

Melvin Georg, Alexander Kahlert, Sascha Kotris, Felix Kunold, Heiko Tröndle

Quantitative Analyse des kognitiven Verständnisses in Bezug auf Risikoberichterstattungen von Finanzinstituten

Forschungsbericht zum lehrintegrierten Forschungsprojekt

Studiengang BWL-Bank

Jahrgang 2015

Version 1.0 vom 05.06.2019

Abstract

Der Risikobericht als ein maßgeblicher Teil des Jahresabschlusses umfasst die Angaben zu eingegangenen Risiken sowie die Risikohöhe, um für die Stakeholder Transparenz herzustellen. Dabei kann der steigende Detaillierungsgrad zu einer mangelnden Übersichtlichkeit des Risikoberichtes führen, so dass auch Grafiken zur Visualisierung eingesetzt werden. Wie diese Grafiken gelesen und verstanden werden, ist noch nicht ausreichend erforscht. In der vorliegenden Studie wurde mittels Eye-Tracking das Leseverhalten von ausgebildeten Probanden erforscht. Es wurde ein Zusammenhang mit Fähigkeiten Risikoinformationen zu verstehen und dem Blickverhalten bzw. der Beantwortung der Fragen getestet. Die Legende, Skalen und Fragestellung wurden besonders fokussiert. Dabei wurde auch ein Zusammenhang mit der Leseschwierigkeit sichtbar. Zudem scheint auch die Geschwindigkeit, mit der auf Zielregionen fixiert wird, je nach Antwort unterschiedlich zu sein.

Schlüsselbegriffe: Risikoberichterstattung Kreditinstitute, Eye-Tracking, Risk Numeracy

1 Einführung

Chance und Risiko sind zwei untrennbar miteinander verknüpfte Begriffe. Besonders das Geschäft von Kreditinstituten basiert auf dem Eingehen von Risiken, die zu erweiterten Ertragschancen führen. Aus diesem Grund ist es für Stake- und Shareholder von besonderer Bedeutung, transparente Informationen bezüglich der Höhe der eingegangenen Risiken der jeweiligen Institute zu erhalten. Diese Risiken werden innerhalb des Jahresabschlusses in einem Risikobericht veröffentlicht. Der Risikobericht ist somit ein maßgeblicher Teil des Jahresabschlusses. Im Zuge der Finanzkrise 2007 ist das von Banken eingegangene Risiko wieder verstärkt in den Fokus des öffentlichen Interesses gerückt. Dies führt dazu, dass sowohl professionelle Investoren als auch Privatanleger sich vermehrt für die Risikoberichte eines Kreditinstitutes interessieren. Dementsprechend stehen die Banken als Herausgeber der Risikoberichte vor der Herausforderung, einen gut verständlichen und trotzdem lückenlosen Risikobericht zu erstellen, der sowohl den professionellen als auch den öffentlichen Interessenten einen Überblick über die eingegangenen Risiken ermöglicht.

Die zunehmenden gesetzlichen Ansprüche wie IFRS 7 sowie die Ausweitung der aufsichtsrechtlichen Anforderungen führen zu einer erhöhten inhaltlichen Komplexität der Risikoberichte, welche unweigerlich mit einem größeren Umfang dieser einhergeht. So müssen die Risikoberichte die eingegangenen Risiken und Risikomanagementprozesse immer detaillierter wiedergeben. Dieser steigende Detaillierungsgrad kann zu einer mangelnden Übersichtlichkeit des Risikoberichtes führen. Damit wird die Möglichkeit einer einfachen Verständlichkeit der Risikoberichte gehemmt. Folglich ist es essentiell wichtig, dass die Inhalte des Risikoberichtes in einer so exakten, eindeutigen und übersichtlichen Weise wie möglich dem Leser präsentiert werden, denn „Informationspflichten müssen sich, um sinnvoll zu sein, an den Informationsinteressen der Informationsadressaten orientieren.“¹ Um dies gewährleisten zu können,

¹ Moxter, A., Grundsätze ordnungsgemäßer Rechnungslegung, Düsseldorf (2003), S.233. Zitiert in Zepp, M. (2007) S.83.

muss ein umfassendes Verständnis über das Leseverhalten der Leser von Risikoberichten vorhanden sein.

Da dies in der einschlägigen Literatur noch nicht ausreichend analysiert wurde, wird in der vorliegenden Studie diese Lücke geschlossen. Hierzu wird das Leseverhalten beispielhafter Probanden beim Lesen von Risikoberichten einer Bank durch das Eye-Tracking-Verfahren beobachtet. Die vorliegende Untersuchung konzentriert sich auf einen Auszug der Untersuchung, den exemplarischen Ausschnitt eines Risikoberichtes der DZ Bank. Auf Basis dieser Beobachtungen werden in dieser Arbeit Hypothesen erstellt und mittels unterschiedlicher Testverfahren auf ihre Gültigkeit geprüft. Mit den daraus gewonnenen Erkenntnissen können zukünftig Risikoberichte adressatenorientierter erstellt werden.

Die Untersuchung ist wie folgt aufgebaut: Zuerst werden die Relevanz und der Stand der Untersuchung in der aktuellen Fachliteratur dargestellt. Darauf folgt eine Erläuterung der Methodik und des durchgeführten Eye-Tracking Experiments. Im Anschluss daran werden die Thesen und Ergebnisse der Untersuchung besprochen. Abschließend wird ein Fazit gezogen und die Studie kritisch reflektiert.

2 Methodische Hintergründe

Als Probanden dienen 29 Studierende der Bankbetriebswirtschaft des sechsten Semesters an der Dualen Hochschule Baden-Württemberg am Campus Stuttgart, die im Bereich Kreditinstitute fachkundig sind. Die Versuchsleitung oblag Thomas Berger und Thorsten Wingenroth.

Der Aufbau des Experiments wird von der Versuchsleitung vorgegeben und an einem Tag durchgeführt, die Auswertung erfolgt durch das Autorenteam. Die Untersuchung wird an zwei Laptops mit 17-Zoll Bildschirmen und jeweils angebrachtem Tobii X3-120 Eye-Tracker mit 120 Hz Abtastrate in abgetrennten Bereichen an einem ruhigen Ort durchgeführt. Die Probanden werden per Zufall auf die beiden Versuchsstationen zugeteilt, die Reihenfolge der Beispiele wird manuell balanciert. Die Lichtverhältnisse sind mit 355 Lux und 239 Lux an beiden

Stationen zwar unterschiedlich, jedoch stabil über den Erhebungszeitraum. Die Darstellung der Risikoberichte erfolgt als Teil eines längeren Experiments, das ca. 10 Minuten dauert. Die Versuchsleitung ist während der Erhebung durchgehend anwesend.

Der Gesamttablauf lässt sich wie folgt zusammenfassen: Die Probanden werden in der Gesamtgruppe über den Ablauf und den Hintergrund aufgeklärt. Danach werden je zwei Personen in den Versuchsraum gebeten und beantworteten zunächst Fragen eines Fragebogens. Der Fragebogen besteht zunächst aus Fragen zur Person und deren bisherigen Kenntnissen und Erfahrungen. Dazu kommen Fragen zum numerischen Risikoverständnis (Berlin Numeracy Test von Cockely et al. 2012) sowie zum graphischen Verständnis (Graph Literacy Test von Garcia-Retamero, R./Galesic, M. 2013). Diese beiden Tests sind in der Literatur weit verbreitet und akzeptiert. Nach der Beantwortung der Fragen werden die Probanden an die beiden Eye-Tracker gebeten. Der Arbeitsplatz wird eingerichtet und eine Kalibrierung durchgeführt. Nach Abnahme der Qualität der Kalibrierung durch die Versuchsleitung startet das Experiment und wird eigenständig von den Probanden durchgeführt sowie beendet. Nachdem die letzte Person teilgenommen hat, wird in der Gruppe eine Nachbesprechung mit Aufklärung zu den Testfragen und Beispielen durchgeführt. Zudem werden einige erste deskriptive Ergebnisse gezeigt. Die Rohdaten werden anschließend durch die Versuchsleitung aufbereitet, Heat Maps sowie Areas of Interest gebildet und in aufbereiteter Form der Gruppe zur Verfügung gestellt. Die Auswertung erfolgte anschließend mit dem Programm JASP in der Version 0.8.

3 Literaturüberblick

Der folgende Überblick über die Literatur zum Thema teilt sich in zwei Bereiche. Zum einen in den, der das Darstellen, Verstehen und Lesen von Risikoberichten untersucht, zum anderen in den Bereich, der sich mit der Forschungsmethode Eye-Tracking beschäftigt.

3.1 Verständnis von Risikoberichten

In der Literatur finden sich unterschiedliche Gründe, warum Risikoberichte untersucht werden. So kann dies geschehen, um die Sicherheit im Rahmen der Unternehmensbewertung für den Investor zu erhöhen.² Ebenso gibt es Arbeiten, die aus der Perspektive von Aufsichtsgremien untersuchen, wie solche Berichte aufgebaut sind, beziehungsweise wie Aufsichtsgremien in ihrer Arbeit unterstützt werden können.³ Des Weiteren gibt es auch eine Untersuchung, die sich ganz konkret mit Anforderungen an Risikoberichte von Kreditinstituten beschäftigt und daraus Handlungsempfehlungen ableitet.⁴ Dies entspricht auch dem Fokus der vorliegenden Forschungsarbeit und ist deswegen von besonderer Bedeutung.

Auch gibt es bereits Forschung, die sich durch eine Befragung damit beschäftigt, wie Geschäftsberichte als Ganzes wahrgenommen werden. Dabei legt diese den Fokus auf aktive und ehemalige Studierende der Wirtschaftswissenschaften als Probanden.⁵ Hieraus ergeben sich erste Leitsätze, welche visuellen Faktoren für die Wahrnehmung des Betrachters von besonderer Bedeutung sind.⁶ Aufgrund der gewählten Probanden und des Schwerpunkts auf visuelle Aspekte ist gerade diese Arbeit interessant im Vergleich zur vorliegenden. Ähnliches gilt für eine Forschungsarbeit aus Großbritannien, welche die Lesbarkeit von Risikoberichten mithilfe des Flesch-Index⁷ untersucht.⁸ Allerdings ist hier das Ziel, Schlüsse darüber zu ziehen, ob bestimmte Berichte bewusst schwer verständlich geschrieben werden. Diese Beispiele könne nur ein kleiner Ausschnitt der Forschung zum Thema Risikoberichte sein, sie zeigen allerdings die vielen möglichen Perspektiven und Methoden, die angewendet werden.

² Vgl. Filipiuk, B. (2008), S. 1.

³ Vgl. Mertens, P. (2016), S. 154 ff.

⁴ Vgl. Zepp, M. (2009).

⁵ Vgl. Grosse, G./Ripperda, K./Voelzkow, A. (2016).

⁶ Vgl. Ebenda (2016), S. 67 f.

⁷ Dieser Index misst die Schwierigkeit von Texten anhand der Anzahl der Wörter und Silben. Vgl. Baumgarth, C./Eisend, M./Evanschitzky, H. (2009).

⁸ Vgl. Linsley, P. M./Lawrence, M. J. (2007).

3.2 Grundlagen zum Eye-Tracking

Um die Komplexität des Verstehens von Risikoberichten zu quantifizieren, kann die Methode des Eye Tracking herangezogen werden. Im Folgenden werden gängige Metriken des Eye Tracking kurz vorgestellt.

Das Auge eines Menschen hat einen kleinen, kreisförmigen Bereich auf der Netzhaut, die sogenannte Fovea. Das Auge bewegt sich ständig zu dem, was es fokussiert (sog. „Eye-Mind-Hypothese“, Just & Carpenter, 1980). Durch die Fovea kann der Mensch den fokussierten Bereich scharf sehen. Zu beachten ist, dass in den meisten Fällen das kognitiv verarbeitet wird, was in der Fovea fokussiert wurde (Eye-Mind-Hypothese).⁹ Das Eye-Tracking-Instrument misst die Bewegungen und Fokussierungen der Fovea, indem infrarotes Licht ausgestrahlt wird und aus den Reflektionen die Position des Auges auf dem Bildschirm bestimmt wird. Dabei können werden die Positionsdaten ständig ermittelt (120-mal pro Sekunde) und in zwei Arten von Bewegungen eingeteilt:

- **Fixationen** (Ruhezeiten und Verarbeitungszeiten der Augen) und
- **Sakkaden** (Zeiten, in denen sich das Auge bewegt).¹⁰

Für den vorliegenden Eye-Tracking-Versuch eignet sich vor allem die Analyse der Fixationen der Probanden. Gemäß der Eye-Mind-Hypothese kann die Dauer einer Fixation als Maß für die kognitive Verarbeitungstiefe und mentale Beanspruchung genutzt werden.¹¹

⁹ Vgl. Link, D. et al. (2008), S. 374.

¹⁰ Vgl. Strobel, B., Lindner, M. A., Saß, S., & Köller, O. (2018).

¹¹ Vgl. Link, D. et al. (2008), S. 374.

4 Ergebnisse

Die Eye-Tracking Ergebnisse basieren auf folgender Aufteilung der Abbildung in verschiedene *Areas of Interest* (AOI):



Abbildung 1: Aufteilung der Areas of Interest (AOI)

Der Bildschirm wird in 11 AOI eingeteilt mit 5 Hauptbereichen: Den Teilgrafiken (DZ und WGZ), den Titeln, den Skalen links, den Zielbereichen sowie den Bereichen der Erläuterung des Value-at-Risk (VaR). Daneben gibt es noch die (Rest-)Fläche der Grafik. Hinzu kommt unten die Frage, die zu beantworten ist. Das Hauptinteresse gilt im Folgenden dem Zielbereich, der fokussiert werden muss, um die Frage beantworten zu können.

4.1 Schwerpunkte der Aufmerksamkeit

Eine erste Auswertung der Schwerpunkte der Aufmerksamkeit in Form einer Heat Map findet sich in Abbildung 2. Dabei sind die transparenten Bereiche die Bereiche, die länger fixiert wurden.

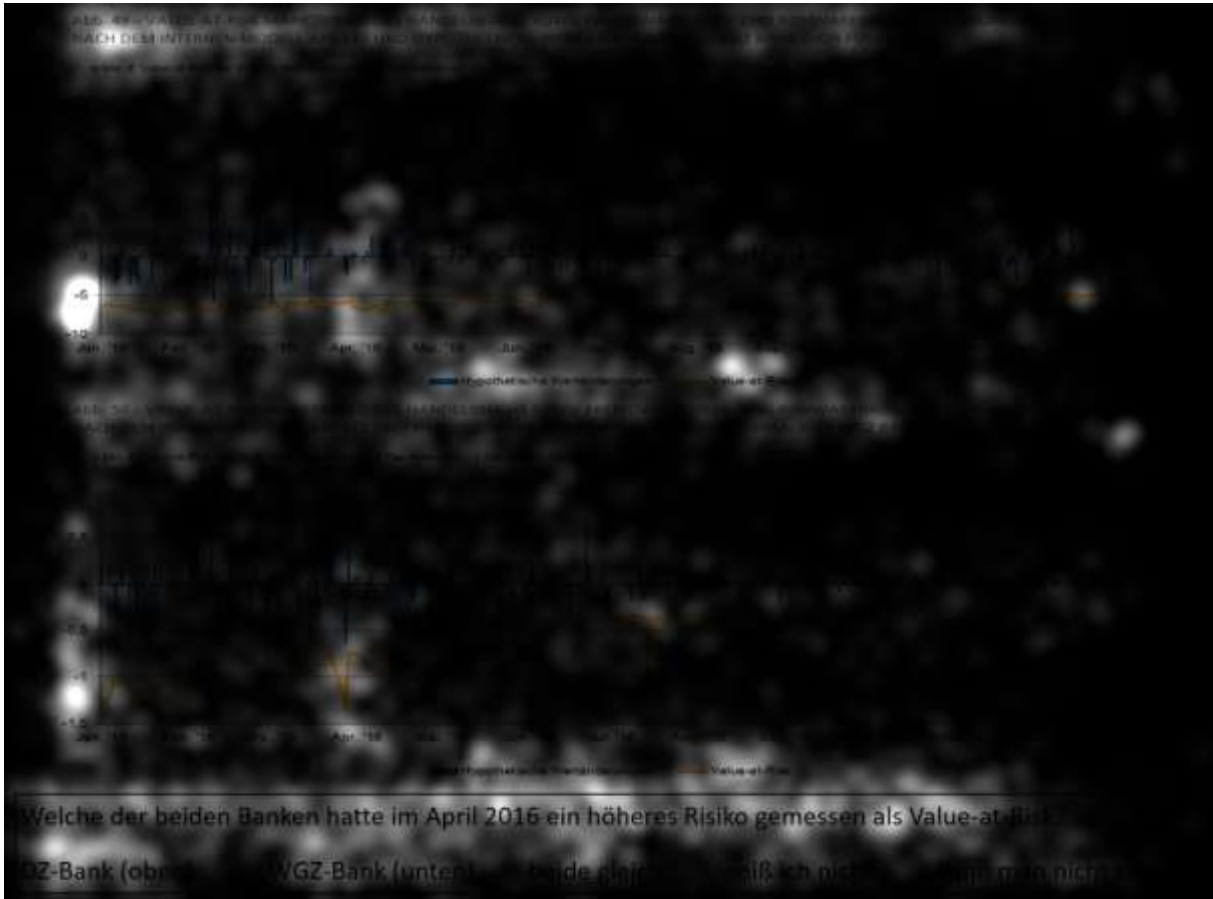


Abbildung 2: Heat Map zur Beispielgrafik (Quelle: Tobii Pro Studio 3.4.8)

Anhand obiger Heat Map lassen sich die Schwerpunkte der Probanden bei Betrachtung der Abbildung entnehmen. Es ist deutlich zu erkennen, wie insbesondere die Skalen, die Fragestellung und die Legende hervortreten. Es scheint also, dass die Probanden bei der Betrachtung die für die Beantwortung der Frage notwendigen Schwerpunkte gesetzt haben. Allerdings fällt auch auf, dass über die gesamte Länge der Abbildungen zahlreiche Betrachtungen stattfinden, die nicht zielführend sind. Es ist zu erkennen, dass insbesondere die Graphen umfangreicher betrachtet werden, als für die Frage nötig. Diese Erkenntnisse bestätigen sich auch

bei Untersuchung der Summe der Fixationen. Hierfür wird die Dauer aller Fixationen innerhalb einer AOI verwendet. Man spricht von der *fixation based dwell-time*, also der Verweildauer ohne Sakkaden.

AOI	Fläche DZ	Fläche WGZ	Frage	Skala DZ	Skala WGZ	Titel DZ
Dwell-time	192,25	163,19	312,53	50,1	47,88	149,92
AOI	Titel WGZ	Var DZ	Var WGZ	Ziel DZ	Ziel WGZ	
Dwell-time	83,59	43,75	33,07	58,06	47,04	

Tabelle 1: Dauer aller Fixationen in den AOI

Anhand der Daten ist deutlich zu erkennen, dass andere Bereiche als die Zielbereiche insgesamt länger betrachtet werden. Hier sind insbesondere die Graphen zu nennen, die in den AOIs „Fläche DZ“ und „Fläche WGZ“ enthalten sind. Allerdings bestätigen sich auch Eindrücke der Heat Map; so etwa, dass die AOI „Frage“ besonders häufig betrachtet wird.

4.2 Time to first Fixation und Antwortverhalten

Wir gehen im Folgenden hypothesen-geleitet vor. Eine erste Hypothese wird aus der Heat Map abgeleitet: *Personen, die schneller den Zielbereich erkennen, beantworten auch eher die Frage richtig.*

Es geht hier also um die Effizienz einer Suche, gemessen als die Zeitdauer vom Start des Experiments bis zur ersten Fixation über 200ms (*Time to First Fixation - TTFF*). Diese zeigt, in welcher Zeitspanne das erste Mal auf eine bestimmte AOI vertieft geblickt wird. Hieraus lässt sich eine Kennzahl – der T50-Wert – ableiten. Dieser drückt aus, wie schnell 50% der Probanden auf einer bestimmten AOI fixierten. In vorliegendem Fall wird der Wert für die AOI „Ziel DZ“ und „Ziel WGZ“ bestimmt. Die Daten werden vorab um die Probanden bereinigt, bei denen keine TTFF für diese Bereiche festgestellt wurde. Zusätzlich wird noch unterschieden

zwischen denjenigen, die die Frage zur Abbildung richtig oder falsch beantworteten. Eine graphische Auswertung findet sich in Abbildung 3.

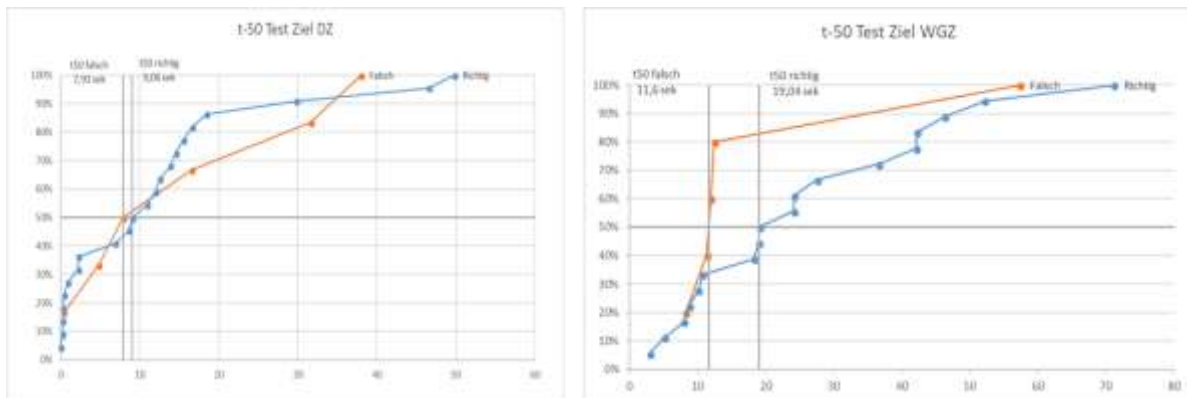


Abbildung 3: TTF für die Zielbereiche ‚DZ‘ und ‚WGZ‘

Es ist erkennbar, dass die AOI „Ziel DZ“ einen geringeren T50-Wert aufweist als die AOI „Ziel WGZ“. Dies könnte ein Indiz dafür sein, dass der erstgenannte Zielbereich mehr Interesse beim Leser geweckt hat. Dieser Unterschied ist deswegen interessant, da die Bereiche sich sehr ähnlich sehen und insbesondere der Graph im AOI „Ziel WGZ“ durch seine Ausschläge auf den ersten Blick als auffälliger bezeichnet werden könnte. Der genaue Grund für den geringeren T50-Wert für das „Ziel DZ“ kann an dieser Stelle nicht geklärt werden. Es wäre allerdings denkbar, dass die Tatsache, dass dieser Bereich über dem „Ziel WGZ“ steht, eine Rolle spielt. Dies ließe sich eventuell mit der gewohnten Leserichtung von links nach rechts und von oben nach unten im Sprachraum der Probanden begründen.

Eine Aussage über den Zusammenhang der T50-Werte und der richtigen Beantwortung der Frage ist schwierig. Dies hängt besonders mit den wenigen Datensätzen zusammen (22 für „richtig“ und fünf für „falsch“), die nach der Bereinigung noch zur Verfügung standen. Nach vorliegender Auswertung haben diejenigen, die die Frage falsch beantwortet haben, schneller den Zielbereich fixiert, als diejenigen, welche die Frage richtig beantwortet haben. Dies ist jedoch nur ein erster Hinweis; weitergehende Analysen sind nicht möglich.

4.3 Tiefe der Informationsverarbeitung

Als weitere Hypothese wird überprüft, *ob Probanden mit einem besseren graphischen Verständnis die Informationen der Zielbereiche tiefer analysieren*, da sie gezielt diese Informationen verarbeiten. Hierzu wird als Kennzahl die *Reading depth (RD)* benötigt, definiert als Fixationsdauer zu Fläche der AOI. Zudem müssen die Probanden in zwei Gruppen mit unterschiedlichem graphischen Verständnis aufgeteilt werden. Beides wird im Folgenden kurz erläutert.

Zur Bestimmung der *Reading depth* werden zunächst die gefilterten Summen aller Fixationen (in Sekunden) auf die Zielbereiche (Ziel WGZ und Ziel DZ) der jeweiligen Probanden summiert. Der Filter wird durch den Versuchsleiter auf 200ms festgelegt. Nach Holmqvist ist davon auszugehen, dass jegliche Fixationen geringer als 200 ms nicht bewusst gelesen werden.¹² Auf dieser Basis werden nur Teilnehmer berücksichtigt, bei denen in beiden Zielbereichen valide Fixationen gemessen werden konnten ($n = 29$). Die Fixationsdauer wird im Anschluss durch die Pixelanzahl der Zielbereiche dividiert (Zielbereich DZ = Zielbereich WGZ = 41,684 Tsd. Px). Somit ist die *Reading depth* pro Proband bestimmt.

Während des Versuchsablaufs mussten die Probanden drei Fragen zum graphischen Verständnis beantworten. Die Fragen wurden in Anlehnung an Garcia-Retamero, R./Galesic, M. gestellt.¹³ Aus den Antworten wird der Median ermittelt. Um die Probanden in zwei Gruppen zu segmentieren, wird mit Hilfe des Medians ein Index für das graphische Verständnis ermittelt: Gruppe eins enthält Probanden mit einem Wert gleich dem Median von eins ($n = 21$), Gruppe zwei enthält Probanden mit einem Wert über dem Median ($n = 8$). Gruppe eins stellt dabei die Gruppe mit einem besseren graphischen Verständnis dar.

¹² Vgl. hierzu ausführlich: Holmqvist et al. (2015), S. 390.

¹³ Vgl. hierzu ausführlich: Garcia-Retamero, R./Galesic, M. (2013), S. 250 ff.

Im Folgenden kann nun die Hypothese überprüft werden. Wird der Mittelwert der beiden Gruppen betrachtet, so kann man feststellen, dass Gruppe 1 die Zielbereiche intensiver verarbeitet hat. Gleichzeitig ist festzustellen, dass bei Gruppe 1 vereinzelt Extremwerte auftreten (siehe Violin-Plot Abbildung 4). Als weitere Metrik kann der Median analysiert werden. Hier ist ebenfalls auffällig, dass die *Reading depth* in den Zielbereichen in Gruppe 1 größer ist.

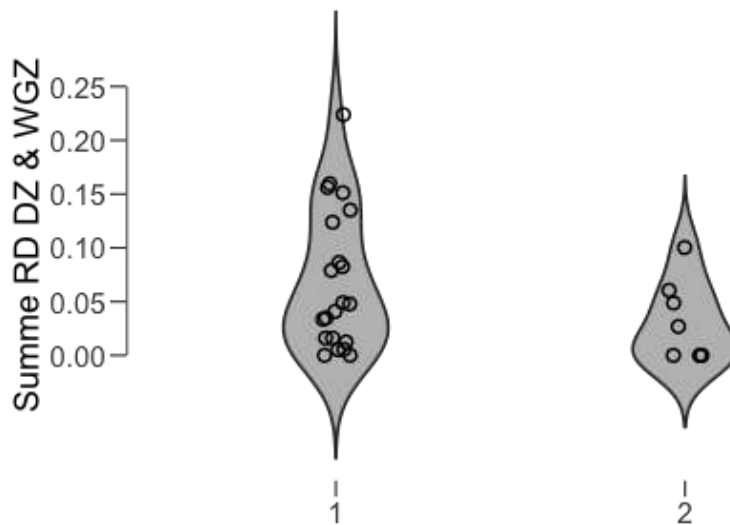


Abbildung 4: Vergleich der Reading Depth (RD) und des graphischen Verständnisses

Deskriptive Statistiken

Gruppe	Summe RD DZ & WGZ	
	1	2
Valide	21	8
Mittelwert	0.069	0.030
Median	0.048	0.013

Durch die deskriptiven Statistiken lässt sich die Hypothese nicht abschließend bestätigen, dass Probanden mit einem besseren graphischen Verständnis die Grafiken intensiver betrachten. Ob diese Beobachtung signifikant ist, soll im weiteren Verlauf mit dem t-Test ermittelt werden.

Nach dem Levene-Test kann Varianzgleichheit beider Gruppen angenommen werden ($p > 0,05$). Nach dem Shapiro-Wilk-Test kann Gruppe 1 jedoch nicht als normalverteilt angenommen werden ($p < 0,05$). Aus diesem Grund wird im weiteren Verlauf mit dem Mann-Whitney-Test weitergerechnet.

Im Rahmen dieser Ausarbeitung wird die Nullhypothese überprüft, ob Probanden mit einem besseren grafischen Verständnis die Grafiken intensiver betrachten ($H_0 = \text{Gruppe 1} > \text{Gruppe 2}$), sodass sich die Alternativhypothese von $H_A = \text{Gruppe 1} < \text{Gruppe 2}$ ergibt.

Mann-Whitney U-Test

	W	p	Rank-Biserial Correlation
Summe RD DZ & WGZ	119.000	0.959	0.417

Mit dem vorliegenden U-Test lässt sich die Nullhypothese mit einem p-Wert von über 95% nicht ablehnen. Somit liegt bei der vorliegenden Stichprobe sehr wahrscheinlich ein Zusammenhang von graphischem Verständnis und *Reading depth* mit mittlerem Effekt ($r=0,417$) vor.

4.4 Leseschwierigkeit

In der nun folgenden Hypothese soll untersucht werden, inwiefern Schwierigkeiten bei der Betrachtung der Zielbereiche der Graphen Einfluss auf die korrekte Beantwortung der Frage haben. Konkret wird folgende These aufgestellt:

Probanden, die die gestellte Frage korrekt beantwortet haben, hatten weniger Schwierigkeiten bei der Betrachtung der Zielbereiche DZ und WGZ (siehe Area of Interests) als falsch antwortende Probanden.

Um die Hypothese untersuchen zu können, ist die Kennzahl *Difficulty* erforderlich. Sie errechnet sich als Fixationsanzahl dividiert durch die Pixelanzahl der jeweiligen AOI (siehe

Holmqvist et al 2015). Es wird unterstellt, dass Probanden mit weniger Schwierigkeiten bei der Beantwortung weniger häufig fixieren müssen.

Um hier eine statistische Untersuchung durchführen zu können, werden die Probanden nach richtiger bzw. falscher Beantwortung der Frage segmentiert (1 = richtige Antwort; 2 = falsche Antwort). Nach Streichung von invaliden Messungen umfasst die Stichprobe, $n = 23$ Probanden, von denen vier die Frage falsch beantwortet haben. Im nächsten Schritt wird pro Proband die *Difficulty* bezogen auf die Zielbereiche berechnet.

	DZ_Ziel		WGZ_Ziel	
	1	2	1	2
Valid	18	5	18	5
Missing	0	0	0	0
Mean	0.2786	0.4030	0.2452	0.4414
Median	0.2639	0.2879	0.1799	0.5278
Std. Deviation	0.2197	0.3888	0.3443	0.2565
Minimum	0.02399	0.09596	0.02399	0.02399
Maximum	0.6957	1.080	1.511	0.6717

Die deskriptive Analyse zeigt auf, dass im Mittel die Probanden mit der falschen Antwort im Zielbereich DZ eine höhere *Difficulty* als die richtig antwortenden Probanden aufweisen ($0,403 > 0,279$). Auch im WGZ-Zielbereich liegen sowohl *Difficulty*-Mittelwert als auch Median der falsch antwortenden Probanden über denen der richtig antwortenden ($0,441 > 0,245$ bzw. Median $0,529 > 0,18$). Die Hypothese wäre demnach, basierend auf den vorliegenden Daten, deskriptiv zu bestätigen. Kritisch anzumerken ist, dass 6 Messungen nicht valide sind, darunter zwei Messungen von Kandidaten, die die Frage falsch beantwortet haben.

Für eine inferenzstatistische Untersuchung werden zunächst die Voraussetzungen für einen t-Test geprüft. Der Levene-Test lässt dabei auf eine Varianzgleichheit schließen. Gemäß Shapiro-Wilk kann jedoch eine Normalverteilung der *Difficulty*-Werte nicht angenommen

werden. Um die Signifikanz der obigen Erkenntnisse zu untersuchen, wird daher der Mann-Whitney-Test angewandt (Signifikanzniveau: 5%). Als Alternativhypothese wird festgelegt, dass Gruppe 1 (richtige Antwort) eine höhere *Difficulty* als Gruppe 2 (falsche Antwort) aufweist. Folgende Tabelle stellt die Ergebnisse gemäß Mann-Whitney-Test dar:

Independent Samples T-Test

	W	p	Rank-Biserial Correlation
Difficulty DZ	37.00	0.737	-0.178
Difficulty WGZ	22.50	0.958	-0.500

Note. Mann-Whitney U test.

Note. For all tests, the alternative hypothesis specifies that group 1 is greater than group 2 .

Es ist zu erkennen, dass beide p-Werte nicht signifikant sind. Die aufgestellte Null-Hypothese (*Difficulty* korrekt antwortender < *Difficulty* falsch antwortender) kann folglich nicht abgelehnt werden. Probanden mit richtiger Antwort scheinen also weniger Schwierigkeiten bei der Betrachtung der Zielbereich zu haben. Die Aussagekraft der Mann-Whitney-Ergebnisse wird über die Rank-Biserial-Korrelation berechnet. Gemäß der Effektstärkenkonvention nach Cohen ist die Effektgröße des Tests als eher schwach bis mittel anzusehen. Der Zusammenhang zwischen *Difficulty*-Wert und korrekter Beantwortung der Frage ist beim DZ-Bild eher schwach ausgeprägt, beim WGZ-Bild hingegen mittelgroß.

4.5 Einfluss der Vorerfahrung im Risikomanagement

Eine weitere Hypothese beschäftigt sich mit der Frage, *ob Probanden, die angaben, bereits Erfahrung im Risikomanagement zu haben (Gruppe 1), die vorgelegte Abbildung schneller bearbeitet haben.* Die korrekte Beantwortung der Frage spielt hierbei ausdrücklich keine Rolle, da ausschließlich die Geschwindigkeit der Bearbeitung untersucht werden soll. Hierzu wird die Zeit betrachtet, die von der Einblendung der Abbildung bis zur Einblendung der

nächsten Abbildung vergeht, die so genannte *Total Visit Duration* (TVD). Die Daten werden um diejenigen Probanden bereinigt, bei denen die Aufzeichnung zur Abbildung nicht vollständig war. Bei den verbleibenden 17 Datensätzen sind die Probanden mit Risikomanagement-Erfahrung im Mittel schneller (62,38 Sek.), als die ohne Erfahrung im genannten Bereich (64,094 s). Die Signifikanz dieser Ergebnisse wird anschließend mit einem t-Test überprüft. Vorab werden Normalverteilung durch den Shapiro-Wilk-Test und Varianzgleichheit durch den Levene-Test überprüft und angenommen werden (Signifikanzniveau 5%).

Es wird die Hypothese geprüft, dass die Gruppe mit Erfahrung eine geringere Zeit für die Bearbeitung braucht.

Independent Samples T-Test					
		t	df	p	Cohen's d
TVD		-0.113	15.000	0.456	-0.060

Tabelle 2: T-Test für Total Visit Duration bei Risikomanagement-Erfahrung

Es zeigt sich, dass bei vorliegendem t-Wert von -0.113 die genannte Hypothese bei einem Signifikanzniveau von 5% nicht angenommen werden kann. Folglich sind die Unterschiede im Mittelwert sehr wahrscheinlich nicht mit einem realen Unterschied begründet. Somit besteht bei vorliegender Stichprobe wohl kein Zusammenhang zwischen der Risikomanagement-Erfahrung und einer schnellen Bearbeitung der Aufgabe.

4.6 Einfluss der Mathematiknote

Als weiteres wird die Hypothese analysiert, ob Probanden mit einer überdurchschnittlich guten Mathenote weniger Zeit zur Beantwortung der Frage benötigen. Im Versuchsablauf sollen die Teilnehmer ihre Abitur-Mathenote sowie die Mathenote im Studium angeben. Um eine bessere Aussagekraft über die Mathefähigkeiten der Probanden treffen zu können, werden beide Noten zu je 50% berücksichtigt. Im Rahmen des Abiturs wird ein breiteres Spektrum an

Themengebieten abgedeckt, wohingegen im Rahmen des Studiums einzelne Themen tiefergehend unterrichtet werden. Um weitere Analysen durchführen zu können, wurden die Versuchsteilnehmer in zwei Gruppen (Gruppe 1 = überdurchschnittliche Mathenote; Gruppe 2 = unterdurchschnittliche Mathenote) eingeteilt. Ausschlaggebend ist der Durchschnitt der individuellen Mathenote im Vergleich zum Gesamtdurchschnitt (bei der vorliegenden Stichprobe $\bar{X} = 2,0$). Es werden ausschließlich die Teilnehmer ($n = 28$) berücksichtigt, bei denen zwei Mathenoten im Fragebogen angegeben sind, die anderen werden nicht bei den Berechnungen berücksichtigt.

Im nächsten Schritt wird die Total Visit Duration *TVD* berechnet. Die *TVD* eines Probanden ergibt sich aus der Summe aller Verweilzeiten auf jedem AOI zuzüglich der Zeit, bis die Frage beantwortet wurde, sofern der Proband nicht auf den Bildschirm geschaut hat ($\sum \text{TVD on AOI} + \text{TVD not on AOI} = \sum \text{TVD}$).

Betrachtet man den Jitter und Violin Plot als erste Indikation, kann man sehen, dass die Gruppe 1 eine tendenziell kürzere *TVD* haben als die Gruppe 2. Allerdings schlagen insbesondere zwei Probanden in der ersten Gruppe mit einer *TVD* über 100 aus. Dies spiegelt sich auch bei der Betrachtung des Durchschnitts wieder. Die Gruppe 1 hat einen Mittelwert von 52,399 wohingegen die Gruppe 2 einen kürzeren Durchschnitt von 44,807 hat. Aufgrund einer möglichen Verzerrung der beiden Probanden mit einer deutlich längeren *TVD* in Gruppe 1 ist eine erweiterte Betrachtung des Medians sinnvoll. Der Median der Gruppe 1 beträgt 40,155 und der Gruppe 2 52,535. Dies würde die Hypothese bestätigen, dass Probanden mit einer besseren Mathenote weniger Zeit zum Bearbeiten der Aufgabe benötigen. Ob diese Beobachtung signifikant ist, soll im weiteren Verlauf mit dem t-Test ermittelt werden.

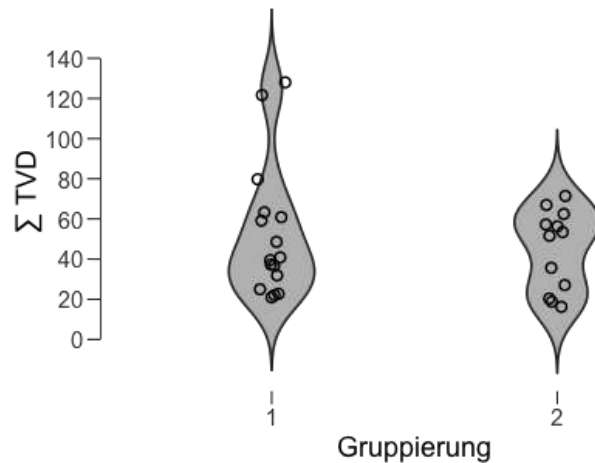


Abbildung y: Vergleich der Mathematiknoten (1 und 2) und Total Visit Duration

Nach dem Shapiro-Wilk Test konnte Gruppe 1 nicht als normalverteilt angenommen werden ($p < 0,05$). Nach dem Levene's Test kann Varianzgleichheit angenommen werden ($p > 0,05$). Aus diesem Grund wurde im weiteren Verlauf mit dem Mann-Whitney-Test weiter gerechnet.

Im Rahmen dieser Ausarbeitung wird die Nullhypothese überprüft, ob Probanden mit einer besseren Mathenote schneller in der Beantwortung der Aufgabe sind ($H_0 = TVD \text{ Gruppe 1} < TVD \text{ Gruppe 2}$), sodass sich die Alternativhypothese von $H_A = TVD \text{ Gruppe 1} > TVD \text{ Gruppe 2}$ ergibt.

	U	p	Rank-Biserial Correlation
Σ TVD	105.000	0.349	0.094

Abbildung 4: U-Test Vergleich Bearbeitungszeit und Mathematiknote

Mit dem vorliegenden U-Test lässt sich die Nullhypothese bei einem Signifikanzniveau von 5% nicht ablehnen. Somit kann bei der vorliegenden Stichprobe ein positiver Zusammenhang zwischen Mathenote und Bearbeitungsdauer unterstellt werden. Der Effekt ist allerdings so klein, dass er kaum praktische Bedeutung hat (Rank Biserial Correlation $< 0,1$).

5 Fazit – Zusammenfassung:

In der vorliegenden Arbeit wurden durch eigene Datenerhebung Hypothesen in Bezug auf die Risikoberichterstattung abgeleitet und getestet, um das Blickverhalten von Anwendern kennenzulernen. Hierbei wurden insbesondere die Fähigkeiten der Probanden im Hinblick auf das graphische Verständnis, Mathefähigkeiten und Erfahrungen im Risikomanagement berücksichtigt.

Analysiert man die Ergebnisse aus der deskriptiven Statistik, kann man einen Zusammenhang mit den oben genannten Fähigkeiten der Probanden und dem Blickverhalten bzw. der Beantwortung der Fragen erkennen. Diese Beobachtungen erscheinen nach weiteren Tests zum Vergleich der Mittelwerte aus den vorliegenden Daten auch als signifikant. Dies gilt beispielsweise für den positiven Zusammenhang von guten Mathematikfertigkeiten und kürzerer Lesedauer. Ähnliches gilt für die tendenziell korrektere Beantwortung bei geringerer Leseschwierigkeit bei dem Leser. Für vorliegende Untersuchung hat sich gezeigt, dass sich subjektiv logisch erscheinenden Zusammenhänge zumindest nach den hier verwendeten Verfahren herstellen lassen. Allerdings unterstreichen die überwiegend geringen Werte der Effektstärke, dass die Bedeutung dieser Ergebnisse nicht überschätzt werden sollte.

Es ist kritisch zu reflektieren, dass nach Bereinigung der Datensätze nur sehr kleine Stichproben zur Überprüfung der Hypothesen vorhanden sind. Generell wurden nur 29 Studierende des 6. Semesters aus dem Studiengang BWL-Bank (mit der Spezialisierung Gesamtbanksteuerung) an der DHBW in Stuttgart in der vorliegenden Studie berücksichtigt. Um weitergehende Aussagen treffen zu können, sollte die Stichprobe deutlich erweitert werden und weitere Studiengänge bzw. Hochschulen eingebunden werden.

6 Literaturverzeichnis

- Baumgarth, C./Eisend, M./Evanschitzky, H. (2009):** Empirische Mastertechniken. Eine anwendungsorientierte Einführung für die Marketing- und Managementforschung. Wiesbaden.
- Cockely et al. (2012):** Measuring Risk Literacy: The Berlin Numeracy Test. In: Judgment and Decision Making, Jg. 7, Nr. 1. S. 25-47.
- Filipiuk, B. (2009):** Transparenz der Risikoberichterstattung: Anforderungen und Umsetzung in der Unternehmenspraxis. Wiesbaden.
- Garcia-Retamero, R./Galesic, M. (2013):** Transparent Communication of Health Risks. Overcoming Cultural Differences. New York.
- Grosse, G./Ripperda, K./Voelzkow, A. (2016):** Visuell berichten. Eine Studie zum Potenzial der Gestaltung in Geschäftsberichten. Wiesbaden.
- Holmqvist et al. (2015):** Eye Tracking: A Comprehensive Guide to Methods and Measures. Oxford.
- Just, M. A., & Carpenter, P. A. (1980).** A theory of reading: From eye fixations to comprehension. Psychological Review, Jg. 87, Nr. 4, S. 329–354.
- Link, D. et al. (2008):** Berührungslose Augen- und Blickbewegungsmessung. In: Schmidt, L./Schlick, C. M./ Grosche, J. (Hrsg.): Ergonomie und Mensch-Maschine-Systeme. Heidelberg, S.371-392.
- Linsley, P. M./Lawrence, M. J. (2007):** Risk reporting by the largest UK companies: readability and lack of obfuscation. In: Accounting, Auditing & Accountability Journal, Jg. 20, Nr. 4, S. 620-627.
- Mertens, P. (2016):** Das digitalisierte Kontrollorgan? In: Bär, C./Gulden, H./Fischer, A. (Hrsg.): Informationstechnologien als Wegbereiter für den steuerberatenden Berufsstand. Berlin, Heidelberg. S. 149-162.
- Strobel, B. et al. (2018):** Task-irrelevant data impair processing of graph reading tasks: An eye tracking study. Learning and Instruction. Jg. 55, Nr. 6, S. 139-147.
- Zepp, M. (2007):** Der Risikobericht von Kreditinstituten: Anforderungen, Normen, Gestaltungsempfehlungen. Berlin

Dieser Bericht entstand in einem studiengangübergreifenden Forschungsprojekt zum Thema *Risikoberichterstattung* an der DHBW Stuttgart.

Er wurde von Studierenden des 6. Semesters BWL-Bank im Vertiefungsfach *Gesamtbanksteuerung* erstellt.

Kontakt:

Studiengang Wirtschaftsingenieurwesen

Prof. Dr. Thomas Berger

Kronenstr. 40

70174 Stuttgart

thomas.berger@dhbw-stuttgart.de

Studiengang BWL-Bank

Prof. Dr. Thorsten Wingenroth

Herdweg 23

70174 Stuttgart

thorsten.wingenroth@dhbw-stuttgart.de