

Zur Paralleltestreliabilität des Corporal Plus

Axel Muttray, Natalie Löb*, Alicia Poplawski, Heiko Hecht, Till Beutel und Dirk-Matthias Rose

Für die Wiederholung psychometrischer Fahreignungstests gibt es parallele Testverfahren. Die Fragestellung lautete, wie hoch die Paralleltestreliabilität des Corporal Plus ist. Je 25 gesunde PKW-Fahrer beiderlei Geschlechts im Alter von 18 bis 35 Jahren absolvierten die Standardtests und am Tag danach die Paralleltests. Zielgrößen waren die Leistungskennwerte der einzelnen Tests, die sich aus den T-Werten (McCall) für mittlere Reaktionszeit und Zahl richtiger Antworten im Verhältnis 1:2 ergeben. Der Grad der Übereinstimmung zwischen Standard- und Paralleltests wurde mit Bland-Altman-Diagrammen untersucht. Diese zeigen große Streuungen der Differenzen der Leistungskennwerte, d. h. geringe Grade an Übereinstimmung. Eine wesentliche Ursache ist die Definition des Leistungskennwerts, die Fehler stärker gewichtet. Es wird empfohlen, künftig die Reaktionszeit stärker zu gewichten. Möglicherweise lassen sich Fehler durch eine Probandeninstruktion reduzieren, die stärker auf die Vermeidung von Fehlern fokussiert. Einige Probanden hatten Verständnisprobleme bei den Orientierungstests. Dies sollte in der Praxis bei der Interpretation von Testergebnissen bedacht werden.

Parallel test reliability of the Corporal Plus

There are parallel test procedures for the repeated psychometric testing of driving aptitude. The objective was to analyze the parallel test reliability of the Corporal Plus. 25 healthy car drivers of both sexes aged 18 to 35 years completed the standard tests of Corporal Plus, and the next day the parallel tests. The target values were the performance indices of the individual tests. They result from the T-values (McCall) for mean reaction time and number of correct answers at a ratio of 1:2. The degrees of agreement between standard and parallel tests were assessed with Bland-Altman plots. They show wide scatters of the differences in the performance indices, i. e. low degrees of agreement. One main reason for this is the definition of the performance index, which gives greater weight to errors. We therefore recommend to give greater weight to the reaction time in the future. Possibly, the number of errors can be reduced by an instruction that focuses more on avoiding errors. Some subjects had problems understanding the orientation tests. This should be taken into account in practice when interpreting test results.

1 Einleitung

In den „Begutachtungsleitlinien zur Kraftfahreignung“ sind die Anforderungen an die psychische Leistungsfähigkeit festgelegt. Diese ist „mit geeigneten, objektivierbaren psychologischen Testverfahren“ zu untersuchen (Gräcman, Albrecht 2018). Als von grundlegender Bedeutung für eine Fahrtätigkeit werden die Bereiche Auffassung und Orientierung, Konzentrations- und Aufmerksamkeitsleistung, Reaktionsverhalten und Belastbarkeit sowie sensomotorische Koordination angesehen (Müller, Brenner-Hartmann et al. 2013). Die unterschiedlichen Bereiche sind nicht als unabhängig vonein-

ander zu betrachten. Vielmehr sind für jeden Test Wahrnehmung, Aufmerksamkeit, zentralnervöse Verarbeitung und ein Mindestmaß an motorischer Reaktion erforderlich (Müller, Brenner-Hartmann et al. 2013). Um das Leistungsvermögen des Probanden festzustellen, werden in der Praxis verschiedene Testverfahren eingesetzt (Schubert, Schneider et al. 2005.; Poschadel, Falkenstein et al. 2009; Müller, Brenner-Hartmann et al. 2013). Im Kommentar zu älteren Begutachtungsleitlinien (Schubert, Schneider et al. 2005) wird eine Mengenleistung im unterdurchschnittlichen Bereich bei überdurchschnittlicher Qualität (d. h. einer geringen Fehlerquote) als durchaus noch ausreichende Testleistung angesehen; nicht aber der umgekehrte Fall. Deshalb wird die Qualität im Vergleich zur Mengenleistung im Corporal-Plus-System stärker gewichtet (Berg, Nädtke 2015). Bei diesem Testsystem kann der Anwender einzelne Tests verwenden oder eine Testbatterie zusammenstellen (Bild 1).

In der Praxis muss manchmal ein Test wiederholt werden. Wegen des Lerneffekts verbietet es sich, exakt denselben Test zweimal kurz nacheinander absolvieren zu lassen. Vielmehr werden sogenannte parallele

Testverfahren benötigt (Müller, Brenner-Hartmann et al. 2013), die eine zeitnahe Wiederholung für jeden Anforderungsbereich ermöglichen. Um das bei einem solchen Paralleltest erzielte Ergebnis bewerten zu können, sollte bekannt sein, wie hoch der Grad der Übereinstimmung mit dem korrespondierenden Standardtest (Bild 1) ist. Bisher lagen keine Daten zum Grad der Übereinstimmung zwischen den üblicherweise eingesetzten Standardtests und den Paralleltests des Corporal Plus vor. Möglich erscheint, dass Paralleltests z. B. einen anderen Schwierigkeitsgrad aufweisen oder Übungseffekte trotz unterschiedlicher Testkonstruktion auftreten (Schermele-Engel, Werner 2012). In einer Pilotstudie mit jüngeren Pkw-Fahrern (Führerscheinklasse B), die von der Vistec AG initiiert wurde, sind wir der Frage nachgegangen, wie hoch der Grad der Übereinstimmung der vom Hersteller ausgegebenen Ergebnisparameter („Leistungskennwerte“) zwischen sieben Standardtests des Corporal Plus und ihren jeweiligen Paralleltests ist. Auf den ersten Blick

■ Dokumentation

Muttray, A.; Löb, N.; Poplawski, A.; Hecht, H.; Beutel, T.; Rose, D.-M.: Zur Paralleltestreliabilität des Corporal Plus, Z. f. Verkehrssicherheit 65, (2019) Nr. 4, S. 264

■ Schlagwörter

Corporal Plus, psychometrischer Test, Paralleltestreliabilität, Fahreignung, Orientierungstest

* Die Publikation enthält Daten aus der Dissertationsschrift von N. Löb, in Vorbereitung.

sind die Leistungskennwerte für manche Anwender eher schwer zu durchschauen. Die Kenntnis des Grades der Übereinstimmung zwischen einem durchgeführten Paralleltest und dem korrespondierenden Standardtests ist für den Anwender bei der Bewertung von Testergebnissen hilfreich.

2 Methoden

2.1 Studiendesign

50 Probanden mit Pkw-Führerschein absolvierten an Tag 1 die Standardtests des Corporal Plus (Bild 1). Am folgenden Tag wurden die Paralleltests in analoger Reihenfolge bearbeitet. Entsprechend der Praxis war der Versuchsplan nicht balanciert. Bei realen Fahreignungstests sind die Prüflinge in der Regel motiviert. Um unseren Probanden einen motivationalen Anreiz zu verschaffen, wurde eine Prämie ausgelobt (Jenkins, Mitra et al. 1998; Robinson, Stevens et al. 2012; Gajda, Sulzenbruck et al. 2016), die grundsätzlich für jeden Teilnehmer erreichbar war. Eine von uns zuvor durchgeführte Befragung von Studierenden hatte ergeben, dass der Betrag von 20 € durchaus als Anreiz empfunden wurde und

nicht zu niedrig war (Gneezy, Rustichini 2000). Die standardisierte Instruktion lautete: „Bei der Studie ist es wichtig, dass Sie bei allen Tests konzentriert arbeiten. Wenn Sie sich besonders viel Mühe geben und voll konzentriert dabei sind, erhalten Sie eine zusätzliche Prämie von 20 €.“ Den Probanden wurde mitgeteilt, dass ihre Ergebnisse erst nach dem zweiten Experimentaltag hinsichtlich der Prämie ausgewertet würden.

2.2 Probanden

Die Probanden wurden mittels Aushängen, über eine Facebook-Gruppe und durch Mund-zu-Mund-Propaganda akquiriert. Einschlusskriterien waren: Alter zwischen 18 und 35 Jahren, keine Erkrankungen und keine Einnahme von Medikamenten und Drogen, die die psychomotorische Leistungsfähigkeit und/oder das Fahrvermögen beeinträchtigen können, kein Schlafdefizit, gültiger Führerschein der Klasse B. Die Versuchspersonen wurden über die Studie aufgeklärt und gaben ihre schriftliche Einwilligung. Die ärztliche Anamnese wurde ggf. durch gezielte Untersuchungen ergänzt. Der beidäugige korrigierte Nah- und Fernvisus (Titmus 2a, MAICO Diagnostic GmbH, Berlin) betrug minimal 0,7 bzw. 0,8. Die

Atemalkoholkonzentration (Alcotest 6510, Draeger, Lübeck) lag stets unter der Nachweisgrenze von 0,05 mg/l. Die neun Raucher konnten ohne Probleme eine Stunde ohne Entzugssymptome durchhalten. Bezüglich des Konsums koffeinhaltiger Getränke wurden die Probanden gebeten, ihre üblichen Gewohnheiten beizubehalten und an beiden Experimentaltagen in etwa vergleichbare Mengen zu konsumieren. Drei Probanden mussten wegen unentschuldigtem Fehlen, Nachtdiensts bzw. eines Infekts ersetzt werden. Das Kollektiv umfasste je 25 Männer und Frauen mit einem medianen Alter von jeweils 26 Jahren (Altersbereiche 19–34 bzw. 18–35 Jahre). 44 Probanden waren Studierende, davon fünf im Fach Psychologie. Die übrigen Probanden waren berufstätig: ein Azubi, ein Arbeiter, drei Angestellte und ein Beamter. Der Wachheitszustand der Probanden wurde vor und nach den Experimenten mittels der Karolinska Sleepiness Scale (Akerstedt, Gillberg 1990) erfragt (Tabelle 1). Die Händigkeit wurde mit dem Edinburgh-Inventar bestimmt (Oldfield 1971). 46 Probanden waren Rechtshänder, die übrigen vier Linkshänder. Die Probanden erhielten für ihre Teilnahme 30 €, 45 Personen erhielten zusätzlich die Prämie von 20 €. Die Ethikkommission der Landesärztekammer

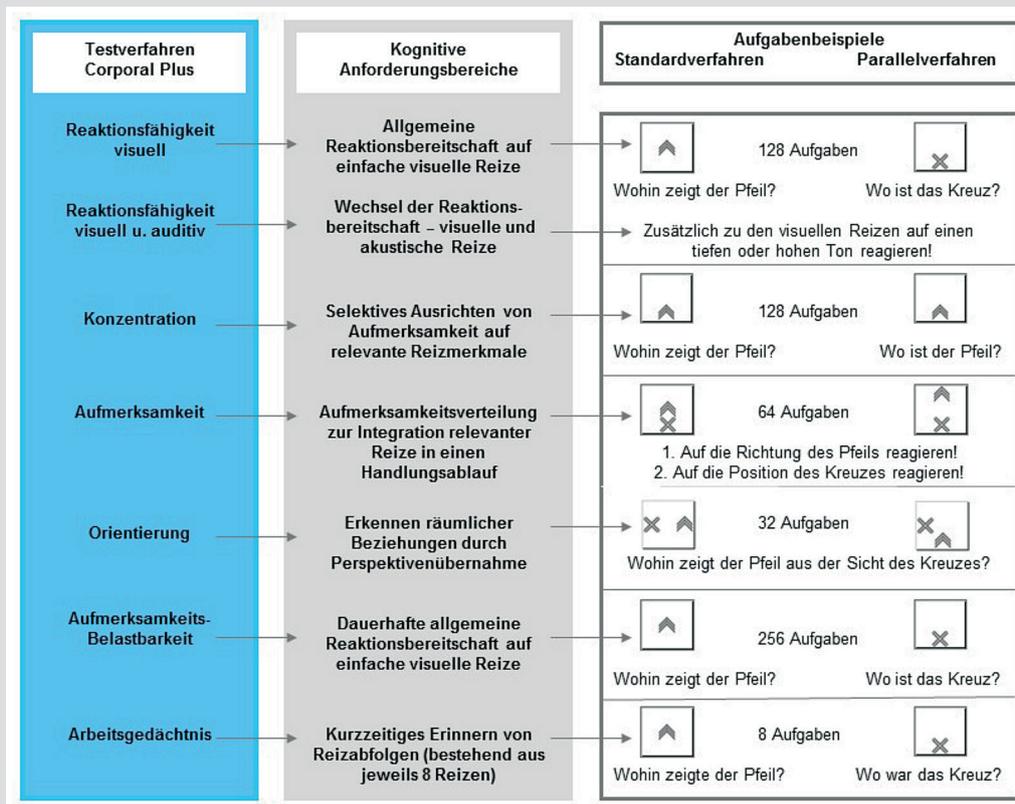


Bild 1: Standard- und Parallelverfahren (modifiziert nach Berg und Nädtke (2015))

	Vor Experiment 1	Nach Experiment 1	Vor Experiment 2	Nach Experiment 2
Minimum	1	1	1	1
10. Perzentil	2	1	2,1	2
1. Quartil	3	2	3	3
Median	3	3	3	3
3. Quartil	4	4	4	4
90. Perzentil	6	5	5	5
Maximum	7	7	7	8

Experiment 1 und 2: am 1. bzw. 2. Versuchstag mit Standard- bzw. Paralleltests. Die Ankerpunkte der Ordinalskala sind: 1 = sehr wach, 3 = wach – normaler Zustand, 5 = weder wach noch müde, 7 müde, aber keine Probleme, wach zu bleiben, sowie 9 = sehr müde, große Probleme wach zu bleiben, mit dem Schlaf kämpfend. n = 50

Tabelle 1: Selbsteinschätzung des Wachheitszustands der Probanden mittels Karolinska Sleepiness Scale

Rheinland-Pfalz erteilte ein positives Votum für die Studie.

2.3 Testpsychologische Untersuchung

Jede Testbatterie enthielt sieben PC-gestützte Tests, die in der dargestellten Reihenfolge bearbeitet wurden (Bild 1). An Tag 1 wurden die Standardtests absolviert, an Tag 2 die Paralleltests. Im Vergleich zur Vorgängerversion des Corporal Plus waren die Zahlen der Items der Orientierungstests halbiert (Update von Vistec vom 19.12.2016). Vor jedem Test wurde auf einem Monitor folgende Instruktion angezeigt: „Vermeiden Sie Fehler, aber vergessen Sie nicht, dass die Zeit gemessen wird.“ Pro Test gab es eine Übungsphase, die laut Manual höchstens dreimal wiederholt werden sollte. Davon musste bei Orientierungstests abgewichen werden, weil sonst einige Messungen wegen Verständnisproblemen unmöglich gewesen wären. Beim Standardtest wurden einmal 4 Wiederholungen benötigt, beim Paralleltest waren es je einmal 4, 5, 8 bzw. 10. (In der Praxis wären diese Tests bei den betreffenden Probanden abgebrochen und als nicht bestanden bewertet worden.) Die Tests wurden tagsüber zu unterschiedlichen Zeiten

durchgeführt. Jeder Proband begann seine Tests an beiden Tagen zur gleichen Uhrzeit. Alle Tests wurden in einem ruhigen Raum durchgeführt. Dabei wurden die Probanden von der Versuchsleiterin beobachtet.

2.4 Statistik

Aus den Rohwertwerten der mittleren Reaktionszeiten und denen der Anzahlen der richtigen Antworten wurden vom System jeweils T-Werte (T) berechnet (McCall 1939; Amelang, Schmidt-Atzert 2006; Bortz, Schuster 2010; Berg, Nädtke 2015). Die T-Werte für die Fehler entsprechen denjenigen für die Anzahlen richtiger Antworten. Bei T-Werten handelt es sich um transformierte Messwerte, um einen gemessenen Rohwert einer Person in Relation zu denen anderer Personen setzen zu können. Durch T-Werte erhält man eine Normskala von 20 bis 80 (entsprechend +/- 3 Standardabweichungen), die einen Mittelwert von 50 und eine Standardabweichung von 10 aufweist. Der Leistungskennwert L ist wie folgt definiert:

$$L = \frac{(2 \cdot T_{\text{Anzahl der richtigen Antworten}} + T_{\text{mittlere Reaktionszeit}})}{3}$$

Leistungskennwerte wurden vom System nicht ausgegeben, wenn T-Werte für die

Zahl der richtigen Antworten < 40 waren. Bei den Reaktionszeiten wurde das Konfidenzintervall von 3 berücksichtigt, sodass der Grenzwert bei 37 lag (Berg, Nädtke 2015). Vom System nicht ermittelte Leistungskennwerte wurden von uns berechnet. Einfluss- und Zielgrößen wurden deskriptiv analysiert. Die Zusammenhänge zwischen den korrespondierenden Leistungskennwerten, den Reaktionszeiten bzw. den Fehlerzahlen von Standard- und Paralleltests wurden mittels Rangkorrelationen nach Spearman explorativ untersucht. Korrelationskoeffizienten sind nicht hinreichend für die Beurteilung des Grades der Übereinstimmung (Bland, Altman 1986, du Prel, Muttray 2011). Die Grade der Übereinstimmung wurden mittels Bland-Altman-Diagrammen (Bland, Altman 1986, Grouven, Bender et al. 2007, du Prel, Muttray 2011) beurteilt. Die Analysen wurden mit Microsoft Excel 2007, IBM SPSS 23 (Ehningen, Deutschland) und R (Wickham 2009, R Core Team 2017) durchgeführt.

3 Ergebnisse

Die Zeit, die für jeweils sieben Tests insgesamt benötigt wurde, variierte stark. Für die Standardtests wurden im Median 19:43 Min. (Bereich 15:50–31:50) gebraucht, für die Paralleltests 16:21 Min. (12:30–42:09). Die Bland-Altman-Diagramme der Leistungskennwerte (Bilder 2a-g) zeigen, dass die Grade der Übereinstimmung zwischen den Standard- und ihren jeweiligen Paralleltests gering sind. Diese Tatsache ist wesentlich durch die Streuungen der Differenzen der Paare der Leistungskennwerte begründet. Die Korrelationskoeffizienten der Leistungskennwerte der Standard- und ihrer korrespondierenden Paralleltests sind bis auf eine Ausnahme niedrig (Tabelle 2). Die Korrelationskoeffizienten für die mittleren Reaktionszeiten

Tabelle 2: Korrelationskoeffizienten von Parametern der Standard- und ihrer jeweiligen Paralleltests

Test	Leistungskennwert	Mittlere Reaktionszeit	Fehler
Reaktionsfähigkeit (visuell)	r = 0,148, p = 0,305	r = 0,658, p < 0,001	r = 0,146, p = 0,312
Reaktionsfähigkeit (visuell & auditiv)	r = 0,224, p = 0,118	r = 0,768, p < 0,001	r = 0,162, p = 0,261
Konzentration	r = 0,308, p = 0,030	r = 0,499, p < 0,001	r = 0,230, p = 0,108
Aufmerksamkeit	r = 0,441, p = 0,001	r = 0,889, p < 0,001	r = 0,267, p = 0,061
Orientierung	r = 0,430, p = 0,002	r = 0,810, p < 0,001	r = 0,228, p = 0,112
Aufmerksamkeitsbelastbarkeit	r = 0,452, p = 0,001	r = 0,739, p < 0,001	r = 0,500, p < 0,001
Arbeitsgedächtnis	r = 0,694, p < 0,001	r = 0,730, p < 0,001	r = 0,646, p < 0,001

r: Rangkorrelationskoeffizient nach Spearman. p: p-Wert. n = 50

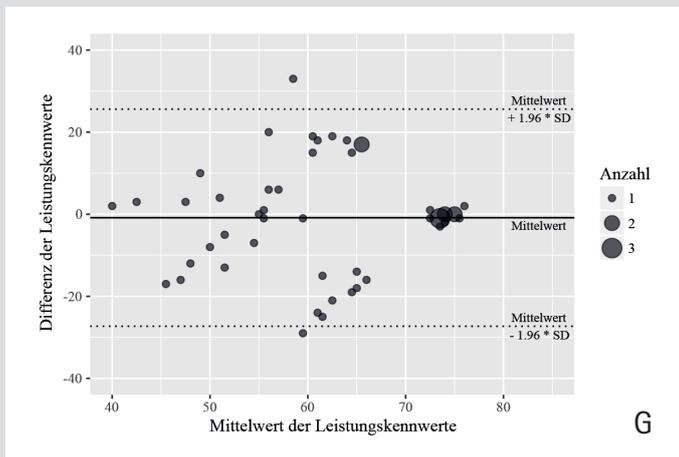
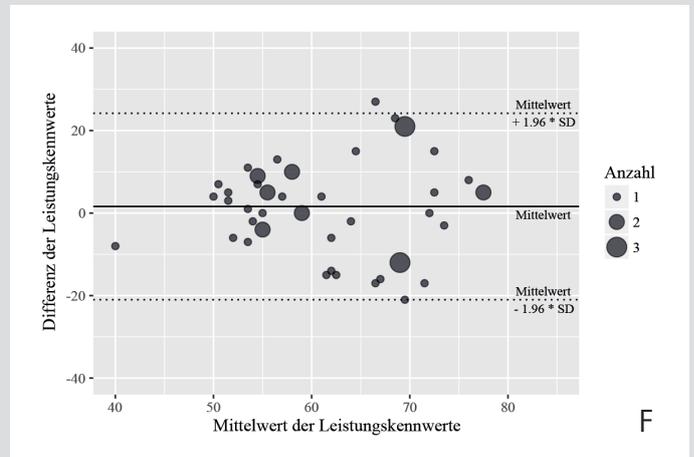
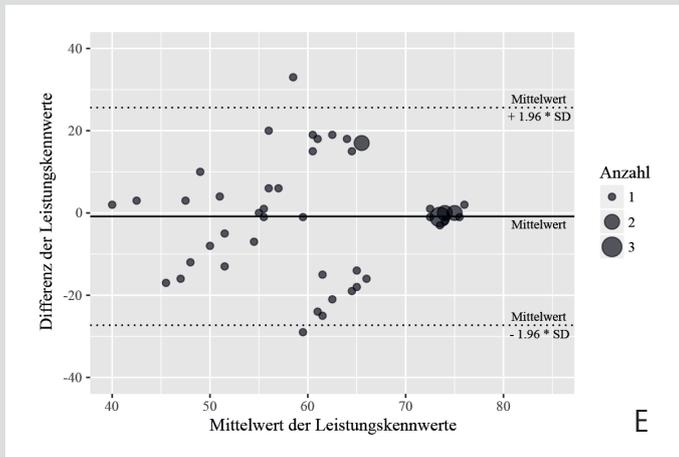
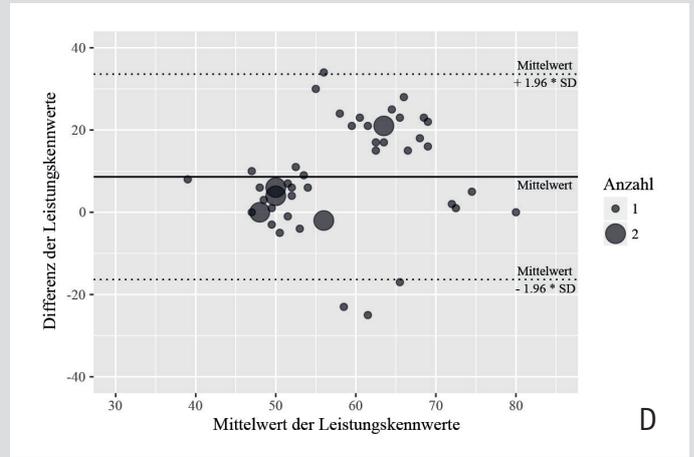
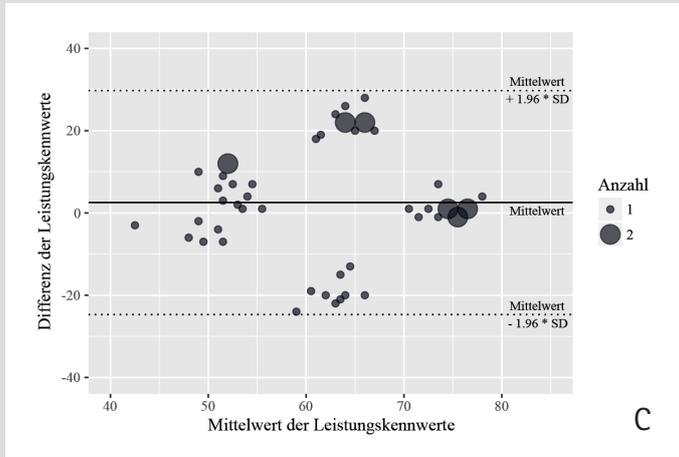
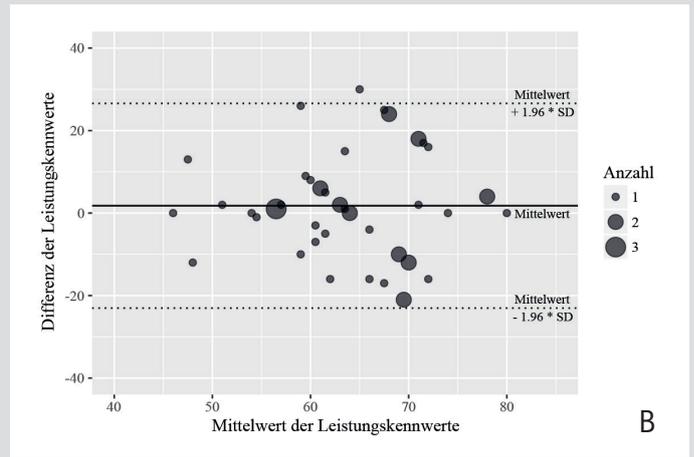
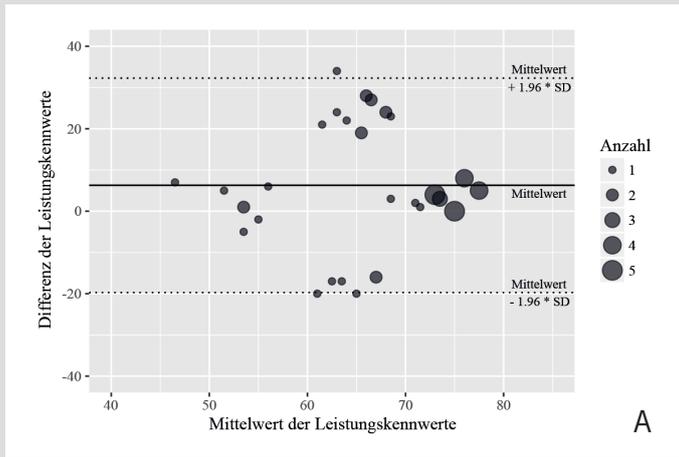


Bild 2: Bland-Altman-Diagramme der Leistungskennwerte

- A: Reaktionsfähigkeit (visuell).
 - B: Reaktionsfähigkeit (visuell Et auditiv).
 - C: Konzentration.
 - D: Aufmerksamkeit.
 - E: Orientierung.
 - F: Aufmerksamkeitsbelastbarkeit.
 - G: Arbeitsgedächtnis. Die durchgezogene Gerade gibt den Mittelwert der Differenzen an.
- Minuend: Leistungskennwert von Tag 2 (Paralleltest).
 Subtrahend: Leistungskennwerte von Tag 1 (Standardtest).
 SD: Standardabweichung.

Bild 3: Reaktionszeiten bei den Orientierungstests
Tag 1: Standardtest
Tag 2: Paralleltest

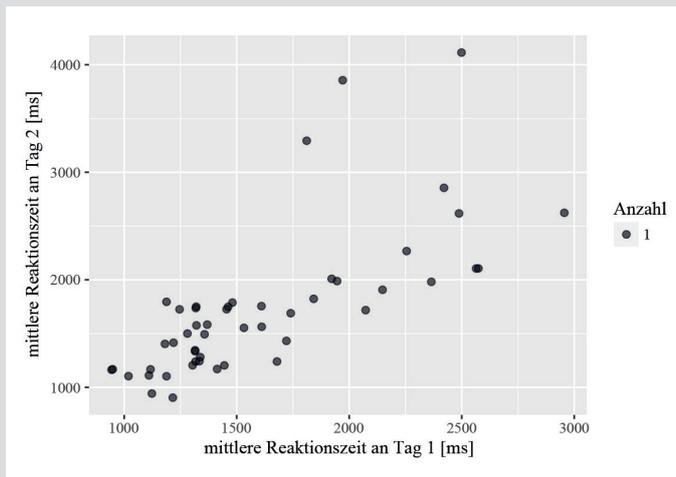


Tabelle 3: Probanden mit Leistungskennwerten ≤ 50 in den Orientierungstests

Leistungskennwert	Standardtest		Paralleltest	
	Männer	Frauen	Männer	Frauen
$x \leq 40$	0	1	0	2
$40 < x \leq 50$	1	5	3	7

von Standard- und Paralleltests sind fast alle deutlich höher als die entsprechenden für die Fehlerzahlen (Tabelle 2).

Bei den Orientierungstests hatten einige Probanden Schwierigkeiten, die sich auch in den Reaktionszeiten und den Leistungskennwerten widerspiegeln (Bild 3 und Tabelle 3).

4 Diskussion

4.1 Grad der Übereinstimmung zwischen Standard- und Paralleltests

Bei den Bland-Altman-Diagrammen unterscheiden sich die Mittelwerte der Differenzen teilweise deutlich von null (Bild 2). Solche systematischen Abweichungen (Verzerrungen) sind für die Praxis unproblematisch, weil sie bei der Bewertung der Ergebnisse von Paralleltests berücksichtigt werden können. Die Streuungen der Differenzen der Leistungskennwerte zwischen den Standard- und ihren jeweiligen Paralleltests sind hoch. Dies entspricht geringen Graden an Übereinstimmung. Dies sollte der Anwender berücksichtigen, wenn er die Ergebnisse von aktuellen Paralleltests bewertet. Wünschenswert wären klare Vorgaben des Ordnungsgebers zum geforderten Grad der Übereinstimmung (Rauch, Moosbrugger 2011, Wellek, Blettner 2012), die aber in den Begutachtungsleitlinien für Kraftfahrerngung leider nicht in einer hinreichend konkreten Form existieren. Daraus resultiert ein Ermessensspielraum für den Anwender, der die Beurteilung bzw. die Begutachtung

nicht erleichtert.

Der Vergleich der Korrelationskoeffizienten für die mittleren Reaktionszeiten und die Fehlerzahlen bei den jeweiligen Testpaaren zeigt, dass die hohen Streuungen der Leistungskennwertdifferenzen wesentlich durch die noch höheren Streuungen der Fehlerzahlen (gleichbedeutend mit denen richtiger Antworten) verursacht wurden. Bereits ein zusätzlicher Fehler kann eine erhebliche Abnahme des T-Werts verursachen. Beispielsweise ergibt ein fehlerfreier Orientierungstest einen T-Wert von 80, hingegen entspricht ein Fehler einem T-Wert von nur 55. Gemäß der Definition des Leistungskennwerts wird der T-Wert für die richtigen Antworten, der mit demjenigen für die Fehlerzahl identisch ist, doppelt so stark wie derjenige für die Reaktionszeit gewichtet. Somit wirken sich Fehler wesentlich stärker aus als ein langsames Arbeiten. Unser Kollektiv ist relativ jung und somit nicht repräsentativ für die Grundgesamtheit aller Prüflinge. Jüngere Probanden arbeiten häufig schnell auf Kosten einer hohen Fehlerrate (Schubert, Schneider et al. 2005). Nach unserem – subjektiven – Eindruck trifft das auch für etliche unserer Probanden zu. Somit stellt sich die Frage, ob die Instruktion hinreichend war, eine mehr ergebnisorientierte Arbeitsweise zu forcieren. Die Probanden wurden bewusst nicht nach ihrer Teststrategie gefragt, um einer Verzerrung der Ergebnisse vorzubeugen. Dies gilt insbesondere für Fragen nach dem ersten Testdurchgang, die das Testergebnis am

Folgetag beeinflussen könnten. Darüber hinaus hätten Teststrategien auch mit künftigen Probanden kommuniziert werden können, wenn die Versuchspersonen auf die Thematik aufmerksam gemacht worden wären. Aus den genannten Gründen schlagen wir eine methodische Modifikation vor. Die Definition des Leistungskennwerts sollte geändert werden, sodass Fehler weniger stark gewichtet werden. Diese Frage wird von den Testentwicklern geprüft (Mitteilung der Vistec AG). Zudem sollte geprüft werden, ob sich Fehler durch eine Probandeninstruktion, die stärker auf die Vermeidung von Fehlern fokussiert, vermeiden lassen. Das relativ junge Kollektiv ist nicht repräsentativ für alle Altersgruppen. Unsere Ergebnisse sind aber so eindeutig, so dass unabhängig von der Frage der Repräsentativität eine Optimierung der Tests empfohlen werden muss. Anschließend ist selbstverständlich eine Untersuchung der Paralleltestreliabilität an einem repräsentativen Kollektiv erforderlich.

Die in unserer Studie gefundenen Korrelationskoeffizienten für die Reaktionszeiten von Standard- und Paralleltests liegen zwischen 0,499 und 0,889 (Tabelle 2). Im Manual (Kapitel 9.1.4) werden Korrelationskoeffizienten für mehrere Tests angegeben. Sie betragen bei jeweils 25 Probanden 0,944 für die Reaktionsfähigkeit visuell, 0,958 für die Reaktionsfähigkeit visuell & auditiv, 0,752 für die Konzentration, 0,911 für die Aufmerksamkeit sowie 0,967 für die Aufmerksamkeitsbelastbarkeit (Berg, Nädtke 2015). Die Frage, worauf die Unterschiede zwischen diesen Korrelationen und unseren Resultaten zurückzuführen sind, lässt sich nicht beantworten. Aufgrund der geringeren Probandenzahl erscheinen die im Manual berichteten Korrelationen weniger belastbar im Vergleich zu den von uns ermittelten Werten.

4.2 Schwierigkeiten bei den Orientierungstests

Vor allem der Paralleltest, aber auch der Standardtest bereitete insbesondere einigen weiblichen Versuchspersonen Probleme. Diese spiegeln sich in der Anzahl benötigter Übungsdurchgänge und in den Ergebnissen wider (Bild 3 und Tabelle 3). Es sei ausdrücklich darauf hingewiesen, dass alle diese Personen gesund und kooperativ waren. Mit den anderen Tests hatten sie keine vergleichbaren Schwierigkeiten. Wegen der kleinen Fallzahlen ist unser Ergebnis nicht repräsentativ. Ein derart hoher Anteil von Personen mit Schwie-

rigkeiten bei den Orientierungstest wurde bei der Erstellung der Normdaten für den Standard- und Paralleltest nicht beobachtet. Es handelte sich um jeweils ca. 460 Personen, statistische Ausreißer waren nicht eingeschlossen (Mitteilung der Vistec AG). Eine mögliche Erklärung für unser Ergebnis könnte sein, dass in unserem Kollektiv überzufällig viele Probanden mit Problemen bei den Orientierungstests enthalten waren. Wir regen an, dieser Fragestellung bei künftigen Studien besondere Beachtung zu schenken. Sollte unser Resultat bestätigt werden, wäre es sinnvoll, die aktuellen Orientierungstests durch leichtere zu ersetzen.

Ein geschlechtsspezifischer Unterschied wurde auch beim Orientierungstest einer früheren Version des Corporal beobachtet: 17,2% der Frauen (n=6), aber nur 1,8% der Männer (n=1) wiesen eine mittlere Reaktionszeit von mehr als 8 Sek. auf (Müller 2010). Auf Gruppenbasis sind Männer bei Aufgaben mit einer mentalen Rotation bei einigen Untersuchungen im Vorteil (Coluccia, Louse 2004, Weiss, Deisenhammer et al. 2005). In einer Studie wurde bei einer hohen interindividuellen Variabilität gefolgert, dass 30 % der Frauen die mentale Rotation langsamer als die übrigen Frauen und die Männer vollzogen (Kail, Carter et al. 1979). Als Ursache wurden unterschiedliche Strategien bei der mentalen Rotation vermutet (Kail, Carter et al. 1979).

Für die Praxis ist von Bedeutung, dass (auch) die Orientierungstests nicht als repräsentativ für spezielle Aspekte der Fahreignung im Sinne einer Prognose über die Fahrleistung angesehen werden und dem-

entsprechend auch nicht dafür anhand einer Fahrprobe validiert sind. Vielmehr dienen Tests aus der Testbatterie dazu, eine Bewertung der psychophysischen Leistung des Probanden zu erlauben (Berg, Reimann et al. 2014). Dieser Tatsache sollte sich der Anwender bei der Beurteilung von Testergebnissen, insbesondere auch von Orientierungstests, bewusst sein. Wenn bestimmte kognitive Leistungsgebiete beeinträchtigt sind, kann das sichere Führen eines Kraftfahrzeugs in Frage gestellt sein.

4.3 Schlussfolgerung

Die Definition des Leistungskennwerts sollte zugunsten einer stärkeren Gewichtung der Reaktionszeit im Vergleich zur Fehlerzahl geändert werden, um einen höheren Grad der Übereinstimmung zwischen Standard- und Paralleltests erzielen zu können. Möglicherweise könnte eine geänderte Probandeninformation dazu beitragen, Fehler aufgrund einer zu stark zeitorientierten Arbeitsweise zu vermeiden. Diese Frage sollte in künftigen Studien geklärt werden.

Interessenskonflikt

Die Studie wurde von der Vistec AG, Olching, gefördert. Die Fragestellung und das Studiendesign wurden gemeinsam mit der Vistec AG festgelegt. Die Durchführung der Studie, die Auswertung der Daten, die Interpretation der Ergebnisse und das Abfassen des Manuskripts erfolgten in alleiniger Verantwortung der Autoren.

Literaturverzeichnis

- Akerstedt, T.; Gillberg, M. (1990): "Subjective and objective sleepiness in the active individual." *Int J Neurosci* 52(1-2): 29-37
- Amelang, M.; Schmidt-Atzert, L. (2006): *Psychologische Diagnostik und Intervention*. Berlin, Springer Verlag, 165-166
- Berg, M.; Nädtke, J. (Hrsg.) (2015): *Psychometrisches Testsystem Corporal Plus*. Testsystem zur Erfassung kognitiver Funktionen im bildlich-räumlichen Bereich. Fachliche Information. Olching, Vistec AG
- Berg, M.; Reimann, C.; Schubert, W. (2014): „Validierung leistungspsychologischer Testverfahren unter Aspekten der Verkehrssicherheit.“ *Z. f. Verkehrssicherheit* 60(3): 2-7. Bonn, Kirschbaum Verlag
- Bland, J. M.; Altman, D. G. (1986): "Statistical methods for assessing agreement between two methods of clinical measurement." *Lancet* 327(8476): 307-310
- Bortz, J.; Schuster, C., Eds. (2010): *Statistik für Human- und Sozialwissenschaftler*. Berlin, Springer Verlag, 35
- Coluccia, E.; Louse, G. (2004): „Gender differences in spatial orientation: A review.“ *J Environ Psychol* 24(3): 329-340
- du Prel, J.-B., Muttray, A. (2011): Validität von Diagnostik- und Screening-Studien. *Statistische Grundlagen. Grundlagen der Klinischen Ökonomik*. F. Porzolt (Hrsg.). Berlin Eigenverlag des PVS Verbandes: 43-56
- Gajda, K.; Sulzenbruck, S.; Heuer, H. (2016): „Financial incentives enhance adaptation to a sensorimotor transformation.“ *Exp Brain Res* 234(10): 2859-2868
- Gneezy, U.; Rustichini, A. (2000): "Pay enough or don't pay at all." *Quarterly J Economics* 115(3): 791-810
- Gräcmann, N.; Albrecht, M. (2018): *Begutachtungsleitlinien zur Kraftfahreignung*, Bundesanstalt für Straßenwesen
- Grouven, U.; Bender, R.; Ziegler, A.; Lange, S. (2007): „Vergleich von Messmethoden.“ *Dtsch Med Wochenschr* 132 Suppl 1: e69-73
- Jenkins, D.; Mitra, A.; Gupta, N.; Shaw, J. D. (1998): "Are financial incentives related to performance? A meta-analytic review of empirical research." *J Appl Psychol* 83: 777-787
- Kail, R., Carter, P., Pellegrino, J. (1979). „The locus of sex differences in spatial ability.“ *Percept Psychophys* 26(3): 182-186
- McCall, W. A. (1939): *Measurement: A revision of how to measure education*, Macmillan

Verlagesanzeige
185x80

Müller, J. (2010): Zur leistungsdiagnostischen Erfassung der Orientierungsfähigkeit in verkehrspsychologischen Kontexten – Untersuchungen zur Validierung des "Take Over B" Unveröffentlichte Diplomarbeit, Potsdam.

Müller, K.; Brenner-Hartmann, J.; Klieschke, U.; Krause, B.; Schmidt-Atzert, L.; Müller, A.; Jacobshagen, W.; Nickel, W.-R.; Schubert, W. (2013): Der Einsatz von psychologischen Testverfahren in der Begutachtung der Fahreignung. Beurteilungskriterien: Urteilsbildung in der Fahreignungsbegutachtung. W. Schubert, V. Dittmann und J. Brenner-Hartmann. Bonn, Kirschbaum Verlag: 273–317

Oldfield, R. C. (1971): "The assessment and analysis of handedness: the Edinburgh inventory." *Neuropsychologia* 9(1): 97–113

Poschadel, S.; Falkenstein, M.; Pappachan, P.; Poll, E.; Wilmes von Hinckeldey, K. (2009): Testverfahren zur psychometrischen Leistungsprüfung der Fahreignung. *Berichte der*

Bundesanstalt für Straßenwesen. Unterreihe Mensch und Verkehr. <http://bast.opus.hbz-nrw.de/volltexte/2011/245/pdf/M203.pdf>

R Core Team (2017): R: A language and environment for statistical computing. Wien. <https://www.r-project.org/>, R Foundation for Statistical Computing

Rauch, W.; Moosbrugger, H. (2011): Klassische Testtheorie. Grundlagen und Erweiterungen für heterogene Tests und Mehrfacettenmodelle. Enzyklopädie der Psychologie: Themenbereich B Methodologie und Methoden, Serie II Psychologische Diagnostik. L. F. Hornke, M. Amelang and M. Kersting (Hrsg.). Göttingen, Hogrefe. 2: 1–87

Robinson, L. J.; Stevens, L. H.; Threapleton, C. J.; Vainiute, J.; McAllister-Williams, R. H.; Gallagher, P. (2012): Effects of intrinsic and extrinsic motivation on attention and memory. *Acta Psychol (Amst)* 141(2): 243–249

Schermelleh-Engel, K.; Werner, C. S. (2012): Methoden der Reliabilitätsbestimmung. Testtheorie und Fragebogenkon-

struktion. H. Moosbrugger and A. Kelava (Hrsg.). Berlin, Springer Berlin Heidelberg: 119–141

Schubert, W.; Schneider, W.; Eisenmenger, W.; Stephan, E. (2005): Begutachtungs-Leitlinien zur Kraftfahreignung. Kommentar. Bonn, Kirschbaum Verlag

Weiss, E. M.; Deisenhammer, E. A.; Hinterhuber, H.; Marksteiner, J. (2005): „Geschlechtsunterschiede kognitiver Leistungen – populärwissenschaftliche Stereotypen oder evidenzbasierte Studienergebnisse?“ *Fortschr Neurol Psychiatr* 73(10): 587–595

Wellek, S.; Blettner, M. (2012): „Klinische Studien zum Nachweis von Äquivalenz oder Nichtunterlegenheit – Teil 20 der Serie zur Bewertung wissenschaftlicher Publikationen.“ *Dtsch Arztebl Int* 109: 674–679

Wickham, H., Ed. (2009): *ggplot2: Elegant graphics for data analysis*. New York, Springer Verlag

■ Verfasser

Prof. Dr. med. Axel Muttray
amuttray@uni-mainz.de



Axel Muttray war Oberarzt des Instituts für Arbeits-, Sozial- und Umweltmedizin der Universitätsmedizin der Johannes-Gutenberg-Universität Mainz und ist ab Juni 2019 im Ruhestand. Er ist Facharzt für Innere Medizin und Arbeitsmedizin und besitzt die Zusatzbezeichnungen Sozialmedizin und Umweltmedizin. 1999 habilitierte er in Arbeits- und Umweltmedizin. 2006 erfolgte die Ernennung zum außerplanmäßigen Professor. 2001 wurde ihm der Franz-Koelsch-Preis des Bayerischen Staatsministers für Arbeit und Sozialordnung verliehen. 2008 erhielt er den Best-Paper-Award der Gesellschaft für Arbeitswissenschaft. Zu seinen Forschungsgebieten zählen die Arbeitstoxikologie, die Arbeitsphysiologie einschließlich Fahrsimulationen sowie das Vibrationsbedingte Vasospastische Syndrom.

Anschrift
Institut für Arbeits-, Sozial- und Umweltmedizin
Universitätsmedizin der
Johannes-Gutenberg-Universität Mainz
Obere Zahlbacher Straße 67
55131 Mainz

Cand. med. Natalie Löb
natalie.loeb321@gmail.com



Natalie Löb hat 2009–2012 eine Ausbildung zur Gesundheits- und Krankenpflegerin absolviert und begann anschließend das Medizinstudium an der Johannes-Gutenberg-Universität in Mainz. Ihre Dissertation zur Paralleltestreliabilität des Corporal Plus wird von Herrn Prof. Muttray betreut.

Anschrift
Institut für Arbeits-, Sozial- und Umweltmedizin
Universitätsmedizin der
Johannes-Gutenberg-Universität Mainz
Obere Zahlbacher Straße 67
55131 Mainz

Dr. Alicia Poplawski
alpoplaw@uni-mainz.de



Alicia Poplawski arbeitet seit 2013 als Biostatistikerin und Bioinformatikerin am Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI) der Universitätsmedizin der Johannes-Gutenberg-Universität Mainz. Im Jahr 2017 Abschluss der Promotion zum Thema „Evaluation von Anwendungen für die Planung und Analyse von RNA-seq Daten“.

Anschrift
Institut für Medizinische Biometrie,
Epidemiologie und Informatik,
Universitätsmedizin der
Johannes-Gutenberg-Universität Mainz
Obere Zahlbacher Straße 69
55131 Mainz

Univ.-Prof. Dr. Heiko Hecht
hecht@uni-mainz.de



Heiko Hecht absolvierte sein Studium an den Universitäten Trier und Virginia, USA. Die Promotion erfolgte 1991 an der Universität von Virginia. Danach war er am Max-Planck-Institut für Psychologische Forschung tätig, an der Ludwig-Maximilians-Universität in München, am NASA Ames Research Center, am Zentrum für interdisziplinäre Forschung der Universität Bielefeld sowie am Center for Space Research (Man-Vehicle Laboratory) am Massachusetts Institute of Technology in Cambridge, MA. Seit 2002 ist Heiko Hecht Professor für Allgemeine Psychologie an der Johannes-Gutenberg-Universität Mainz. Er forscht zu grundlegenden und angewandten Themen der Kognitionspsychologie, u. a. zu Fahrsimulation, Wahrnehmen und Handeln in extremen Umgebungen, Kontaktzeitschätzung, Virtuelle Realität und künstliche Schwerkraft.

Anschrift
Psychologisches Institut der
Johannes-Gutenberg-Universität Mainz
Binger Straße 14–16
55122 Mainz

Dipl.-Psych Till F. Beutel
till.beutel@unimedizin-mainz.de



Till F. Beutel ist seit 2017 leitender Psychologe und Wissenschaftlicher Mitarbeiter am Institut für Lehrer- und Gesundheitsmedizin des Instituts für Arbeits-, Sozial- und Umweltmedizin der Universitätsmedizin Mainz. An diesem Institut ist er seit 2013 tätig. Zuvor war er in der Klinik für Psychiatrie sowie der Klinik für Psychosomatische Medizin beschäftigt. Er schloss sein Psychologiestudium im Jahr 2011 mit dem Diplom ab.

Anschrift
Institut für Arbeits-, Sozial- und Umweltmedizin
Universitätsmedizin der Johannes-Gutenberg-Universität Mainz
Obere Zahlbacher Straße 67
55131 Mainz

Univ.-Prof. Dr. med. Dirk-Matthias Rose
dirk-matthias.rose@unimedizin-mainz.de



Dirk-Matthias Rose ist wissenschaftlicher Leiter des Instituts für Lehrer- und Gesundheitsmedizin und stellvertretender Direktor des Instituts für Arbeits-, Sozial- und Umweltmedizin der Universitätsmedizin Mainz. Er ist Facharzt für Arbeits- und Allgemeinmedizin und hat die Zusatzbezeichnungen Flug-, Notfall-, Sozial- und Umweltmedizin. 2017 erhielt er die Joseph-Rutenfranz-Medaille der Deutschen Gesellschaft für Arbeits- und Umweltmedizin. Forschungsschwerpunkte sind die Entwicklung und Implementierung von E-Health-Produkten im Bereich der Arbeitsmedizin und des betrieblichen Gesundheitsmanagements.

Anschrift
Institut für Arbeits-, Sozial- und Umweltmedizin
Universitätsmedizin der Johannes-Gutenberg-Universität Mainz
Obere Zahlbacher Straße 67
55131 Mainz