

Development and Validation of a German Version of the Player Experience Inventory (PXI)

LINDA GRAF*, University of Duisburg-Essen, Germany

MAXIMILIAN ALTMAYER, German Research Center for Artificial Intelligence (DFKI), Germany

KATHARINA EMMERICH, University of Duisburg-Essen, Germany

MARC HERRLICH†, German Research Center for Artificial Intelligence (DFKI), Germany

ANDREY KREKHOV, University of Duisburg-Essen, Germany

KATTA SPIEL, TU Wien, Austria

The Player Experience Inventory (PXI), initially developed by Abeele et al. (2020), measures player experiences among English-speaking players. However, empirically validated translations of the PXI are sparse, limiting the use of the scale among non-English speaking players. In this paper, we address this issue by providing a translated version of the scale in German, the most widely spoken first language in the European Union. After translating the original items, we conducted a confirmatory factor analysis ($N = 506$) to validate the German version of the PXI. Our results confirmed a 10-factor model—which the original authors of the instrument suggested—and show that the German PXI has valid psychometric properties. While model fit, internal consistency and convergent validity were acceptable, there was room for improvement regarding discriminant validity. Based on our results, we advocate for the German PXI as a valid and reliable instrument for assessing player experiences in German-speaking samples.

CCS Concepts: • **Human-centered computing** → **HCI design and evaluation methods**; *Empirical studies in HCI*; • **General and reference** → *Validation*.

Additional Key Words and Phrases: player experience, PX, questionnaire, game experience, games user research, GUR, scale translation, scale validation, measurement instrument

ACM Reference Format:

Linda Graf, Maximilian Altmeyer, Katharina Emmerich, Marc Herrlich, Andrey Krekhov, and Katta Spiel. 2022. Development and Validation of a German Version of the Player Experience Inventory (PXI). In *Mensch und Computer 2022 (MuC '22)*, September 4–7, 2022, Darmstadt, Germany. ACM, New York, NY, USA, 16 pages. <https://doi.org/10.1145/3543758.3543763>

1 INTRODUCTION

Games and play are fundamental human activities that can facilitate learning, social relationships and recreation [16]. Over the past decades, computer and video games have matured from a niche activity to a global mainstream mass media and multi-billion dollar industry with huge socio-cultural impact [11, 20, 23]. Hence, games are studied and used throughout many scientific disciplines from developing new algorithms improving technological aspects and

* All authors contributed equally to this research.

† Also with University of Kaiserslautern, Serious Games Engineering.

Permission to make digital or hard copies of all or part 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 components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

utilizing games as test or research environments to exploring the sociological and psychological impacts of games in various ways [22, 28, 29, 35]. Games User Research (GUR) in particular brings together aspects of game development and human-computer interaction (HCI) [10]. One main research direction in GUR is how to measure, understand, and improve player experiences, which describe the subjective feelings of players during and immediately after playing a game [37].

Over the past decades, researchers have developed and investigated a number of tools and procedures to measure and quantify player experiences, ranging from relatively simple playtest procedures, interviews, and observation protocols to complex bio-signal measurements and advanced data mining techniques [10, 11, 20, 23]. Still, subjective player feedback is one of the most important methods to understand how specific game aspects or design changes impact player experiences. Efforts have been made to develop scales for measuring player experiences [6, 17–19] based on different psychological theories, for example, the Player Experience of Need Satisfaction (PENS) [21] based on Self-Determination Theory (SDT) [31–33]. While some of these scales have been successfully used to find general insights, for example, on how games motivate players, others continue to be used even though they have not been strictly validated and sometimes scales are created “on-the-fly” to get more specific feedback [20, 24, 27]. To address this need for a more specific validated instrument, Abeele et al. have recently introduced the Player Experience Inventory (PXI) [1].

The PXI is a self-report questionnaire containing ten subscales that measure different dimensions of player experiences [1]. Thereby, five dimensions measure the functional consequences of player experiences, which means the direct experiences that result from the design of the game: *Ease of Control*, *Goals and Rules*, *Challenge*, *Progress Feedback*, and *Audiovisual Appeal*. Five additional dimensions measure the psychosocial consequences of game experiences, which means the secondary emotional experiences: *Meaning*, *Curiosity*, *Mastery*, *Immersion*, and *Autonomy*. All subscales consist of three items scored on a 7-point Likert scale (-3 to 3). Based on the two types of consequences measured, the PXI can be used to assess diverse facets of player experiences and investigate how game design choices are linked to players’ emotional responses.

The German speaking countries in Europe possess a very active GUR community and they are also important markets for the international games industry. For instance, the size of the commercial games market in Germany alone was estimated with about 4.6 billion Euro for the first half of 2021¹. In addition, German is the most widely spoken first language in the European Union². This illustrates that a consistently translated version of the PXI is needed to measure player experiences for games research with German-speaking samples. Furthermore, cross-cultural comparisons are a relevant research topic in their own right and would further benefit from a comparable validated scale.

We address this gap by presenting a validated German translation of the PXI. To the best of our knowledge, this work is among the first validated German language player experience scales overall. We conducted the translation and validation following an iterative process involving nine GUR experts with (different dialects of) German as their first language over several rounds of discussion and pre-studies. The resulting translation was then validated with an external sample ($N = 506$) following the procedure of the original PXI validation study [1] as closely as possible for direct comparison.

We found that, overall, the German PXI achieved very good and similar results in comparison to the English PXI version for most items and we can recommend it as a measurement tool for researchers and industry practitioners working with German speaking samples. However, there are also some edge cases that raise relevant points for discussion not only for the German PXI but also for the English scale.

¹<https://www.game.de/marktdaten/deutscher-games-markt-erstes-halb-jahr-2021/>, last accessed August 8, 2022

²<https://www.deutschland.de/de/topic/kultur/deutsche-sprache-ueberraschende-zahlen-und-fakten>, last accessed August 8, 2022.

2 RELATED WORK

In our discussion of related works, we focus on the most relevant scales and articles related to the translation of questionnaires. Finally, we refer our readers to the original paper [1] for an in-depth discussion of theories and scales for measuring player experiences in general.

Besides different qualitative and quantitative and subjective and objective measurements to assess player experiences, there are various questionnaires. For example, the Player Experience in Need Satisfaction (PENS) [33], the Game Experience Questionnaire (GEQ) [18], the Game Engagement Questionnaire [6], or the Immersive Experience Questionnaire (IEQ) [19]. When considering questionnaires related to player experiences, it is noticeable that there is a lack of validity in the measurement instruments. Empirical validation for the PENS by its authors is still pending [33]. The GEQ similarly exhibits issues with its validity. Accordingly, further validity studies could not confirm the proposed factor structure [7, 24]. Authors of the IEQ report an initial validation study that indicates five underlying factors of the questionnaire; however, when interpreting IEQ results, only a total score for immersion is calculated, and sub dimensions are not explicitly considered [19]. A nuanced analysis of different aspects of player experiences is, therefore, not readily possible with the mentioned questionnaires.

Moreover, the original PXI paper authors argue that most player experience questionnaires lack questions that can measure player experiences at the functional level. Furthermore, the PXI is increasingly used and accepted in the GUR community [2, 5, 30, 36]. There is also a short version of the PXI that consists of ten instead of thirty items, the Mini Player Experience Inventory (mPXI) [14]. Therefore, we decided to translate the PXI, as it is a valid tool for measuring player experiences and should subsequently be made available with the same degree of confidence in its reliability to the German-speaking GUR community.

The PXI measures two different consequences of player experiences: *functional consequences* and *psychosocial consequences* [1]. Hence, the PXI can be used to measure not only direct experiences that result from the design of a given game, but also the secondary emotional experiences while playing a game. Comparing the PXI to other PX measurements, it does not explicitly consider aspects like the influence of narrations or storytelling, or social aspects tied to specific game types and genres. In this way, the PXI can be used for a wide variety of games and players while enabling researchers to ask situated and nuanced questions through the instrument. There is no subscale measuring *enjoyment* while playing in the originally proposed instrument. Nevertheless, since its initial publication, the authors of the PXI recommend three further items measuring enjoyment on their website³.

There are no valid German translations of common player experience questionnaires, which suggests that researchers in German-speaking countries seem to resort to self-translated versions for their research. Due to the lack of consistency between these different versions it becomes methodologically difficult within the field to compare or even reproduce outcomes from different studies [20]. Especially, as player experiences present a complex construct that is not directly observable. Therefore, it depends on operationalization that occurs through a questionnaire. Accordingly, it is important that the operationalization is reliable and consistent [11]. Denisova et al. add that there is no need of one tool measuring all facets of PX, but each tool should accurately measure the aspects it claims to measure [9]. Furthermore, research has shown that constructs or concepts are presented differently in different cultures or do not have a similar meaning [3]. Therefore, it is important to validate the factor structure of subjective questionnaires and, in our case, the distinction of different aspects of player experiences measured by the PXI in each language and the respectively connected culture. Wilde et al. translated the Intrinsic Motivation Inventory (IMI) by Deci and Ryan [8] into a German version and

³<https://playerexperienceinventory.org/>, last accessed August 8, 2022.

shortened it simultaneously [38]. Their study not only focuses on validating the questionnaire, both translated and shortened, but also investigates whether it can be used to measure students' intrinsic motivation after a trip to the museum. Unfortunately, there are no details on the translation process given. Even, if they present an evaluation of validity by conducting a confirmatory factor analysis, their work underlines the importance of having a validated questionnaire to reach comparable results. Especially for questionnaires that ask subjective feeling, the correct wording is relevant [25]. When translating questionnaires, it is important to check whether the constructs occur equally in the respective culture. A 1:1 translation is not always necessary, since a word in one language may represent the respective construct, but the literal translation of this word in another language may say something different.

3 TRANSLATION PROCESS

For the process of translating the original PXI into a German version, we used the translation technique 'Modified Direct Translation' that includes translators and experts discussing the wording until reaching agreement [4]. First, nine researchers and experts from the GUR area translated all items independently. In a second step, all translations were compared and each item that received differing translations was noted. Thirdly, in a further round, discrepancies were discussed in the group and a final version was agreed upon. Finally, the questionnaire was checked for comprehensibility by two independent researchers who were not involved in the translation process, as well as by one person not from the GUR field. Again, each discrepancy was discussed until reaching agreement. Behling and Law value the used translation method as medium-high in terms of translation quality compared to other techniques. The best technique according to the authors is 'Translation/Back-Translation' [4], where bilingual individuals are involved who translate the questionnaire from source language to the target language and back, until reaching agreement. No back-translation into English was done since the main goal was that the constructs are represented correctly in German. Therefore, to make sure that our translation was comprehensible, we conducted a first online study to investigate the comprehensiveness of our translation directly with the target group. We asked participants to name the last game they played and to keep that in mind while filling in the PXI, like they did in the original PXI validation study [1]. Our instructions were as follows: "*Bitte denken Sie nun an das Spiel, das Sie vorhin angegeben haben (also das Spiel, welches Sie zuletzt gespielt haben). Denken Sie dabei daran, wie Sie sich während des Spielens gefühlt haben. Geben Sie für jede der folgenden Aussagen an, inwieweit Sie dieser zustimmen.*"⁴ In addition, there was the hint: "*Wenn Ihnen bei einer der Aussagen nicht klar ist, was gemeint ist, dann können Sie auch die Option "Ich verstehe nicht, was hier gemeint ist" wählen.*"⁵. In case participants chose this option, a further query on why they found the statement difficult to understand was presented for the respective item, so that we could improve the wording of problematic items. All subjects consented to our approach to data protection and indicated they understood the provided information regarding their participation. Additionally, all data were collected and stored anonymously.

In total, 73 people (29 female, 44 male, 1 non-binary, 3 non-disclosed) between 17 and 56 years ($M = 27.7$, $SD = 7.01$) participated in the preliminary study. Most participants did not report any problem of comprehension. However, four times the "I don't understand what is meant here"-option was selected for different items by different participants (one participant made comments on two items). Based on participants' explanations for their choice, we were able to determine that in two cases there was no fundamental problem of misunderstanding. The two remaining items are listed in Table 1. The first comment was on the *mea3* item: "*Wertvoll ist für mich gleich wie wichtig. Und das ist es nicht.*

⁴transl.: "Now please think about the game you indicated earlier (i.e., the game you played last). Think about how you felt while playing the game. For each of the following statements, indicate the extent to which you agree with it."

⁵transl.: "If it is unclear to you what is meant by one of the statements, you can also select the option 'I don't understand what is meant here'."

Table 1. Item translations that were changed after the pre-study.

Item	Original Item	Initial Translation	Modified Translation
mea3	Playing this game was valuable to me.	Dieses Spiel zu spielen war wertvoll für mich.	Das Spiel zu spielen war <i>ein wertvoller Zeitvertreib</i> für mich.
mas3	I felt a sense of mastery playing this game.	Ich hatte das Gefühl das Spiel zu beherrschen.	Ich hatte das Gefühl das Spiel <i>gut zu meistern</i> .

⁶(ID 35). This comment suggests that the word ‘wertvoll’ (valuable) can be perceived as a synonym for ‘important’ in German, which does not correspond to the intended meaning. The item is more about playing as a valuable activity at that moment, not necessarily whether it is described as important for the person beyond that. Therefore, we changed the word ‘wertvoll’ (valuable) into ‘wertvoller Zeitvertreib’ (valuable pastime) to distinguish this difference. The other comment was on the *mas3* item: “*Beherrschen im Sinne von Können; oder Beherrschen im Sinne von: Über das Spiel bestimmen?*”⁷ (ID 75). Here, the participant seemed to be unsure what the meaning of *sense of mastery* (German: Gefühl [das Spiel] zu beherrschen) was. Therefore, we changed the German word ‘beherrschen’ (to master) to ‘meistern’ (also to master, but more associated with skill) as this is more appropriate for performing in the game and not directing the game. The final version of the German PXI can be found in Appendix A and on <https://playerexperienceinventory.org/>.

4 STUDY

We conducted an online study to evaluate the validity of the German PXI. Here, we describe our preparatory strategies and describe the population sample engaging with the translated questionnaire.

4.1 Procedure, Study Design, and Recruitment

We decided to recruit participants online, as, on the one hand, our study took place during the Covid-19 pandemic, which affects the possibility of inviting participants to laboratories; on the other hand, this procedure is in line with the method of the original PXI paper [1] and provides the chance to recruit many participants. The recruitment was conducted via Reddit, Facebook, and traditional online channels (mailing lists and Moodle groups) of all involved universities. The survey link was posted and the study’s objectives and information about the responsible researchers were described. Participants agreed to the privacy policy that all data were collected and stored anonymously by placing a check mark. For the study design, we followed the approach of the original PXI validation study [1]: We asked participants to recall their last gaming session and to rate their respective experience by filling out the PXI. Hence, every participant related the questions to their very own game, for example, while one participant thought of a role-playing game, the other referred to a first-person shooter while filling out the questionnaire. The choice of game was thus left to the participant. Participants could refer to any digital game played on a smartphone, tablet, computer, or console.

In the survey, participants had to name the game they were referring to when filling out the PXI, indicate when they had last played the game, state how often they are playing digital games in general, and provide demographic information if they wanted to. Furthermore, there was a freeform text field at the end of the survey in case they had any questions or comments. The survey was online for 18 weeks. During this period, discussions on the various online platforms were monitored and moderated. The study was approved by an institutional ethics committee.

⁶transl.: “To me, valuable is equal to important. And it isn’t.”

⁷transl.: “Master meaning ‘skill’; or master meaning ‘rule over the game’?”

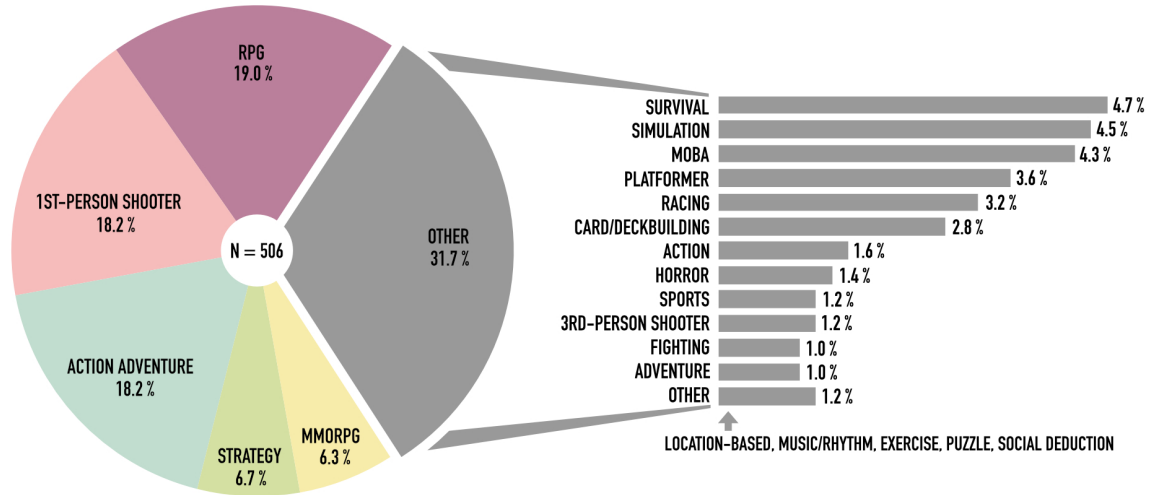


Fig. 1. Percentage distribution of genres of all the games participants referred to in our survey ($N = 506$).

4.2 Sample Description

In total, 1379 persons participated in the online study. 873 data sets were excluded because they were incomplete. Hence, 506 participants were included in the analysis. The participants (72 self-identifying as female, 414 self-identifying as male, 6 self-identifying as non-binary, 6 not specified, 8 not disclosed) were aged between 16 and 56 ($M = 28.0$, $SD = 5.43$). Participants reported playing digital games 18.08 hours per week on average ($SD = 12.71$). On a scale ranging from 1 to 7, they rated their gaming affinity ($M = 6.37$, $SD = 0.92$), frequency of play ($M = 6.17$, $SD = 1.12$), and their general passion for digital games ($M = 6.28$, $SD = 1.07$). Hence, the majority of participants described themselves as having a passion for digital games and playing such games often.

When completing the PXI questionnaire, participants referred to a total of 267 different games. Figure 1 provides an overview of the games' genres that the participants indicated. The three most represented game genres, which comprise more than 50 % of the sample, were role-playing games (19.0 %), first-person shooters (18.2 %), and action adventures (18.2 %). Further genres with more than ten mentions were strategy (6.7 %), MMORPG (6.3 %), survival (4.7 %), simulation (4.5 %), MOBA (4.3 %), platformer (3.6 %), racing (3.2 %), and card games (2.8 %). 37 % of the games were pure single-player games, whereas 28 % were multi-player only. The other 35 % of the games featured both single-player and multi-player modes.

5 RESULTS

The following presents the results, divided into the Confirmatory Factor Analysis and an analysis of qualitative data.

5.1 Confirmatory Factor Analysis

We conducted a Confirmatory Factor Analysis (CFA) with maximum likelihood estimation to test whether the translated items of the PXI in German fit the hypothesized model of the original version. The full list of non-standardized (B) and standardized (β) factor loadings as well as standard errors (SE) for each of the scale items can be found in Table 2.

Table 2. Factor loadings for survey items of the German PXI, Internal reliability (Cronbach's α), Composite Reliability (CR) estimates, and Average Variance Extracted (AVE). Bold entries represent values meeting the established cutoff criteria.

Factor	Item	B	β	SE	Cronbach's α	CR	AVE
Ease of Control	ec1	.98	.75	.06	.74	.75	.50
	ec2	.71	.66	.05			
	ec3	1.02	.71	.06			
Goals and Rules	gr1	.63	.67	.04	.78	.79	.55
	gr2	.94	.78	.05			
	gr3	.75	.77	.04			
Challenge	ch1	1.15	.77	.07	.79	.79	.56
	ch2	1.07	.71	.07			
	ch3	1.01	.76	.06			
Progress Feedback	pf1	1.18	.74	.07	.71	.74	.50
	pf2	.61	.46	.06			
	pf3	1.43	.86	.07			
Audiovisual Appeal	aa1	.60	.64	.04	.81	.82	.61
	aa2	.82	.63	.04			
	aa3	.94	.85	.04			
Meaning	mea1	1.47	.89	.06	.83	.84	.65
	mea2	1.51	.89	.06			
	mea3	.81	.61	.06			
Curiosity	cur1	1.49	.87	.06	.88	.88	.71
	cur2	1.34	.80	.07			
	cur3	1.50	.86	.07			
Mastery	mas1	1.23	.87	.05	.87	.87	.70
	mas2	1.13	.79	.06			
	mas3	1.12	.84	.05			
Immersion	imm1	.94	.51	.08	.75	.79	.56
	imm2	1.15	.88	.05			
	imm3	1.12	.82	.06			
Autonomy	aut1	1.39	.90	.06	.89	.89	.74
	aut2	1.35	.84	.06			
	aut3	1.19	.84	.05			

5.1.1 *Model Fit.* The CFA reaches significance ($\chi^2(360) = 742.33, p < .001$) and the chi-square test indicates a good to acceptable fit ($\frac{\chi^2}{df} = \frac{742.33}{360} = 2.06 < 3$ [34]). Because the chi-squared is sensitive to sample size and easily reaches significance for larger samples, we analyzed model fit indices to bypass this issue. The root mean square error of approximation (RMSEA = .05), with a threshold of <.06 [12], indicates good model fit. This is supported by looking at the comparative fit index (CFI = .95), which is $\geq .95$ and thus also represents a good model fit as well as the and the Tucker-Lewis index (TLI = .94), which is slightly below this threshold [15]. Overall, these results support that the model derived from the original PXI scale fits the data of the German PXI well.

5.1.2 *Internal Reliability.* The internal reliability of the translated factors was assessed by looking at Cronbach's α . As can be seen in Table 2, scale reliability is acceptable for all translated factors of the German PXI ($\alpha \geq 0.7$) [26]. This indicates that the sets of items of each factor of the German PXI are sufficiently related as a group.

Table 3. Discriminant validity of the German PXI: AVE and shared variances of the scales. Bold entries mark the highest AVE/covariance of each scale.

	EC	GR	CH	PF	AA	MEA	CUR	MAS	IMM	AUT
Ease of Control	.50									
Goals and Rules	.63	.55								
Challenge	.26	.27	.56							
Progress Feedback	.49	.65	.27	.50						
Audiovisual Appeal	.17	.13	.24	.26	.61					
Meaning	.12	.14	.19	.15	.46	.65				
Curiosity	.04	.02	.23	.12	.35	.50	.71			
Mastery	.34	.38	.35	.28	.30	.42	.27	.70		
Immersion	.00	.12	.17	.13	.37	.65	.42	.36	.56	
Autonomy	.07	.06	.24	.11	.27	.32	.27	.38	.26	.74

5.1.3 Convergent and Discriminant Validity. To conclude our analysis, we looked at convergent validity (i.e., whether items which should be related, actually are related) and discriminant validity (i.e., whether items that are not supposed to be related, are actually unrelated). To assess convergent validity, we considered composite reliability (CR), with a recommended criterion of 0.7 [39], and the average variance extracted (AVE), with a recommended criterion of 0.5 [13]. To analyze discriminant validity, we compared the AVE with the shared variance between the factors to check whether the AVE of each factor is higher than the covariance of any of the other factors. The CR estimates of all factors were above the recommended criterion of 0.7 [39] (see Table 2). Also, all AVE estimates met the recommended threshold of ≥ 0.5 . However, it should be noted that the *Ease of Control* and *Progress Feedback* factors were just on the border of this cutoff value. Since the AVE is considered a rather conservative measure for establishing convergent validity and a good CR is considered sufficient [13], we conclude that convergent validity is adequate, with room for improvement regarding the AVE of the aforementioned factors. In terms of discriminant validity, as shown in Table 3, the AVE is higher than the shared variance for six out of ten factors. This indicates that the explanatory power of the items of these factors is always higher than the explanatory power of these items on different factors, meaning that these factors can be discriminated well. However, it can be seen that the factors *Meaning*, *Ease of Control*, *Progress Feedback*, and *Goals and Rules* cannot be considered distinct enough to establish discriminant validity. For instance, the *Immersion* factor roughly explains as much variance of the *Meaning* factor as the factor itself. Also, the *Goals and Rules* factor explains more variance of the *Ease of Control* as well as the *Progress Feedback* factors themselves and has a higher shared variance to these factors than its own AVE. Based on these results, discriminant validity cannot be established for the whole scale.

5.2 Qualitative Analysis of Participants' Feedback

Apart from our main goal to validate the German translation of the PXI, we were also interested in how participants perceived the questionnaire. Hence, we now present their qualitative feedback in more detail in the following. 74 participants used the freeform text field at the end of the survey to provide feedback. We read through all comments, extracted the main message(s) and clustered them accordingly. We excluded 36 answers from our analysis, because they do not refer to the PXI questionnaire (e.g., "Viel Erfolg bei der Forschung!"⁸ (ID 779)). Based on the remaining 38

⁸transl.: "Good luck with your research!"

responses, we identified two main issues raised by our participants: (1) the repetitive nature of the questions and (2) the impression that certain items were not applicable to the game in question. The first aspect, repetitive questions, was mentioned by twelve participants in very similar statements, such as “Fragen in der Umfrage wiederholen sich teils mehrfach”⁹ (ID 625), “Fragen fühlen sich redundant an”¹⁰ (ID 422), and “Viele Fragen waren so minimal anders formuliert, meist [...] nur durch recht schwammige Begriffe, dass es für mich irritierend war”¹¹ (ID 145).

The second aspect is a bit more nuanced. In total, 29 participants expressed the feeling that the questionnaire or at least parts of it do not adequately capture the experiences they made with the last game they had played. However, the comments reveal different reasons for that impression. 15 participants referred to a lack of applicability to competitive games or multi-player games in general, e.g., “Für kompetitive Multiplayer Spiele passen manche Fragen nicht besonders gut”¹² (ID 733). Some participants further specified which items they found inappropriate, like in the following statement: “Da das von mir beschriebene Spiel ein PvP Multiplayerspiel ist, machen einige Fragen zum Spielfortschritt und zur Schwierigkeit wenig Sinn.”¹³ (ID 280). Here, the PXI dimensions *Progress Feedback* and *Challenge* are characterized as incongruous for competitive games. Another participant pointed out that questions regarding mastery and curiosity did not fit well for competitive games: “Die Fragen sind für kompetitive Spiele eher semi-optimal. Wenn das Ranking System gut ist, fühlt man sich nie als würde man das Spiel meistern. Auch eine Welt kann man nicht entdecken oder ‘sehen wie es weiter geht’.”¹⁴ (ID 335). The statement “Bei online-multiplayer spielen sind die Fragen danach, wie sehr man sich für den weiteren Verlaufe des Spiels interessiert, nicht wirklich sinnvoll. Es geht ja nicht um eine Story, etc.”¹⁵ (ID 929) also seems to refer to the *Curiosity* dimension of the PXI, which deals with players’ interest in how the game evolves. Some participants seem to instinctively associate the wording of the *Curiosity* items, such as “Ich wollte herausfinden, wie das Spiel verläuft”¹⁶ (cur2), with a narrative or explicit storyline. The following comments underline that: “Fragen sind stark auf Story / Einzelspieler Games ausgelegt”¹⁷ (ID 703) and “Zu Beginn der Umfrage wäre der Hinweis sinnvoll, dass die Fragen auf Spiele mit einer Handlung/Story ausgelegt sind.”¹⁸ (ID 333).

A lack of story or progress also seems to be the main reason why eleven participants perceived some items of the questionnaire as inappropriate for certain game genres. Apart from competitive games, participants mention sandbox games (IDs 774, 1115), strategy games (ID 774, 1115), e-sports (ID 333), and simulations (ID 161) in this context, e.g., “Ich fand die Formulierung ‘ich wollte wissen, wie das Spiel weitergeht’ oder Ähnliches schwierig, da dies eigentlich nur auf Spiele mit Fokus auf die Story zutrifft. Bei Aufbauspielen oder Simulatoren z.B. fände ich diese Aussage schwierig zu bewerten.”¹⁹ (ID 161).

Finally, three participants assessed some questions to be inapplicable for games that they have already played a lot: “Die Fragen sind z.T. sehr unpassend, wenn man ein Spiel schon länger spielt.”²⁰ (ID 774). One person pointed out that the result of the PXI might be biased due to the fact that the survey only refers to the last game session, particularly

⁹transl.: “Questions in the survey are sometimes repeated several times”

¹⁰transl.: “Questions feel redundant”

¹¹transl.: “Many questions were formulated so minimally differently, mostly with only somewhat vague terms, that it was irritating for me.”

¹²transl.: “For competitive multi-player games some questions do not fit very well”

¹³transl.: “Since the game I described is a PvP multiplayer game, some questions about game progress and difficulty make little sense.”

¹⁴transl.: “The questions are rather semi-optimal for competitive games. When the ranking system is good, you never feel like you are mastering the game. Also, you can’t discover a world or ‘see what happens next’.”

¹⁵transl.: “In online multiplayer games, the questions about how interested you are in the further course of the game don’t really make sense. After all, there is no storyline, etc. involved.”

¹⁶Original item: “I wanted to explore how the game evolved.”

¹⁷transl.: “Questions are heavily weighted towards story / single-player games.”

¹⁸transl.: “At the beginning of the survey, it would be useful to note that the questions are designed for games with a plot/story.”

¹⁹transl.: “I found the phrase ‘I wanted to know how the game would continue’ or similar difficult, since this really only applies to games with a focus on the story. For example, I would find this statement difficult to evaluate for construction games or simulations.”

²⁰transl.: “Some of the questions are very inappropriate if you have been playing a game for a while.”

when the game was a competitive game: “Die Fragen sind für Multiplayer Spiele schlecht/unpassend gewählt. Ich hab bsp über 2k spielstunden in League of Legends. Heißt ich mag das Spiel anscheinend doch genug um entsprechend viel Zeit zu investieren. Da aber nach meiner letzten erfahrung gefragt wurde musste ich überwiegend negativ eingeben, weil dass die erfahrung in diesem einem match war. Somit nicht repräsentativ für das gesamttempfinden über das Spiel.”²¹ (ID 35).

In addition to the direct feedback provided at the end of the survey, we tracked the evolving discussions about our survey on the online platforms that we used to promote the study. While there were no comments regarding our posts in any of the Facebook groups, we observed that lively discussions evolved in some sub-reddits (in particular, the sub-reddits “r/zocken” and “r/rocketbeans”). These forums were also the ones where the majority of participants could be recruited. In such sub-reddits, users can write comments and use so-called *upvotes* to express their agreement with an existing comment. Overall, the comments in the sub-reddits reflect the feedback from the freeform text of the survey. The most popular comment (receiving 27 upvotes) points out that the questions were not well-suited for certain game types, such as competitive games or sandbox games: “Die Fragen sind jetzt nicht besonders gut für kompetitive Spieler oder Sandbox Spieler geeignet. Das bezieht sich alles zu sehr auf story-lastige Spiele die man durch spielen kann. Oder sowas wie Jump’n’Runs.”²² Again, the commenter seems to attribute this problem to items of the PXI which relate to progress or story. Similarly, Reddit users highlighted the problem of playtime: The comment “Die Fragen sind alle so formuliert, als würde davon ausgegangen, ich hätte das Spiel gerade zum ersten Mal gespielt”²³ received 22 upvotes, and the same user further stated: “[...] wenn einfach nur nach dem Spiel gefragt wird, das man als Letztes gespielt hat, war das in meinem Fall ein Spiel, das ich schon so ca. 400 Stunden gespielt hab. Da sind dann Fragen darüber, ob ich mit der Steuerung zurecht komme oder mir das Spielziel klar ist, nicht besonders aussagekräftig. [...]”²⁴ Finally, the repetitiveness of the questions was criticized several times, in four individual comments, which received six, six, three, and one upvotes, respectively.

Overall, the qualitative comments to the survey indicated that participants expressed difficulties with repetitive questions and, occasionally, with translating specific questions to their individual game of reference.

6 DISCUSSION

As presented above, the goal of this work was to create and validate a German version of the PXI to be used for studies with German speakers. We followed an iterative approach that involved several repetitions of letting nine German and English speaking GUR experts create translations and then identify and resolve potential issues through group-based discussion. Then, we conducted a pre-study with a smaller sample of external participants and finally a larger validation study. The statistical analysis revealed that our translation meets well accepted standards with respect to the specific measures and the results are also very close to those that Abeele et al. reported for English speaking samples using the English language PXI [1]. However, we discuss some specific results leave room for improvement and some relevant insights from the qualitative feedback we received in the following paragraphs.

²¹ transl.: “The questions are poorly chosen for multiplayer games. For instance, I have over 2k hours of play in League of Legends. That means I apparently like the game enough to invest a lot of time. But since my last experience was asked, I had to enter mostly negative, because that was the experience in this one match. Thus not representative for the overall feeling about the game.”

²² transl.: “The questions are not very well suited for competitive players or sandbox players. It’s all too much about story-heavy games that you can play through. Or something like jump’n’runs.”

²³ transl.: “The questions are all phrased as if it were assumed that I had just played the game for the first time”

²⁴ transl.: “If you’re just asked about the last game you played, in my case it was a game I’d already played for about 400 hours. Questions about whether I can handle the controls or whether the game’s objective is clear to me are not very meaningful.”

Overall, the empirical validation revealed that the German PXI has valid psychometric properties: We found that the proposed model fitted the data well, meeting established thresholds for RMSEA and CFI. Reliability and convergent validity was similarly given, since Cronbach's α , CR and AVE were all within acceptable thresholds. However, it should be noted that some factors (especially *Ease of Control* and *Progress Feedback*), were just above threshold regarding AVE. This indicates that the respective items could be refined such that more variance is explained by them on the factor they are supposed to measure. It should be considered that these factors had similar issues in the original scale validation study (the AVE of *Ease of Control* even fell below the threshold of .50, as the authors reported an AVE value of .47 [1]). This suggests that this issue is not necessarily related to our translation, but could be a general problem of the constructs of the PXI.

Similarly, we found issues related to the discriminant validity of the German PXI, i.e., some factors could not be discriminated well enough from other factors. This issue especially concerned the *Ease of Control*, *Goals and Rules*, and *Meaning* factors. To better understand whether these issues are related to the German translation or are already evident in the original scale, we compared the covariances between factors with those reported in the original scale validation study. Although no covariances were greater than the AVE in the original version of the PXI, we could see that there are substantial covariances on the factors that could not be well discriminated in our study. For instance, the AVE of the *Meaning* factor was as high as its covariance to the *Immersion* factor in our study. In the original version, this covariance is not greater than the AVE of the *Meaning* factor, but still represents the greatest covariance to another factor. Similarly, there is substantial covariance between *Goals and Rules*, *Ease of Control*, and *Progress Feedback*. Although these covariances are not as high as in our data, we can find similar patterns. Overall, we can conclude that the issues related to discriminant validity regarding the aforementioned factors are already prevalent in the original scale, but less pronounced. To improve our scale, we suggest to carefully consider the problematic factors and take action such that they can be better distinguished from each other. This could be achieved by changing the focus of items, merging factors or even removing or adding new factors and/or items. It could further be worthwhile to consider such improvements for the original version of the PXI to improve the discriminant validity there as well.

Overall, we have shown that our German translation of the PXI is a valid instrument in terms of formal criteria such as model fit, reliability, and convergence. However, we have also identified some issues regarding certain factors. The qualitative analysis of both survey responses and discussions on Reddit provides further insights about how participants perceived the questionnaire. Such insights might reveal reasons for the emergence of the identified issues and, moreover, point towards additional potential limitations of the PXI.

First of all, participants rated some questions to be inapplicable with respect to the game they had in mind. Therefore, they were unsure how to answer them. According to the reasons mentioned in the comments, this problem is not related to the German translation but to the questions' content, indicating a more general issue of the questionnaire. A lack of fit of certain questions was mainly perceived by participants who evaluated their player experience in a competitive game or another type of game that does not contain a storyline or a progress-based structure (e.g., sandbox games). In such cases, participants reported difficulties with answering questions related to the factors *Progress Feedback*, *Challenge* and *Curiosity*. Considering that about one third of our sample referred to multi-player games, the problem may have contributed to a lack of convergent and discriminant validity, especially with regard to the factor *Progress Feedback*.

Interestingly, participants seem to have an implicit understanding of items of the *Curiosity* subscale as referring to a narrative, even though the items do not directly relate to a story, narrative, plot, or similar. This is particularly remarkable, since during the scale construction of the original PXI the authors decided to exclude constructs that have a strong focus on narration, with the reason that they would not apply to certain game genres [1]. Our results indicate

that despite these considerations some participants tend to instinctively interpret the *Curiosity* items as referring to a narrative—at least in our German translation. We suggest to further investigate participants' understanding of this factor in future studies to figure out if the unintentional association is a specific feature of the German translation or already occurs in the original scale.

Furthermore, we have to consider that many participants in our sample referred to a game they had played frequently before. Accordingly, several participants stated that questions regarding the understanding of controls and rules or the game's progression seemed pointless to them. This reaction is comprehensible, because one would expect most players to quit playing a game if they do not understand the rules or think the controls are too complicated. The perceived inappropriateness of the respective items may have biased our results regarding the PXI factors *Goals and Rules*, *Ease of Control* and *Progress Feedback* and might account for their substantial covariance to some degree.

In addition, some participants felt that rating a game just based on the last gaming session is not representative for their average experience with the game. The question arises whether the PXI is applicable for assessing long-term player experiences or is limited to novel game experiences. It seems that at least parts of the questionnaire, more precisely the constructs *Goals and Rules* and *Ease of Control*, make less sense for evaluating frequently played games. In other contexts, such as prototype playtesting or laboratory studies, where participants play an unknown game for the first time, the factors are much more relevant and should not cause any confusion.

Apart from criticizing some questionnaire items regarding their content, participants also commented on the structure of the questionnaire. More precisely, they pointed out that several questions seemed redundant. While some participants framed this as a rather neutral observation, others expressed irritation or even annoyance.

Considering our recruitment method, we can assume that most participants were unfamiliar with scientific surveys or the principles of questionnaire construction (in contrast to samples drawn among university students). Hence, there is a lack of understanding for repetitive questions, which might lead to negative feelings and disapproval, which in turn might bias the results or even cause participants to abandon the survey before completion. We therefore recommend to include an explicit statement at the beginning of the PXI, which explains that some questions in the following might feel redundant and that this is intended (our suggestion for a complete instruction text is provided in Appendix A). Such a note will prevent irritation and presumably increase acceptance, though some participants might still question the purpose of repetitive items.

Another obvious solution would be to completely remove repetitive items, i.e., the development of a short version of the PXI that includes only one item for each dimension. While a short version would have the added benefit of being less time-consuming and more flexible to use (e.g., in repeated measurement study designs), items should not be removed on the fly without proper validation. Hence, we see the development and validation of a short version PXI as one promising direction for future work. In this context, work-in-progress by Harteveld et al. presents initial advances [14].

7 LIMITATIONS

As games user researchers, we should not forget how diverse the gaming community has become due to the ubiquitous character of games. Playing is not limited by age, gender or other individual characteristics and can range from infrequent, casual engagements to intense, daily sessions. Hence, study-oriented research in this area must consider the underlying population sample when interpreting and generalizing the results.

We relied on acquisition methods and social media channels that are very popular among frequent players. Our sample description reflects this acquisition method: most of the participants are frequent male players, 28 years old on

average, with a very pronounced passion for games. It is not surprising that the three most popular genres—namely role-playing games, first-person shooters, and action adventures—are the ones that usually target frequent players and less so casual players. Hence, the applicability of the PXI might differ for underrepresented genres in our sample. Similarly, the outcomes could vary if we consider a sample that differs greatly in age and, most importantly, gaming behavior. For the future, we suggest a repetition of our procedure with an explicit focus on specific target groups and application areas, such as games for children, casual mobile games, or gaming experiences focusing on elderly adults.

It should be considered that the drop-out rate was comparably high in the online study. This can be explained by how and where participants were recruited. The participants were not compensated and the fact that the survey was publicly posted on Reddit might have attracted many people being curious but not necessarily motivated to complete the survey.

We would like to additionally emphasize one specific characteristic of our validation procedure: Similarly to the original PXI validation, the questions of our questionnaire referred to the last gaming session. This design decision needs to be addressed because it differs from the “real life” application procedure where such questionnaires are ideally administered right after playing the—usually new—game. For instance, the last gaming session in our case could have been several weeks ago, which can lead to biased answers due to the ability to forget memories selectively. Furthermore, the last game of our participants was often one they had already played before. This nuance can render specific factors, for example, *Ease of Control*, obsolete, and related questions might lead to slight confusion, as we already debated in section 6.

8 CONCLUSION

This paper presents the development and validation of a German version of the PXI. By providing a validated translation, we aim at supporting games researchers and practitioners who work with German speaking target groups. Until now, there was only the option of using ‘on-the-fly’ translations when incorporating the PXI into GUR studies. Such untested translations carry the risk that constructs are not measured correctly and diminish the comparability of results across studies. Our translation achieves good values for model fit, internal reliability, and convergent validity, and the results are comparable to those from the original PXI. Hence, we conclude that our German PXI is valid and recommend its use in future studies with German-speaking participants.

So, besides the practical use, how could our insights contribute to future research on player experience questionnaires? Our results reveal potential limitations and shortcomings of the PXI, which should be considered in follow-up investigations, for example, the applicability to long-term player experiences. We argue that some of these findings are not just related to the linguistic aspects of our translation. In this regard, we rather classify this manuscript as a further iteration of the original PXI research. Our insights pave the way for an even more fine-tuned version of the PXI questionnaire in the future.

ACKNOWLEDGMENTS

We want to thank all the people who participated in our surveys and also discussed our project with us. We also like to thank all GUR experts contributing to the translation, specifically Stefan Liszio, Kathrin Gerling, and Vero Vanden Abele.

REFERENCES

- [1] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, Daniel Johnson, and Kathrin Gerling. 2020. Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human-Computer Studies* 135 (2020), 102370. <https://doi.org/10.1016/j.ijhcs.2019.102370>
- [2] Lena Fanya Aeschbach, Sebastian AC Perrig, Lorena Weder, Klaus Opwis, and Florian Brühlmann. 2021. Transparency in measurement reporting: A systematic literature review of chi play. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 1–21. <https://doi.org/10.1145/3474660>
- [3] Dominique Banville, Pauline Desrosiers, and Yvette Genet-Volet. 2000. Translating questionnaires and inventories using a cross-cultural translation technique. *Journal of teaching in Physical Education* 19, 3 (2000), 374–387.
- [4] Orlando Behling and Kenneth S Law. 2000. *Translating questionnaires and other research instruments: Problems and solutions*. Vol. 133. sage.
- [5] João Bosco Borges, Carmen Li Juy, Izac Sidarta de Andrade Matos, Paulo Vitor Angelo Silveira, and Ticianne de Gois Ribeiro Darin. 2020. Player Experience Evaluation: a Brief Panorama of Instruments and Research Opportunities. *Journal on Interactive Systems* 11, 1 (2020), 74–91. <https://doi.org/10.5753/jis.2020.765>
- [6] Jeanne H. Brockmyer, Christine M. Fox, Kathleen A. Curtiss, Evan McBroom, Kimberly M. Burkhart, and Jacquelyn N. Pidruzny. 2009. The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology* 45, 4 (2009), 624–634. <https://doi.org/10.1016/j.jesp.2009.02.016> Publisher: Elsevier.
- [7] Florian Brühlmann and Gian-Marco Schmid. 2015. How to measure the game experience? Analysis of the factor structure of two questionnaires. In *Proceedings of the 33rd annual acm conference extended abstracts on human factors in computing systems*. 1181–1186. <https://doi.org/10.1145/2702613.2732831>
- [8] Edward L Deci and Richard M. Ryan. 2013. *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.
- [9] Alena Denisova, A Imran Nordin, and Paul Cairns. 2016. The convergence of player experience questionnaires. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. 33–37. <https://doi.org/10.1145/2967934.2968095>
- [10] Anders Drachen, Pejman Mirza-Babaei, and Lennart E. Nacke. 2018. *Games user research*. Oxford University Press.
- [11] Katharina Emmerich, Natalya Bogacheva, Mareike Bockholt, and Viktor Wendel. 2016. Operationalization and measurement of evaluation constructs. In *Entertainment Computing and Serious Games*. Springer, 306–331.
- [12] Leandro R. Fabrigar, Duane T. Wegener, Robert C. MacCallum, and Erin J. Strahan. 1999. Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods* 4, 3 (sep 1999), 272–299. <https://doi.org/10.1037/1082-989X.4.3.272>
- [13] Claes Fornell and David Larcker. 1981. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research* 18, 1 (1981), 39–50. <https://doi.org/10.1177/002224378101800104>
- [14] Casper Hartevelde, Nithesh Javvaji, Tiago Machado, Yevgeniya V. Zastavker, Victoria Bennett, and Tarek Abdoun. 2020. Preliminary Development and Evaluation of the Mini Player Experience Inventory (MPXI). In *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play*. Association for Computing Machinery, New York, NY, USA, 257–261. <https://doi.org/10.1145/3383668.3419877>
- [15] Li Tze Hu and Peter M. Bentler. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling* 6, 1 (1999), 1–55. <https://doi.org/10.1080/10705519909540118>
- [16] Johan Huizinga. 2020. *Homo ludens*. Editora Perspectiva SA.
- [17] Johan Högberg, Juho Hamari, and Erik Wästlund. 2019. Gameful Experience Questionnaire (GAMEFULQUEST): an instrument for measuring the perceived gamefulness of system use. *User Modeling and User-Adapted Interaction* 29, 3 (2019), 619–660. <https://doi.org/10.1007/s11257-019-09223-w> Publisher: Springer.
- [18] Wijnand A. IJsselstein, Yvonne AW de Kort, and Karolien Poels. 2013. The game experience questionnaire. *Eindhoven: Technische Universiteit Eindhoven* (2013), 3–9.
- [19] Charlene Jennett, Anna L Cox, Paul Cairns, Samira Dhoparee, Andrew Epps, Tim Tijs, and Alison Walton. 2008. Measuring and defining the experience of immersion in games. *International journal of human-computer studies* 66, 9 (2008), 641–661. <https://doi.org/10.1016/j.ijhcs.2008.04.004>
- [20] Daniel Johnson, M John Gardner, and Ryan Perry. 2018. Validation of two game experience scales: the player experience of need satisfaction (PENS) and game experience questionnaire (GEQ). *International Journal of Human-Computer Studies* 118 (2018), 38–46. <https://doi.org/10.1016/j.ijhcs.2018.05.003>
- [21] Dominik Kayser, Sebastian Andrea Caesar Perrig, and Florian Brühlmann. 2021. Measuring Players' Experience of Need Satisfaction in Digital Games: An Analysis of the Factor Structure of the UPEQ. In *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play*. 158–162. <https://doi.org/10.1145/3450337.3483499>
- [22] Pascal Klein, Nicole Burkard, Larissa Hahn, Merten Nikolay Dahlkemper, Kevin Eberle, Tina Jäger, and Marc Herrlich. 2021. Coordinating vector field equations and diagrams with a serious game in introductory physics. *European Journal of Physics* (2021). <https://doi.org/10.1088/1361-6404/abef5c>
- [23] Patri Lankoski, Staffan Björk, et al. 2015. *Game research methods: An overview*. Lulu. com.
- [24] Effie L.-C. Law, Florian Brühlmann, and Elisa D. Mekler. 2018. Systematic review and validation of the game experience questionnaire (geq)-implications for citation and reporting practice. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. 257–270.
- [25] Tara T Lineweaver and Jessica W Broolsma. 2014. How you ask matters: An experimental investigation of the influence of mood on memory self-perceptions and their relationship with objective memory. *Memory* 22, 8 (2014), 1103–1115. <https://doi.org/10.1080/09658211.2013.870209>

- [26] Robert R McCrae, John E Kurtz, Shinji Yamagata, and Antonio Terracciano. 2011. Internal Consistency, Retest Reliability, and Their Implications for Personality Scale Validity. *Personality and Social Psychology Review* 15, 1 (2011), 28–50. <https://doi.org/10.1177/1088868310366253>
- [27] Elisa D. Mekler, Julia Ayumi Bopp, Alexandre N. Tuch, and Klaus Opwis. 2014. A systematic review of quantitative studies on the enjoyment of digital entertainment games. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 927–936. <https://doi.org/10.1145/2556288.2557078>
- [28] Julia Müller, Max Sprenger, Tobias Franke, Paul Lukowicz, Claudia Reidick, and Marc Herrlich. 2020. Game of TUK: deploying a large-scale activity-boosting gamification project in a university context. In *Proceedings of the Conference on Mensch und Computer (MuC '20)*. Association for Computing Machinery, New York, NY, USA, 169–172. <https://doi.org/10.1145/3404983.3410008>
- [29] Anke Verena Reinschluessel, Joern Teuber, Marc Herrlich, Jeffrey Bissel, Melanie van Eikeren, Johannes Ganser, Felicia Koeller, Fenja Kollasch, Thomas Mildner, Luca Raimondo, Lars Reising, Marc Ruedel, Danny Thieme, Tobias Vahl, Gabriel Zachmann, and Rainer Malaka. 2017. Virtual Reality for User-Centered Design and Evaluation of Touch-free Interaction Techniques for Navigating Medical Images in the Operating Room. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2001–2009. <https://doi.org/10.1145/3027063.3053173> event-place: Denver, Colorado, USA.
- [30] Katja Rogers, Maximilian Milo, Michael Weber, and Lennart E Nacke. 2020. The Potential Disconnect between Time Perception and Immersion: Effects of Music on VR Player Experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 414–426. <https://doi.org/10.1145/3410404.3414246>
- [31] R.M. Ryan, V. Mims, and R. Koestner. 1983. Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of Personality and Social Psychology* 45, 4 (1983), 736–750. <https://doi.org/10.1037/0022-3514.45.4.736>
- [32] Richard M. Ryan and Edward L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68. <https://doi.org/10.1037/0003-066X.55.1.68>
- [33] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The motivational pull of video games: A self-determination theory approach. *Motivation and emotion* 30, 4 (2006), 344–360. <https://doi.org/10.1007/s11031-006-9051-8>
- [34] Karin Schermelleh-Engel, Helfried Moosbrugger, Hans Müller, et al. 2003. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online* 8, 2 (2003), 23–74.
- [35] Jan David Smeddinck, Marc Herrlich, and Rainer Malaka. 2015. Exergames for Physiotherapy and Rehabilitation: A Medium-term Situated Study of Motivational Aspects and Impact on Functional Reach. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 4143–4146. <https://doi.org/10.1145/2702123.2702598> event-place: Seoul, Republic of Korea.
- [36] April Tyack and Peta Wyeth. 2021. "The Small Decisions Are What Makes it Interesting" Autonomy, Control, and Restoration in Player Experience. *Proceedings of the ACM on Human-Computer Interaction* 5, CHI PLAY (2021), 1–26. <https://doi.org/10.1145/3474709>
- [37] Josef Wiemeyer, Lennart Nacke, Christiane Moser, and Florian 'Floyd' Mueller. 2016. *Player Experience*. Springer International Publishing, Cham, 243–271. https://doi.org/10.1007/978-3-319-40612-1_9
- [38] Matthias Wilde, Katrin Bätz, Anastassiya Kovaleva, and Detlef Urhahne. 2009. Überprüfung einer Kurzsкала intrinsischer Motivation (KIM). *Zeitschrift für Didaktik der Naturwissenschaften* 15 (2009). <https://pub.uni-bielefeld.de/record/2404161>
- [39] Adriana Zaiț and PSPE Berteau. 2011. Methods for Testing Discriminant Validity. *Management & Marketing Journal* 9, 2 (2011), 217–224.

A GERMAN PXI

Instruction: Bitte denken Sie nun an das Spiel, das Sie vorhin gespielt haben. Denken Sie dabei insbesondere daran, wie Sie sich während des Spielens gefühlt haben. Geben Sie für jede der folgenden Aussagen an, inwieweit Sie dieser zustimmen. Ihnen wird auffallen, dass einige Aussagen sich sehr ähneln. Dies ist beabsichtigt. Lassen Sie sich davon bitte nicht irritieren und bewerten Sie jede einzelne Aussage.

Scale: 7-stufige Likert-Skala, von -3 (= Stimme überhaupt nicht zu) bis +3 (= Stimme voll und ganz zu)

Constructs & Items

Benutzbarkeit - Ease of Control

1. (ec1) Es war einfach zu verstehen, wie Aktionen im Spiel auszuführen sind.
 2. (ec2) Die Aktionen zur Spielsteuerung waren für mich klar.
 3. (ec3) Ich fand das Spiel war leicht zu bedienen.
-

Ziele & Regeln - Goals & Rules

1. (gr1) Ich habe das Gesamtziel des Spiels verstanden.
 2. (gr2) Die Ziele des Spiels waren für mich klar verständlich.
 3. (gr3) Ich habe meine Aufgaben im Spiel verstanden.
-

Herausforderungen - Challenge

1. (ch1) Das Spiel war weder zu einfach noch zu schwierig.
 2. (ch2) Das Spiel war herausfordernd, aber nicht zu sehr.
 3. (ch3) Der Schwierigkeitsgrad des Spiels war für mich angemessen.
-

Feedback zum Fortschritt - Progress Feedback

1. (pf1) Das Spiel informierte mich über meinen Spielfortschritt.
 2. (pf2) Ich konnte meine Leistung im Spiel leicht einschätzen.
 3. (pf3) Das Spiel gab mir klares Feedback über meinen Fortschritt auf dem Weg zu den einzelnen Zielen.
-

Audiovisuelles Erscheinungsbild - Audiovisual Appeal

1. (aa1) Mir gefiel die Art und Weise, wie das Spiel gestaltet war.
 2. (aa2) Mir gefiel das Aussehen (Look and Feel) des Spiels.
 3. (aa3) Ich schätzte die Ästhetik des Spiels.
-

Bedeutsamkeit - Meaning

1. (mea1) Das Spiel zu spielen war für mich bedeutsam.
 2. (mea2) Das Spiel fühlte sich für mich wichtig an.
 3. (mea3) Das Spiel zu spielen war ein wertvoller Zeitvertreib für mich.
-

Neugierde - Curiosity

1. (cur1) Ich wollte entdecken, wie sich das Spiel entwickelt.
 2. (cur2) Ich wollte herausfinden, wie das Spiel verläuft.
 3. (cur3) Ich war gespannt darauf, wie das Spiel weitergeht.
-

Können/Abschneiden - Mastery

1. (mas1) Ich hatte das Gefühl gut in dem Spiel zu sein.
 2. (mas2) Ich fühlte mich kompetent, während ich das Spiel spielte.
 3. (mas3) Ich hatte das Gefühl das Spiel gut zu meistern.
-

Immersion - Immersion

1. (imm1) Ich war mir meiner direkten Umgebung nicht mehr bewusst, während ich spielte.
 2. (imm2) Ich war in das Spiel vertieft.
 3. (imm3) Ich war voll auf das Spiel konzentriert.
-

Autonomie - Autonomy

1. (aut1) Ich hatte das Gefühl, das Spiel auf meine eigene Art spielen zu können.
 2. (aut2) Ich hatte das Gefühl, die Wahl zu haben, wie ich das Spiel spielen wollte.
 3. (aut3) Ich hatte das Gefühl, das Spiel so spielen zu können, wie ich es wollte.
-