten bekannte und in der neuropsychologischen Praxis häufig eingesetzte Trail Making Test (TMT) (Reitan, 1958, 1992), dessen Entstehungsgeschichte weit in die Vergangenheit zurück reicht. Der selbst konstruierte, via Internet lauffähige, Zahlenreihenfolgetest ZRF_20_5 (Jacobs 2013a, 2015b) basiert auf derselben Grundidee wie der TMT A. Er sollte daher auch Vergleichbares messen, was ja bereits vom ZVT, empirisch begründet, behauptet werden kann (Jacobs (2013b)). Zum einen sollte geprüft werden, ob die unterschiedliche Anzahl der zu vergleichenden Zahlen in TMT A und ZRF_20_5 überhaupt eine Relevanz besitzt, zum anderen sollte eine internetfähige TMT B-Variante erstellt und die Beziehungen der Trail Making Varianten untereinander sowie zu sonstigen Tests geprüft werden.

Als Validitätskriterien für den ZRF_20_5 bzw. die Trail Making Varianten kamen 2 weitere Konzentrationstests zum Einsatz. Beide erfordern als eine Teilleistung die Beachtung von Reihenfolgen. Darüber hinaus verlangen sie aber auch häufiges Umdenken bzw. kognitive Flexibilität. Insofern ließen sie einen ansprechenden Zusammenhang mit dem Trail Making Test, insbesondere Form B vermuten. Schließlich war geplant, den früher bereits nachgewiesenen recht bedeutsamen Zusammenhang des Zahlenreihenfolgetests mit einer Internetvariante des bekanntesten deutschen Konzentrationstests, dem d2 (Brickenkamp, R. Schmidt-Atzert, L., Liepmann, D. (2010), zu replizieren.

Die Studie erforderte aus mehrfachen Gründen 2 Erhebungszeitpunkte, die 3 bis 4 Wochen auseinander lagen. Aus Anhang A geht hervor, zu welchem Zeitpunkt welche Testverfahren zum Einsatz kamen. Als Probanden dienten SeminarteilnehmerInnen (Alter 23,5, 70% weiblich) aus 4 Seminaren des Verfassers im SS 2015. Sie wurden via Email auf Passwort geschützte Seiten im Internet verwiesen und bearbeiteten die Tests zu Hause am Desktop-Computer oder Laptop mit Mausanschluss. Der Erhebungsaufwand wurde ihnen als eine Seminarsitzung angerechnet und etliche Erkenntnisse sowie Daten aus der Erhebung für didaktische Lehrzwecke in den Seminaren genutzt. Um die Testmotivation der Studierenden zu fördern, erhielten diese nach jeder Testbearbeitung Rückmeldung über ihr Ergebnis (etwa Bearbeitungszeit und Fehler), allerdings zunächst ohne soziale Vergleichsmöglichkeiten. Ihnen wurde jedoch versichert, nach Abschluss jeder Erhebung Einblick in die Leistungen aller Studierenden nehmen zu können, um ihre eigene Leistung später anhand ihrer relevanten sozialen Bezugsgruppe bewerten zu können.

Was sind nützliche Anordnungen für die Konstruktion von Reihenfolgen?

Bei einer Reihenfolgeaufgabe muss die Testperson ungeordnete Elemente in die korrekte Reihenfolge bringen. Soll auf diese Weise Konzentration gemessen werden, so kommt es darauf an, fehlerfrei möglichst schnell die zutreffende Reihenfolge herzustellen. So hat die Testperson im Kurzzeitkonzentrationstest Zahlenreihenfolge ZRF_20_5 die Aufgabe, die zufällig angeordnete Zahlen 1 bis 20 in der ansteigenden Reihenfolge anzuklicken. Nach Salthouse et al. (2000) bestimmen 4 Prozesse die Leistung in einem simplen Reihenfolgetest (Zahlen oder Buchstaben).

Tabelle 1:
Prozessmodell für simplen Reihenfolgetest nach Salthouse et al. (2000)

Nr.	Prozess	Beispiel
1.	target identifikation	Jetzt muss die Zahl 3 angeklickt werden
2.	search-comparison	Suche die Zahlen ab, bis die Zahl 3 gefunden ist
3.	response	Klicke auf die Zahl 3
4.	sequence update	Erhöhe die Zahl 3 um 1; = 4

Je nach speziellem Test können die Anforderungen an die einzelnen Prozesse mehr oder weniger schwierig ausfallen und dadurch bedingt Bearbeitungszeit und/oder Fehleranfälligkeit variieren.

Zur Konzeption von Reihenfolgetests sind arabische Zahlen am besten geeignet, weil fast alle Menschen diese Zahlendarstellung kennen und häufig mit ihr umgehen. Arabische Zahlen können sehr schnell wahrgenommen und gut miteinander verglichen werden. Die zutreffende Zahlenanordnung der ersten 20 oder 25 Zahlen steht automatisiert zur Verfügung. Deshalb beansprucht die Regel "Ordne die Zahlen der Größe nach" in keiner ernsthaften Weise das Gedächtnis. Mithin beschränkt sich die eigentliche Konzentrationsleistung ausschließlich auf ein triviales Hochzählen der jeweils nächsten Zahl und ein stetes Vergleichen der anvisierten Zahlen im Hinblick auf Übereinstimmung mit der Zielzahl.

Als nächst beste Form einer Anordnungsaufgabe eignen sich vermutlich die Buchstaben des Alphabets. Allerdings schränkt diese Präsentationsform den universellen Einsatz eher ein, weil die Reihenfolge des Alphabets mitunter von der Sprache abhängt -etwa: deutsches vs. [griechisches Alphabet](#)- sowie etliche Personen die korrekte Anordnung nicht zuverlässig beherrschen. Zudem erfordert die alphabetische Anordnung teilweise eine mentale Rekonstruktion aus dem Gedächtnis, (z. B: "Was kommt nach F?" *A B C D E F G*). Sie erschwert den Prozess "sequence update", weil die korrekte Reihenfolge nicht so hoch automatisiert zur Verfügung steht wie bei den arabischen Zahlen. Trotz dieser Einschränkungen lieferten zumindest Studierende im Buchstabenfolge- und Zahlenfolgetests sehr ähnliche Ergebnisse, auch wenn der Buchstabenfolgetest etwas mehr Bearbeitungszeit beanspruchte.

Bisherigen Ergebnissen von Jacobs (2014b, 2015a) zufolge messen nachfolgende Tests, vornehmlich die beiden ersten, in hohem Maße dasselbe und können somit als Parallelversionen aufgefasst werden.

1. [ZRF_20_5](#) : Zahlenreihenfolge ansteigend von 1 nach 20
2. [RZRF_20_5](#) : Zahlenreihenfolge abfallend von 20 nach 1
3. [BRF_20_5](#): Buchstabenreihenfolge ansteigend von A nach T

Es erweist sich aber als recht schwierig, nach Zahlen oder Buchstaben weitere zwingende Folgen eines größeren Umfangs bis mindestens 20 Elementen zu finden, welche den Personen mit vergleichbarer Evidenz bekannt sind und für Vergleichsoperationen ähnlich problemlos zur Verfügung stehen. Statt nach sonstigen Folgen zu suchen, könnte man auf die Idee kommen, Zahlen in einem anderen Darstellungsmodus heranzuziehen, etwa als [römische Zahlen](#). Diese müssten allerdings von etlichen Personen zunächst erlernt oder gründlich eingeübt werden und eignen sich deshalb nicht. Hier wurde der durchaus problematische Versuch unternommen, Zahlen als Menge von Punkten oder als Zahlworte darzustellen.

Die eingesetzten Konzentrationstests

ZRF_20_5

Der **Zahlenreihenfolgetest 20_5 (ZRF_20_5)** ist ein Kurzzeit-Konzentrationstest, der eine schnelle Erfassung der Konzentrationsfähigkeit über Internetbrowser ermöglicht. Hierbei hat die Testperson die Aufgabe, die zufällig angeordneten Zahlen 1 bis **20** insgesamt **5** mal fehlerfrei anzuklicken. (siehe Näheres unter Jacobs 2015b - Testmanual bei Psychometrikon). Als einziger Konzentrationsmesswert gilt die durchschnittliche Bearbeitungszeit in Sekunden.

Bei der Testbearbeitung realisierten 2 Probanden in bisher ungewohnter Weise 6 Fehlversuche, was normalerweise zum Ausschluss führt. Bei diesen beiden Probanden wurde der Durchschnitt aus den verfügbaren Zeiten (mindestens 3 Durchgänge) gewählt. Tabelle 2 veranschaulicht die deskriptiven Ergebnisse. Sie entsprechen weitgehend den bisherigen Befunden. Lediglich der Mittelwert fällt etwas besser als in der studentischen Normenstichprobe ($M=20.2$, $s=4,4$) aus. Allerdings hatten 22 Studierende (ca. 1/5 der Probanden) den Test bereits in früheren Semestern mindestens einmal bearbeitet.

Tabelle 2: Bearbeitungszeiten (sec) des ZRF_20_5 (N=101)

	α	M	s	Md	Schiefte	Kurtosis
ZRF_20_5	.89	19.4	4.5	19	.75	.20

Entsprechend bisherigen Erfahrungen sind Bearbeitungszeit und Anzahl der Fehlversuche unabhängig voneinander ($r_{\text{Spearman}}=0.04$). Ebenso es gibt keine signifikanten Zusammenhänge des ZRF_20_5 mit Alter, Geschlecht und Abiturnotendurchschnitt. Wenn man ziemlich konservativ für jede Testperson einen Fehlversuch hinzurechnet (Mittelwert = 0.75, Median=0 Fehlversuche), so umfasst die reine Bearbeitungszeit des ZRF_20_5 im Durchschnitt ca. 2 min bzw. genau 112 sec.

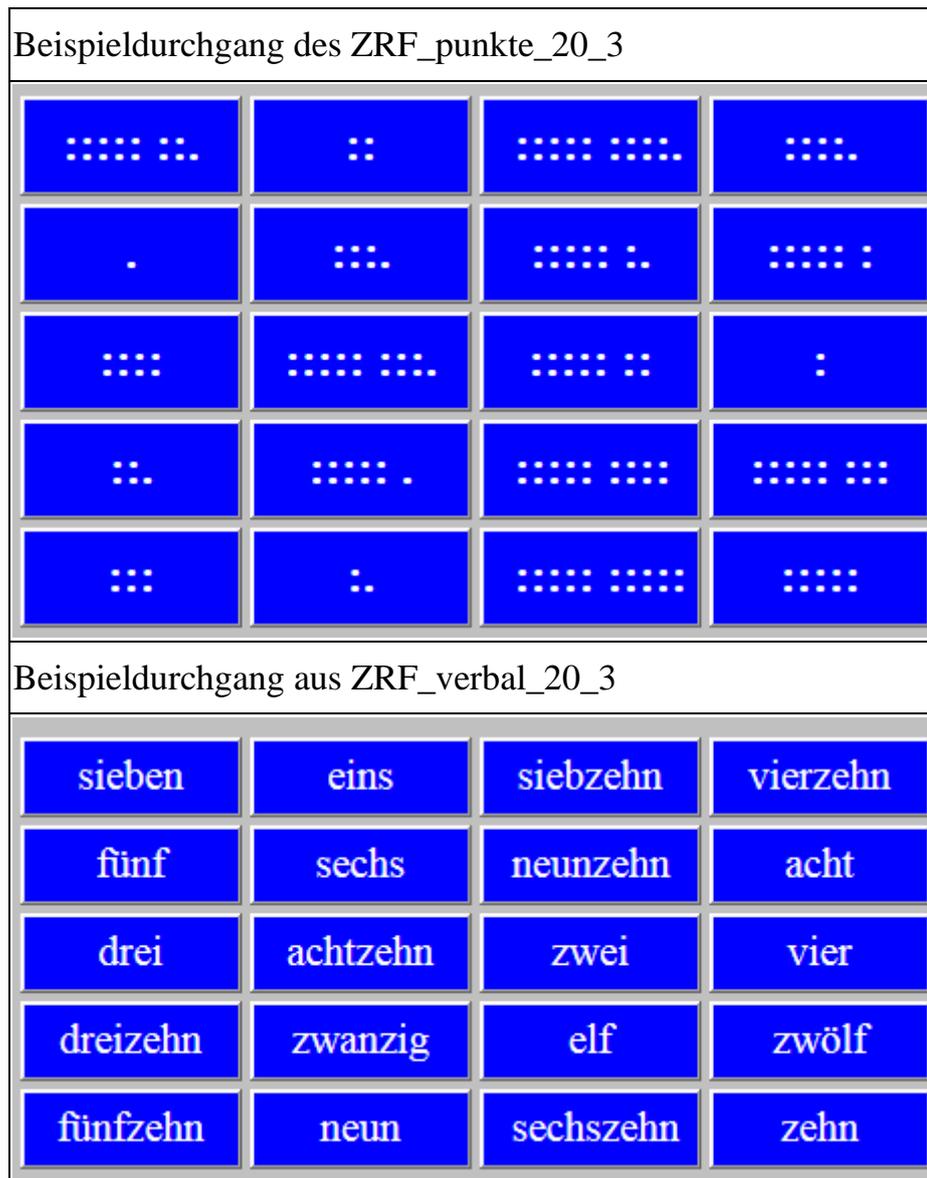
Zahlen als Mengen oder Zahlworte

Um analog zum ZRF_20_5 einen figuralen und verbalen Reihenfolgetest zu verwenden, wurden weitere Tests konstruiert, bei denen statt arabischer Zahlen die Menge von Punkten oder Zahlwörter in die korrekte Reihenfolge gebracht werden sollten. Hierbei sind der Testperson zwar die korrekten Anordnungen ohne aufwendigen Rückgriff auf das Gedächtnis unmittelbar evident. Allerdings lagen bereits vor dieser Konstruktion erhebliche Bedenken vor, auf diese Weise ähnlich wie bei den arabischen Zahlen besonders gut Konzentration messen zu können, weil es sich bei diesen Reihenfolgevarianten eher um komplexe Reize handelt, was vornehmlich den Prozess 3 im Prozessmodell der Tabelle 1 deutlich erschwert. Mithin werden die Ergebnisse vermutlich auch wesentlich mitbestimmt durch weitere Fähigkeiten der Probanden wie Wahrnehmungsgenauigkeit, Lesefähigkeit und kognitive Verarbeitungskapazität.

Die Verwendung dieser Tests diene somit vornehmlich Erkundungszwecken. Wegen der ungewohnten Darstellung von Zahlen als Zahlworte oder Punkte lag die eindeutige Erwartung zugrunde, beide Alternativversionen benötigten im Vergleich zum ZRF_20 erheblich mehr Bearbeitungszeit. Das wurde auch bereits zuvor durch Eigenversuche des Autors überdeutlich belegt. Es war jedoch von Interesse genauer zu erfahren, in welcher Größenordnung Ähnlichkeiten zwischen den Leistungen der unterschiedlichen Präsentationsmodi ausfallen, um abschätzen zu können, ob und wie stark die komplexeren Varianten auch von der Konzentrationsfähigkeit abhängen.

ZRF_Punkte_20_3 und ZRF_verbal_20_3

ZRF_Punkte_20_3 und ZRF_verbal_20_3 wurden analog der Prinzipien des ZRF_20_5 konstruiert. Die Anordnung wurde jeweils nach Zufall bestimmt und nur korrekte Durchgänge kamen in die Testwertung. Beiden Tests ging eine gemeinsame Testanweisung sowie ein Probedurchgang für die Punkteversion voraus, da das generelle Prozedere bereits aus der Bearbeitung des ZRF_20_5 bekannt war. Um die Gesamtbearbeitungszeit in Grenzen zu halten, waren in beiden Tests lediglich 3 fehlerfreie Durchgänge verlangt worden. Abbildung 1 zeigt jeweils einen möglichen Testdurchgang.

Abbildung 1: Beispiele für ZRF_punkte und ZRF_verbal

Insgesamt wurden im ZRF_Punkte_20_3 sehr viele Fehler gemacht (durchschnittlich 1,5 Fehlversuche für 3 notwendige korrekte Durchgänge). 7 Probanden lieferten durch 6 Fehlversuche gänzlich ungültige Ergebnisse ab. 2 weitere Probanden mussten als Ausreißer aus dem Datensatz genommen werden. Im ZRF_verbal_20_3 erbrachten hingegen lediglich 2 Probanden ungültige Daten und einer musste als Ausreißer ausgeschlossen werden.

Tabelle 3: Bearbeitungszeiten (sec) für ZRF_Punkte_20_3 und ZRF_verbal_20_3

	α	M	s	Md	Schiefe	Kurtosis	N
ZRF_Punkte_20_3	.91	53.3	19.2	48.8	1.67	3.53	91
ZRF_verbal_20_3	.75	39.1	8.8	38.4	.37	.04	96

Der ZRF_Punkte_20_3 weicht etwas von der NV ab. Er lässt auch eine gewisse Übungsanfälligkeit erkennen, da bereits im zweiten und dritten Testdurchgang signifikante Zeitvorteile gegenüber dem ersten Durchgang erzielt wurden. Beim ZRF_Punkte_20_3 ergab sich zudem eine für Speedtests ungewöhnliche, signifikant **positive** Korrelation

zwischen Bearbeitungszeit und Anzahl der Fehlversuche (Spearman Rangkorrelation $r_s = .33$; $p_z = .002$). Je mehr Zeit eine Person benötigte, die Punkte der Menge nach zu ordnen, desto mehr Fehler traten zu Tage. Dieser positive Zusammenhang sowie die lange durchschnittliche Bearbeitungszeit weisen auf die hohen, für manche zu hohen Anforderungen der Aufgabenstellung hin. Im Gegensatz zur Punktevariante waren die Bearbeitungszeiten und Fehlversuche beim ZRF_verbal_20_3 unabhängig voneinander und keine Übungsgewinne während der Testung zu finden. Eine korrekte Anordnung der Zahlworte erforderte aber bereits ca. doppelt so viel Bearbeitungszeit wie die einer numerischen Reihenfolge (siehe Ergebnisse zu ZRF_20_5 in Tabelle 2), so dass die beiden alternativen Reihenfolgetests schon aus ökonomischen Gründen dem ZRF_20 deutlich unterlegen sind.

Varianten des Trail-Making-Tests

Der Trail Making Test (Reitan 1958, 1992) umfasst die zwei Testformen A und B. In seiner Gesamtheit soll der Trail Making Test etliche Prozesse erfassen, wie z.B. „visual attention, visual scanning, visuospatial sequencing, speed of processing, mental flexibility, executive function, cognitive-set shifting.“ Mittlerweile existieren etliche Varianten des TMT sowohl in einer Papier- wie Computerfassung (siehe z.B. einige dieser Varianten in Buck et al., 2008).

Der ZRF_20_5 kann als eine Variante des Trail Making Test Part A (TMT A) aufgefasst werden, da die Testperson in beiden Tests ungeordnete Zahlen in eine aufsteigende Reihenfolge bringen soll. TMT A und ZRF_20_5 unterscheiden sich hauptsächlich in der Anzahl der Zahlen (25 vs. 20), der Anordnung der Zahlen (über einen großen Bereich verteilt vs. nach Zufall in Tabellenform) und der Art und Weise, die Zahlen der Reihenfolge nach zu bestimmen (mit Bleistift verbinden vs. der Reihenfolge nach mit der Maus anklicken). Während der TMT A sich mit einem Versuch begnügt, muss die Testperson im ZRF_20_5 insgesamt 5 fehlerfreie Durchgänge absolvieren. Dadurch wird eine Konsistenzbestimmung ermöglicht, die beim ZRF_20_5 in der Nähe von $\alpha=90$ liegt.

Wegen der häufigen Anwendung des **TMT**, vornehmlich in der neuropsychologischen Forschung und Praxis, wurde deshalb eine weitere Annäherung an diesen Test vorgenommen und analog der Konzeption des ZRF_20_5 (Jacobs 2015b) zwei weitere Online-Varianten eingeführt:

TMT A_v: bzw. ZRF_25_3

TMT B_v: bzw. ZBRF_25_3

Die Bezeichnung **_v** soll klarstellen, dass es sich um eine Variante und nicht das Original des TMT handelt. Die alternativen Spezifikationen präzisieren das nähere Vorgehen. Beim ZRF_25_3, hier auch TMT A_v [Variante von TMT A] genannt, muss die Testperson die ersten **25** Zahlen insgesamt **3** mal fehlerfrei in der korrekten Reihenfolge anklicken. Im ZBRF_25_3, hier auch TMT B_v genannt, hat sie die Aufgabe, entsprechend der Vorgehensweise im TMT B, jeweils zwischen Zahlen- und Buchstabenfolge abzuwechseln (siehe Abbildung 2). Dadurch erweitert sich Prozess 4 (sequence update). Denn es gibt nun zwei Anordnungen (Zahlen und Buchstaben). Zunächst muss die bisherige Anordnung gewechselt werden und dann der Zähler der neu anstehenden Anordnung um 1 hoch gesetzt werden.

Abbildung 2: Beispieldurchgang für den ZBRF_25_3 bzw. TMT B_v

10	2	A	1	3
13	I	7	D	H
E	G	C	B	12
K	6	4	9	L
J	5	11	8	F

Korrekte Anordnung: 1 A 2 B 3 C 4 D 5 L 13)

Wegen der im Vergleich zum ZRF_20_5 höheren Anforderungen und dem dadurch bedingten größeren Zeitbedarf, wurde die Anzahl der korrekten Durchgänge jeweils auf 3 reduziert, um noch hinreichend reliable Kurzzeittests mit ökonomisch akzeptablen Bearbeitungszeiten zu ermöglichen. Bei jedem Durchgang werden die Zeichen nach Zufall angeordnet, so dass Abbildung 2 eine der 25! möglichen Anordnungen darstellt. Bei einem Fehler wird die Bearbeitung abgebrochen und ein neuer Durchgang vorgelegt. Als Konzentrationstestwert dient ausschließlich die durchschnittliche Bearbeitungszeit aus drei korrekten Durchgängen.

Tabelle 4: Deskriptive Ergebnisse zu TMT A_v und TMT B_v

TMT A_v bzw. ZRF_25_3 (N=98)

	α	M	s	Md	Schiefe	Kurtosis
Zeit (sec)	.82	31.3	6.8	30.2	.7	.5
Fehlversuche		0.4	.8	0.0	4.0	22.0

TMT B_v bzw. ZBRF_25_3 (N=97)

	α	M	s	Md	Schiefe	Kurtosis
Zeit (sec)	.88	44.0	9.6	42.9	.6	.7
Fehlversuche		1.1	1.5	1.0	1.4	1.2

Wie aus Tabelle 4 hervorgeht, erzielten beide Tests durchaus ansprechende Konsistenzreliabilitäten. Diese liegen im erwarteten Rahmen und fallen beim TMT B_v sogar etwas höher als vermutet aus. Diesem Test liegt im Vergleich zum TMT A_v aber auch mehr Gesamtbearbeitungszeit zugrunde. Aus den Daten lassen sich solide Schätzungen für die durchschnittliche reine Bearbeitungszeit ableiten. Sie beträgt beim TMT A_v ca. 106 sec und beim TMT B ca. 180 sec. Während der TMT A_v und ZRF_20_5 sehr ähnliche Gesamtbearbeitungszeiten aufweisen, ist die Reliabilität des ZRF_20_5 mit ca. .90 höher einzuschätzen. Da der TMT B_v durch den permanenten Wechsel zwischen Zahlen - und Buchstabenfolge deutlich anspruchsvollere Leistungen verlangt als der

TMT A_v, verwundert es nicht, dass bei diesem signifikant mehr Fehlversuche absolviert wurden (Wilcoxon $z=4.9$). Während Fehlversuche und Bearbeitungszeiten, wie aus den konsistenten Befunden bisheriger Studien zum ZRF_20_5 bekannt, auch beim TMT A_v unabhängig voneinander sind, findet man beim TMT B_v einen positiven Zusammenhang von $r_s = .26$ ($p_z = .009$). Dieser deutet, ähnlich wie bei der oben geschilderten Punkteversion, auf die besondere Schwierigkeit des TMT B_v hin. Da wenig leistungsfähige Personen sowohl mehr Zeit benötigen wie auch mehr Fehler machen.

In allen 3 Durchgängen des TMT A_v wurden hoch vergleichbare Mittelwerte der Bearbeitungszeiten festgestellt. Es findet somit während des Testens kein Lernen statt. Im TMT B_v erzielten die Probanden jedoch im Verlauf der Testung kontinuierlich leichte Übungsgewinne. So fällt die Bearbeitungszeit des dritten Durchgangs hochsignifikant in einem Ausmaß von ca. $d = 1/3$ Effektstärke geringer aus als die des ersten Durchgangs. Das Ergebnis deutet in die Richtung, durch weitere gezielte Übungen bedeutsame Leistungsfortschritte im TMT B_v erreichen zu können. Höchstwahrscheinlich werden die Übungsgewinne durch eine zunehmend bessere Repräsentation der gesamten Zeichenanordnung erlangt (=verbessertes update sequencing) und enden erst bei einer sicheren und hoch verfügbaren Automatisierung der verlangten Zeichenanordnung.

Beziehungen der Reihenfolgetests untereinander

Der interessierte Leser findet im Anhang C eine Interkorrelationsmatrix aller eingesetzten Konzentrationstests. Im Folgenden werden thematisch zusammengehörige Teile daraus getrennt behandelt und noch einige Ergänzungen und Differenzierungen vorgenommen. Zur besseren Übersicht sind in Tabelle 5 die Interkorrelationen zwischen ZRF_20_5 und den TMT-Varianten in Fettdruck dargestellt und die übrigen Beziehungen kursiv gehalten. Wie aus Tabelle 5 ersichtlich, korreliert der ZRF_20_5 erwartungsgemäß mit dem TMT A_v etwas höher als mit dem TMT B_v, was insofern nicht verwundert, als die beiden erstgenannten Tests sich letztlich nur in der Testlänge unterscheiden. Man kann den Zusammenhang zwischen ZRF_20_5 und TMT A_v von $r = .79$ auch als eine Art Retest deuten. Denn der ZRF_20_5 wurde 3 bis 4 Wochen früher erhoben. Jedenfalls entspricht der Wert $r = .79$ in hohem Maße den bisher ermittelten Retestkoeffizienten des ZRF_20_5. Bereinigt man die Messfehler beider Tests durch doppelte Minderungskorrektur, so ergibt sich ein "wahrer Zusammenhang" von r doppelt-minderungskorrigiert = $.92$. Auch die Korrelationen beider Zahlenreihenfolgetests mit dem TMT B_v in der Höhe von r ca. $.70$ fallen recht ähnlich aus. Mithin spielt die unterschiedliche Gesamtanzahl von ZRF_20_5 und TMT A_v keine wesentliche Rolle.

Tabelle 5: Interkorrelationen der Reihenfolgetests (N=90-93)

	ZRF_20_5	ZRF_25_3	ZBRF_20_3	ZRF_
		TMT A_v	TMT B_v	Punkte
ZRF_20_5	-			
ZRF_25_3 [TMT A_v]	.79	-		
ZBRF_25_3 [TMT B_v]	.67	.72	-	
ZRF_Punkte_20_3	<i>.11</i>	<i>.15</i>	<i>.10</i>	-
ZRF_verbal_20_3	<i>.26</i>	<i>.23</i>	<i>.21</i>	<i>.55</i>

Hinweis: TMT A_v und TMT B_v wurden ca. 3 bis 4 Wochen später als die übrigen Tests erhoben

Obgleich der TMT B_v komplexere Anforderungen als der TMT A_v stellt, gehen die Leistungen in beiden Tests offenbar zu einem hohen Anteil auf dieselbe Fähigkeit zurück. Dies ist möglicherweise der Grund dafür, warum beim TMT häufig auch abgeleitete Kennwerte wie B-A gebildet werden. Man will offenbar den vom TMT A bereinigten Anteil an TMT B erfassen.

Ganz anders sieht die Lage aus, wenn man die schwachen Beziehungen der komplexeren Zahlenvarianten in Form von Punkten oder Zahlworten mit den übrigen Reihenfolgetests vergleicht. Beide Varianten korrelieren zwar noch untereinander in ansprechender Höhe von $r=.55$ ($rs=.60$), stehen aber in keinem hinreichenden Zusammenhang mit den übrigen Reihenfolgetests sowie auch mit den weiter unten noch vorgestellten weiteren Konzentrationstests. Insbesondere der hoch reliable Punktetest erfasst definitiv etwas ganz anderes als Konzentrationsfähigkeit. Schon jetzt ist offensichtlich, dass die Punkte- und Zahlwort-Reihenfolgevarianten als Konzentrationstests ungeeignet sind.

Weitere Konzentrationstests als Kriteriumsvariablen für die Reihenfolgetests

Als Kriteriumsvariablen zur Bestimmung der konvergenten Validität kamen insgesamt 3 weitere Konzentrationstests zum Einsatz. Alle diese Tests waren als Internet fähige Varianten konzipiert worden. Neben dem allseits bekannten d2 wurden zwei bisher noch nicht genutzte Tests herangezogen, die in irgendeiner Weise auch die Beachtung von Reihenfolgen verlangten. Zunächst werden die Tests näher beschrieben sowie die deskriptiven Ergebnisse mitgeteilt und anschließend die Beziehungen zu den Reihenfolgetests aufgezeigt.

d2c

Zum Erhebungszeitpunkt 2 wurde eine Internetversion des d2 (Brickenkamp, Schmidt-Atzert, Liepmann, 2010), im Folgenden d2c (Computervariante) genannt, zur Bearbeitung vorgelegt (siehe Genaueres bei Jacobs 2014a, S. 13; ein Beispiel im Anhang B). Bei diesem Test muss die Testperson möglichst schnell alle d mit zwei Strichen anklicken und etliche Distraktoren wie d mit einem oder mehr als 2 Strichen sowie alle p mit beliebig vielen Strichen ignorieren. Diese d2-Version entspricht keiner exakten Kopie des originären d2, soll aber in Form eines Kurztests hoch Vergleichbares messen. In Erweiterung von Jacobs (2014a) wurde vor der eigentlichen Testung noch ein nicht gewerteter kurzer Probedurchgang hinzugefügt.

Zunächst mussten 18 Probanden ausgeschlossen werden, da sie insgesamt mehr als 25% Auslassungsfehler realisierten. 8 dieser Probanden lieferten aber einen gültigen ersten Durchgang ab, weswegen dieser Durchgang dann als Test gewertet wurde. Schließlich wiederholten 5 Studierende den d2c in gültiger Form. Ein nach Zufall bestimmter Teil der Studierenden hatte den Test unter Beschallung einer Rede von Angela Merkel bearbeitet, während die übrigen Studierenden den Test unter Normalbedingungen absolvierten (siehe [Jacobs 2015c](#)). Da zwischen beiden Gruppen keine Unterschiede zwischen den wesentlichen Konzentrationsmaßen festzustellen waren, wurden alle Ergebnisse zusammengefasst.

Als entscheidende Konzentrationsmaße gelten

1. **d2c_zeit** = Arbeitstempo: die Bearbeitungszeit in Sekunden
2. **d2c_leistung** = Fehleradjustierte Anzahl korrekter Zeichen pro Minute:
 $(R - AF - VF) / \text{Bearbeitungszeit in Minuten}$
[R= Anzahl richtige Zeichen; AF=Anzahl Auslassungsfehler VF=Anzahl Verwechslungsfehler]
3. **d2c_VF** = Anzahl der Verwechslungsfehler

Tabelle 6: Ergebnisse zum d2 (N=85-91)

	α	M	s	Md	Schiefte	Kurtosis
d2c_zeit (sec)	.94	145	23	142	.7	.9
d2c_richtig	.72	102	4.1	104	-1.7	3.5
d2c_VF	.72	0.4	0.8	0	2.8	7.8
d2c_leistung	.89	41.5	7.1	41.3	0.0	0.1

Hinweis: d2c_Auslassungsfehler = 106-d2c_richtig; α von AF deshalb ebenfalls .72.

Bei der d2c_leistung werden AF quasi doppelt gewichtet, da sie bereits in d2c_richtig enthalten sind. Die doppelte Gewichtung bewirkt allerdings nur minimale Effekte, da die Korrelation zwischen d2c_leistung mit einer Leistungsvariante ohne Abzug der AF $r = .95$ beträgt.)

Bearbeitungszeit und Genauigkeit (d2c_richtig) entsprechen in hohem Maße den bisherigen Befunden (Jacobs 2014a). In die d2c_leistung fließt vorwiegend das Arbeitstempo ein, da Bearbeitungszeit und Leistung $r = -.85$ miteinander korrelieren.

Temekko

Eine ausführliche Beschreibung des originären TEMEKKO findet der interessierte Leser in der Dissertation von Verena Schmid (2010). Der Test wurde theoriefundiert auf der Basis etlicher Literaturanalysen zur Konzentration nach einer rationalen Methode konstruiert. Im Fazit ihrer Validierungsstudien kommt Schmid (2010, S. 73) zu der Einschätzung. "Aufgrund der sehr guten Reliabilitäten des TEMEKKO sowie der belegten Konstruktkonvergenz mit Konzentration kann das Verfahren als Markertest zur Erfassung von Konzentration bzw. exekutiver Kontrolle bezeichnet werden". Basierend auf den bei der Autorin in Kapitel 3 dargelegten Konstruktionsprinzipien wurde hier unter Nutzung etlicher Freiheiten eine sehr schlanke Internetvariante des numerischen Teils von Temekko entwickelt. Sie greift vorwiegend aus Ökonomiegründen lediglich auf das Grundprinzip zurück, die Reihenfolge von drei Zahlen im Hinblick auf strikten Anstieg oder Abfall zu bewerten und verzichtet auf alle weiteren, zielirrelevanten Darstellungsvarianten. Diese verfolgten vornehmlich den Zweck, Ablenkung zu erzeugen und auf diese Weise die Aufgabenschwierigkeit zu erhöhen, um so Fehler zuverlässiger erfassen zu können. Somit fehlt hier vollständig der durchaus wichtige Punkt 5 bei Schmid 2010, S. 36. Aber auch die Auswahl der Distraktoren berücksichtigt nicht alle möglichen Kombinationen.

Testkonstruktion des Temekko_v

Bei der numerischen Internetvariante des Temekko, im folgenden Temekko_v genannt, sehen die Testpersonen in 2 Testdurchgängen jeweils 96 Anordnungen mit 3 Ziffern hintereinander und sollen jeweils entscheiden, ob die Ziffern strikt ansteigen oder strikt abfallen. Unmittelbar vor Testbeginn werden pro Durchgang nach etlichen Zufallsprozessen aus einer großen Menge möglicher korrekter und falscher Reihenfolgen 28 zutreffende und 68 falsche Zuordnungen erzeugt und in jeweils zufälliger Itemreihenfolge

dargeboten. Abbildung 3 stellt somit eine der möglichen Testvorgaben für den ersten Testdurchgang dar.

Abbildung 3: Ein Beispiel für den ersten Testdurchgang des Temekko_v

Testdurchgang 1

9 6 8	1 7 8	9 3 7	9 8 8	8 7 9	7 9 7	9 9 5	9 4 9
1 3 1	1 4 2	6 6 1	5 6 2	9 8 9	2 7 8	3 1 1	7 8 9
9 6 3	3 6 3	9 4 3	1 3 2	9 7 1	5 9 9	7 4 9	7 8 8
3 6 5	5 3 3	8 8 8	3 3 1	6 7 7	8 8 8	9 8 1	4 6 8
4 4 2	7 5 4	9 5 9	4 5 3	6 7 7	6 7 9	9 9 8	2 3 8
3 3 6	1 6 5	8 5 4	2 5 7	8 7 5	8 5 9	8 6 7	1 1 4
7 7 9	7 4 8	7 2 2	4 4 4	5 7 5	9 6 9	2 4 1	6 8 4
6 1 1	1 3 2	6 6 7	1 5 3	6 6 9	9 5 4	3 4 2	5 8 1
9 7 3	3 8 9	6 8 7	5 7 3	4 5 9	8 8 6	1 1 3	4 6 1
6 8 1	4 5 1	2 3 4	9 8 6	5 5 5	9 9 9	9 4 8	6 4 3
3 4 1	2 3 2	1 5 8	3 5 6	9 4 2	6 6 7	5 5 6	4 9 5
9 9 5	3 7 9	1 3 2	8 7 2	5 4 1	2 5 8	1 4 9	1 1 4

bestätigen

Testanweisung

Zunächst erhielten die Testpersonen eine ausführliche Erklärung, wie der Ausschnitt aus nachfolgende Angaben der Testanweisung belegt.

Hier ein paar Beispiele:

richtig ansteigend	1 3 7	2 4 5	5 7 9
richtig abfallend	4 3 1	9 8 2	7 2 1

falsch	1 1 7	2 5 5	3 4 3	5 7 4
falsch	4 3 3	9 5 7	8 8 8	7 2 3

Richtig liegt demnach vor, wenn von den 3 Ziffern...

- a) die nachfolgende Ziffer immer höher als die vorherige Ziffer ausfällt, oder
- b) die nachfolgende Ziffer immer niedriger als die vorherige Ziffer ausfällt.

Alle anderen Anordnungen sind falsch.

Ihre Aufgabe besteht darin, **so schnell wie möglich, aber dennoch korrekt**, alle richtigen Anordnungen anzuklicken. Vor dem eigentlichen Test absolvierten die Probanden eine Probetestung mit 2 Durchgängen von jeweils 16 Items und erhielten Rückmeldung zu ihrem Ergebnis in Form des Prozentsatzes der korrekten Lösungen.

Testergebnisse

3 Probanden wurden wegen extremer Fehler in beiden Durchgängen (vornehmlich Auslassungsfehler) vollständig aus dem Datensatz genommen. Bei zwei Probanden wurde wegen zu hoher Fehleranzahl in einem Durchgang dieser gestrichen und als Testwert der Durchgang mit akzeptabler Fehleranzahl als Testergebnis gewertet.

Tabelle 7 zeigt die deskriptiven Ergebnisse zur Bearbeitungszeit in Sekunden, den beiden Fehlerarten Auslassungsfehler (AF) und Verwechslungsfehler (VF), sowie einem Leistungsmaß, das die zuvor genannten Kennwerte nach folgender Formel zu einer fehleradjustierten Anzahl korrekter Reaktionen pro Minute verarbeitet.

$\text{Temekko_v_leistung} = (\text{Anzahl richtig} - \text{Anzahl (VF)}) / (\text{Bearbeitungszeit in Minuten})$

Auf eine Berücksichtigung der Auslassungsfehler wurde verzichtet, da diese sich bereits in der korrekten Anzahl manifestieren und nicht wie beim d2c doppelt gewichtet werden sollten.

Tabelle 7: Ergebnisse zum Temekko_v (N= 93-96)

	α	M	s	Md	Schiefte	Kurtosis
Temekko_v_zeit (sec)	.94	237	58	235	.61	.58
Temekko_v_AF	.67	5.5	4.4	4	1.24	1.16
Temekko_v_VF	.52	5.4	3.0	5	1.03	2.27
Temekko_v_leistung	.89	12.2	3.4	11.7	.92	1.32

Die augenscheinlich normal verteilte Temekko_v -Bearbeitungszeit steht in keinem signifikanten Zusammenhang mit beiden Fehlerarten. Zügige Bearbeitung wird somit nicht auf Kosten einer größeren Fehleranzahl erkaufte. Auch die beiden Fehlerarten korrelieren nicht untereinander, sodass ihre Zusammenfassung im Leistungsmaß auch unabhängige Beiträge beisteuern. Allerdings erfassen Temekko_v_bearbeitungszeit und -leistung in hohem Maße Ähnliches, weil beide Variablen $r = -.83$ miteinander korrelieren. Wenngleich das Leistungsmaß somit hauptsächlich durch die Bearbeitungszeit bestimmt erscheint, gilt mit geringerem Einfluss jedoch auch. "Wer eine hohe Temekko_v_leistung erbringt, macht insgesamt weniger Fehler (r_s mit AF = $-.27$, r_s mit VF = $-.44$).

Die Reliabilität der Bearbeitungszeit fällt sehr gut aus, die des Auslassungsfehlers von $\alpha = .67$ erreicht für einen Kurztest eine durchaus noch ansprechende Höhe, fällt allerdings gegenüber der Originalversion deutlich geringer aus, was höchstwahrscheinlich dem Fehlen der irrelevanten Reize anzulasten ist. Die Reliabilität der Fehler erhöht sich nicht, wenn man beide Fehler zusammenfasst.

Buchstabe, Sonderzeichen, Zahl (BSZ-Konzentrationstest)

Beim selbst konstruierten Konzentrationstest "Buchstabe Sonderzeichen Zahl" (BSZ) hat die Testperson die Aufgabe, jeweils zu überprüfen, ob die Vorgabe einer Abfolge von 3 möglichen Begriffen (Buchstabe, Sonderzeichen oder Zahl bzw. Ziffer) nachfolgend auch in gleicher Abfolge durch konkrete Realisationen dieser Begriffe eingehalten wird oder nicht.

Dabei werden jeweils Behauptungen aufgestellt, deren Zutreffen die Testperson durch Anklicken bestätigen soll. Auf falsche Behauptungen soll gar nicht reagiert werden. An zwei Beispielen wird klar, was gemeint ist.

- 1.) **BZS = C3&** ist richtig
 2.) **ZSB = 8EF** ist falsch

Auf der linken Seite der Gleichung steht jeweils die verlangte Abfolge der Zeichen in Form von Abkürzungen: **B** für Buchstabe, **S** für Sonderzeichen und **Z** für Zahl. Die erste Behauptung ist offensichtlich richtig, weil rechts vom Gleichheitszeichen zunächst ein Buchstabe, dann eine Zahl und am Ende ein Sonderzeichen steht. Die zweite Behauptung ist falsch, weil in der Mitte kein Sonderzeichen, sondern der Buchstabe E steht. Für ein korrektes Anklicken muss folglich alles (Zeichenart und Reihenfolge) richtig sein. Ansonsten ist die Behauptung falsch und muss ignoriert werden.

Von den insgesamt 27 möglichen Anordnungen, 3 Begriffe dreimal hintereinander vorzugeben (Kombinatorik, Variation mit Wiederholung: 3^3) wurden die 12 Anordnungen ausgewählt, welche eine eher aufwändige Verarbeitung versprachen. Dazu gehören die 6 möglichen (3!), besonders schwierigen Anordnungen B, S und Z in eine Reihenfolge zu bringen, sowie weitere 6 Anordnungen mit Verdopplung eines Begriffs und Fehlen eines anderen Begriffs (z.B. BSS, SZS).

Zu jeder der zutreffenden Aufgaben wurden 3 mögliche Distraktoren konstruiert. Für SBZ etwa $SBZ = SBS$; $SBZ = BSZ$; und $SBZ = SZB$; Während auf der linken Seite der Gleichung grundsätzlich nur Abkürzungen stehen, wurden die konkreten Zeichen stets nach Zufall aus der Menge möglicher Buchstaben (A..Z), Zahlen (1..9) oder Sonderzeichen (insgesamt 14) ausgewählt.

Von den insgesamt 84 Aufgaben (Behauptungen) des BSZ treffen 30 Behauptungen zu, was einem Prozentsatz von 36% korrekter Zielaufgaben entspricht. Zu Testbeginn generiert der Computer zunächst die 84 Aufgaben entsprechend den festgelegten Regeln einschließlich Zufallsprozeduren und präsentiert diese dann in einer zufälligen Reihenfolge auf dem Bildschirm. Dieses Prinzip führt dazu, dass jede Testung in der konkreten Ausgestaltung etwas anders aussieht als Abbildung 4.

Abbildung 4: Eine mögliche Realisation des BSZ

ZSB = U;8	SZB = ±A6	ZBS = 2E§	SZB = +2I	SSB = ±~D	SSB = +QQ
ZSB = 9T~	SSB = -9H	ZBB = 7LB	SBZ = &F&	ZZB = 34X	BZS = O8N
BZS = 1X€	ZBB = 7B6	ZBB = 1OQ	BSZ = M+5	ZSB = 7KZ	ZZB = 4AL
SBZ = ±I4	ZZB = 6©S	SZZ = ++6	ZSB = M;4	ZBS = 2§E	BBS = TD8
SZB = ±7%	BSZ = I6%	SBZ = &V6	BZS = 9Q:	BSZ = U:2	BZS = W99
BBS = KZ§	SZZ = €5P	ZBS = 1&S	SZZ = €58	BZS = W9?	SSB = +*Q
ZSB = 9€T	ZSB = 7§K	SZB = @P2	BZS = X3S	SBB = 8TT	SSB = ±3Y
SBB = WW&	ZSB = 9FV	SZZ = ;;7	SZB = ±7I	ZBS = 2R@	SBZ = :9N
ZBS = 2R7	SZB = ~4X	ZZB = 3XZ	SZZ = ~~6	SBB = +LM	SZZ = ;45
SBZ = ±4I	SSB = ±F±	ZBB = 6K@	SBZ = L*4	BBS = M8%	BBS = K2§
ZZB = 54B	SBZ = &V+	BSZ = €D2	SZB = ~AX	BSZ = D€2	BZS = Q9:
BBS = MZ%	ZBB = 2J*	ZZB = 55+	ZSB = 4;M	SBB = 3XM	BSZ = +M5
SZB = +2§	BZS = O8©	ZBS = 4N©	SBB = +3M	SBB = @KT	BSZ = U2:
ZBB = 1Q3	SBZ = *L6	ZBS = 4N7	BSZ = %F1	ZBS = Q9!	BBS = QP9
<input type="button" value="Bestätigen"/>					

Der Test hat in Ansätzen gewisse Ähnlichkeit mit einem Zahlensymboltest. Von diesem unterscheidet er sich jedoch dadurch, dass einer Bezeichnung mehrere Elemente zugeordnet werden können. Zudem ist er insofern komplizierter, da nicht nur eine, sondern eine Abfolge von 3 Zuordnungen auf Korrektheit überprüft werden muss. Insofern erfordert der BSZ ein stetes Umdenken im Hinblick auf die verlangte Reihenfolge. Während im Temekko_v und ZRF_20_5 stets eine für alle Aufgaben bereits festgelegte Regel (ansteigend, bzw. strikt ansteigend oder strikt abfallend) angewandt werden muss, ist die Testperson beim BSZ gezwungen, bei jeder Aufgabe zunächst zu erfassen, welche Reihenfolge verlangt wird.

Beim BSZ spielt die Reihenfolge allerdings keine so tragende Rolle wie beim Temekko. Nur bei den korrekten Items muss die gesamte Reihenfolge durchgearbeitet werden, während die falschen Items teilweise bereits bei Sichtung eines Zeichens abgelehnt werden können.

Testanweisung- und -durchführung (Genaue Anweisung):

Zunächst erhält die Testperson eine ausführliche Beschreibung mit 2 konkreten Beispielen einschließlich Erklärung. Sie hat sodann die Möglichkeit, 7 Behauptungen durch Anklicken auf Korrektheit oder Falschheit zu überprüfen. Auf einen klassischen Probestettdurchgang mit Ergebnisrückmeldung wurde aus Zeitgründen verzichtet. Mit einem deutlichen Hinweis wird darauf aufmerksam gemacht, nach der Aufgabenbearbeitung sofort den Button "Bestätigen" zu aktivieren. Die Anweisungsseite schließt ab mit nachfolgender Zielsetzung:

"Ihre Aufgabenbearbeitung hat zwei Ziele

- 1.) **Genauigkeit:** nur die korrekten Behauptungen anklicken
- 2.) **Arbeitstempo:** möglichst schnell und zügig arbeiten.

Ihre Aufgabe besteht also darin, so schnell wie möglich nur die zutreffenden Behauptungen anzuklicken."

Bei der Bearbeitung selbst werden angeklickte Behauptungen ganz schwach markiert. Die Testperson wurde zuvor jedoch darauf hingewiesen, diese Markierung nicht weiter zu beachten. Mehrfaches Anklicken auf dieselbe Behauptung wird nicht gewertet, sondern kostet nur Bearbeitungszeit.

Testauswertung

Die Studierenden bearbeiteten den BSZ zum ersten Erhebungszeitpunkt und wiederholten die Testung in der Erhebung 3 bis 4 Wochen später. Nachfolgende Daten beziehen sich auf die erste Testung.

6 Probanden mussten wegen extrem hoher Fehleranzahl (z.B. 50% Auslassungen) aus dem Datensatz ausgeschlossen werden. Tabelle 8 zeigt die deskriptiven Ergebnisse für etliche Messvariablen. Die BSZ-Leistung erfasst die fehleradjustierte Anzahl korrekter Antworten pro Minute.

[BSZ-Leistung =(Anzahl richtig - Anzahl VF)/(Bearbeitungszeit in Minuten)]

Tabelle 8: Ergebnisse zum BSZ (N= 93: Retest N=86)

	r _{tt}	M	s	Md	Schiefe	Kurtosis
BSZ_zeit (sec)	.69	247.6	37.9	245	.5	1.4
BSZ_AF		0.6	0.9	0	1.7	2.8
BSZ_VF		1.4	1.2	1	.9	.3
BSZ_leistung	.59	7.0	1.1	6.9	.2	.6

Der Test beansprucht eine mittlere Bearbeitungszeit von ca. 4 Minuten. Rechnet man für die Testinstruktion mindestens eine Minute ein und gibt auch den schwächeren Testpersonen eine Teilnahmechance, so liegt der realistische Zeitbedarf für den Test bei Studierenden in einem Intervall von ca. 4 bis 7 Minuten.

Wenn man mal von den krassen Ausreißern absieht, welche vermutlich die Testanweisung nicht sorgfältig durchgelesen oder richtig verstanden haben, wurden mit jeweils ca. 2% der möglichen Falschantworten wenige Auslassungs- und Verwechslungsfehler (AF und VF) gemacht.

Bearbeitungszeit, AF- und VF-Fehler sind unabhängig voneinander. In die BSZ-Leistung geht jedoch die Bearbeitungszeit mit hoch dominantem Gewicht ein ($r_{s} = -.93$), während die Fehler eine ziemlich untergeordnete Bedeutung spielen. Von beiden Fehlerarten korreliert lediglich der Verwechslungsfehler (VF) signifikant mit der BSZ_leistung ($r_{s} = -.36$). Hohe Leistung geht demnach auch mit etwas geringen Verwechslungsfehlern einher. Letztlich erfasst der BSZ, wie die meisten Konzentrationstests, im Wesentlichen die Bearbeitungsgeschwindigkeit. Bedingt durch das Konstruktionsprinzip (zufällige Anordnung der Itemreihenfolge) und die Präsentation aller Items auf einer Bildschirm-

seite, konnte keine Reliabilität im Sinne der Konsistenz ermittelt werden. Die Retestreliabilitäten von .69 für die Bearbeitungszeit und .59 für die Konzentrationsleistung erreichen leider nicht das für Konzentrationstests wünschenswerte Niveau von mindestens $r_{tt}=.8$. Wie bei den meisten Konzentrationstests führte die zweite Messung zu einer hochsignifikanten Verbesserung der Bearbeitungszeit ($t(85)=-7.9$; $d=.68$ und der Leistung $t(85)=5.7$; $d=.56$) im Bereich mittlerer Effektstärke d , wobei anfangs langsame Probanden mehr Zeitgewinne erzielten ($r_{BSZ1.(BSZ2-BSZ1)}=.50$). Die relativ schwache Retestreliabilität lässt zwar einige Zweifel an der Erfassung einer sehr stabilen Fähigkeit zu, spricht aber nicht zwingend für eine schwache Messgenauigkeit, sondern lässt auch Raum für die Interpretation, die Testpersonen hätten vielleicht in unterschiedlichem Ausmaße echte Übungsgewinne erzielt.

Beziehungen zwischen den Reihenfolgetests und sonstigen Konzentrationstests

Da die Reihenfolgetests mit Ausnahme der Erkundungsvarianten beanspruchen können, im Wesentlichen Konzentrationsfähigkeit zu erfassen, waren deutliche Zusammenhänge mit allen Kriteriumsvariablen erwartet worden. In den nachfolgenden Tabellen sind die Erhebungszeitpunkte der aufgeführten Konzentrationstests zu beachten. Nur beim BSZ beruhen alle Korrelationen auf Erhebungen zum gleichen Zeitpunkt, weil dieser zu beiden Zeitpunkten erhoben wurde und der jeweils zeitnahe Konzentrationswert herangezogen wurde. Als bedeutsamstes Konzentrationsmaß gelten die Bearbeitungszeiten. Leistungsdaten werden nur der Vollständigkeit wegen mitgeteilt.

Tabelle 9: Korrelationen der Reihenfolgetests mit sonstigen Konzentrationstests (N=80-93)

a) alle Tests: Bearbeitungszeiten;

b) Reihenfolgetests: Bearbeitungszeiten, sonstige Tests: Konzentrationsleistung

a)		Temekko ¹	BSZ	d2c ²
	ZRF_20_5 ¹	.50	.50	.62
ZRF_25_3	[TMT A] ²	.40	.42	.50
ZBRF_25_3	[TMT B] ²	.46	.53	.42
b)		Temekko ¹	BSZ	d2c ²
	ZRF_20_5 ¹	-.45	-.51	-.57
ZRF_25_3	[TMT A] ²	-.32	-.40	-.52
ZBRF_25_3	[TMT B] ²	-.48	-.52	-.42

¹= erhoben zum Zeitpunkt 1; ²= erhoben zum Zeitpunkt 2

Der BSZ wurde zu beiden Zeitpunkten erhoben und jeweils der zeitnahe Wert verwendet

Alle Reihenfolgetests korrelieren mit den 3 sonstigen Konzentrationstests in ansprechender Höhe. Ursprünglich war vermutet worden, die Zusammenhänge der Reihenfolgetests würden mit den Konzentrationsleistungswerten etwas höher als mit den Bearbeitungszeiten der Kriteriumstests zusammenhängen. Denn die Bearbeitungszeiten der Reihenfolgetests basieren auf fehlerfreien Durchgängen und bei den Leistungswerten der sonstigen Konzentrationstests sind die Fehler quasi herausgerechnet. Wie aus Tabelle 9 hervorgeht, erwies sich diese Vermutung jedoch als falsch. Im Nachhinein erscheint sie auch sehr gewagt, weil Bearbeitungszeiten und Leistungswerte eines Konzentrationstests eben sehr hoch miteinander korrelieren.

Von allen Reihenfolgetests erzielt der ZRF_20_5 den höchsten Zusammenhang mit dem d2c. Denn die Korrelation .62 fällt signifikant höher aus als .5 und .42 (nach Korrelationsdifferenztest für abhängige Stichproben) und dies, obwohl der ZRF_20_5 bereits 3 bis 4 Wochen früher erhoben wurde. Auch in der Studie von Jacobs (2014a), bei der beide Tests zeitgleich erhoben wurden, ergab sich eine Korrelation von $r = .62$ zwischen d2c und dem ZRF_20<=5, einer Vorgängerversion des ZRF_20_5. Man kann somit von einer sehr gelungenen Replikation sprechen.

Zugleich deutet sich an, der TMT B_v hänge gegenüber dem TMT A_v etwas höher mit den komplexeren Tests zusammen, dem Temekko_v und insbesondere dem BSZ. Dies deutet darauf hin, der TMT B_v würde auch Aspekte messen, die nicht im TMT A_v erfasst würden. Um diese These weiter zu unterstützen, wurde überprüft, ob der um den TMT A bereinigte Anteil am TMT B noch signifikante Zusammenhänge aufweist. Als relevante Variable diente hierbei die Differenz der Bearbeitungszeiten von TMT B_v und TMT A_v. Zur Reliabilitätsbestimmung dieser Differenz wurden die drei Zeitdifferenzen B-A aus den 3 Durchgängen beider Tests herangezogen und ein Cronbachs $\alpha = .54$ ermittelt.

Tabelle 9: Korrelationen zwischen der Differenz der TMT v_Tests (B-A) und den sonstigen Konzentrationstests (N= 78-86)

	Temekko_v ¹	BSZ ²	d2c ²
a) TMT_B_v ² - TMT_A_v ²	.27	.35	.10
b) TMT_B_v ² - TMT_A_v ²	-.38	-.34	-.07

Anmerkung: a): nur Bearbeitungszeiten, b) Leistungswerte für sonstige Tests; Fettdruck: sign. 5%
¹ erhoben zum Zeitpunkt 1; ² = erhoben zum Zeitpunkt 2

Wie man der Tabelle 9 entnimmt, erzielten Temekko_v und BSZ auch mit der Differenz zwischen B- und A-Variante des TMT_v noch signifikante Zusammenhänge. Die Differenzmethode führte im Übrigen nachweislich zu sehr hoch vergleichbaren Werten wie Korrelationsberechnungen von TMT B_v mit den Konzentrationstests unter Auspartialisierung des TMT A_v. Somit vermitteln die Werte in Tabelle 9 die Beziehung zwischen TMT B_v und den Konzentrationstests bei konstanten TMT A_v-Werten. Sie sollten ja auch den Anteil messen, den die B-Variante über die A_Variante hinaus noch erfasst. Sowohl der Temekko wie der BSZ verlangen von der Testperson ähnlich wie der TMT_B ein häufiges Shifting bzw. Switching und vielleicht ist die entsprechende Fähigkeit, sich zügig mental umzustellen, für diese Zusammenhänge zumindest mitverantwortlich. In die gleiche Richtung passt auch der Befund, dass der Leistungszuwachs [genauer die Reduktion der Bearbeitungszeit] im BSZ von Erhebung 1 nach 2 umso höher ausfiel, je geringer die Differenz B-A im TMT_v ausgeprägt war ($r = -.24$; $p_z = 0.032$). Ähnliches gilt für den Temekko_v. Je schneller dieser bearbeitet wurde, um höher war der Leistungszuwachs im BSZ ($r = -.24$; $p_z = 0.027$). Allerdings wird der Übungsgewinn im BSZ auch vom Ausgangslevel bestimmt. Deshalb wurden Partialkorrelationen berechnet, um ein konstantes Bearbeitungsniveau bei der ersten Testung des BSZ zu simulieren. Hier zeigt sich nun, dass die partielle Korrelation beim Temekko_v zusammenbricht [$r_{\text{temekko}_v, (\text{BSZ2}-\text{BSZ1})}$, $\text{BSZ1} = .09$], bei der Differenz B-A im TMT_v jedoch von ursprünglich $r_{(B-A), (\text{BSZ2}-\text{BSZ1})} = -.24$ auf $r_{(B-A), (\text{BSZ2}-\text{BSZ1})}$, $\text{BSZ1} = -.33$ ($p_z = 0.003$) dem Betrage nach ansteigt. Wer große Schwierigkeit mit der Umstellung von Zahlen - und Buchstabenfolge hat, kann sich im BSZ auch nicht so wirksam verbessern.

Insgesamt deuten die Ergebnisse der anspruchsvolleren Tests in die Richtung, neben klassischer Konzentrationsfähigkeit, zu einem kleineren Teil auch komplexere Verarbeitungsprozesse mit zu erfassen. (siehe dazu auch Studien zum Temekko von Schmid, 2010).

Nachfolgende Tabelle 10 wird hier nur der Vollständigkeit wegen präsentiert. Sie verdeutlicht die Zusammenhänge der 3 für die Validitätsprüfung der Reihenfolgetests herangezogenen Konzentrationstests untereinander sowohl im Hinblick auf die reine Bearbeitungszeit, wie auch für die Leistungswerte.

Tabelle 10: Interkorrelationen der sonstigen Konzentrationstests. (N=81- 93)

	Temekko ¹	BSZ	d2 ²
Temekko ¹	-	.53	.53
BSZ	.46	-	.47
d2 ²	.38	.51	-

Hinweis:

Oberhalb der Diagonalen Bearbeitungszeiten, unterhalb Konzentrationsleistung:

¹= erhoben zum Zeitpunkt 1; ²= erhoben zum Zeitpunkt 2; BSZ jeweils zeitnah

Faktorenanalyse mit allen Konzentrationstests

Die eingesetzten Tests weisen teilweise hohe Ähnlichkeiten, teilweise aber auch Unterschiede auf. TMT B_v, Temekko_v und BSZ erscheinen z.B. komplizierter als die übrigen Tests. Von den Testverfahren der ersten Erhebung liegen Anstrengungseinschätzungen nach jeweiliger Testbearbeitung vor. Danach empfanden die Probanden den Temekko_v und BSZ in vergleichbarer Höhe ziemlich anstrengend, den ZRF_20_5 aber in sehr bedeutsamem Ausmaß weniger anstrengend (Effektstärke d jeweils deutlich über 1). Auch wenn einzelne der hier getestete Konzentrationstests in geringem Umfang neben Konzentration sonstige Facetten messen sollten, so gilt doch im Grundsatz. Alle Konzentrationstests (ZRF_20_5, TMT A_v, TMT B_v sowie die 3 sonstigen Konzentrationstests) erfassen in hohem Maße einen gemeinsamen Aspekt der Konzentrationsfähigkeit. Denn eine Faktorenanalyse mit den Bearbeitungszeiten aller 6 Konzentrationstests ergab unmissverständlich nur einen einzigen Faktor, der 62 % der Gesamtvarianz erklärte sowie Ladungen im Bereich von .72 bis .86 hervorbrachte. Fügt man den 6 Konzentrationstests noch den ZRF_punkte_20_3 und den ZRF_verbal_20_3 hinzu, so laden diese beiden Tests .75 und .68 auf einem zweiten Faktor, ein deutlicher Hinweis darauf, ZRF_Punkte und ZRF_verbal erfassten hauptsächlich etwas anderes als Konzentration.

Zusammenfassung und Diskussion

Beim ZRF_20_5 und den beiden Erkundungsvarianten steht die verlangte Anordnung der Elemente in vergleichbarer Weise automatisiert zur Verfügung. Es bedarf keiner großen Überlegung, nach "zwei" die "drei" zu suchen, oder nach 17 Punkten anschließend nach 18 Punkten Ausschau zu halten. Die Unterschiede zwischen ZRF_20_5 und den Erkundungsvarianten liegen im Zeitbedarf, die einzelnen Elemente spontan wahrzunehmen, eindeutig zu identifizieren oder problemlos miteinander vergleichen zu können. Wie eine zusätzliche Erhebung im Anschluss an die Tests ergab, wurde die Bearbeitung von ZRF_punkte_20_3 und ZRF_verbal_20_3 auch als wesentlich anstrengender empfunden als die Testung des ZRF_20_5 sowie aller sonstigen Tests. Da Punkte

und Zahlworte keine hinreichenden Zusammenhänge mit allen sonst erhobenen Konzentrationstests erbrachten und in der Faktorenanalyse einen zweiten Faktor bildeten, sind sie in der hier realisierten Form zur Messung der Konzentrationsfähigkeit schlichtweg ungeeignet. Es bleibt unklar, was diese Erkundungsvariablen überhaupt messen. Da sie jedoch substantziell untereinander korrelieren, erfassen beide irgendeinen gemeinsamen Anteil an der Fähigkeit, komplexere Reize schnell erfassen und vergleichen zu können. Der ZRF_verbal_20_3 wäre eventuell ein nützlicher Kurztest, um Legasthenie zu erfassen.

Arabische Zahlen sind in der Schnelligkeit der Reizwahrnehmung und dem Vergleichen „Zahl x ist größer, kleiner, ungleich oder gleich Zahl y“ ohne ernsthafte Alternative. Die Nützlichkeit sonstiger Zahlendarstellungen nimmt mit wachsender Anzahl der zu vergleichenden Elemente rapide ab. Als grundsätzliche Alternative bietet sich an, mehrmals eine überschaubare Anzahl von Elementen der Reihenfolge nach anordnen oder bewerten zu lassen. Im Temekko wird dieses Prinzip anhand von 3 Elementen erfolgreich angewandt. Hier kam zwar nur die numerische Variante mit Zahlen zum Einsatz. In der Originalfassung muss die Testperson aber noch eine visuelle (Balken) und verbale Fassung (Wörter; Größe der Objekte) bearbeiten. Für Größenrelationen mit wenigen Elementen lassen sich sehr viele Darstellungsvarianten konstruieren, letztlich alle Variablen, die sich mindestens als Ordinalskala anordnen und deren Elemente sich von der Wahrnehmung her sehr gut unterscheiden lassen, z.B. auch wenige Daten aus etlichen Diagrammen wie Säulen, Balken, Punkte, Kreissegmente.

Zahlenreihenfolgetests als Varianten des TMT

Eine Erhöhung der Zahlen von 20 auf 25, wie sie im TMT A_v realisiert ist, erhöht zwar die augenscheinliche Ähnlichkeit zum TMT A, verbessert jedoch keineswegs die Testeigenschaften gegenüber dem ZRF_20_5. Da die Erhöhung der Zahlenanzahl um 1/5 mit der Zunahme der Bearbeitungszeit um ca. 1/3 Drittel einherging, war es angemessen, den TMT A_v auf 3 Durchgänge zu reduzieren. Bei vergleichbarer Gesamtbearbeitungsdauer beider Zahlenreihenfolgetests verspricht der ZRF_20_5 eine höhere Zuverlässigkeit und teilweise sogar höhere Validitätskoeffizienten. Zudem zeigt er einen vergleichbar hohen Zusammenhang mit dem TMT_B_v. auf. Die offensichtliche Ähnlichkeit der Anforderungen von ZRF_20_5 und TMT A, sowie die hier gefundenen Ergebnisse deuten klar in die Richtung, der ZRF_20_5 messe in hohem Maße Vergleichbares wie der TMT A. Auch der Zahlen-Verbindungs-Test (ZVT) gehört nachweislich hierzu, wie der hohe Zusammenhang zwischen ZRF_20_5 und einer Computervariante des ZVT von $r = .81$ in der Studie von Jacobs (2013b) belegt. Der ZRF_20_5 wäre somit zumindest in den Forschungsbereichen, die keine Normen benötigen, eine echte Alternative zum TMT A, zumal seine Handhabung etliche ökonomische Vorteile aufweist.

Beim TMT B_v ist die Ähnlichkeit zum TMT B in gleichem Maße gegeben. Der wesentliche Unterschied zu der A-Version ist zu vermuten in der weniger evidenten Repräsentation der Buchstabenreihenfolge und dem ständigen Umschalten zwischen Zahlen- und Buchstabenfolge. Dadurch wird der Prozess 4 in Tabelle 1 erschwert und eine höhere kognitive Flexibilität eingefordert. Das dafür erforderliche Fähigkeitspotenzial lässt sich vielleicht teilweise durch Differenzbildung beider TMT-Varianten erfassen. Im Vergleich zu einer deutschen Normenstichprobe mit modifizierten Trail Making Tests (Rodewald et al 2012) fielen der Unterschied in der Bearbeitungszeit zwischen den A und B Varianten (B-A) hier geringer und die Korrelationen zwischen beiden Varianten etwas höher aus.

Wie bereits die Übungsgewinne beim TMT B_v während einer Testung mit 3 Durchgängen nahelegen, dürften mehrfache Messungen typische Testeffekte nach sich ziehen und gezielte Übungen der Zahlen-Buchstabenpaare (1:A .. 6:F ..11:K) bis zur Automatisierung klare 'Leistungsvorteile' bewirken.

Im Gegensatz zum TMT B_v ist der TMT A_v weniger übungsanfällig. Das konnte auch Woods et al. (2015) mit einer speziellen Computervariante des C-TMT belegen. In die gleiche Richtung deuten die Ergebnisse etlicher TMT-Varianten in Tabelle 1 von Buck et al. (2008). Dort wurden auch deutliche Testeffekte für den TMT A ermittelt, die aber wesentlich höher ausfallen als Übungsgewinne im ZRF_20_5 (Jacobs 2014a). Bisherigen Erfahrungen zufolge bewirkte der erste Retest im ZRF_20_5 einen Testeffekt von höchstens $d=.2$ und 10 Testungen einen Übungsgewinn in der Nähe von $d=.5$.

Man kann sich leicht weitere komplizierte TMT-Anordnungen ausdenken, etwa "1 B 3 D 5 F 7..." oder den TMT-G von Amodio et al. (2002), entfernt sich damit aber immer mehr von dem Konstruktionsprinzip eines Konzentrationstests. Hierbei sind der Testperson Anforderungen zu stellen, die sie ohne Rückgriff auf das Langzeitgedächtnis und etliche, teilweise fähigkeitsbedingte Denkopoperationen, nur auf der Basis einfacher, direkt ausführbarer Regeln, bewältigen kann. Je komplexer und geistig anspruchsvoller die Anforderungen, desto eher wird die kognitive Verarbeitungskapazität bzw. fluide Intelligenz erfasst. Es ist daher auch theoretisch nachvollziehbar, wenn der TMT B im Vergleich zum TMT A höhere Zusammenhänge mit kognitiv anspruchsvolleren Tests erbrächte. Hierfür gibt es bereits schon frühe empirische Belege (z.B. Goulet & Brown, (1970). Vernon (1993) erfasste diverse Varianten des Zahlenverbindungstests (ZVT) und kommt zu dem Ergebnis " there was a trend for more complex ZVT versions to correlate more highly with IQ scores than less complex ZVT versions." Auch Analysen von Salthouse (2011 a,b) weisen in die Richtung, alternierende Reihenfolgetests (Zahlen und Buchstaben im Wechsel) würden gegenüber einfachen Reihenfolgetests (nur Zahlen oder nur Buchstaben) höher mit fluider Intelligenz zusammenhängen.

Insgesamt bestätigen bisherige Befunde sowie die Daten dieser Studie, der ZRF_20_5 erfasse in hohem Maße Ähnliches wie der TMT A und er sei hinreichend geeignet, Konzentrationsfähigkeit bzw. Informationsverarbeitungsgeschwindigkeit zu messen.

Literatur

- Amodio, P., Wenin, H., Del Piccolo, F., Mapelli, D., Montagnese, S., Pellegrini, A. & Umiltà, C. (2002). Variability of trail making test, symbol digit test and line trait test in normal people. A normative study taking into account age-dependent decline and sociobiological variables. *Aging clinical and experimental research*, 14(2), 117-131.
- Brickenkamp, R. Schmidt-Atzert, L., Liepmann, D. (2010). d2 –Revision
Aufmerksamkeits-und Konzentrationstest - Manual. Hogrefe
- Buck, K. K., Atkinson, T. M., & Ryan, J. P. (2008). Evidence of practice effects in variants of the Trail Making Test during serial assessment. *Journal of clinical and experimental neuropsychology*, 30(3), 312-318.
- Goul, W. R., & Brown, M. (1970). Effects of age and intelligence on Trail Making Test performance and validity. *Perceptual and Motor Skills*, 30(1), 319-326.
- Jacobs, B. (2013a). Erprobung zweier Online-Konzentrationstests mit Zahlen an Studierenden des Lehramts.
<http://bildungswissenschaften.uni-saarland.de/personal/jacobs/diagnostik/tests/konzentration/konzentrationstests.html>
- Jacobs, B. (2013b). Ähnlichkeit zwischen Zahlen-Verbindungs-Test (ZVT) und Zahlenreihenfolgetest_20 (ZRF_20).
http://bildungswissenschaften.uni-saarland.de/personal/jacobs/diagnostik/tests/konzentration/zrf_20_vs_zvt.html
- Jacobs, B. (2014a): Analyse von Testgütekriterien und Übungseffekten zweier Online-Konzentrationstests.
URN: urn:nbn:de:bsz:291-psydok-49678
URL: <http://psydok.sulb.uni-saarland.de/volltexte/2014/4967/>
- Jacobs, B. (2014b). Konzentrationstests durch Ordnen der zufällig angeordneten ersten 20 Zahlen oder Buchstaben
http://bildungswissenschaften.uni-saarland.de/personal/jacobs/diagnostik/tests/konzentration/zrf_und_br/zrf_und_br.html
http://bildungswissenschaften.uni-saarland.de/personal/jacobs/diagnostik/tests/konzentration/zrf_und_br/zrf_und_br.pdf
- Jacobs, B. (2015a). Unkonzentriert trotz hoher Konzentrationsfähigkeit?
URN: urn:nbn:de:bsz:291-psydok-52336
URL: <http://psydok.sulb.uni-saarland.de/volltexte/2015/5233/>
- Jacobs, B. (2015b) Der Zahlenreihenfolgetest 20_5
[Psychometrikon](http://psychometrikon.de) - psychologisch medizinisches Testportal -
<http://psychometrikon.de/inhalt/suchen/test.php?id=c98cf246e31e3cd376cdf94ad04df78>
DOI: 10.6099/1000211
- Jacobs, B. (2015c). Beeinträchtigt eine Rede während der Bearbeitung eines Konzentrationstests die Leistung in diesem Test?
http://bildungswissenschaften.uni-saarland.de/personal/jacobs/diagnostik/tests/konzentration/laerm/d2_und_laerm.html
- Oswald W. & Roth, E: (1987). Der Zahlen-Verbindungs-Test (ZVT) –Handanweisung. Göttingen, Hogrefe.
- Reitan, R. (1958). Validity of the trail making test as an indicator of organic brain damage. *Perceptual Motor,Skills*, 8, 271–276. Zitiert nach Buck et al. (2008)

Reitan, R.M. (Ed.). (1992). Trail Making Test. Tucson, AZ:Reitan Neuropsychology Laboratory. Zitiert nach: Rodewald, K. et al. (2012)

Rodewald, K., Bartolovic, M., Debelak, R., Aschenbrenner, S., Weisbrod, M., & Roesch-Ely, D. (2012). Eine Normierungsstudie eines modifizierten Trail Making Tests im deutschsprachigen Raum. *Zeitschrift für Neuropsychologie*, 23(1), 37-48.

Salthouse, T. A. (2011a). Cognitive correlates of cross-sectional differences and longitudinal changes in trail making performance. *Journal of clinical and experimental neuropsychology*, 33(2), 242-248.

Salthouse, T. A. (2011b). What cognitive abilities are involved in trail-making performance?. *Intelligence*, 39(4), 222-232.

Salthouse, T. A., Toth, J., Daniels, K., Parks, C., Pak, R., Wolbrette, M., & Hocking, K. J. (2000). Effects of aging on efficiency of task switching in a variant of the trail making test. *Neuropsychology*, 14(1), 102.

Schmid, V. (2010). TEMEKKO (Doctoral dissertation, lmu).
<http://nbn-resolving.de/urn:nbn:de:bvb:19-117420>
http://edoc.ub.uni-muenchen.de/11742/1/Schmid_Verena.pdf

Vernon, P. A. (1993). Der Zahlen-Verbindungs-Test and other trail-making correlates of general intelligence. *Personality and Individual Differences*, 14(1), 35-40.

Woods, D. L., Wyma, J. M., Herron, T. J., & Yund, E. W. (2015). The Effects of Aging, Malingering, and Traumatic Brain Injury on Computerized Trail-Making Test Performance. *PloS one*, 10(6), e0124345.

Anhang

A) Zeitpunkt der erhobenen Konzentrationstests.

Zeitpunkt 1: ZRF_20_5, ZRF_punkte_20_3, ZRF_verbal_20_3, BSZ, Temekko_v
 Zeitpunkt 2: ZRF_25_3, ZBRF_25_3, d2c, BSZ

Abstand zwischen Zeitpunkt 1 und 2 beträgt 3 bis 4 Wochen.

