

Der Einsatz von Eyetracking bei der Entwicklung der Online-Version des d2-R

Kontaktangaben

Dr. Herbert A. Meyer
artop - Institut an der Humboldt-Universität zu Berlin
Christburger Str. 4
10405 Berlin
meyer@artop.de

Bastian Funken
Hogrefe Verlag GmbH & Co. KG
Merkelstraße 3
37085 Göttingen
bastian.funken@hogrefe.de

Abstract

Der d2 ist ein seit 50 Jahren bewährter psychologischer Papier-und-Bleistift-Test zur Messung der Konzentrationsfähigkeit und Aufmerksamkeitsleistung. Bei der Entwicklung der Online-Version der Revision des d2 mussten wichtige Gestaltungsentscheidungen getroffen werden. Wie kann erreicht werden, dass das Testmaterial wie in der Papierform zeilenweise bearbeitet wird? Wie ist die Instruktion zu gestalten, wenn kein Testleiter zur Verfügung steht? Wie ist das Zeitlimit zu setzen, damit besonders leistungsstarke Personen an ihre Grenze stoßen? In einem Usability-Test mit einem High-Fidelity-Prototypen wurden Antworten auf diese Fragen erarbeitet, indem Screencasts mit synchronen Blickortüberlagerungen qualitativ analysiert wurden. Die Ergebnisse werden anhand von Highlightvideos präsentiert und zur Diskussion gestellt.

Keywords

Usability, Usability-Test, Eyetracking, Psychometrie, Testentwicklung

1. Einleitung

Der Test d2 wurde ursprünglich im Institut für Sicherheit im Bergbau, Industrie und Verkehr des Technischen Überwachungs-Vereins Essen entwickelt, um anhand des Merkmals „Aufmerksamkeit/Unaufmerksamkeit“ zwischen „guten“ und „schlechten“ Kraftfahren bzw. sicherheitsbewussten Arbeitern und sogenannten „Unfällen“ zu differenzieren. Die erste Version wurde 1962 als „Aufmerksamkeits-Belastungs-Test d2“ veröffentlicht (Brickenkamp, 1962). Mit der aktuellen 10. Auflage wurde das Testverfahren grundlegend überarbeitet und in d2-R (d2-Revision) umbenannt. Inhaltlich wird der d2-R nun als „Aufmerksamkeits- und Konzentrationstest“ bezeichnet, um kenntlich zu machen, welche psychologischen Konzepte er tatsächlich misst (Brickenkamp, Schmidt-Atzert & Liepmann, 2010). Das Testverfahren erfreut sich einer großen Beliebtheit in Forschung und Praxis. Aufgrund der vielfältig abgesicherten Testgütekriterien, der einfachen Anwendung und des geringen Aufwands an Zeit und Material gehört er zu den am häufigsten verwendeten psychodiagnostischen Verfahren in Deutschland.

Der vorliegende Beitrag gibt einen Einblick in einen Teil der Arbeiten zur Überführung der aktuellen Fassung des d2-R von der Papierform in das digitale Format (Online-Test). Hierbei musste die Herausforderung bewältigt werden, sowohl die psychometrische Qualität als auch die Einfachheit der Anwendung zu erhalten. Zudem sollte der Test optional über Maus- oder Fingereingaben durchgeführt werden können. Weiterhin sollte die Durchführung im Gegensatz zum Papier-und-Bleistift-Verfahren ohne die Anwesenheit eines Testleiters möglich sein.

2. Fragestellungen und Untersuchungsmaterial

Um die bearbeiteten Fragestellungen besser nachvollziehen zu können, wird die Art und Weise der Durchführung des d2-R kurz vorgestellt. Abbildungen 1 und 2 zeigen Ausschnitte aus der Kurzanleitung und dem Testbogen der aktuellen Papierversion.

Jedes d, das zwei Striche hat, durchstreichen!

"	"	"
d	d	d
"	"	"
2 Striche oben	2 Striche unten	1 Strich oben 1 Strich unten

Nicht durchstreichen:

d mit weniger oder mehr als 2 Strichen:	d	d	d	d	d
p, egal mit wie vielen Strichen:	p	p	p	p	p

Abb. 1: Ausschnitt aus der Kurzanleitung der Papierversion des d2-R, © Hogrefe Verlag

Auf der linken Seite von Abbildung 1 werden alle Varianten der Zielobjekte dargestellt, die beim Test durch manuelles Durchstreichen zu markieren sind (Targets). Auf der rechten Seite sind alle Zeichen aufgeführt, die nicht zu markieren sind (Distraktoren). Die offensichtliche Schwierigkeit des Testverfahrens besteht darin, Targets und Distraktoren beim schnellen und möglichst fehlerfreien Arbeiten voneinander zu unterscheiden.

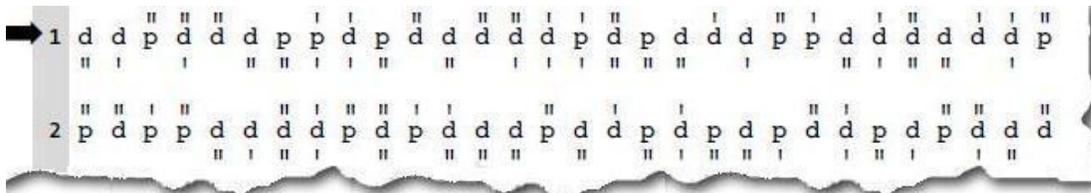


Abb. 2: Ausschnitt aus dem Testbogen der Papierversion des d2-R, © Hogrefe Verlag

Abbildung 2 zeigt einen Ausriss aus dem Testbogen des d2-R. Auf dem Testbogen finden sich insgesamt 14 Zeilen mit jeweils 57 Zeichen (in Abbildung 2 werden jeweils die ersten 30 Zeichen der ersten beiden Zeilen dargestellt). Der Testbogen muss zeilenweise von links nach rechts bearbeitet werden. Für jede Zeile gibt es eine Bearbeitungszeit von 20 Sekunden. Gibt der Testleiter nach Ablauf der 20 Sekunden das mündliche Signal „Halt! Nächste Zeile“ muss sofort mit der nächsten Zeile angefangen werden. Bevor der Test beginnt wird das Instruktionsverständnis durch zwei Übungsaufgaben auf dem Bogen mit der Kurzanleitung überprüft.

In Abbildung 3 wird ein Teil der Schritte bei der iterativen Entwicklung (vgl. Perfetti, 2007) der Online-Version des d2-R dargestellt. Der vorliegende Beitrag

konzentriert sich auf die dort erwähnte Eyetracking-Studie. Bei der Erzeugung des Interaktionskonzepts mussten unter Berücksichtigung der Anforderungen weittragende Designentscheidungen getroffen werden. Da einige Entscheidungen unsicher blieben, wurden bestimmte Gestaltungsvorschläge in einem voll-funktionsfähigen PC-Prototyp umgesetzt. Dieser Prototyp wurde in der Usability-Studie mit Eyetracking eingesetzt.

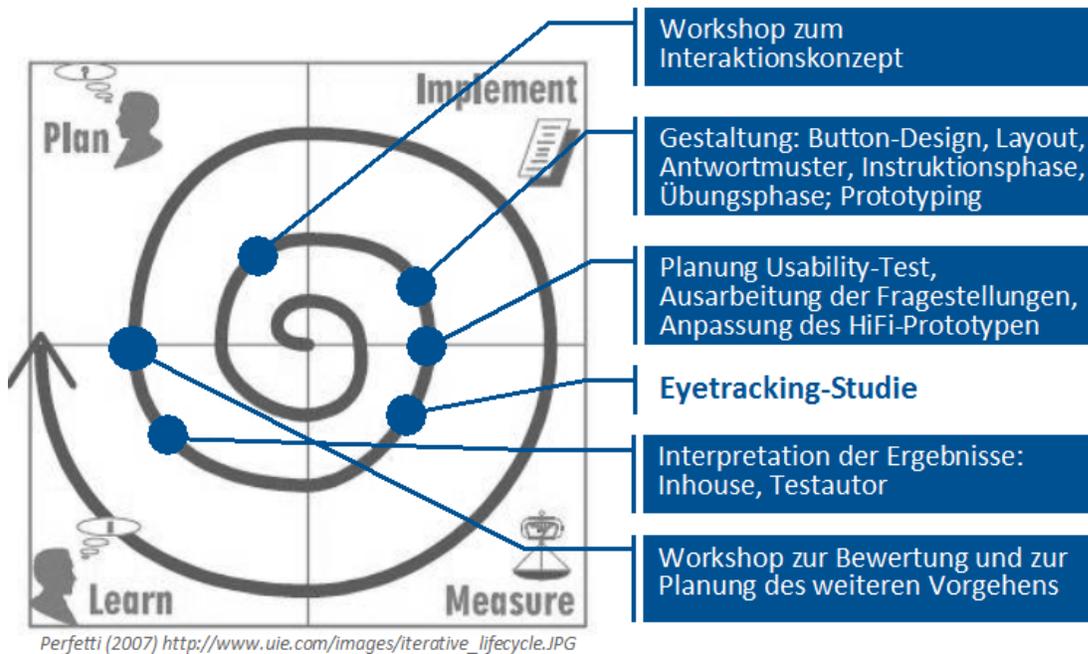


Abb. 3: Iterativer Zyklus bei der Entwicklung der d2-R Online-Version (Ausschnitt, Abbildung in Anlehnung an ein Modell von Perfetti, 2007)

Im Rahmen der Eyetracking-Studie sollte hauptsächlich die Frage geklärt werden, ob beim Online-Test wie bei der Papier-und-Bleistift-Version die zeilenweise Bearbeitung des Testmaterials gewährleistet ist (Fragestellung 1). Bei der Papier-und-Bleistift-Version kontrolliert ein Testleiter vor Ort mit Start- und Stoppsignalen, dass eine Zeile mit jeweils 57 Zeichen innerhalb von 20 Sekunden bearbeitet wird. Bei dem neuen Interaktionskonzept bekam jede Zeile einen eigenen Screen, d.h. sie wird nicht mehr als eine Zeile, sondern mit 6x10 Zeichen (Buttons) auf einem Screen dargestellt. Wir sprechen daher bei den Bearbeitungseinheiten nicht mehr von Zeile, sondern von Durchgang.

Abbildung 4 zeigt die für den PC-Prototyp neu entwickelte Anordnung mit 60 Testreizen (Buttons) für einen Durchgang. Es stellte sich die Frage, ob durch die gewählten Abstände zwischen den Buttons zuverlässig und quasi unbewusst erreicht wird, dass das Testmaterial zeilenweise bearbeitet wird (was nach dem „Gesetz der Nähe“ zu erwarten ist, vgl. dazu Abbildung 5). Eine zeilenweise Bearbeitung ist für die Normierung des Testverfahrens von großer Bedeutung. Zudem kann dadurch das technische Problem der Handballenaufgabe bei Touchscreens umgangen werden (Aktivierung von Folgezeilen erst nach einem bestimmten Zeitmuster).



Abb. 4: Darbietung des Testmaterials für einen Durchgang beim PC-Prototyp im Hogrefe TestSystem 5, © Hogrefe Verlag

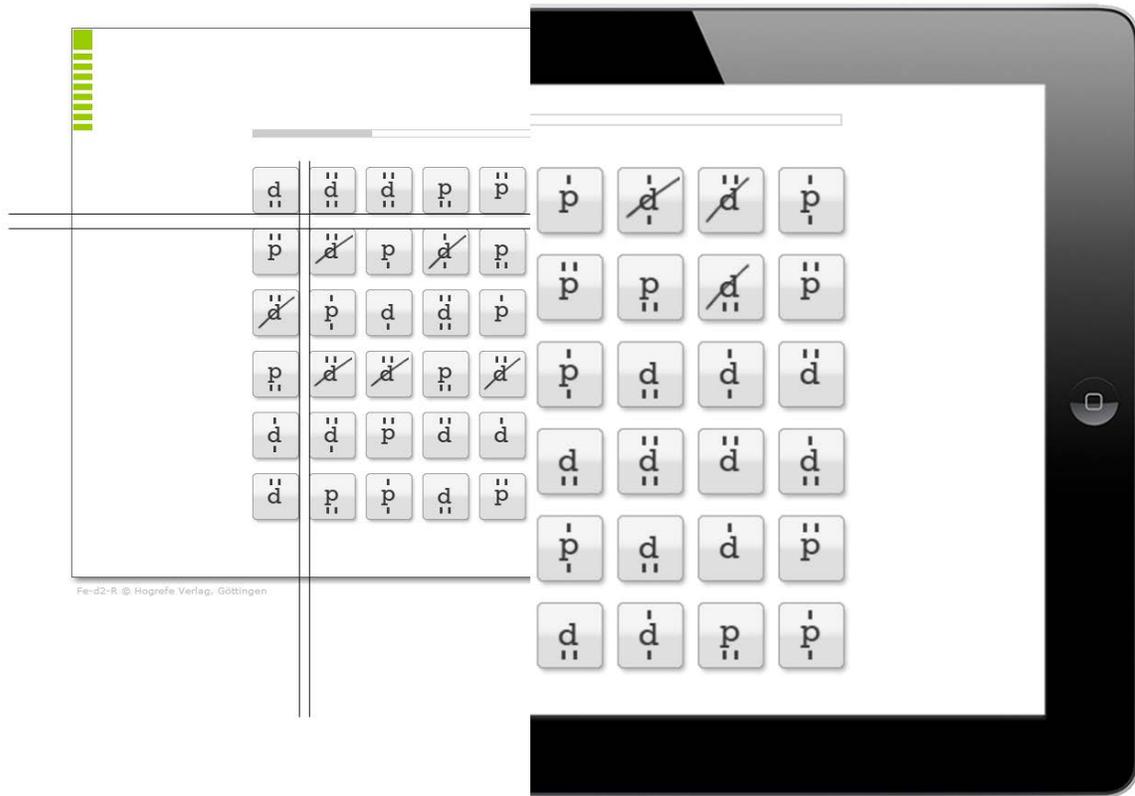


Abb. 5: Veranschaulichung der Anwendung des „Gesetz der Nähe“ zur Forcierung der zeilenweisen Bearbeitung beim PC-Prototyp im Hogrefe TestSystem 5, © Hogrefe Verlag

Da im Vorfeld überlegt wurde, ob beim Online-Test auf eine Übungsphase verzichtet werden kann, sollte bei der Untersuchung gezielt beobachtet werden, ob bei der Anleitung des Online-Tests eine Übungsphase von Vorteil ist (Fragestellung 2). Zusätzlich sollte erkundet werden, ob das Zeitintervall zur Bearbeitung eines Durchgangs richtig gewählt ist (Fragestellung 3). Die Testreize werden wie bei der Papier-und-Bleistift-Version 20 Sekunden pro Durchgang dargeboten. Es besteht die Gefahr von Deckeneffekten, wenn das Intervall für die Online-Version zu kurz gewählt ist.

3. Methode

Die Studie wurde mit insgesamt acht Testpersonen durchgeführt. Vier Testpersonen bekamen eine Instruktion und eine Übungsphase, bevor sie den

vollständigen Test mit 14 Durchgängen bearbeiteten. Drei weitere Testpersonen bekamen nur die Instruktion und keine Übungsphase. Als achte Testperson wurde zur Prüfung der dritten Fragestellung ein erfahrener Computerspieler mit einer ausgeprägten psychomotorischen Begabung rekrutiert. Diese Testperson durchlief zunächst eine Trainingsphase. Danach sollte sie versuchen, bei möglichst vielen der 14 Durchgänge das vorgegebene Zeitintervall zu unterlaufen.

Die acht Testsitzungen wurden als Screencast mit Blickortüberlagerung aufgezeichnet (mit hervorgehobenen Maus-Ereignissen und einem Webcam-Bild der Testperson). Das in den Videos dokumentierte Verhalten wurde qualitativ analysiert.

4. Ergebnisse

Die erste Fragestellung konnte eindeutig beantwortet werden. Die gewählte Anordnung der Button-Matrix sorgt zuverlässig dafür, dass Testbenutzer die einzelnen Durchgänge zeilenweise bearbeiten. Es wurde kein einziges gegenläufiges Verhalten beobachtet.

Die zweite Fragestellung konnte mehr oder weniger eindeutig beantwortet werden. Es sollte keinesfalls auf eine Übungsphase verzichtet werden. Zwei der drei Testpersonen ohne Übungsphase zeigten Verständnis- bzw. Orientierungsprobleme bei dem ersten Durchgang. Ähnliche Probleme wurden bei Testpersonen mit Übungsphase nicht beobachtet. Die inhaltliche Analyse der Leseaktivitäten gab wertvolle Hinweise dazu, welche Passagen der Instruktion modifiziert bzw. gekürzt werden können.

Auch die dritte Fragestellung wurde erfolgreich bearbeitet. Trotz intensiver Bemühung konnte die trainierte Testperson das Zeitintervall von 20 Sekunden bei ihren 14 Durchgängen nicht fehlerfrei unterlaufen. Bei einer vertiefenden Frame-by-frame-Analyse der Blickbewegungen zeigten sich als Nebenergebnis Einzelheiten der sehr schnellen Informationsverarbeitung, die zu der überdurchschnittlich guten Performanz führten. Es konnte sichtbar gemacht werden, wie bei hoher Verarbeitungsgeschwindigkeit schon parallel zur

motorischen Reaktion ein folgendes Zeichen fokussiert wird. Dabei wird in einem Blicksprung eine bestimmte Anzahl von Zeichen überflogen und das fokussierte Zeichen verarbeitet und gegebenenfalls motorisch darauf reagiert. Wenn über parafoveale Wahrnehmung unter den übersprungenen Zeichen ein Target entdeckt wurde, findet ein rückwärtsgerichteter Blicksprung statt, wenn nicht, ist der nächste Blicksprung vorwärtsgerichtet.

5. Diskussion

Die dargestellten Ergebnisse wurden bei der weiteren Entwicklung des d2-R konsequent berücksichtigt. Die Übungsphase wurde fest verankert und die Instruktion verbessert, so dass die Testbearbeitung zügig und ohne Rückfragen ablaufen kann. Die überprüfte Buttonmatrix wurde 1:1 übernommen. Das Zeitintervall von 20 Sekunden pro Durchgang wurde beibehalten. Das Verfahren befindet sich seit Mitte 2013 in der Normierungsphase für Desktop- und Tablet-PCs und wird aller Voraussicht nach 2014 auf den Markt kommen.

Bei der Präsentation der Ergebnisse soll die qualitative Analyse von Blickbewegungsaufzeichnungen als Erfolgsmethode zur Diskussion gestellt werden. Dazu werden Screencasts mit synchroner Blickortüberlagerung in Echtzeit und in Zeitlupe gezeigt. Anhand der Anschauungsbeispiele wird beiläufig ein nachhaltiges Verständnis davon erzeugt, wie der Mensch visuelle Informationen über Fixationen und Sakkaden verarbeitet. Es wird demonstriert, dass über die qualitative Interpretation von Blickbewegungsaufzeichnungen Effekte aufgeklärt werden können, indem kognitive Aktivitäten bei der Mensch-Computer-Interaktion verstanden werden. Es wird gezeigt, wie Eyetracking auch ohne quantitative Auswertungen (z.B. in Form von Heatmaps und Areas of Interest) beim Usability-Testing gewinnbringend eingesetzt werden kann.

Literatur

Brickenkamp, R. (1962). Aufmerksamkeits-Belastungs-Test (Test d2). Göttingen: Hogrefe.

Brickenkamp, R., Schmidt-Atzert, L. & Liepmann, D. (2010). d2-R - Test d2 - Revision (Aufmerksamkeits- und Konzentrationstest). Göttingen: Hogrefe.

Perfetti, C. (2007, August). Five survival techniques for creating usable products. Abgefragt am 16.06.2014, von https://uie.com/articles/product_survival_techniques/

Vita

Dr. Herbert A. Meyer war von 1988 bis 2000 als Experimentalpsychologe in der Forschung tätig. Seit 2001 arbeitet er bei artop im Arbeitsbereich Usability/User Experience und ist u.a. Dozent bei der Ausbildung zum Usability Consultant mit dem Schwerpunkt Methoden. Seit 2010 betreut er als Lehrbeauftragter ein Blickbewegungslabor an der Hochschule für Technik und Wirtschaft Berlin.



Bastian Funken ist als Diplom-Psychologe in den Bereichen Softwareentwicklung und User Experience Design tätig. Seit 1998 gehört computer- und internetbasierte Diagnostik im Bereich der Kognitionspsychologie zu seinen Schwerpunkten. Seit 2009 leitet er die Abteilung "E-Assessment" der Hogrefe Verlagsgruppe und koordiniert die Entwicklung eines Testsystems an neun Standorten in Europa.

