

Gamification: Eine experimentelle Untersuchung der Spielelemente Abzeichen und Story

Athanasios Mazarakis

Arbeitsgruppe Web Science, Institut für Informatik, Christian-Albrechts-Universität zu Kiel

Zusammenfassung

Wissenschaftliche Untersuchungen zu Gamification verwenden häufig nur Ergebnisse aus Umfragen. Außerdem werden oftmals Spielelemente nicht isoliert, sondern in Kombination mit anderen Spielelementen untersucht. In der vorliegenden Feldstudie wird diese Lücke geschlossen. Mit 364 Probanden wurde untersucht, ob die Spielelemente Story und Abzeichen zu signifikant mehr Partizipation in einem Quiz führen als bei einer Kontrollgruppe ohne diese Elemente. Die Spielelemente wurden hierzu einzeln untersucht. Zusätzlich wurden Geschlechterunterschiede analysiert. Die Ergebnisse zeigen einen statistisch signifikanten Vorteil für die Nutzung von Abzeichen und Geschlechtsunterschiede im Vorwissen.

1 Einleitung und verwandte Arbeiten

Motivation und ihre Steigerung ist ein viel diskutierter Bereich und hängt von vielen Faktoren ab, da sich Personen in ihren Eigenschaften sehr unterscheiden können (Ryan & Deci, 2000b). Eine besonders häufig verwendete Unterteilung ist die intrinsische und extrinsische Motivation (Ryan & Deci, 2000a). Intrinsische Motivation bezieht sich auf die Durchführung einer Tätigkeit, wenn diese wegen der Tätigkeit selbst durchgeführt wird und keine andere Belohnung gegeben wird. Im Gegensatz dazu ist die extrinsische Motivation durch äußere Anreize gekennzeichnet, zum Beispiel durch Geld oder eine andere materielle Entlohnung (Ryan & Deci, 2000a). Obwohl allerdings monetäre Anreize oftmals notwendig sind für die Rekrutierung von Probanden, so sind für eine längerfristige Motivation zwingend auch nicht-monetäre Anreize in Betracht zu ziehen (Smadja, 2009). Ein solcher nicht-monetärer Anreiz um eine Motivationssteigerung zu erreichen, ist Gamification. Unter Gamification wird allgemein die Verwendung von Spielelementen in spielfremden Kontexten verstanden (Deterding et al., 2011). Als Spielelemente werden zum Beispiel Punkte oder Ranglisten angesehen (Hamari et al., 2014).

Die Definition von Deterding et al. (2011) wurde allerdings bereits mehrfach kritisiert. So führen zum Beispiel Hamari & Huotari (2016) an, dass das Erleben der Spielerfahrungen wichtiger sein sollte als die Erreichung eines Ziels. Der bloße Einsatz von Spielelementen kann nicht zwingend zu einer positiven Spielerfahrung führen, womit auch gleichzeitig keine

Motivationssteigerung erreicht wird. Gamification wird in dieser Darstellung eher als Prozess angesehen und nicht nur als Einsatz von Spielelementen (Huotari & Hamari, 2016).

Gamification kann als Versuch angesehen werden, intrinsische Motivation zu generieren oder zu steigern (Hamari et al., 2014). Dieser Versuch ist aber diskussionswürdig, zumal Gamification extrinsische Motivation aktivieren soll bzw. extrinsische Elemente verwendet. Trotzdem können Ansatzpunkte zur Verknüpfung von intrinsischer und extrinsischer Motivation in Kombination mit Gamification konstruiert werden (Denny, 2014).

Zwei Spielelemente, welche in der vorliegenden Studie genauer betrachtet werden, sind Abzeichen (Badges) und Story. In einer zweijährigen Studie von Hamari (2017) wurde gezeigt, dass eine statistisch signifikante Aktivitätssteigerung durch den Einsatz von Abzeichen generiert werden kann. Abzeichen können hierbei für bestimmte Aufgaben oder für die allgemeine und unspezifische Nutzung vergeben werden (Hamari, 2017). Aber auch das Spielelement Story wird als motivationssteigernd beschrieben, da es einer Tätigkeit eine (höhere) Bedeutung geben kann und damit positive Emotionen fördert (Kucwaj, 2016).

Geschlechterunterschiede sind im Bereich Gamification ebenfalls ein nicht zu unterschätzender Faktor. So zeigt die Studie von Koivisto & Hamari (2014), dass Frauen sich mehr soziale Vorteile, wie zum Beispiel Anerkennung, von der Nutzung von Gamification versprechen. Eine weitere Studie zeigt, dass zum Beispiel Abzeichen von männlichen Probanden bevorzugt werden, weil diese eher als motivierender empfunden werden als von Frauen (Jent & Janneck, 2016). Allen Studien gemein ist allerdings der Nachteil, dass nicht die tatsächliche Nutzung auf Geschlechtsunterschiede untersucht wird, sondern immer nur Umfrageergebnisse. Außerdem werden oftmals nicht Spielelemente isoliert untersucht, sondern in Kombination mit anderen Spielelementen, was zu einer Konfundierung der Ergebnisse führen kann (Mekler et al., 2017). In der vorliegenden Studie wird diese Lücke geschlossen, da mittels eines experimentellen Designs die Nutzung gemessen wird und die Spielelemente isoliert analysiert werden.

2 Rahmenbedingungen des Experiments

Die vorliegende Studie basiert auf den Ergebnissen einer bereits ähnlich durchgeführten Studie (Mazarakis, 2013), allerdings mit gravierenden Unterschieden. Das grundlegende Ziel beider Studien ist es herauszufinden, wie man die Anzahl beantworteter Fragen in einem Quiz erhöhen kann. Während allerdings die Studie von Mazarakis (2013) nur ein allgemeines Setting hatte, wird in der vorliegenden Studie der Schwerpunkt auf spezifische Länderfragen gelegt. Damit werden auch andere Fragen gestellt als es in der Studie von Mazarakis (2013) der Fall war. Außerdem ging es im Experiment von Mazarakis (2013) um die Untersuchung der Spielelemente Feedback und Rangliste. Dies ist in der vorliegenden Studie nicht der Fall, sondern der Schwerpunkt liegt bei den beiden Spielelemente Abzeichen und Story. Eine detaillierte Beschreibung zum Quiz findet sich im folgenden Abschnitt 2.1.

Das vorliegende Experiment besteht aus drei Versuchsbedingungen: eine Kontrollgruppe (KG) ohne weitere Manipulation, eine Experimentalbedingung mit dem Spielelement

Abzeichen und eine Experimentbedingung mit dem Spielelement Story. In allen drei Versuchsbedingungen wurde sofort Feedback über die korrekte oder falsche Beantwortung einer Frage gegeben. Die korrekte Antwortmöglichkeit wurde grün markiert, eine inkorrekt ausgewählte Antwort wurde rot angezeigt. Aus den Rahmenbedingungen des Experiments und den verwandten Arbeiten in Abschnitt 1 können folgende Hypothesen abgeleitet werden:

- Hypothese 1 (H1): Die Gruppe in der Experimentalbedingung Abzeichen (Badges) beantwortet mehr Fragen als die Kontrollgruppe (Abzeichen > KG).
- Hypothese 2 (H2): Die Gruppe in der Experimentalbedingung Story beantwortet mehr Fragen als die Kontrollgruppe (Story > KG).
- Hypothese 3 (H3): Weibliche Probanden beantworten mehr Fragen in der Experimentalbedingung Story als männliche Probanden (Story: $W > M$).
- Hypothese 4 (H4): Weibliche Probanden beantworten weniger Fragen in der Experimentalbedingung Abzeichen als männliche Probanden (Abzeichen: $W < M$).

2.1 Beschreibung des Quiz

Die Grundidee des Quiz ist es, dass man „solange spielen kann, wie man will bzw. solange wie man Spaß daran hat“ – mit diesem Hinweis wurde das Quiz beworben. Dies wird auch z. B. zu Beginn auf der Einstiegsseite erklärt. Die Entscheidung für ein Quiz wurde aufgrund der Annahme getroffen, dass dies gut in eine Lernsituation hineinpasst und z. B. im Kontext von Learning Analytics sinnvoll wäre (Mazarakis, 2013). Das Quiz besteht aus insgesamt 170 Fragen, wobei die Gesamtanzahl der Fragen zu keiner Zeit mitgeteilt wird. Dadurch soll verhindert werden, dass man als Proband versucht, alle Fragen zu beantworten. Die Fragen bestehen aus jeweils zehn Fragen zu einem Kontinent bei insgesamt sieben Kontinenten. Außerdem werden zu jedem Kontinent vier souveräne Länder dargestellt, zu denen nochmals fünf weitere Fragen zu beantworten sind. Für die Kontinente Antarktika und Australien ist dies nicht möglich, weswegen dort nur zehn Kontinentfragen abgefragt werden. Die 170 Fragen setzen sich damit aus $5*((4*5)+10) + (2*10)$ Fragen zusammen. Folgende Länder wurden je Kontinent zusätzlich abgefragt (die Reihenfolge entspricht auch der Reihenfolge im Quiz): Europa: Deutschland, Griechenland, Italien und das Vereinigte Königreich; Asien: China, Indien, Japan und Russland; Australien: Keine Länder abgefragt; Afrika: Ägypten, Nigeria, Südafrika und Madagaskar; Nordamerika: Kanada, USA, Mexiko und Karibik; Südamerika: Brasilien, Ecuador, Peru und Argentinien; Antarktika: Keine Länder abgefragt.

Die Fragen selbst sind keiner Kategorie zugeordnet, sondern sind aus dem Bereich Allgemeinwissen. So stammen die Fragen z. B. aus Politik, Kultur, Film, Mythen, Wirtschaft oder Musik, wobei aber immer der entsprechende Kontext im Quiz (Land oder Kontinent) beachtet wird.

Eine Bedingung beschäftigt sich mit der Wirkung von Abzeichen, welche nach Hamari (2017) motivierend sind. Hierzu wurden 14 unterschiedliche Abzeichen erstellt. Acht der 14 Abzeichen werden nur bei korrekter Beantwortung einer bestimmten Frage vergeben. Die restlichen Abzeichen werden für eine hohe Antwortgeschwindigkeit von richtigen Antworten

vergeben oder für mehrere korrekte Antworten in Folge. Für beide Kategorien werden jeweils drei Abstufungen verwendet, und zwar von leicht zu schwer (zum Beispiel drei, fünf oder zehn richtige Fragen hintereinander). In Abbildung 1 sind alle erspielbaren Abzeichen dargestellt. Diese werden zuerst nur grau hinterlegt. Es ist nicht ersichtlich, welche Bedingungen man erfüllen muss, um ein Abzeichen zu erhalten. Dadurch soll nur das Verhalten belohnt werden, das Quiz zu spielen und nicht, um auf ein bestimmtes Ziel hinzuarbeiten. Nachdem ein Abzeichen erspielt wurde, konnte man durch die Ansteuerung mit dem Mauszeiger erfahren, wofür das Abzeichen vergeben wird. Zusätzlich wurde man durch eine Bildschirmblendung informiert, falls man gerade ein Abzeichen erhalten hat.



Abbildung 1: Darstellung aller erspielbaren Abzeichen

Die zweite Experimentalbedingung verwendet das Spielelement Story, welches nach Kucwaj (2016) ebenfalls sehr motivationssteigernd sein soll. Als Hintergrundgeschichte wurde eine Weltreise gewählt. In dieser Experimentalbedingung wurde für jedes Land über der jeweiligen Frage ein Bild der Landesflagge angezeigt, wie in Abbildung 2 links dargestellt. Für Kontinentfragen wurde zum Beispiel ein Dschungel für Südamerika gezeigt. Außerdem gab es im Storymodus Bilder, welche den Fortschritt der Weltreise anzeigen. Abbildung 2 rechts zeigt eine solche Zwischengrafik.



Abbildung 2: Links eine Länderfrage, hier konkret Deutschland. Rechts eine Zwischengrafik im Storymodus

3 Die Feldstudie

Die Laufzeit der Feldstudie betrug zwei Wochen im März 2016. Die Probanden wurden online über soziale Netzwerke wie Facebook und Twitter akquiriert. Es wurde keine Entlohnung gezahlt und auch nicht damit geworben. Insgesamt haben 364 Probanden einmalig teilgenommen, was durch den Vergleich der IP-Adresse und der Session-ID überprüft wurde. 148 Probanden haben als Geschlecht männlich angegeben. Diese sind im Schnitt 27.15 Jahre alt (Standardabweichung: 8.32 Jahre), der Median liegt bei 25 Jahren mit 17 Jahren als niedrigstes und 62 Jahren als höchstes angegebenes Alter. Außerdem haben 175 Frauen teilgenommen, welche im Schnitt 27.51 Jahre alt sind (Standardabweichung: 9.14 Jahre), der Median liegt bei 24 Jahren mit 17 Jahren als niedrigstes und 58 Jahren als höchstes angegebenes Alter. 41 Probanden haben kein Alter und kein Geschlecht angegeben. Die Unterschiede im Alter zwischen den Frauen und Männern sind nicht statistisch signifikant. Umgesetzt wurde das Experiment mit PHP und einem MySQL-Server.

Beim Klick auf den Link zur Webseite wurde man im Hintergrund randomisiert einer der drei Versuchsbedingungen zugewiesen. Ein Wechsel zwischen den Versuchsbedingungen war nicht möglich. Gleichzeitig wurde eine Informationsseite angezeigt, wo unter anderem mitgeteilt wurde, dass das Quiz je Frage vier Antwortmöglichkeiten besitzt und so lange gespielt werden kann, wie man will. Außerdem wurden die Probanden darüber aufgeklärt, dass alle Daten anonym sind und dass man sich so viel Zeit bei der Beantwortung der Fragen lassen kann, wie man möchte. Schließlich wurden die Probanden darüber informiert, dass man bei der Beendigung des Spiels unbedingt den „Quiz beenden“-Button drücken muss, da ansonsten die Daten nicht gespeichert werden. Ein Überspringen einer Frage war nicht möglich.

Falls der Proband mit diesen Informationen einverstanden war, dann wurde er auf der nächsten Seite nach persönlichen Daten wie Geschlecht und Alter befragt, welche freiwillig zu beantworten waren. Danach konnte man sofort versuchen, die erste Frage zu beantworten. Sobald alle 170 Fragen beantwortet waren oder aber man vorher auf „Quiz beenden“ geklickt hatte, wurde im Anschluss eine Statistik mit der Anzahl der richtigen und falschen Antworten angezeigt. Außerdem hat man sich beim Probanden für die Teilnahme bedankt. Schließlich gab es noch ein Freitextfeld zur freiwilligen Abgabe von Kommentaren.

4 Ergebnisse und statistische Auswertungen

Es wird zuerst auf systematische Unterschiede in der Stichprobe getestet. Hierbei stehen zunächst Geschlechtsunterschiede im Vordergrund. Tabelle 1 zeigt, dass es keine Geschlechtsunterschiede in der Anzahl der beantworteten Fragen gibt. Allerdings unterscheiden sich diejenigen, die kein Geschlecht angegeben haben, von Personen, welche ihr Geschlecht angegeben haben. Dieser Unterschied ist mit einer Varianzanalyse statistisch signifikant mit $F(2, 361) = 6.45, p = .002$. Der Post-Hoc-Test nach Tukey wird hierbei nur statistisch signifikant für die Unterschiede zwischen keiner Angabe und Männern ($p = .019$) bzw. Frauen (p

=.001) und nicht zwischen Männern und Frauen ($p = .402$). Damit existieren keine allgemeinen Geschlechtsunterschiede bei der Anzahl der beantworteten Fragen. Weitere Analysen zeigen, dass alle inferenzstatistischen Ergebnisse ebenfalls signifikant bzw. nicht signifikant ausfallen, wenn man nur die Teilstichprobe aus Männern und Frauen verwendet. Aus Platzgründen wird auf eine gesonderte Darstellung verzichtet.

Außerdem werden in Tabelle 1 die Geschlechterunterschiede in Bezug auf das Vorwissen deskriptiv dargestellt (Mittelwert und Standardabweichung richtiger Antworten). Diese sehen zuerst unauffällig aus, sodass hier auch von keinem Unterschied zwischen Männern und Frauen ausgegangen werden kann. Allerdings müssen diese Ergebnisse standardisiert werden, da Frauen mehr Fragen als Männer beantwortet haben. Die deskriptiven Ergebnisse aus Tabelle 1 werden daher normiert, und zwar wird ein Quotient aus der Anzahl der korrekt beantworteten Fragen und der gespielten Fragen gebildet.

Geschlecht	Anzahl	Mittelwert beantwortete Fragen	Mittelwert richtiger Antworten	Standardabweichung beantwortete Fragen	Standardabweichung richtiger Antworten
Männer	148	79.18	50.52	60.09	39.93
Frauen	175	88.19	49.98	63.72	37.79
k. Angabe	41	49.29	32.41	66.45	43.75

Tabelle 1: Geschlechterunterschiede in Bezug auf die Anzahl der gespielten Fragen

Der Post-Hoc-Test nach Tukey wird hierbei statistisch signifikant für den Quotient-Unterschied zwischen Männern und Frauen ($p = .001$), allerdings nicht bei Männern ($p = .376$) bzw. Frauen ($p = .724$) und bei keiner Angabe des Geschlechts. Damit existieren statistisch signifikante Unterschiede im Vorwissen zwischen Männern und Frauen. Darauf wird in Abschnitt 5 in der Diskussion nochmals Bezug genommen. Zwischen den Versuchsbedingungen mit dem normierten Ergebnis aus korrekten und gespielten Fragen finden sich hingegen keine statistisch signifikanten Unterschiede mit $F(2, 343) = 0.16$, $p = .854$. Dies bedeutet also, dass es in der vorhandenen Stichprobe nur systematische Unterschiede zwischen dem Vorwissen der Geschlechter gibt, aber dass diese sich nicht auf die randomisierte Verteilung zu den Versuchsbedingungen ausgewirkt haben. Dies zeigt sich auch bei der Anzahl der erspielten Abzeichen, da hier ebenfalls zwischen Männern und Frauen keine Geschlechtsunterschiede mit $t(305.63) = 0.43$, $p = .669$ existieren.

Die Verteilung zwischen den drei Versuchsbedingungen wird in Abbildung 3 dargestellt. Wie leicht ersichtlich ist, handelt es sich hierbei in jeder Bedingung um eine bimodale Verteilung. 81 Probanden (22.2 %) beantworteten alle Fragen unabhängig von der Versuchsbedingung. Zusätzlich kann festgehalten werden, dass 19 Probanden in der Kontrollgruppe 170 Fragen beantwortet haben, dies entspricht 14 %. Bei den Abzeichen waren es hingegen 32 Probanden (27 %) und im Storymodus 30 Probanden (28 %). Interessant ist hierbei, dass mit 47 weiblichen Probanden fast doppelt so viele Frauen die Maximalanzahl der Fragen beantwortet haben, als die männlichen Probanden mit 26. Auch sonst haben die weiblichen Probanden in jeder Versuchsbedingung ca. acht bis neun Fragen länger gespielt als die männli-

chen Probanden. Dieser Unterschied ist allerdings nicht statistisch signifikant und wird genauer analysiert, insbesondere für die Beantwortung der H3 und H4.

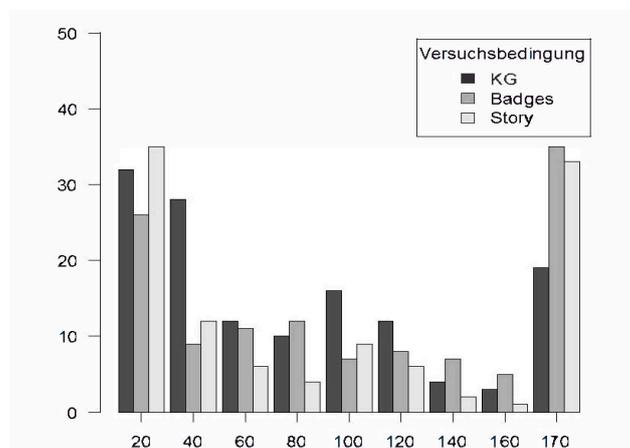


Abbildung 3: Verteilung der Häufigkeit der gespielten Fragen in Abhängigkeit von der Versuchsbedingung

In Tabelle 2 werden Mittelwert und Standardabweichung je Versuchsbedingung der beantworteten Fragen angegeben. Zusätzlich wird die Anzahl der Probanden angegeben, welche keine einzige Frage beantwortet haben. Diese können nicht von der Auswertung ausgeschlossen werden, da in den Versuchsbedingungen mit den Abzeichen und der Story schon sofort zu Beginn Hinweise auf die Versuchsbedingung gegeben werden. Diese Hinweise können eventuell zu einem vorzeitigen Abbruch führen. Außerdem ist die Anzahl der Probanden angegeben, welche alle Fragen beantwortet haben.

Versuchsbedingung	Anzahl	Mittelwert	Standardabweichung	Anzahl mit keinen gespielten Fragen	Anzahl mit allen gespielten Fragen
Kontrollgruppe	136	69.60	57.07	9	19
Abzeichen	120	91.30	64.54	5	32
Story	108	81.03	68.27	4	30
Gesamt	364	80.15	63.51	18	81

Tabelle 2: Anzahl, Mittelwert, Standardabweichung und die Anzahl ohne/mit allen gespielte Fragen je Bedingung

Aus den deskriptiven Ergebnissen wird bereits ersichtlich, dass in der Kontrollgruppe im Schnitt die wenigsten Fragen gespielt bzw. beantwortet wurden. Die Experimentalgruppe mit den Abzeichen hat mehr Fragen absolviert als die Gruppe im Storymodus. Um nun diese Ergebnisse auch belastbar interpretieren zu können, werden inferenzstatistische Analysen durchgeführt. Im vorliegenden Fall ist eine Varianzanalyse indiziert, um die Mittelwerte der unterschiedlichen Gruppen miteinander vergleichen zu können. Dies soll der robusten Inter-

pretation der Ergebnisse dienen. Außerdem werden Kontrastanalysen durchgeführt, welche die aufgestellten Hypothesen unterstützen sollen (Field, 2013).

Der Test zur Homogenität der Varianzen (Levene-Test) für die Anzahl der beantworteten Fragen wird statistisch signifikant mit $p = .001$, wobei die Levene-Statistik 6.78 beträgt. Bei allen folgenden Ergebnissen wird daher von nicht gleichen Varianzen ausgegangen und entsprechend konservativ korrigiert. Dadurch erklären sich auch krumme Werte bei den Freiheitsgraden.

Die Varianzanalyse zeigt einen statistisch signifikanten Unterschied zwischen den einzelnen Versuchsbedingungen an, $F(2, 361) = 3.79$, $p = .023$. Da die Homogenität der Varianzen nicht gegeben ist, muss durch den Welch-Test entsprechend korrigiert werden (Field, 2013). Das korrigierte Ergebnis der Varianzanalyse lautet nun 4.06 mit $p = .019$. Es liegt damit ein statistisch signifikanter Unterschied in der Anzahl der beantworteten Fragen zwischen den einzelnen Versuchsbedingungen vor. Daher können alle aufgestellten Hypothesen nun statistisch untersucht werden.

Die Probanden in der Experimentalbedingung mit den Abzeichen haben statistisch signifikant mehr Fragen beantwortet als die Kontrollbedingung $t(239.40) = 2.83$, $p = .003$ (einseitig getestet). Damit kann die H1 erfolgreich unterstützt werden. Dieses Ergebnis trifft allerdings nicht zu beim Vergleich zwischen der Kontrollgruppe und den Probanden im Storymodus mit $t(207.95) = 1.40$, $p = .083$ (einseitig getestet), da die statistische Signifikanz knapp verfehlt wird. Dadurch kann keine Unterstützung für die H2 gefunden werden.

Für die H3 und H4 müssen die Geschlechterunterschiede genauer betrachtet werden. Frauen haben zwar mit 88.19 (Standardabweichung = 63.72) beantworteten Fragen im Schnitt mehr gespielt als Männer mit 79.18 (Standardabweichung = 60.09), dieses Ergebnis wird allerdings nicht statistisch signifikant mit $t(317.18) = 1.31$, $p = .193$.

Frauen beantworten nicht mehr Fragen als Männer im Storymodus mit $t(75.65) = 0.83$, $p = .206$ (einseitig getestet), H3 kann nicht unterstützt werden. Der Mittelwert der Männer im Storymodus für die beantworteten Fragen beträgt 74.95 mit einer Standardabweichung von 68.15. Bei den Frauen ist der Mittelwert 86.70 und die Standardabweichung 66.47.

Ebenfalls kann die H4 nicht unterstützt werden, da bereits in der deskriptiven Auswertung klar wird, dass Frauen in der Bedingung mit den Abzeichen mehr Fragen beantwortet haben (Mittelwert: 100.75, Standardabweichung: 63.82 vs. Männer mit einem Mittelwert von 92.12 und einer Standardabweichung von 61.17) als Männer. Eine inferenzstatistische Testung der gerichteten Hypothese verbietet sich laut Field (2013).

Interessant ist schließlich noch die Überprüfung, ob die Information, ob man eine Frage richtig oder falsch beantwortet hat, einen Einfluss auf die Anzahl der beantworteten Fragen hat. Hierzu wird auf individueller Ebene zuerst das Verhältnis zwischen richtigen und falschen Antworten bestimmt. Im Schnitt liegt dies in der vorliegenden Stichprobe bei 57 %. Danach werden alle Probanden mit 0 % korrekten Antworten aus der Stichprobe entfernt. Dies ist notwendig, da diese im Schnitt nur 1.31 Fragen beantwortet haben, wobei die maximale Anzahl bei drei Fragen liegt. Es wird angenommen, dass eine demotivierende Wirkung in so kurzer Zeit nicht stattfinden kann. Mit den verbliebenen Probanden führt eine Korrela-

tion zwischen der Anzahl der beantworteten Fragen und dem Verhältnis von korrekt beantworteten und insgesamt beantworteten Fragen zu $r_s = .06$, $p = .321$. Falsche Antworten führten damit nicht zu statistisch signifikant weniger beantworteten Fragen bzw. richtige Antworten führten nicht zu statistisch signifikant mehr beantworteten Fragen.

5 Fazit, Diskussion und Ausblick

Die vorliegende Studie mit 364 Probanden hat zu neuen Erkenntnissen im Bereich der Nutzer motivation mit Gamification geführt. Hierbei konnte gezeigt werden, dass Probanden in der Experimentalbedingung mit den Abzeichen statistisch signifikant mehr Fragen beantwortet haben als die Kontrollgruppe. Dieses Ergebnis konnte nicht für den Storymodus wiederholt werden. Auch konnte die angenommene Dominanz des Spielelements Story, wie es teilweise wegen seiner Bedeutsamkeit erwartet wird (Kucwaj, 2016), nicht bestätigt werden. Eine Interpretation wäre, dass die konkrete Wahl der Story nicht fesselnd genug war und deswegen nicht überzufällig die Motivation steigern konnte. Zusätzlich konnten Geschlechterunterschiede nicht statistisch abgesichert werden, obwohl diese in anderen Studien berichtet werden (Koivisto & Hamari, 2014; Jent & Janneck, 2016).

Die Verwendung von Abzeichen zur Steigerung der Motivation kann aufgrund dieser Studie empfohlen werden. Auch wenn es nicht das Ziel war, Abzeichen im Detail zu untersuchen, so scheint die Kombination aus intransparenter Realisierung des Erlangens eines Abzeichens sowie die Vergabe für konkrete Aufgaben als auch für regelmäßige Schnelligkeit und Korrektheit sehr motivierend zu wirken. Bei der Story war dies hingegen nicht der Fall. Eine inhaltliche Adaption der Story mit einer erneuten Experimentdurchführung könnte hier die notwendige Datengrundlage für eine Neuinterpretation liefern.

Interessant ist es auch die Ergebnisse vorheriger Studien mit dieser zu vergleichen. So konnte Mazarakis (2013) zeigen, dass schon Feedback alleine als Spielelement ausreicht, um die Motivation zu steigern. Da in der durchgeführten Studie jede Versuchsbedingung dieses Spielelement besaß, könnte es durchaus zu einer Konfundierung der Spielelemente gekommen sein. Allerdings ist durch die konsequente Nutzung des Feedbacks, die *ceteris paribus*-Regel nicht verletzt. Dies bedeutet, dass die hier dargestellten Ergebnisse sicher interpretiert werden können, da die Effekte zwischen den einzelnen Versuchsbedingungen unabhängig voneinander berechnet werden können.

Die Unterschiede zwischen den Geschlechtern in Bezug auf den Anteil der korrekt beantworteten Fragen lässt die Interpretation zu, dass die gestellten Fragen leichter für Männer zu beantworten sind. Dies lässt sich allerdings nur in weiteren und größeren Studien gesichert bestätigen. Eine solche Studie ist bereits in Planung und wird den Forschungsstand im Bereich Gamification und Geschlechterunterschiede weiter voranbringen.

Mit dieser Studie kann die Lücke geschlossen werden, dass Erkenntnisse zu Anreizen und Gamification oftmals nur durch Umfragen erlangt werden. Durch eine experimentelle Feldstudie ist es gelungen, dem Element Abzeichen eine besondere Bedeutung im Bereich der Nutzer motivation attestieren zu können. Auch, dass Frauen sich eher von Gamification moti-

vieren lassen, konnte gezeigt werden, wenn auch knapp nicht die notwendige statistische Signifikanz erreicht wurde. Durch die isolierte Betrachtung der Spielelemente ist keine Konfundierung der Ergebnisse eingetreten, was sonst üblich ist im Bereich der Gamificationforschung. Eine Interpretation der einzelnen Spielelemente ist dadurch möglich.

Natürlich kann Gamification in unterschiedlichen Kontexten bzw. Systemen andere Effekte aufweisen und auch die konkrete Umsetzung der Spielelemente und ihre Eignung für den Kontext bzw. das System sind entscheidende Faktoren. So hätte man zum Beispiel die beiden Spielelemente auch anders gestalten können. Diese Einschränkungen existieren aber in allen wissenschaftlichen Studien. Durch die Anwendung der *ceteris paribus*-Klausel wurde versucht, diesen Einschränkungen entgegenzuwirken. Trotzdem wird natürlich keine Allgemeingültigkeit der Ergebnisse der online durchgeführten Feldstudie angenommen.

Danksagung: Birger Hein, Jörn Hauke und Anna-Lena Janßen wird für die Unterstützung der durchgeführten Studie im Rahmen des Seminars Incentives für Collective Intelligence und User Generated Content gedankt.

Literaturverzeichnis

- Denny, J. (2014). Gamification: Intrinsic Motivation for Lasting Engagement. Abgerufen 5. April 2017, von <https://elearningindustry.com/gamification-intrinsic-motivation-lasting-engagement>
- Deterding, S., Dixon, D., Khaled, R. & Nacke, L. (2011). From Game Design Elements to Gamefulness: Defining Gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments - MindTrek '11*. New York, New York, USA: ACM, S. 9–15.
- Field, A. (2013). *Discovering Statistics Using SPSS*. London: SAGE.
- Hamari, J. (2017). Do Badges Increase User Activity? A Field Experiment on the Effects of Gamification. *Computers in Human Behavior*, 71, 469–478.
- Hamari, J., Koivisto, J. & Sarsa, H. (2014). Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *Proceedings of the 49th Annual Hawaii International Conference on System Sciences (HICSS)*. Washington, DC, USA: IEEE, S. 3025–3034.
- Huotari, K. & Hamari, J. (2016). A Definition for Gamification: Anchoring Gamification in the Service Marketing Literature. *Electronic Markets*, 27(1), 21–31.
- Jent, S. & Janneck, M. (2016). Gender- und Altersaspekte beim Einsatz von Gamification. In W. Prinz, J. Borchers & M. Jarke (Hrsg.), *Mensch und Computer 2016 - Tagungsband*. Aachen: Gesellschaft für Informatik e.V., S. 1–4.
- Koivisto, J. & Hamari, J. (2014). Demographic Differences in Perceived Benefits from Gamification. *Computers in Human Behavior*, 35, 179–188.
- Kucwaj, H. (2016). Why Do We Like Stories? Storytelling in Gamification. Abgerufen von <http://www.sointeractive.co/blog/why-do-we-like-stories-storytelling-in-gamification/>

- Mazarakis, A. (2013). Like Diamonds in the Sky – How Feedback Can Boost the Amount of Available Data for Learning Analytics. *International Journal of Technology Enhanced Learning*, 5(2), 107–116.
- Mekler, E. D., Bruehlmann, F., Tuch, A. N. & Opwis, K. (2017). Towards Understanding the Effects of Individual Gamification Elements on Intrinsic Motivation and Performance. *Computers in Human Behavior*, 71, 525–534.
- Ryan, R. M. & Deci, E. L. (2000a). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1), 54–67.
- Ryan, R. M. & Deci, E. L. (2000b). Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being. *American Psychologist*, 55(1), 68–78.
- Smadja, F. (2009). *Mixing Financial, Social and Fun Incentives for Social Voting*. 1. International Workshop on Motivation and Incentives on the Web at the 18th International World Wide Web Conference, WWW2009, Madrid.

Autor



Mazarakis, Athanasios

Athanasios Mazarakis studierte Psychologie an der Universität Mannheim. Im Anschluss forschte er am FZI Forschungszentrum Informatik in mehreren nationalen und internationalen Projekten in den Bereichen Wissensmanagement und Anreize. 2013 promovierte er im Bereich Wirtschaftswissenschaften am Karlsruher Institut für Technologie (KIT). Aktuell ist er als Postdoc für Web Science an der Christian-Albrechts-Universität zu Kiel tätig. Zu seinen Hauptaufgaben zählen die Forschung und Mitarbeit in Projekten zu Gamification, Anreizgestaltung, Altmetrics und Open Science.