

Gamification und die Wahrnehmung von Punkten – Eine Umfragestudie

Athanasios Mazarakis

Web Science

ZBW – Leibniz Information Centre for Economics

Kiel, Germany

a.mazarakis@zbw.eu

Paula Bräuer

Web Science

Kiel University

Kiel, Germany

p.braeuer@zbw.eu

ZUSAMMENFASSUNG

Punkte zählen zu den am meisten verwendeten Spieldesignelementen im Bereich von Gamification. Allerdings existieren keine Studien dazu, welche konkreten Punktwerte sinnvoll sind und wann zu wenig oder zu viele Punkte vergeben werden könnten. In der vorliegenden Umfragestudie wurden die Daten von 160 Teilnehmenden analysiert, wobei 154 Teilnehmende eingehend untersucht wurden, um die konkrete Fragestellung zu beantworten, wie Punkte wahrgenommen werden. Die Studie führte hierbei zu Erkenntnissen, welche Punktezahlen als angebracht angesehen werden und welche nicht. Insbesondere die Erkenntnis, dass durch einen gegebenen Kontext sich die Wahrnehmung von Punkten stark verändern kann, bildet eine Grundlage für weitere Forschungsmöglichkeiten.

SCHLÜSSELWÖRTER

Gamification, Punkte, Umfrage, Motivation, Wahrnehmung

1 EINLEITUNG UND VERWANDTE ARBEITEN

Die Umsetzung von Gamification findet noch immer sehr stark fokussiert durch die sogenannte PBL-Triade statt, also der Verwendung von Punkten (points), Abzeichen (badges) und Ranglisten (leaderboards), um die Motivation für bestimmte Tätigkeiten zu erhöhen [7, 9, 20]. Dies ist auch nicht weiter verwunderlich, da zum Beispiel Punkte mehrfach als am meisten verwendetes Spieldesignelement identifiziert wurden [2, 5, 14].

Das Spieldesignelement Punkte kann allgemein definiert werden als eine numerische Einheit zum Anzeigen des Fortschritts [15]. Punkte können aber auch als eine numerische Darstellung und einhergehende Belohnung von Spieler:innen für in einem Spiel durchgeführte Aktivitäten charakterisiert werden [18]. Genauso wie Ranglisten, können Punkte als extrinsische Anreize angesehen werden [14], wobei man Gamification und die Verwendung der entsprechenden Spieldesignelemente, auch allgemein als von außen kommend und daher als extrinsisch definieren kann [10].

Punkte unterscheiden sich stark in ihrer Ausgestaltung und Wahrnehmung. So können Punkte als Erfahrungspunkte, Fähigkeitspunkte, einlösbare Punkte, Karmapunkte oder auch Rufpunkte designt werden [13, 21]. Punkte können aber als externer Anreiz auch einen Spielstand festhalten, einen möglichen Sieg oder Fortschritt anzeigen oder allgemein Feedback geben [20]. Meistens werden Punkte einzeln für erfüllte Aufgaben vergeben und dabei aufsummiert (manchmal auch von einem Punktwert abgezogen), um so eine Gesamtpunktzahl zu berechnen [1]. Auch Bonuspunkte werden häufig vergeben, um eine (besondere) Leistung zu belohnen [8]. Schließlich können Punkte auch als Scoring-System verstanden werden, um eine Rangordnung oder Bestenliste zu strukturieren, als ein Instrument zur Zielsetzung angewendet zu werden oder als (virtuelle) Währung auftreten [11].

Allerdings ist die Punktsetzung im konkreten Fall, also der Gamifizierung einer Tätigkeit oder Anwendung, nicht trivial. In der Tat ist es häufig eher so, dass man sich zum entsprechenden Design keine Gedanken macht, beziehungsweise dies mindestens nicht in den Forschungsarbeiten kommuniziert wird. So haben Kolpondinos und Glinz [3] zwei unterschiedliche Punktesysteme für eine gemeinsame Aufgabe untersucht. Hierbei ging es den Autor:innen darum, eine Optimierung von Prioritäten von Aufgaben und Bedürfnissen zu gewährleisten. In einem Punktesystem, welches als Kontrollgruppe fungierte, wurden linear Punkte vergeben im Verhältnis zu den erfüllten Aufgaben. In der Experimentalgruppe hingegen wurde die volle Punktzahl vergeben, wenn alle Aufgaben erledigt waren. War dies nicht der Fall, wurden in einem binären Vorgehen keine Punkte vergeben. Als Maximalpunktzahl wurden 100 Punkte vergeben. Die Experimentalgruppe führt zu deutlich besseren Ergebnissen [3]. Allerdings wurde nicht geklärt, wieso ausgerechnet die Zahl 100 gewählt wurde und wieso nicht 10, 913 oder 9999999 Punkte.

Aus dieser Herleitung ergibt sich die praktische Fragestellung: *Wie werden Punkte wahrgenommen?* Das Ziel ist es, anhand einer Umfragestudie herauszufinden, wie Punkte primär bei Spieler:innen wirken und daraus Schlüsse zu ziehen, ob und wie eine Unterscheidung in der Höhe von Punktezahlen stattfinden kann, zum Beispiel in *angebrachte*, *zu viele* und *zu wenige* Punkte. Gleichzeitig soll herausgefunden werden, ob es eventuell Unterschiede zwischen geraden oder ungeraden Zahlen geben könnte und ob einfache (zum Beispiel Zahlen wie 1, 2 oder 12) oder eher komplexere Zahlen, wie zum Beispiel 3,31 oder 112,89, bevorzugt wurden.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Veröffentlicht durch die Gesellschaft für Informatik e.V.

in K. Marky, U. Grünefeld & T. Kosch (Hrsg.):

Mensch und Computer 2022 – Workshopband, 04.-07. September 2022, Darmstadt

© 2022 Copyright held by the owner/author(s).

<https://doi.org/10.18420/muc2022-mci-ws08-417>

2 METHODE

Zur Beantwortung der Fragestellung wird eine explorative Umfragestudie online durchgeführt. Neben demografischen Angaben wurde primär die Wahrnehmung von Punkten erfragt. Der Fragebogen ist in drei Abschnitte unterteilt.



Abbildung 1: Beispiel für ein Output-Agreement Game, im vorliegenden Fall aus der Originalstudie von von Ahn und Dabbish [19].

Zu Beginn wurden die Teilnehmenden begrüßt und demografische Daten abgefragt, wie zum Beispiel das Geschlecht oder die ausgeübte Tätigkeit. Es wurden nur Aspekte abgefragt, welche einen Einfluss auf die Wahrnehmung von Punkten nehmen könnten, zum Beispiel ob häufig Videospiele gespielt werden. Das Alter wurde explizit aus Gründen der Datensparsamkeit nicht erfasst. Alle Fragen waren optional und die Fragen waren hauptsächlich angelehnt an eine Studie, mit der das Spielverhalten von Kindern und Jugendlichen untersucht wurde [16].

Im zweiten Abschnitt wurde abgefragt, wie viele Stunden pro Tag Videospiele gespielt werden. Analog zu Simons et al. [16], allerdings feingranularer mit einer vierstufigen Kategorisierung, wurden Teilnehmende mit weniger als eine Stunde pro Tag an Spielzeit als Wenig-Spieler:innen kategorisiert. Falls man eine bis zwei Stunden Videospieldauer pro Tag angegeben hatte, so wurde man „nur“ als Spieler:innen charakterisiert. Spielzeiten über zwei Stunden pro Tag führten zu einer Kategorisierung als Viel-Spieler:innen. Zeiten mit weniger als 30 Minuten mündeten schließlich zu einer Kategorisierung als Nicht-Spieler:innen. Diese Einteilung deckt sich auch mit den Ergebnissen anderer Studien [6, 16] und wurde nur für die

Auswertung der Daten vorgenommen, den Teilnehmenden aber nicht kommuniziert.

Im letzten Abschnitt wurden die Teilnehmenden explizit zu Ihrer Wahrnehmung von Punkten befragt. Da der Kontext bei der Entwicklung von gamifizierten Anwendungen und Experimenten immer mitbedacht werden muss [9], wurden die Teilnehmenden für die Umfrage randomisiert einem von drei Kontexten zugewiesen. Anschließend bekamen sie eine beispielhafte Aufgabe gestellt. Die Teilnehmenden wurden dann in jedem ihnen zugewiesenen Kontext befragt, wie viele Punkte sie für die erfolgreiche Bearbeitung einer Aufgabe als *angebracht* ansehen, wie viele Punkte *zu wenig* und wie viele Punkte *zu viel* sind. Diese Wahrnehmung hatte jeweils als Eingabe einer Zahl zu erfolgen.

Die erste Gruppe hatte als Kontext kein Szenario und nur die Aufgabe in Textform. Diese Gruppe diente als Kontrollgruppe. Die zweite Gruppe hatte als Kontext ein Szenario mit einem Bild eines Output-Agreement Game. Ein Output-Agreement Game ist eine Spezialform eines Game with a purpose und kann als Miniaufgabe oder Minispiel betrachtet werden, bei denen die Spieler:innen in Paare eingeteilt werden und versuchen zu einem Stimulus (zum Beispiel eine Abbildung) einen ähnlichen Begriff ohne gemeinsame Absprache zu erzeugen [19]. Abbildung 1 zeigt ein Beispiel für ein solches Szenario. Die letzte Gruppe hatte als Szenario den Kontext mit einer Abbildung eines Online-Quiz, ähnlich wie in der Studie von Mazarakis und Bräuer [10]. In der folgenden Auflistung werden zu jedem Szenario beziehungsweise Kontext die konkreten Fragen angegeben:

- Ohne Szenario (KG): „Stellen Sie sich vor, Sie erhalten eine Aufgabe. Als Belohnung für die erfolgreiche Erledigung der Aufgabe sollen Sie Punkte erhalten.“
- Szenario Output-Agreement-Game (OG) mit bildhafter Darstellung (Abbildung 2): „Stellen Sie sich vor, Sie erhalten eine Aufgabe. Die Aufgabe ist ein Spiel, in dem Ihnen zunächst ein:e anonym:e:r Mitspieler:in zugewiesen wird. Innerhalb dieses Spiels wird Ihnen ein Bild präsentiert und Begriffe, die das Bild beschreiben könnten. Sie sollen dann die Begriffe als passend auswählen, die Ihr Mitspieler wahrscheinlich auch ausgewählt hat. Als Belohnung für übereinstimmende Begriffe (Begriffe, bei denen Sie und Ihr:e Partner:in der gleichen Meinung waren) sollen Sie Punkte erhalten.“
- Szenario Online-Quiz (QG) mit bildlicher Darstellung (Abbildung 3): „Stellen Sie sich vor, Sie erhalten eine Aufgabe, in der Sie in einem Choice-Quiz eine zufällige Frage zu allgemeinen Themen beantworten müssen. Als Belohnung für die korrekte Antwort sollen Sie Punkte erhalten.“

Die Umfrage selbst wurde weder mit einer Kompensation beworben, noch wurde irgendeine Belohnung gewährt. Teilnehmende sind in ihrer Gruppe verblieben, welcher Sie randomisiert zu Beginn zugewiesen wurden und konnten somit nur jeweils eines der drei Szenarios bearbeiten.

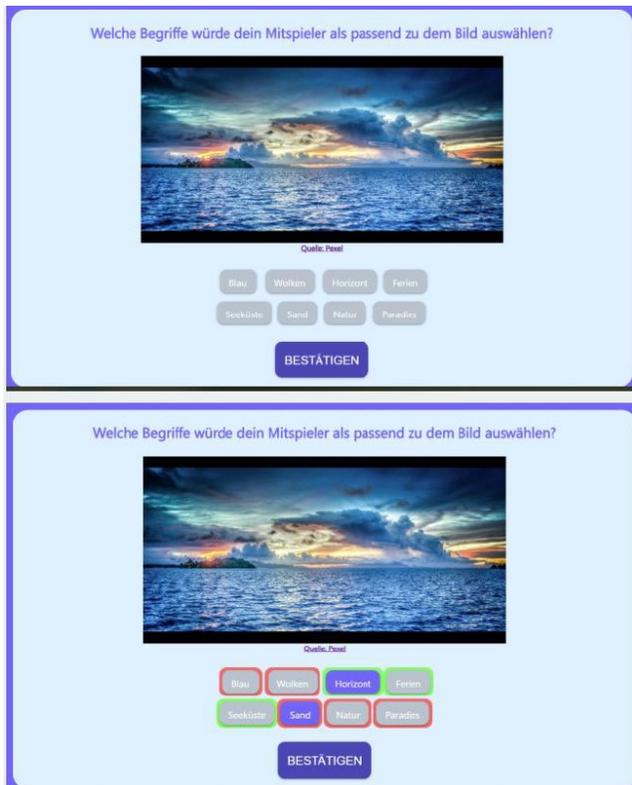


Abbildung 2: Verwendetes Bild in der Umfrage zum Szenario OG.

Seit wie vielen Jahren gibt es Menschen auf der Erde?

1,4 Millionen Jahre
 2,8 Millionen Jahre
 7,6 Millionen Jahre
 3 Milliarden Jahre

Weiter

Quiz beenden

Abbildung 3: Verwendetes Bild in der Umfrage zum Szenario QG.

3 ERGEBNISSE

In diesem Abschnitt folgt zuerst die Beschreibung der Stichprobe. Im Anschluss werden die Ergebnisse deskriptiv und inferenzstatistisch ausgewertet.

3.1 Beschreibung der Stichprobe

Insgesamt haben 264 Teilnehmende die Studie gestartet. Mehrfachteilnahmen sollten anhand von Cookies und IP-Adressen identifiziert werden, wobei keine Mehrfachteilnahme identifiziert wurden. Allerdings wurden 104 Teilnehmende aus der weiteren Analyse entfernt, falls eine der folgenden Bedingungen erfüllt war:

- Befragung abgebrochen, ohne die Wahrnehmung von Punkten anzugeben (96 Teilnehmende).
- Es wurden mehr Punkte vergeben bei der Frage nach zu *wenigen* Punkten im Vergleich zu *angebrachten* Punktezahlen beziehungsweise es wurden mehr Punkte angegeben bei *angebrachten* Punktwerten als bei *zu vielen* Punkten (8 Teilnehmende).
- Aufgrund einer Analyse des Boxplots wurden weitere 6 Teilnehmende von der Stichprobe entfernt. Dabei wurden im Einzelfall Extremwerte zwischen 25000 und 1000000001 als zu viel in der KG angegeben (4 Teilnehmende). Jeweils eine Person gab als zu viele Punkte 10000000000000000 Punkte in der OG und 1000001 in der QG an.

Diese sechs Personen wurden aus Gründen der statistischen Datenanalyse aus der Stichprobe entfernt. Die Angabe solch hoher Punktwerte ist jedoch konform mit der Aufgabenstellung der Studie. Sie werden daher für die Interpretation der Ergebnisse gesondert betrachtet. Somit verbleiben für die weitere Analyse 154 Teilnehmende.

109 der Teilnehmenden sind Studierende unterschiedlichster Studienrichtungen, 43 sind keine Studierenden und 2 Personen haben keine Antwort gegeben. Darüber hinaus waren 63 Teilnehmende erwerbstätig, 83 haben diese Frage verneint. Auch wurde abgefragt, ob die Teilnehmenden regelmäßig Videospiele (auch am Smartphone) spielen (60,40 % Zustimmung), wobei 6,50 % weniger als 30 Minuten pro Tag spielen, 8,45 % zwischen 30 und 60 Minuten, 25,32 % zwischen 1 und 2 Stunden und 20,13 % mehr als 2 Stunden pro Tag spielen.

3.2 Deskriptive und inferenzstatistische Ergebnisse

In Tabelle 1 wird ein erster Überblick über die deskriptiven Ergebnisse zwischen den drei Szenarien und den Punktgruppen gegeben. Im Anschluss werden die Gesamtergebnisse dargestellt.

Insgesamt liegt bei den 154 Teilnehmenden der Mittelwert bei 14,90 Punkten (SD 27,25) in der Bedingung, wo die Punkteanzahl für *angebracht* gehalten wird (Minimum 1 Punkt, Maximum 100 Punkte). Für *zu wenig* Punkte ist der Mittelwert bei 4,08 (SD 11,07), wobei hier als Minimum 0 Punkte und als Maximum 99 Punkte angegeben wurden. Schließlich liegt der Mittelwert in der Bedingung, wo *zu viele* Punkte vergeben wurden bei 231,96 (SD 1157,13), mit einer Minimalpunktzahl von 2 und einer Maximalpunktzahl von 10000.

Tabelle 2 stellt jeweils die fünf am meisten genannten Punktzahlen für *angebrachte*, *zu wenige* oder *zu viele* Punkte allgemein dar. Das Ziel war es hierbei, mindestens 50 % an kumulierten Prozent zu erreichen. Die kumulierten Prozent erreichen nicht 100 %, da auch

Tabelle 1: Deskriptive Ergebnisse.

Punktegruppe	Kontrollgruppe			Output Agreement Game Gruppe			Quizgruppe		
	N	M	SD	N	M	SD	N	M	SD
Angebracht	55	29,75	38,52	56	5,04	4,41	43	8,77	17,74
Zu wenig	55	8,29	17,26	56	1,66	2,10	43	1,83	4,59
Zu viel	55	370,29	1367,57	56	203,68	1333,48	43	91,84	313,45

Tabelle 2: Die jeweils fünf meistgenannten Punktezahlen je Punktegruppe.

Platz	N	Punktegruppen								
		Angebracht			Zu wenig			Zu viel		
		Punkte	Kumulierte Prozent	N	Punkte	Kumulierte Prozent	N	Punkte	Kumulierte Prozent	
Platz 1	44	10,00	28,58	68	1,00	44,16	19	20,00	12,34	
Platz 2	27	1,00	46,11	28	0,00	62,34	18	5,00	24,03	
Platz 3	25	2,00	62,34	15	5,00	72,08	17	100,00	35,07	
Platz 4	17	5,00	73,38	12	2,00	79,88	13	10,00	43,51	
Platz 5	13	100,00	81,82	6	10,00	83,77	12	2,00	51,30	

weitere Angaben gemacht wurden. Für *zu viele* Punkte wurden auch für den Wert 3,00 insgesamt 12 Nennungen getätigt, womit die kumulierte Prozentzahl bei 59,09 % liegen würde.

Bei der Durchführung einer Varianzanalyse werden signifikante statistische Unterschiede zwischen den Szenarien für eine *angebrachte* Anzahl an Punkten berichtet mit $F(2, 151) = 15,35, p < 0.001$ und für *zu wenige* Punkte mit $F(2, 151) = 6,67, p = 0.002$. Für *zu viele* Punkte wird kein statistisch signifikantes Ergebnis mit $F(2, 151) = 0,72, p = 0.487$ berichtet. Die Korrekturen nach Welch beziehungsweise Brown-Forsythe führen zu ähnlichen Ergebnissen und werden aus Redundanzgründen an dieser Stelle nicht wiedergegeben. Eine Post-hoc-Analyse zeigt hierbei statistisch signifikante Ergebnisse zwischen den Szenarien KG und OG und KG und QG für *angebrachte* oder *zu wenige* Punkte.

Für alle drei Bedingungen lassen sich keinerlei statistisch signifikanten Geschlechterunterschiede finden. Auch in Bezug auf die individuelle Häufigkeit, wie viel Teilnehmende pro Tag spielen, konnten keine statistisch signifikanten Unterschiede als Einflussfaktoren gefunden werden.

3.3 Weiterführende statistische Analysen

Da die Unterschiede zwischen der Kontrollgruppe und den beiden Experimentalgruppen so groß sind und in dieser Studie damit der Kontext mindestens einen medierenden Effekt hat, soll noch überprüft werden, ob ohne die Daten der Kontrollgruppe sich die beiden anderen Gruppen unterscheiden. Hierbei zeigen sich bei der Durchführung der Varianzanalysen keine statistisch signifikanten Effekte für eine *angebrachte* Anzahl an Punkten mit $F(1, 97) = 2,30, p = 0.133$, für *zu wenige* Punkte mit $F(1, 97) = 0,06, p = 0.805$ und für *zu viele* Punkte mit $F(1, 97) = 0,29, p = 0.592$.

Die weiterführende statistische Analyse zeigt, dass sich die Unterschiede in allen drei Szenarien zwischen den Experimentalgruppen nicht unterscheiden. Auch hier könnte die Kontextabhängigkeit

eine Rolle spielen. Aber auch andere Erklärungsansätze wie eine zu spezifische Aufgabenstellung oder eine zu kleine Stichprobe sind denkbar.

In Tabelle 3 werden die Ergebnisse der sechs entfernten Personen (aufgrund der Boxplotanalyse) dargestellt. Dabei wird ersichtlich, dass zum Beispiel Person 1 die Punktwerte in Relation gesetzt hat, während es bei den anderen Personen schwierig ist, diese Relation zweifelsfrei für alle drei Punktegruppen festzustellen. Allen Personen gemein ist eine hohe Punktezahl für *zu viele* Punkte.

4 DISKUSSION

Die vorliegende Studie mit 160 Teilnehmenden, wobei die Ergebnisse von 154 Teilnehmenden eingehend bezüglich der Fragestellung, wie Punkte wahrgenommen werden, statistisch untersucht wurden, offenbart einige interessante Erkenntnisse und noch viel mehr Forschungsbedarf. Würde man sich nur die Mittelwerte insgesamt anschauen, könnte man zu dem Schluss kommen, dass ein Punktwert von ca. 15 Punkten wohl ideal wäre, um ihn in einer gamifizierten Anwendung zu verwenden. In der Tat ist es so, dass dieser Wert nur ein einziges Mal als *angebrachter* Punktwert genannt wurde. Unabhängig vom Kontext werden unkomplexe, einfache und häufig gerade Zahlen bevorzugt wie 10 Punkte, 1 Punkt oder 100 Punkte.

Das eindeutigste Ergebnis ist die Abwesenheit von Punkten für eine erfolgreiche Tätigkeit, für die fast die Hälfte der Teilnehmenden 1 Punkt als *zu wenig* Punkte genannt hat, gefolgt von 0 Punkten (siehe Tabelle 2). Dies kann so interpretiert werden, dass die Vergabe von keinen Punkten oder nur sehr wenigen Punkten (1 Punkt) für eine erfolgreiche Bewältigung einer Aufgabe als nicht ausreichend angesehen wird. Die Varianz bei *zu vielen* Punkten ist hingegen sehr hoch (SD 1157,13), was leider keine eindeutigen Schlussfolgerungen zulässt, ab wann genau eine bestimmte Grenze überschritten ist. Wenn man anhand der einzelnen Platzierungen in Tabelle 2 das Verhältnis zwischen *angebrachten* und *zu vielen* Punkten vergleicht, dann scheint der Faktor 2 (zum Beispiel Platz

Tabelle 3: Darstellung der sechs entfernten Personen aufgrund der Boxplotanalyse.

Person	Szenario/Gruppe	Punktegruppen der statistischen Ausreißer		
		Angebracht Punkte	Zu wenig Punkte	Zu viel Punkte
1	KG	12500,00	5000,00	25000,00
2	KG	1000,00	49,00	1000000,00
3	OG	2,00	1,00	10000000000000000,00
4	KG	10,00	1,00	300000,00
5	KG	1000000000,00	1,00	1000000001,00
6	QG	1000000,00	0,00	1000001,00

1 und 4) häufig vorzukommen. In der Tat ist dies bei 24,00 % der Teilnehmenden der Fall, wohingegen der Faktor 10 bei 12,30 % und der Faktor 1,5 in 13,00 % der Fälle vorkommt.

Eine Vermutung, welche sich angesichts der Ergebnisse aufdrängt hat, ist dass der Kontext, welcher in der Gamificationforschung an sich schon relevant ist [9], auch für die Wahrnehmung einer adäquaten Bewertung durch Punkte eine Rolle spielt. Während die Kontrollgruppe ohne Szenario deutlich höhere Punktwerte vergibt, sind die beiden anderen Gruppen als homogen in der Bewertung zu betrachten.

Auch die Anzahl der Teilnehmenden, welche 0 Punkte als *wenig* kategorisiert haben, ist ein Indiz in diese Richtung. Während es in der KG nur 4 (7,27 %) Teilnehmende sind, welche 0 Punkte als *wenig* finden, sind es in der OG 13 (23,21 %) Teilnehmende und in der QG 11 (25,58 %) Teilnehmende.

Es spricht vieles dafür, dass die Vergabe und damit einhergehend auch die Wahrnehmung von Punkten nicht ohne Kontext passieren. Auch wenn diese Erkenntnis auf den ersten Blick vielleicht trivial erscheinen mag, ist in der Realität das sukzessive und wissenschaftliche Testen adäquater Punktzahlen in der Gamificationforschung nicht bekannt. Im Gegenteil, auf einschlägigen Plattformen wird damit argumentiert, dass man Punkte so auswählen sollte, dass diese sich „gut anfühlen und gut aussehen“ sollten [4]. Ratschläge tendieren eher zu kleineren und runden Zahlen, ohne diese Ratschläge genauer zu begründen. Stattdessen wird bei der Präferenz von hohen und krummen Zahlen, welche auch als vorteilhaft beschrieben werden, der Ratschlag gegeben diese mit der Zahl 100 zu multiplizieren [4]. Damit einhergehend bleibt offen, ob überhaupt eine kontextunabhängige Untersuchung von Spieldesignelementen sinnvoll oder möglich ist.

Ein weiterer Faktor der die Ergebnisse beeinflusst haben kann, ist die Wahl des verwendeten Punktsystems. Für eine Erweiterung der Fragestellung könnte eine Unterscheidung zum Beispiel nach Bonuspunkten, Ranglistenpunkten oder Fähigkeitspunkten vorgenommen werden. Insbesondere eine systematische Untersuchung der letzten beiden Punktearten stellte eine weitere Herausforderung dar, denn eine individuelle Betrachtung dieser Punkttypen ohne eine Kombination mit weiteren Spieldesignelementen ist schwierig bis gar nicht umsetzbar.

Eine weitere Erklärungsmöglichkeit könnte der Ankereffekt sein [17], wo eine selbst gewählte Zahl als Anker agieren könnte und die vorhergehende Zahl damit die nachfolgenden Zahlen beeinflusst. Im Kontext dieser Studie könnten also zum Beispiel die Werte, welche als *zu hoch* eingeschätzt werden durch die zuvor abgegebene Einordnung für einen passenden Wert beeinflusst worden sein. Aber auch die Prototypentheorie [12] beziehungsweise die Prototypensemantik könnte eine mögliche Erklärung liefern. Dieser Theorie folgend ist anzunehmen, dass bestimmte Zahlen als prototypischer und damit gewöhnlicher wahrgenommen werden, als andere Zahlen. Insgesamt ist weitere Forschung notwendig, um diese Erklärungsansätze auf ihre Plausibilität zu überprüfen.

Es kann ebenfalls in Betracht gezogen werden, dass das Output-Agreement Game und die Quiz-Bedingung für sich alleine bereits eine (implizite) Referenz vorgeben. So könnte man zum Beispiel bei einem Quiz fragen "Welche Punkte haben Sie typischerweise bei Quizaufgaben erhalten?", um daran anschließend die Frage zu stellen, ob diese Punkte in ihrer Höhe angemessen waren. Diese Umformulierung der Frage verschiebt den Fokus von der Wahrnehmung hin zu einer Wiedererkennung, Rückerinnerung und Interpretation von Punktwerten. Eventuell könnte man so praktikablere Werte erhalten, als zum Beispiel bei den weniger gängigen Output-Agreement Games.

Für die vorliegende Studie sind weitere Analysen ohne Mitbetrachtung der Kontrollgruppe denkbar, insbesondere im Bereich der *angebrachten* Punkte. Auch eine Analyse der individuellen Antwortkombinationen, wie bereits mit Tabelle 3 im kleinen Maßstab durchgeführt, könnte eventuell weiteren Aufschluss zur Beantwortung der Forschungsfrage liefern, zum Beispiel mit der Analyse von Häufungen.

Zukünftige Forschung zu diesem Gebiet sollte auch die Untersuchung von qualitativen Daten in Betracht ziehen. Dies könnte zum Beispiel durch Interviews oder Fokusgruppen in einem Laborexperiment geschehen. Hierbei könnte dann auch die Frage beantwortet werden, welchen Effekt *zu viele* oder *zu wenige* Punkte auf die Motivation und Leistung haben. Dies würde den Fokus weg von der passenden Höhe von *angebrachten* Punkten lenken.

Weitere Untersuchungen könnten aber auch den Effekt der Vergabe von negativen Punkten betrachten, ähnlich dem Effekt von zurückgezogener Gamification [14]. Auch wird die Höhe einer Punktzahl wahrscheinlich anders wahrgenommen, wenn Punkte nicht nur erworben werden, sondern auch wieder verloren gehen können. Diese spannenden Forschungslücken sollten ebenfalls bei zukünftigen Studien mitbetrachtet werden.

Für konkretere Schlussfolgerungen oder Handlungsempfehlungen ist allgemein zur Wahrnehmung von Punkten im Gamificationsbereich weitere Forschung notwendig. Die grundlegende Frage, ob Spieldesignelemente kontextfrei untersucht werden können, ist dabei nicht von der Hand zu weisen.

DANKSAGUNGEN

Wir danken Marius Leka für die initiale Datensammlung im Rahmen seiner Masterarbeit.

LITERATUR

- [1] Paula Bräuer and Athanasios Mazarakis. 2022. "Alexa, Can We Design Gamification Without a Screen?" - Implementing Cooperative and Competitive Audio-Gamification for Intelligent Virtual Assistants. *Computers in Human Behavior* 135 (2022), 107362. <https://doi.org/10.1016/j.chb.2022.107362>
- [2] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work? – A Literature Review of Empirical Studies on Gamification. In *Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS)*. 3025–3034.
- [3] Martina Z. Huber Kolpondinos and Martin Glinz. 2017. Behind Points and Levels – The Influence of Gamification Algorithms on Requirements Prioritization. In *2017 IEEE 25th International Requirements Engineering Conference (RE)*. 332–341. <https://doi.org/10.1109/RE.2017.59>
- [4] Imthepfessordammit. 2021. How to Determine the Scale of Points in a Game? www.reddit.com/r/BoardgameDesign/comments/neyc8j/how_to_determine_the_scale_of_points_in_a_game/
- [5] Jonna Koivisto and Juho Hamari. 2019. The Rise of Motivational Information Systems: A Review of Gamification Research. *International Journal of Information Management* 45 (2019), 191–210. <https://doi.org/10.1016/j.ijinfomgt.2018.10.013>
- [6] Lisa A. Kort-Butler. 2021. Gamers on Gaming: A Research Note Comparing Behaviors and Beliefs of Gamers, Video Game Players, and Non-Players. *Sociological Inquiry* 91, 4 (2021), 962–982. <https://doi.org/10.1111/soin.12363>
- [7] De Liu, Radhika Santhanam, and Jane Webster. 2017. Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems. *MIS Quarterly* 41, 4 (2017), 1011–1034.
- [8] Athanasios Mazarakis. 2015. Using Gamification for Technology Enhanced Learning: The Case of Feedback Mechanisms. *Bulletin of the IEEE Technical Committee on Learning Technology* 4, 17 (2015), 6–9.
- [9] Athanasios Mazarakis. 2021. Gamification Reloaded: Current and Future Trends in Gamification Science. *i-com* 20, 3 (2021), 279–294. <https://doi.org/10.1515/icom-2021-0025> Publisher: Oldenbourg Wissenschaftsverlag.
- [10] Athanasios Mazarakis and Paula Bräuer. 2022. Gamification is Working, but Which One Exactly? Results from an Experiment with Four Game Design Elements. *International Journal of Human-Computer Interaction* (2022), 1–16. <https://doi.org/10.1080/10447318.2022.2041909> Publisher: Taylor & Francis.
- [11] Ganit Richter, Daphne R. Raban, and Sheizaf Rafaeli. 2015. Studying Gamification: The Effect of Rewards and Incentives on Motivation. In *Gamification in Education and Business*, Torsten Reiners and Lincoln C. Wood (Eds.). Springer International Publishing, Cham, 21–46. https://doi.org/10.1007/978-3-319-10208-5_2
- [12] Eleanor Rosch. 1975. Cognitive Reference Points. *Cognitive Psychology* 7, 4 (1975), 532–547. [https://doi.org/10.1016/0010-0285\(75\)90021-3](https://doi.org/10.1016/0010-0285(75)90021-3)
- [13] Michael Sailer. 2016. *Die Wirkung von Gamification auf Motivation und Leistung: Empirische Studien im Kontext manueller Arbeitsprozesse*. Springer, Wiesbaden, Germany.
- [14] Katie Seaborn. 2021. Removing Gamification: A Research Agenda. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–7. <https://doi.org/10.1145/3411763.3451695>
- [15] Katie Seaborn and Deborah I. Fels. 2015. Gamification in Theory and Action: A Survey. *International Journal of Human-Computer Studies* 74 (2015), 14–31. <http://linkinghub.elsevier.com/retrieve/pii/S1071581914001256>
- [16] Monique Simons, Emely de Vet, Johannes Brug, Jaap Seidell, and Mai J. M. China-paw. 2014. Active and Non-Active Video Gaming Among Dutch Adolescents: Who Plays and How Much? *Journal of Science and Medicine in Sport* 17, 6 (2014), 597–601. <https://doi.org/10.1016/j.jsams.2013.10.250>
- [17] Amos Tversky and Daniel Kahneman. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* 185, 4157 (1974), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124> Publisher: American Association for the Advancement of Science.
- [18] Elena Vitkauskaitė and Rimantas Gatautis. 2018. Points for Posts and Badges to Brand Advocates: The Role of Gamification in Consumer Brand Engagement. In *Proceedings of the 51st Hawaii International Conference on System Sciences*. Hawaii, USA, 1148–1157. <http://hdl.handle.net/10125/50030>
- [19] Luis von Ahn and Laura Dabbish. 2008. Designing Games with a Purpose. *Commun. ACM* 51, 8 (2008), 58–67. <https://doi.org/10.1145/1378704.1378719>
- [20] Kevin Werbach and Dan Hunter. 2012. *For the Win: How Game Thinking Can Revolutionize Your Business*. Wharton Digital Press, Philadelphia, PA, USA. OCLC: 911049737.
- [21] Gabe Zichermann and Christopher Cunningham. 2011. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. O'Reilly Media, Sebastopol, California, USA.