Hexad-12: Developing and Validating a Short Version of the Gamification User Types Hexad Scale

Jeanine Krath*
University of Koblenz-Landau
Koblenz, Rhineland-Palatinate, Germany
jkrath@uni-koblenz.de

Maximilian Altmeyer*
German Research Center for Artificial Intelligence (DFKI), Saarland Informatics Campus
Saarbrücken, Saarland, Germany
HCI Games Group, Games Institute, University of Waterloo
Waterloo, ON, Canada
maximilian.altmeyer@dfki.de

Gustavo F. Tondello
HCI Games Group, Games Institute, University of Waterloo
Waterloo, ON, Canada
gustavo@tondello.com

Lennart E. Nacke
Stratford School of Interaction Design and Business, University of Waterloo
Waterloo, ON, Canada
HCI Games Group, Games Institute, University of Waterloo
Waterloo, ON, Canada
lennart.nacke@acm.org

ABSTRACT

The Hexad scale is a crucial tool for personalized gamification in user experience (UX) design. However, completing a 24-item questionnaire can increase dropout rates and screen fatigue within online surveys. When included in larger surveys, scale brevity makes a difference. To reduce the time required for the assessment process, we developed and validated a 12-item version of the Hexad scale. To create it, we carried out an exploratory factor analysis on an existing data set to identify appropriate items (n = 882). To validate the 12-item version, we conducted a confirmatory factor analysis on a new data set (n = 1,101). Our results show that Hexad-12 outperforms the original Hexad scale regarding model fit, reliability, convergent, and discriminant validity. Therefore, Hexad-12 resolves issues found in studies using the original Hexad scale and provides a suitable and swift instrument for concisely assessing Hexad user types in tailored gamification design.

CCS CONCEPTS

• **Human-centered computing** → **User models**; **HCI theory, concepts and models**; • **Social and professional topics** → *User characteristics.*

*Both authors contributed equally to this research.

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).
CHI '23, April 23–28, 2023, Hamburg, Germany
© 2023 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9421-5/23/04.
https://doi.org/10.1145/3544548.3580968

KEYWORDS

Gamification, Personalization, Hexad, User Types, Player Types, Tailored Gamification, Adaptive Gamification

ACM Reference Format:
Jeanine Krath, Maximilian Altmeyer, Gustavo F. Tondello, and Lennart E. Nacke. 2023. *Hexad-12:* Developing and Validating a Short Version of the Gamification User Types Hexad Scale. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3544548.3580968

1 INTRODUCTION

Gamification—*the use of game elements in non-game contexts* [24]—has been researched and applied in many contexts, including health and well-being, education, and crowdsourcing [36, 77]. Early work primarily focused on the design of gamified systems [65] and studying *if* they lead to beneficial outcomes [58]. However, latest studies have also pointed out mixed or even negative effects [17, 90], for example, in gamified applications for physical activity [5]. Therefore, recent works aimed to better understand *how and why* gamification works by studying different moderating factors [44, 58].

Most of these works focused on personal factors [44] and found that there are interpersonal differences in how certain gamification elements (such as points, badges, or leaderboards) are perceived [7, 16]. Thus, personalization of gamified systems is an important issue for successful gamification design [46, 74]. To explain these interpersonal differences and guide gamification design, Marczewski [53] proposed the Gamification User Types Hexad model. In contrast to other player typologies such as Bartle [13] or BrainHex [57], the Hexad model is the only model which has been specifically developed to conceptualize and explain user preferences in gamified systems, rather than full games [60, 85]. To enable using Marczewski's Hexad model for gamification design, Tondello et al.

[86] developed a 24-item questionnaire for the assessment of the six Hexad user types (four items per factor), refined it, and demonstrated its reliability as well as validity [84]. As a result—despite its novelty—the Hexad model has become one of the most frequently used models to personalize gamified systems [44]. It has already been used in various domains, such as physical activity [7], education [56], health [60], and energy conservation [45]. Furthermore, it was shown to be the most appropriate user typology to explain user preferences in gamified systems, compared to other factors and models [34]. Because of its popularity, it was translated into many languages already, such as Turkish [2, 81], Dutch [59], German [47], Spanish [84], and Brazilian Portuguese [69].

However, the 24-item length of the Hexad scale could be a limiting factor for both academics and practitioners. In particular, there are four reasons why we believe that an abbreviated version of the Hexad questionnaire would be a valuable contribution to both research and practice. (1) First, *assessment time is a relevant factor for UX researchers* [64] to decrease the dropout rate of participants, the chance of random responding, and to prevent negative effects on data quality [38, 66]. Researchers typically rely on extensive questionnaire sets in user studies [87], which makes the length of each questionnaire a crucial factor. While it may not take much time to answer the 24 Hexad items in isolation, that time may add up significantly if participants have to answer other questionnaires in the same session. Therefore, a decrease of even one or two minutes to complete the questionnaire is potentially beneficial. Bansak et al. [12] showed that *data quality and completion rates decrease significantly* with longer questionnaires, which may be explained by a lower cognitive load on the participants [80]. Therefore, developing and validating short versions of scales used in HCI research, such as the player experience inventory [32] or the game user experience satisfaction scale [40], has recently gained importance. (2) Second, the length of the Hexad questionnaire may *prevent a wider uptake in the industry*. In industry settings, iteration cycles of UX design are typically rapid and a short turnaround time is vital. The relevance of a short but valid instrument in this context is further supported by Andrzej Marczewski, stating that "[...] a shorter questionnaire, if provably as accurate as the standard questionnaire, will simplify the process of using the HEXAD in predictive analysis for the success of gamified designs making it more likely to be used outside the core field of gamification professionals and academics."[1] (3) Also, long questionnaires can be particularly *cumbersome to fill out on mobile devices*, which have become the most prominent mode of accessing the web as a result of "mobile-first" development paradigms. (4) Lastly, dynamic personalization of gamified systems demands *less invasive instruments* [10] to not interrupt immersion and compromise the gameful experience.

To address these issues and enable efficient use of the Hexad questionnaire, the main goal of this work is to *develop and empirically validate a short version of the Hexad questionnaire*. First, we analyzed an existing dataset (n = 882) and selected two items for each Hexad user type from the existing 24-item scale ("Hexad-24") based on an exploratory factor analysis, reducing the set of items to twelve. In a second survey study (n = 1101), we investigated the psychometric properties of the short version of the Hexad questionnaire ("Hexad-12") by conducting a confirmatory factor analysis.

We present three main results that contribute to HCI research. First, our results show that the *Hexad-12 has good psychometric properties*, meaning good model fit and acceptable reliability as well as convergent and discriminant validity. Second, comparing the Hexad-12 with the Hexad-24, we found that the *short version is advantageous with respect to all of the above properties*. Finally, using bivariate and canonical correlation analyses, we show that the *Hexad-12 represents the Hexad-24 exceptionally well*. Thus, we contribute a concise instrument for assessing Hexad user types for researchers and practitioners that meets established criteria for reliability and validity and addresses existing problems with the original Hexad scale identified in previous research (e.g., model fit, reliability of specific factors, and discriminant and convergent validity [47, 59, 84]). Thus, the Hexad-12 can facilitate using the Hexad model for statically or dynamically adapting gamified systems in research and practice, by providing an *improved* and *faster* assessment tool compared to the Hexad-24 that mitigates the detrimental effects of long questionnaires on player experience.

2 RELATED WORK

In this section, we discuss related work from tailored gamification, focusing on prior work using the Hexad model.

2.1 Tailored Gamification

In line with the trend of increasingly focusing on how and why gamification works, the investigation of adaptive approaches to gamification design is one of the major directions in the current scientific debate about gamification [73, 74]. Under the terms personalized, adaptive, or tailored gamification, previous studies have examined different approaches to modify aspects of gamification with appropriate solutions to meet specific user needs [31]. These include personalization based on gender, personality traits, age, behavior, culture, different user motivations (utilitarian, hedonic, or social) for using gamification services [35, 44], or goal orientations and their association with different game elements [11, 33].

However, the most commonly used approach to personalizing gamification is player typologies [44]. There are a variety of typologies that classify players based on their psychographic characteristics, behaviors, motivations, or needs, and that share common dimensions such as achievement, sociability, exploration, domination, and immersion [37]. The most popular typologies that have been used to personalize gamification [44] are Bartle's player types [13], the Gamification User Types Hexad model [53], and BrainHex archetypes [57]. For instance, Akasaki et al. [1] investigated whether the perception of gamification elements differs across Bartle's player types in the context of a sharing economy service. They found that Achievers and Killers preferred collecting and badges, while Explorers preferred collecting and narratives. Regarding BrainHex, Lavoué et al. [48] conducted a study in which gamification elements were adapted to users based on BrainHex player types in a web-based learning environment teaching French spelling and grammar. The gamification elements were randomly assigned to one of three conditions (adapted gamification elements,

[1] https://bit.ly/3PuOTwd, last accessed March 10, 2023

counter-adapted gamification elements, and no gamification elements). They found that among users who used the platform on a regular basis, those receiving adapted gamification elements spent significantly more time on the platform and participants receiving counter-adapted gamification elements reported higher levels of amotivation.

Although both Bartle's and BrainHex's typologies have been used in past research for personalization purposes, they have several drawbacks that make them less suitable as a basis for personalizing gamified systems. Bartle's typology is based on the motivations and preferences of multi-user dungeon players. This limits its generalizability to other games and gamification [14], especially because users might experience game elements differently in a non-game context than in games [85]. The lack of empirical validation of the model is also a concern, as it jeopardizes the use of the model for scientific purposes [14, 19]. As for BrainHex, researchers found problems regarding its psychometric properties [20, 83]. Busch et al. [19] found that only two types—Socializer and Achiever—could be discriminated. Also, when using the BrainHex model to predict the game experience, no significant predictions could be found [20].

Therefore, a user typology that specifically targets gamified systems (rather than games) and has a solid empirical basis is needed for personalization purposes. The Hexad user types model meets this need for several reasons. First, it is the only model that targets gamification [60]. Second, there is an instrument for assessing Hexad user types that has been empirically validated [84]. Third, Hexad user types have been shown to be an appropriate and reliable factor to explain preferences for gamification elements [47, 86] in several domains, including physical activity [6, 21], healthy nutrition [9, 60], energy consumption [18, 45], warehouse management [62], and education [49, 56]. Finally, the Hexad model has been shown to have advantages over the use of personality traits or BrainHex player types in explaining preferences for gamification elements [34].

2.2 The Gamification User Types Hexad

In light of the limitations of applying previous player typologies from games research to the gamification context, the gamification user type Hexad typology was developed explicitly for gamification [53, 60] and builds on insights from Pink's four drives theory [63] and self-determination theory [68]. Accordingly, Marczewski [53] distinguishes between six user types that differ in terms of their need for autonomy, relatedness, competence, and purpose. *Philanthropists* are primarily purpose-driven, are considered altruistic, and want to support other users, while *Socializers* primarily seek relatedness and interaction with others. *Achievers* are driven primarily by competence needs and striving to improve in the face of challenges, while *Free Spirits* prefer autonomy and freedom to create and explore. In addition to these intrinsically motivated types, the *Player* type is described as a primarily extrinsically motivated user type, i.e., *Players* seek rewards for their actions. The last type, the *Disruptor*, is characterized by a lack of motivation to use the system and is mainly concerned with testing the boundaries of the system [53].

Based on this primarily conceptual typology, researchers have made efforts to develop and validate a reliable scientific instrument to capture the different Hexad user types so that the Hexad typology can be used for tailored gamification design. Tondello et al. [86] were the first to systematically construct and develop a scale to measure the six Hexad types based on collaborative expert workshops [25] to generate questions, followed by a quantitative factor analysis study with 133 participants (mostly graduate and undergraduate students) resulting in a scale with 24 items, four for each of the Hexad types. This preliminary version has been used in many studies to date and was translated into Turkish by Akgün and Topal [2] in 2018 (N=452, freshmen students). In a subsequent validation that included three studies, Tondello et al. [84] modified the original Hexad scale, particularly with regard to the Free Spirit and Achiever items, which improved the factor loadings compared to the first version. In the first study, they considered 196 participants for the English validation, in the second study, 1,073, and in the third study, 152 participants. This final scale has so far been translated and validated by Taşkın and Çakmak in Turkish [81] (N= 330, university students), by Ooge et al. in Dutch [59] (N= 293, adolescents), by Krath and von Korflesch in German [47] (N= 380, mostly between 21 and 30 years old), by Tondello et al. [84] (N= 360, in the first study, N= 255 in the second study, mostly between 18 and 39 years old) and Manzano-León et al. [52] (N= 1, 345, adolescents) in Spanish, and by Santos et al. [69] (N= 421, between 10 and 60 years old) in Brazilian Portuguese.

These efforts paved the way for a variety of studies building on Hexad types to assess the gamification element preferences of different user types in gamified systems and design tailored gamification accordingly. Self-reported differences between different Hexad types have been repeatedly found in terms of preferred gamification elements [47, 50]. In the first validation study of the Hexad questionnaire, Tondello et al. [86] found significant relationships between *Socializers* and elements such as teams, social networks, social comparison, and competition, while *Free Spirits* preferred exploratory tasks, nonlinear gameplay, Easter eggs, unlockable content, creativity tools, and customization. *Achievers*, on the other hand, liked challenges, certificates, learning, progress, and quests, while *Players* were mainly motivated by points, rewards, leaderboards, and badges [86]. Later, Mora et al. [55] confirmed that *Socializers* and *Philanthropists* found teams to be motivating, and *Players* especially preferred leaderboards. However, their results also showed that almost all user types desired challenges and that *Socializers* and *Philanthropists* preferred exploratory tasks more than *Free Spirits* [55]. In the context of fitness systems, Altmeyer et al. [7] underlined that *Socializers* particularly preferred social collaboration, while *Philanthropists* were keener on knowledge sharing. Also, *Achievers* preferred several elements related to goal-setting (custom goals, personalized goals, challenges, points), whereas *Disruptors* only liked one element: cheating. In a recent large-scale study, Krath and von Korflesch [47] replicated the study design of Tondello et al. [86] and identified significant relationships between *Philanthropists* and gifts, knowledge sharing, teams, and administrative tasks; *Socializers* and all social elements such as teams, social networks, competition, social comparison, and social discovery; *Free Spirits* and creativity tools, exploratory tasks, challenges, and learning; and *Disruptors* and anarchic gameplay and

innovation platforms. *Achievers* and *Players*, in turn, liked a wide variance of different game elements, with *Achievers* more involved with intrinsically motivating progression elements (competition, leaderboards, challenges, learning, levels) and Players more motivated by extrinsic rewards for progress (rewards, achievements, points, leaderboards, competition, certificates) [47] .

In addition, studies have examined the relationships between Hexad types and appropriate persuasion strategies [18, 60]. For example, Orji et al. [60] found that *Philanthropists* were best persuaded by simulation, while competition and reward strategies were better suited for *Players*. *Socializers* reported high persuasiveness across all persuasive strategies studied, while *Achievers* and *Free Spirits* did not show significant associations with any of the strategies, which is similar to findings from the study of Böckle et al. [18]. *Disruptors*, in particular, showed many negative responses to a variety of persuasive strategies, underscoring their nature as rebellious types who are usually not motivated to use the system at all [60]. Although the results of these studies differed to some extent with respect to specific preferences, the general relevance of Hexad types to perceptions of gamification design, and thus the value of considering them in a tailored gamification design, was confirmed in these previous efforts.

In fact, the Hexad typology has been shown to be an appropriate approach for tailoring gamification design that supports desirable psychological and behavioral outcomes and outperforms previous typologies [34]. For example, using the Hexad typology to personalize gamification designs increased affective experiences [6], motivation, and satisfaction [92] in the fitness context. In an experimental design with physical tasks, Lopez and Tucker [51] showed that participants performed better in an adapted gamification design based on Hexad types than in a non-adapted gamification design. In addition, participants who were exposed to counter-adapted gamification in terms of their Hexad types performed worse than the other groups [51]. Similarly, Passalacqua et al. [62] studied the effects of tailored gamification design based on Hexad types in a warehouse management environment and found that personalized gamification design significantly outperformed the general design in terms of task completion time and errors.

Hence, the Hexad typology constitutes a useful basis for tailored gamification design not only in research studies but also in practice. Actually, the original intent of the Hexad typology was to help gamification designers think about the types of people who might use their system and thereby assist in considering gamification features in design decisions that appeal to each of the different types [25, 53].

However, previous studies have also pointed out the limitations of the scale used for analysis. Personalization in its current form requires users to complete a long questionnaire with 24 items, which can interrupt immersion and player experience [10]. Also, dynamic personalization during use is limited, which poses a particular challenge because user types can be considered dynamic and change over time [67, 70]. Therefore, researchers have experimented with new approaches to identify Hexad types. For example, by predicting Hexad types from smartphone data [8] or mobile banking data [42]. In a recent study, Altmeyer et al. [10] used a gameful application called "cloud clicker", with short statements about each user type to avoid the long questionnaire. Although these approaches were very promising, their applicability to other studies and contexts is rather limited and less suitable for scientific purposes—which is why a short, validated version of the Hexad scale that is as universally applicable as the original Hexad scale would be of great value for the successful personalization of gamified systems.

3 FIRST STUDY: IDENTIFICATION OF ITEMS FOR THE HEXAD-12

The aim of the first study was to identify suitable items for a short version of the 24-item validated Gamification User Types Hexad questionnaire in English (Hexad-24) [84]. To do this, we merged existing datasets from two previous studies (see Altmeyer et al. [10], Krath and von Korflesch [47]) that used the Hexad-24 and conducted an exploratory factor analysis and scale reliability analyses to identify appropriate items of each scale. Following previous developments of short versions of scales, such as the 10-item short version of the Big Five Personality Inventory [64] or the short version of the User Experience Questionnaire[4], we attempted to reduce each scale to half of its items, resulting in a 12-item version (Hexad-12).

3.1 Procedure

In both previous studies, online surveys were conducted in which participants were asked to respond to the final validated version of the Hexad-24 by Tondello et al. [84] after giving informed consent and providing demographic data such as age and gender. The questionnaire consisted of 24 items, four for each of the six user types, which were rated on a 7-point Likert scale from "strongly disagree" to "strongly agree" (see Table 15 in the appendix for all items). Hexad user type scores were calculated as the sum of the four respective items of each scale. In [10], the online platform Prolific was used to solicit participants, who received compensation of £2 GBP. In [47], the survey was advertised on Facebook and participants received no compensation other than the display of their own Hexad type at the end of the survey.

To determine suitability for merging, we reviewed the demographic characteristics of both samples. The age distribution was quite similar (*M = 33.20, SD = 11.60, MD = 30.00* in the sample of [10] and *M = 28.00, SD = 7.50, MD = 28.00* in the sample of [47]), whereas the gender distribution was more balanced in the sample of [10] (49. 7% female, 48.4% male, 0.02% other) than in the sample of [47] (22.6% female, 55.6% male, 0.03% other, 18.7% did not want to answer the question). However, the distribution of Hexad user types was also quite similar (Achiever type had the highest average scores, followed by Philanthropist, Free Spirit, and Player in the middle, and Socializer and Disruptor with the lowest scores, as shown in Table 1). Therefore, we checked whether gender had a significant effect on Hexad types in either sample using one-way ANOVAs with gender as the grouping variable and Hexad types as the dependent variable in both datasets. Levene's test was not significant ($p > .05$) for all user types in either data set, so variance homogeneity was met as a requirement for the ANOVA. Yet, the Shapiro-Wilk test was significant for all user types ($p < .01$) except Disruptor ($p = .19$) in the sample of [10] and for all user types ($p < .05$) in the sample of [47]. Because ANOVA is relatively robust to non-normality as long as the variances are homogeneous [15], we decided to conduct

the ANOVA nevertheless and found that the effect of gender was non-significant for all user types in the sample of [10] and for all user types except Achiever and Disruptor in the sample of [47] (please refer to Table 16 in the appendix to see the results of the ANOVA for both samples). Therefore, we did not consider the effect of the difference in gender distribution to be a critical factor militating against merging the datasets. To create a merged dataset, we selected all responses that completed the questionnaire in English and merged them in random order.

3.2 Participants

In total, the summary data set consisted of 882 participants (153 from [10] and 729 from [47]). Of them, 54.3% reported themselves as male, 27.3% as female, 2.9% as other than male or female, and 15.5% preferred not to answer this question. The mean age was *M= 29.00, SD = 8.68, MD = 28.00*. In the distribution of Hexad user types, the Achiever type had the highest mean scores (*M = 23.8, SD = 3.47*), followed in descending order by Philanthropist (*M = 23.3, SD = 3.57*), Free Spirit (*M = 22.7, SD = 3.51*), Player (*M = 21.8, SD = 4.16*), Socializer (*M = 19.8, SD = 5.34*), and Disruptor as the least represented type (*M = 15.8, SD = 4.84*).

3.3 Results

The following section presents the results of the first study in terms of identifying appropriate items for a 12-item short version of the Hexad-24.

3.3.1 Internal Reliability and Correlations. In the first step, we checked the internal reliability of the six sub-scales to investigate which scales were already working well in the Hexad-24 and which scales were still causing problems, as problems with individual scales were repeatedly found in previous studies that validated the Hexad-24 in multiple languages [59, 84, 86]. In addition, we examined how much each item contributes to scale reliability in order to identify problematic items that might be better omitted in a short version of the Hexad-24. As can be seen in Table 2, scale reliability is acceptable ($\alpha \geq 0.7$) [54] for the Philanthropist, Socializer, and Achiever scales and is slightly below threshold for the Player and Disruptor scales, while the Free Spirit scale causes the most problems. From the detailed analysis, it appears that items R3, R1, D1, F2, and F4, in particular, do not contribute highly to the reliability of the respective scales (as reliability remains relatively high or even increases when they are dropped) and thus may be problematic items that would be better left out of the short version.

In order to test whether there are intercorrelations between Hexad types that were common in previous validation studies [84, 86], and thus to determine whether oblique rotation is required in the subsequent exploratory factor analysis [26], we used Kendall's τ_b correlation because of the non-parametric Likert scales of the Hexad questionnaire. As shown in Table 3, we found partial overlap between Hexad types, consistent with the results of the original validation study [84]. In particular, we found a large overlap between the Philanthropist scale and the Socializer scale (τ_b = .36), and medium overlaps between the Achiever scale and the Player (τ_b = . 29), Free Spirit (τ_b = .29), and Philanthropist (τ_b = .25) scales, and the Free Spirit scale and the Disruptor scale (τ_b = .28).

3.3.2 Exploratory Factor Analysis. Following the reliability and correlation analyses, we conducted an exploratory factor analysis to identify the most appropriate items of each scale. The Kaiser-Mayer-Olkin test (KMO = 0.85) and Bartlett's test for sphericity (χ^2 = 6516, $p < .001$) demonstrate the suitability of the data for factor analysis [79]. For factor extraction, we used the maximum likelihood method in combination with an oblique promax rotation due to the intercorrelations of Hexad types [26], as reported in the previous section, forcing an extraction of six factors. Table 4 shows the factor loadings of the individual items, with the two highest loading items marked in bold. As can be seen, two items loading most strongly on each of the Socializer (s4 & s2), Philanthropist (p4 & p1), Achiever (a2 & a4), and Player (r4 & r2) scales, respectively, are relatively easy to identify, with the other items loading significantly lower on the factor or even loading on other factors (such as s1 & s3). However, item f2 loads on its own factor and causes the other items of the Free Spirit and Disruptor scales to converge on one factor, on which two Disruptor items (d3 & d4) load highest.

3.3.3 Identification of Items. Based on the previous analyses, we created a short version of the Hexad-24. In line with prior developments of short versions of scales, such as the 10-item short version of the Big Five Personality Inventory [64] or the short version of the User Experience Questionnaire[4], we aimed to reduce each scale to half of its items. In general, there are several approaches that can be used to identify appropriate items for short scales [89]. On the one hand, selection can be based on purely statistical approaches, such as factor analyses (e.g., [4, 41, 76]), and on the other hand, combinations of statistical approaches with theoretical aspects can be used to retain facets of a given construct [89]. Because some of the Hexad user types are determined from a theoretical point of view by the expression of two different facets (e.g., the Achiever type is assumed to be interested in both improving skills and overcoming obstacles [84]), we decided to consider both statistical and theoretical aspects when selecting the items.

Thus, for item selection, we were guided by the following criteria:

Factor loadings. We considered factor loadings from the previous exploratory factor analysis (see Table 4) to determine which items loaded strongly on each factor and consequently should be selected to represent the factor in the short version. At the same time, low factor loadings or cross-loadings on other factors indicated lower suitability for representing the factor, so we considered them as exclusion criteria.

Contribution to scale reliability. We used scale reliability assessment (see Table 2) as an indicator to determine how much a particular item contributes to the overall reliability of the respective Hexad scale. If the reliability would decrease significantly if an item was omitted, we took this as an indication that the item should be retained, whereas a small decrease in reliability or even an increase in reliability (as in the case of item d1) due to omission indicated exclusion.

Theoretical facets of the Hexad types. We considered content aspects of the items to select two items that are not highly redundant in meaning and thus measure different core aspects of each Hexad user type (e.g., p1 "It makes me happy if I am able to help others" and p2 "I like helping others to orient themselves in new situations" [84]

Table 1: Distribution of age and Hexad user types in the two datasets

Source	Philanthropist	Socializer	Achiever	Player	Free Spirit	Disruptor	Age
Mean [10]	22.80	18.70	23.50	22.70	22.30	14.80	33.20
Mean [47]	23.50	20.10	23.90	21.70	22.80	16.00	28.00
SD [10]	3.72	5.13	3.52	3.96	4.00	4.71	11.60
SD [47]	3.53	5.35	3.45	4.18	3.40	4.84	7.50

Table 2: Internal reliability scores for each Hexad user type scale in total and if individual items were omitted. Bold entries mark problematic items of scales that fall below the acceptable threshold ($\alpha \geq$ 0.7).

Philantropist		Socializer		Achiever		Player		Disruptor		Free Spirit	
Cronbach's α		Cronbach's α		Cronbach's α		Cronbach's α		Cronbach's α		Cronbach's α	
$\sum$	**.75**	$\sum$	**.85**	$\sum$	**.77**	$\sum$	.68	$\sum$	.65	$\sum$	.59
	if omitted		if omitted		if omitted		if omitted		if omitted		if omitted
p2	.66	s2	.78	a2	.67	r2	.56	d3	.49	f1	.47
p1	.68	s1	.79	a1	.71	r4	.57	d4	.57	f3	.48
p3	.71	s4	.81	a4	.73	**r1**	**.65**	d2	.59	**f4**	**.55**
p4	.72	s3	.84	a3	.75	**r3**	**.67**	**d1**	**.67**	**f2**	**.57**

Table 3: Bivatriate correlation coefficients (Kendall's τ_b) between the Hexad types (p < .001). Bold entries mark values $\geq$ .25.**

Hexad Type	Philanthropist	Socializer	Free Spirit	Achiever	Player
Socializer	**.36****				
Free Spirit	.18**	.05			
Achiever	**.25****	.20**	**.29****		
Player	.10**	.16**	.13**	**.29****	
Disruptor	-.01	.04	**.28****	.13**	.08**

both refer to the willingness to help others, whereas p4 "The well-being of others is important to me" [84] refers to empathy toward others and thus reflects a different aspect of the Philanthropist user type, so we were careful not to select both item p1 and p2 for the short version).

Consequently, Table 6 lists all items and the explanation for their inclusion and exclusion in the short version of the Hexad scale. Based on the statistical and theoretical criteria, the items selected for the Philanthropist, Socializer, Achiever, Player, and Disruptor scales were congruent with the highest loading items highlighted in Table 4. Only the Free Spirit scale was an challenging exception, as the problematic item f2 caused an entirely new factor that resulted in the other items loading on the Disruptor factor instead. In addition, item f4 showed cross-loadings with the Philanthropist factor and, similar to f2, was problematic in terms of scale reliability (see Table 2), so we decided to exclude f2 and f4 for the short version, although we made the trade-off of losing the theoretical facet of self-expression by excluding f3, a limitation reflected in our limitations section.

In this way, we developed our 12-item short version of the Hexad scale (Hexad-12) as shown in Table 5.

4 SECOND STUDY: CONFIRMATORY ANALYSIS OF THE HEXAD-12 ON A NEW DATASET

Following the first study and the resulting exploratory identification of suitable items for a 12-item short version of the Hexad scale, we tested the suitability and potential of the short version by conducting a confirmatory factor analysis on a new data set. To ensure that any effects found were not solely due to the different data set, we chose to ask participants to complete all 24 items of the Hexad scale in order to compare reliability, validity, and model fit between the Hexad-24 and the Hexad-12 in the new sample.

We assessed model fit by conducting a confirmatory factor analysis and inspecting the results of the chi-square test, complemented by inspecting established model fit indices (as the chi-square test is influenced by sample size), such as the root mean square error of approximation (with a cutoff at <.06 [27]), the comparative fit index, and the Tucker-Lewis index (both with a cutoff at >.95 [39]). We evaluated internal reliability by inspecting Cronbach's α (with a cutoff at >.70 [54]). We assessed convergent validity by looking at the composite reliability (with a cutoff at >.70 [91]) and average variance extracted (with a cutoff at >.50 [30]). Lastly, we assessed discriminant validity by checking whether the average variance extracted is higher than the shared variance for each factor [28].

Table 4: Rotated factor loadings for the Hexad survey items (cutoff = 0.3). Bold entries mark the two highest loading items on each factor.

Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
s4	**.84**					
s2	**.78**					
s1	.64	.33				
s3	.42	.43				
p4		**.70**				
p1		**.66**				
p2		.60				
p3		.50				
a2			**1.08**			
a4			**.62**			
a1			.55			.31
a3			.49			
d3				**.79**		
d4				**.62**		
d2				.56		
f4		.32		.40		
d1				.40		
f3	-.31			.38		
f1				.36		
r4					**.82**	
r2					**.77**	
r3					.44	
r1					.44	
f2						**.46**

Table 5: 12-item short version of the Hexad scale (Hexad-12)

Hexad type	Item	English Version (based on [84])
Philanthropist	p1	It makes me happy if I am able to help others.
	p4	The well-being of others is important to me.
Socializer	s2	I like being part of a team.
	s4	I enjoy group activities.
Achiever	a2	I like mastering difficult tasks.
	a4	I enjoy emerging victorious out of difficult circumstances.
Player	r4	If the reward is sufficient, I will put in the effort.
	r2	Rewards are a great way to motivate me.
Free Spirit	f1	It is important to me to follow my own path.
	f3	Being independent is important to me.
Disruptor	d3	I see myself as a rebel.
	d4	I dislike following rules.

We have provided the collected dataset as supplementary material to facilitate replication and to allow fellow researchers to build upon our findings.

4.1 Procedure

To obtain a diverse sample of participants, we recruited participants from both Facebook (without compensation) and Prolific (with compensation of £0.50 GBP for the task, which corresponded to an hourly rate of approximately £10 GBP), focusing on native English speakers to avoid misunderstandings due to language proficiency. After providing informed consent and demographic information on age, gender, and nationality, they completed an online survey consisting of the same questionnaire used in the surveys of the first study [10, 47], i.e., the final version of the Hexad-24, with four items for each of the six Hexad types, rated on a seven-point Likert scale from "strongly disagree" to "strongly agree" by Tondello et al. [84] (see Table 15 in the appendix for all items).

Table 6: Exclusion and inclusion criteria of items for the Hexad-12

HEXAD type	Item	Criteria for item selection	Decision
Philanthropist	p1	High factor loading (> .60), high contribution to reliability (drop: .07), theoretical facet: willingness to help others	*Inclusion* due to factor loading, reliability contribution and content aspect
	p4	High factor loading (> .60), low contribution to reliability (drop < .05), theoretical facet: empathy	*Inclusion* due factor loading and content aspect not captured by the other Philanthropist items
	p2	Medium factor loading (> .50), high contribution to reliability (drop: .09), theoretical facet: willingness to help	*Exclusion* due to content similarity with p1 and comparatively lower factor loading
	p3	Low factor loading (≤ .50), low contribution to reliability (drop < .05), theoretical facet: willingness to help	*Exclusion* due to low factor loading, low reliability contribution and content similarity to p1
Socializer	s2	High factor loading (> .60), high contribution to reliability (drop: .07), theoretical facet: belonging	*Inclusion* due to factor loading, reliability contribution and content aspect
	s4	High factor loading (> .60), low contribution to reliability (drop < .05), theoretical facet: interaction	*Inclusion* due to factor loading and content aspect with comparatively higher loading than s1 and without cross loading
	s1	High factor loading (> .60), cross-loading on a different factor, high contribution to reliability (drop: .06), theoretical facet: interaction	*Exclusion* due to cross loading on different factor and content similarity with s4, which has no cross loading
	s3	Low factor loading (≤ .50), cross-loading on different factor, low contribution to reliability (drop < .05), theoretical facet: belonging	*Exclusion* due to low factor loading, cross-loading on different factor, low reliability contribution and content similarity to s2
Achiever	a2	High factor loading (> .60), high contribution to reliability (drop: .10), theoretical facet: skill improvement	*Inclusion* due to factor loading, reliability contribution and content aspect
	a4	High factor loading (> .60), low contribution to reliability (drop < .05), theoretical facet: overcoming obstacles	*Inclusion* due to factor loading and content aspect with comparatively higher loading than a1 and without cross loading
	a1	Medium factor loading (> .50), cross-loading on a different factor, high contribution to reliability (drop: .06), theoretical facet: overcoming obstacles	*Exclusion* due to cross loading on different factor and content similarity to a4 with comparatively lower factor loading
	a3	Low factor loading (≤ .50), low contribution to reliability (drop < .05), theoretical facet: skill improvement	*Exclusion* due to low factor loading, low reliability contribution and content similarity to a2
Player	r4	High factor loading (> .60), high contribution to reliability (drop: .11), theoretical facet: cost-benefit ratio	*Inclusion* due to factor loading, reliability contribution and content aspect
	r2	High factor loading (> .60), high contribution to reliability (drop: .12), theoretical facet: rewards	*Inclusion* due to factor loading, reliability contribution and content aspect
	r3	Low factor loading (≤ .50), low contribution to reliability (drop < .05), theoretical facet: cost-benefit ratio	*Exclusion* due to low factor loading, low reliability contribution and content similarity to r4
	r1	Low factor loading(≤ .50), low contribution to reliability (drop < .05), theoretical facet: rewards	*Exclusion* due to low factor loading, low reliability contribution and content similarity to r2
Free Spirit	f1	Factor loading not determinable, high contribution to reliability (drop: .12), theoretical facet: independence	*Inclusion* due to reliability contribution and content aspect
	f3	Factor loading not determinable, negative cross loading on different factor, high contribution to reliability (drop: .11), theoretical facet: independence	*Inclusion* due to reliability contribution despite the content similarity to f1
	f4	Factor loading not determinable, cross-loading on different factors, low contribution to reliability (drop < .05), theoretical facet: self-expression	*Exclusion* due to cross loading on different factors and low reliability contribution
	f2	Loads only on an individual factor of its own, low contribution to reliability (drop < .05), theoretical facet: curiosity	*Exclusion* due to missing relation with the other Free Spirit items, causing an own factor, and low reliability contribution
Disruptor	d3	High factor loading (> .60), high contribution to reliability (drop: .16), theoretical facet: provocation	*Inclusion* due to factor loading, reliability contribution and content aspect
	d4	High factor loading (> .60), high contribution to reliability (drop: .08), theoretical facet: disobedience	*Inclusion* due to factor loading, reliability contribution and content aspect
	d2	Medium factor loading (> .50), high contribution to reliability (drop: .06), theoretical facet: disobedience	*Exclusion* due to content similarity to d4 with comparatively lower factor loading
	d1	Low factor loading (≤ .50), low contribution to reliability (increase: .02), theoretical facet: provocation	*Exclusion* due to low factor loading, low reliability contribution and content similarity to d3

Hexad user type scores were calculated as the sum of the four respective items of each scale. Participants were encouraged to complete the survey honestly to reveal their personal Hexad user type at the end of the survey.

4.2 Participants

Altogether, the final data set consisted of 1,101 participants (500 from Prolific and 601 from Facebook), of whom 42.2% identified themselves as male and 35.9% as female, 3.5% indicated a gender other than male or female, and 18.4% did not respond to this question. The mean age was *M=31.70, SD=12.60, MD=28.00*, so overall the gender and age distribution is comparable to the data set of the first study, with women slightly more represented in the sample. In terms of nationality, most participants were from the United Kingdom (n=451), followed by Australia (n=159), Canada (n=149), the United States (n=134), New Zealand (n=56), and Ireland (n=23). The remaining 118 participants hailed from a variety of countries (n=26) or preferred not to answer this question (n=92). Regarding the distribution of Hexad user types, Philanthropist had the highest mean scores (*M = 23.3, SD = 3.48*), followed by Achiever (*M = 23.2, SD = 3.45*), Player (*M = 22.6, SD = 3.74*), Free Spirit (*M = 22.2, SD = 3. 39*), Socializer (*M = 19.2, SD = 5.01*), and Disruptor (*M = 15.3, SD = 4.66*), which is consistent with the Hexad distribution of the dataset from the first study, except for Player achieving a higher mean than in the first sample.

4.3 Results

In the following, we report our results regarding confirmatory factor analysis and model fit of the Hexad-12, internal scale reliability, and convergent as well as discriminant validity of our short scale.

4.3.1 Confirmatory Factor Analysis. To test whether the items in the Hexad-12 adequately explain the variance in the six Hexad user types and how well the data fit the hypothesized model, we conducted a confirmatory factor analysis using the maximum likelihood method. The six Hexad user types were modeled as factors, with the corresponding two items as indicators. Table 7 shows the standardized (β) and unstandardized (B) estimates, as well as standard errors (SE) for each of the scale items. Regarding model fit, the chi-square test ($\chi^2(39) = 107, p < .001$) indicates an acceptable fit ($\frac{\chi^2}{df} = \frac{107}{39} = 2.74 < 3$ [72]). Since the chi-square test is sensitive to sample size (with increasing sample size, the χ^2 value increases), we analyzed model fit indices to bypass this sample size issue inherent in the χ^2 test. The fit index of the root mean square error of approximation (RMSEA = .04) indicates good model fit, as it is less than .06 [27], as do the comparative fit index (CFI = .98) and the Tucker-Lewis index (TLI = .97), which are greater than .95 and thus represent good model fit [39].

4.3.2 Internal Reliability. Second, we examined the internal reliability of the new 2-item subscales of each Hexad user type, to see to what extent shortening the original scale affected scale reliability. As can be seen in Table 8, scale reliability is acceptable for the Philanthropist, Socializer, Player, Free Spirit, and Disruptor subscales ($\alpha \geq 0.7$) [54] and slightly below the threshold for the Achiever ($\alpha = .67$) scale. Thus, although we observe a slight decrease in reliability for the Achiever scale, we see a significant increase in reliability for the previously problematic Player, Disruptor, and Free Spirit scales (see Table 2 for the reliability of the scales in the Hexad-24 in the first study and Table 8 for the reliability of the scales in the Hexad-24 in the second study).

4.3.3 Convergent and Discriminant Validity. Finally, we analyzed 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) of the Hexad-12 constructs. We analyzed convergent validity by looking at composite reliability (CR), with a recommended criterion of 0.7 [91], and the average variance extracted (AVE), with a recommended criterion of 0.5 [30]. We assessed discriminant validity by comparing the AVE with the shared variance between the user types to check whether the AVE of each user type is higher than the explained variance of any of the remaining user types.

The CR estimates were above the recommended criterion of 0.7 [91] for all user types but the Achiever and Free Spirit (see Table 8). For these two user types, the CR estimates fell slightly below 0.7 (.68 for the Achiever and .69 for the Free Spirit). However, as can be seen in Table 8, the AVE estimates of all user types exceed the recommended threshold of 0.5. Based on these results, we conclude that convergent validity is adequate, with slight room for improvement in the CR estimates of Achiever and Free Spirit.

In terms of discriminant validity, as shown in Table 11, the AVE is higher than the shared variance for each of the six user types. Thus, the explanatory power of the items of each factor is always higher than the explanatory power of these items on different factors. This means that the factors are sufficiently distinct so that our data shows discriminant validity for all user types [28].

4.4 Comparison between the Hexad-12 and the Hexad-24

To conclude our analysis, we compared Hexad-12 to the existing Hexad-24 regarding the general model fit, reliability, as well as convergent and discriminant validity. We also investigated how well the Hexad-12 scales generalize to the Hexad-24 scales they were developed to represent.

4.4.1 Model Fit. Regarding general model fit, we found that the Hexad-12 outperforms the Hexad-24 on all measures (see Table 10). The $\frac{\chi^2}{df}$ value is smaller in the 12-item version, which indicates that the Hexad-12 has a better model fit than the Hexad-24 [3]. When looking at the fit indices (RMSEA, CFI, TLI), this result is supported. While the Hexad-12 meets the cutoff of .06 for the RMSEA [27], the Hexad-24 does not meet this criterion. Similarly, the Hexad-12 exceeds the .95 criterion for both CFI and TLI [39], while the Hexad-24 has much lower values, not meeting this criterion.

4.4.2 Internal Reliability. Regarding internal reliability, the Hexad-12 has good reliability on five scales ($\alpha \geq 0.7$) [54], while the Hexad-24 meets the cutoff value of 0.7 only on three of its factors (see Table 9). Interestingly, Cronbach's α decreased slightly in the Hexad-12 on those factors, which had the highest Cronbach's α values in the Hexad-24 (Philanthropist, Socializer, and Achiever), while it increased on factors which were problematic in the original scale (Player, Disruptor, and Free Spirit).

Table 7: Estimated factor loadings as well as Covariance ("Cov.") and standardized Covarianze ("St. Cov.") for survey items of the Hexad-12

Factor	Indicator	B	β	SE	Cov.	St. Cov.	SE
Philanthropist	p1	.80	.83	.04	.30	.32	.04
	p4	.87	.72	.04	.72	.49	.06
Socializer	s2	1.35	.90	.05	.44	.20	.10
	s4	1.20	.75	.05	1.16	.44	.09
Achiever	a2	.82	.68	.04	.80	.54	.06
	a4	.82	.75	.04	.51	.43	.05
Player	r2	.85	.76	.06	.54	.42	.09
	r4	.83	.71	.06	.68	.50	.08
Free Spirit	f1	.96	.79	.05	.54	.37	.07
	f3	.83	.67	.04	.83	.55	.06
Disruptor	d3	1.23	.70	.07	1.59	.51	.15
	d4	1.32	.78	.07	1.10	.39	.16

Table 8: Internal reliability, Composite Reliability (CR) estimates, and Average Variance Extracted (AVE) for each Hexad-12 scale. Bold entries mark acceptable values.

User type	Philantropist	Socializer	Achiever	Player	Disruptor	Free Spirit
Cronbach's α	**.73**	**.80**	.67	**.70**	**.71**	**.70**
CR estimate	**.74**	**.80**	.68	**.70**	**.71**	.69
AVE	**.60**	**.69**	**.51**	**.54**	**.55**	**.54**

Table 9: Internal reliability scores, CR and AVE for the Hexad-12 and Hexad-24. Bold entries mark acceptable values.

User type	Philantropist	Socializer	Achiever	Player	Disruptor	Free Spirit
Cronbach's α Hexad-12	**.73**	**.80**	.67	**.70**	**.71**	**.70**
Cronbach's α Hexad-24	**.77**	**.83**	**.75**	.68	.66	.60
CR estimate Hexad-12	**.74**	**.80**	.68	**.70**	**.71**	.69
CR estimate Hexad-24	**.78**	**.83**	**.76**	**.70**	.68	.63
AVE Hexad-12	**.60**	**.69**	**.51**	**.54**	**.55**	**.54**
AVE Hexad-24	.47	**.55**	.44	.39	.36	.30

4.4.3 Convergent and Discriminant Validity. In the Hexad-24, the CR estimates for all user types except the Disruptor and the Free Spirit were above the recommended criterion of 0.7 [91]. While the Disruptor is only slightly below the threshold of 0.7, the Free Spirit is more clearly below this threshold in the Hexad-24. In the Hexad-12, both user types that do not reach the cut-off value of 0.7 are only slightly below this threshold. With regard to AVE, all factors in the Hexad-12 explain more than 50% of the variance on average, while this cutoff value is not reached for five out of six factors in the Hexad-24. Taking CR and AVE into account, we can conclude that the Hexad-12 has a more adequate convergent validity than the Hexad-24 (see Table 9). In terms of discriminant validity, we can see that there are several problems with the Hexad-24 (see Table 12). While for the Hexad-12 (see Table 11) we found that the AVE of each user type is higher than the shared variances, this conclusion can be drawn in the case of the Hexad-24 for only two user types. In the Hexad-24, the Socializer factor explains more variance of the Philanthropist factor than the Philanthropist factor itself. Similarly, the Free Spirit factor explains more variance of the Achiever factor than the Achiever items themselves. A similar problem exists on the Free Spirit scale, where the Disruptor factor explains more variance than the Free Spirit items. Overall, we can conclude that the proposed Hexad-12 distinguishes the six user types better than the Hexad-24 and thus is advantageous in terms of discriminant validity.

4.4.4 How Well Does the Hexad-12 represent the Hexad-24? To provide answers to this question, we analyzed bivariate correlations between the score of each user type in the Hexad-12 and the score of each user type in the Hexad-24. To complement this, we also conducted a canonical correlation analysis ("CCA") using the score of the six user types of the Hexad-12 as predictors of the six Hexad user types measured by the Hexad-24. Regarding the bivariate correlations, we found that the Hexad-12 factors are strongly correlated to the Hexad-24 items (with coefficients > .8), indicating that the

Table 10: Model fit measures for the Hexad-12 and the Hexad-24. Bold entries mark the better model fit value.

Model fit measure	χ^2	p	df	$\frac{\chi^2}{df}$	RMSEA	CFI	TLI
Hexad-12	**107**	<.001	39	**2.74**	**.04**	**.98**	**.97**
Hexad-24	1652	<.001	237	6.97	.07	.83	.80

2-item scales of the Hexad-12 represent the corresponding 4-item scales of the Hexad-24 well (see Table 13).

Next, we report the results of the CCA, which can be used to assess the multivariate shared variance between two sets of items [78]. It combines the set of predictor and criterion variables into latent variables, whereas the canonical correlation is defined as the correlation between these latent variables. The pairs of latent variables are called canonical functions ("CF"). Canonical functions are similar to principal components in principal component analyses. Thus, CCA is also considered "a double-barreled principal components analysis" [82]. Although CCA does not strictly rely on multivariate normality [88], we assessed multivariate normality by inspecting the skewness and kurtosis of each variable included in the CCA. They were all within the acceptable thresholds of skewness < 3 and kurtosis < 8 [43] (the maximum absolute values of skewness and kurtosis were 1.45 and 3.13 respectively). Thus, the CCA could be conducted.

The full model across all CFs was statistically significant using the Wilks's λ=.00018 criterion, $F(36, 4784.89) = 809.84$, $p < .001$. This means that the model is able to explain 99.98% of the shared variance between the Hexad-12 and the Hexad-24. This result clearly shows that the two variable sets are not independent. Thus, we can continue analysing the results of the dimension reduction analysis to check whether the predictor variables load on the same CF as the criterion variables. This is important to investigate whether the Hexad-12 sufficiently represents the original version.

As a result of the dimension reduction analysis, six canonical functions (CF1–CF6) were established. All of these CFs explain a statistically significant, considerable amount of shared variance between the variable sets (all $p < .001$), after the extraction of the prior functions. The squared canonical correlations are .89, .86, .76, .71, .65, and .51 each. Figure 1 presents the structure coefficients for CF1–CF6 being stronger than |.5|. All standardized canonical function coefficients and structure coefficients can be found in Table 14. Most predictor and criterion variables have large structure coefficients loading substantially (i.e., > |.5| according to [22]) on the same canonical functions. This is supported by the symmetry of the relationships, which can be seen in Figure 1.

Based on the bivariate correlations indicating that the factors of the Hexad-12 are strongly correlated to the respective factors of Hexad-24 as well as the CCA revealing that the shared variance between both sets of variables is higher than 99% and that the factors of both Hexad-12 and Hexad-24 load on the same canonical functions, we conclude that Hexad-12 represents the original Hexad-24 well.

5 DISCUSSION AND IMPLICATIONS

The results of our first study based on an existing dataset indicated that despite the Hexad scale's suitability to measure the six Hexad types and thus serve as the basis for a tailored gamification design, the scale reliability and factor loadings in the Hexad-24 could still be improved. Based on exploratory factor analysis, scale reliability analysis, and item content analysis, we identified two items for each scale that we considered most appropriate for inclusion in the Hexad-12. In the second study, based on a new data set, we assessed the Hexad-12 through confirmatory factor analysis, scale reliability analysis, and convergent and discriminant validity analysis.

Comparison with the Hexad-24 shows that the Hexad-12 outperforms the Hexad-24 in terms of model fit (it achieves a very good model fit of CFI = .98, TLI = .97, RMSEA = .04), convergent validity (see Table 9) and discriminant validity (see Table 11 versus Table 12), which means that the Hexad-12 captures the Hexad types more accurately than the Hexad-24. To ensure that the Hexad-12 still measured the same constructs as the Hexad-24 and truly captured the types better (and not just different types), we conducted bivariate and canonical correlation analyses. Both correlation analyses showed that the types measured by Hexad-12 and Hexad-24 are highly and significantly related, as indicated by high correlation coefficients and high factor loadings on the same canonical functions. Previous studies validating the Hexad-24 have found problems with the scale in terms of model fit and discriminant and convergent validity [47, 59, 84]. For example, the model fit (RMSEA = .06) in the original validation study by Tondello et al. [84] in English was right at the boundary of the recommended threshold (< .06) [27], which may have been caused in particular by problematic items of the Free Spirit, Player, and Achiever scales [84] that were omitted from the Hexad-12, thereby contributing to the improved model fit. With a larger sample, Krath and von Korflesch [47] reported model fit measures (RMSEA = .09, CFI = .72 for the English version and RMSEA = .08, CFI = .73 for the German version) that were not indicative of good model fit and identified items f2 (Free Spirit) and r3 (Player) as particularly problematic that were not included in Hexad-12. In the Dutch version (RMSEA = .09, CFI = .77), similar problems occurred with several items of the Free Spirit scale [59]. Therefore, we assume that by discarding two items of each scale with low factor loadings and problematic contributions to scale reliability, we have improved the capture of Hexad types in the Hexad-12, resulting in better model fit and validity while still measuring the same constructs as the Hexad-24.

In addition, reliability analysis showed that the scale reliability of the Player, Disruptor, and Free Spirit scales improved compared to the Hexad-24 and can now be considered acceptable ($\alpha \geq 0.7$). Similar to model fit, we believe that omitting items that proved problematic both in our initial study and in the previous studies [47, 59, 84] contributed to improving the reliability of these scales. However, it must also be noted that scale reliability of previously well-performing scales (Philanthropist, Socializer, and Achiever) declined in the Hexad-12 scale. While it is to be expected

Table 11: Discriminant validity of the Hexad-12, AVE and shared variances of the short scales. Bold entries mark the highest AVE/covariance of each scale.

	Philantropist	Socializer	Achiever	Player	Disruptor	Free Spirit
Philantropist	**.60**					
Socializer	.51	**.69**				
Achiever	.23	.25	**.51**			
Player	.21	.14	.35	**.54**		
Disruptor	.18	.21	.17	.05	**.55**	
Free Spirit	.03	.16	.47	.20	.45	**.54**

Table 12: Discriminant validity of the Hexad-24, AVE and shared variances of the scales. Bold entries mark the highest AVE/covariance of each scale.

	Philantropist	Socializer	Achiever	Player	Disruptor	Free Spirit
Philantropist	.47					
Socializer	**.66**	**.55**				
Achiever	.37	.33	.44			
Player	.23	.20	.35	**.39**		
Disruptor	.07	.13	.23	.02	.36	
Free Spirit	.21	.01	**.58**	.29	**.64**	.30

Table 13: Pearson's correlations between user types of the Hexad-12 and the Hexad-24. All p <.001. Bold entries mark the highest correlation in each scale.

	PH-Short	SO-Short	AC-Short	PL-Short	DI-Short	FS-Short
Philanthropist	**.88**					
Socializer	.48	**.91**				
Achiever	.21	.26	**.90**			
Player	.14	.12	.29	**.85**		
Disruptor	-	-	.22	.04	**.89**	
Free Spirit	-	.01	.36	.21	.40	**.81**

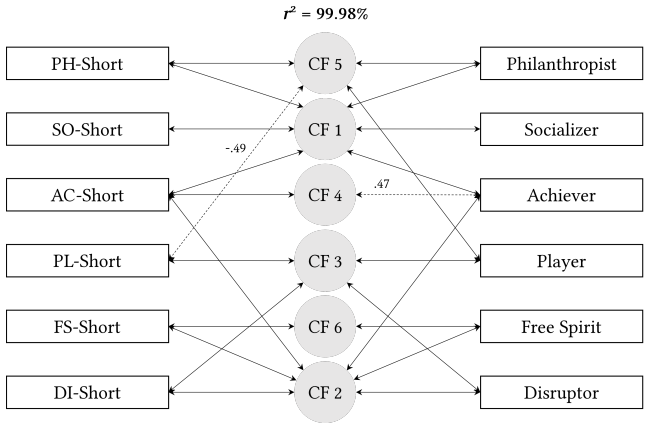

Figure 1: Structure coefficients for CF1–CF6 stronger than |.5|. Dotted lines indicate relationships slightly falling below the |.5| threshold.

Table 14: Structure coefficients (rs) and standardized canonical function coefficients (co) for predictor variables (user type scores assessed with the Hexad-12) and criterion variables (user type scores assessed with the Hexad-24) for the canonical functions. Bold entries represent loads higher than |.50|.

	CF 1		CF 2		CF 3		CF 4		CF 5		CF 6	
Predictor	*co*	*rs*	*co*	*rs*	*co*	*rs*	*co*	*rs*	*co*	*rs*	*co*	*rs*
PH-Short	-.42	**-.72**	-.15	-.17	-.14	-.10	-.42	-.36	.85	**.54**	.29	.14
SO-Short	-.57	**-.82**	-.18	-.24	.59	.36	-.04	-.01	-.63	-.36	-.42	-.12
AC-Short	-.34	**-.53**	.49	**.63**	-.19	-.18	.84	**.51**	.13	.03	.34	.10
PL-Short	-.08	-.29	.20	.30	-.52	**-.57**	-.55	-.45	-.59	-.49	.34	.25
FS-Short	-.05	-.07	.30	**.67**	-.15	-.15	-.27	-.21	.10	.18	-1.05	**-.67**
DI-Short	.08	.16	.51	**.70**	.71	**.59**	-.38	-.30	.04	.08	.48	.20
Criterion	*co*	*rs*	*co*	*rs*	*co*	*rs*	*co*	*rs*	*co*	*rs*	*co*	*rs*
Philanthropist	-.36	**-.79**	-.17	-.09	-.18	-.07	-.44	-.30	.93	**.50**	.54	.16
Socializer	-.55	**-.86**	-.29	-.21	.63	.32	-.02	-.09	-.72	-.29	-.39	-.13
Achiever	-.34	**-.61**	.47	**.62**	-.18	-.17	.99	.47	.06	.04	.20	.03
Player	-.07	-.33	.19	.33	-.59	**-.57**	-.52	-.39	-.58	**-.50**	.36	.24
Free Spirit	.01	-.25	.29	**.68**	-.18	-.03	-.39	-.31	.16	.21	-1.12	**-.58**
Disruptor	.10	.07	.49	**.75**	.73	**.56**	-.29	-.27	-.02	.03	.65	.21

that reliability deteriorates when a scale with four appropriate items is reduced to two items [75], scale reliability for the Philanthropist scale (α = .73) and the Socializer scale (α = .80) remains acceptable, so we do not consider the decrease as problematic. In contrast, the reliability of the Achiever scale fell slightly below the acceptable threshold (α = .67). Because we relied on both statistical measures and content in selecting the items, we assume that the Hexad-12 scale still captures both aspects considered important for Achiever types (i.e. mastery through item a2 "I like to master difficult tasks" and competence through item a4 "I like to emerge victorious from difficult situations"), but in this case, two items may not be sufficient to capture all facets of these aspects (e.g., item a3 "it is important to me to continuously improve my skills" refers to continuous improvement of skills, which is related to mastery but also focuses on learning and self-improvement). As a result, there is still room for improvement of the Hexad-12, particularly the Achiever scale, for example by introducing new or adapted items that better capture the variety of mastery and competence motivations associated with the Achiever type [53].

Although the Hexad-12 outperforms the Hexad-24 in a number of psychometric properties, the decision of whether to use the Hexad-12 or the Hexad-24 should still take into account the application context and the specific topic of interest. The Hexad-12 is a good choice when researchers and practitioners are interested in studying the behavior of different user types or tailoring their gamification solution to the full range of different user types, as the overall model fit and discriminant validity of the Hexad-12 surpasses that of the Hexad-24. In addition, the Hexad-12 is particularly suited for assessing user types under time-constrained conditions in the rapid iteration cycles of UX design [64] or for use on mobile devices, where longer questionnaires can easily cause screen fatigue and increase dropout rates. Similarly, the reduced participant burden of the Hexad-12 may be an advantage for certain target groups that struggle with longer questionnaires, such as children or the elderly (although the applicability of the Hexad model for these target groups should be treated with caution, as discussed in the Limitations section). Because the reliability of the Philanthropist, Socializer, and Achiever scales is higher in the Hexad-24, research and application contexts that focus particularly on socially oriented user types and their behaviors should consider favoring the Hexad-24 to capture the facets of the Philanthropist and Socializer types in more detail. Also, in application contexts of learning and education, the more nuanced capture of the different facets of the Achiever type in the Hexad-24 could be advantageous over the Hexad-12. Finally, with respect to retest reliability, it should be noted that the Hexad-24 has been used in a variety of studies to date, whereas the test-retest reliability of the Hexad-12 has yet to be determined.

Overall, based on our findings, we consider the Hexad-12 to be a suitable, sound, and concise tool for capturing Hexad types, allowing for static or dynamic adaptation of gamified systems *faster* and *more reliably* than the Hexad-24. It should be noted here, however, that the use of validated questionnaires is only one of several approaches to personalizing gamification, and other approaches developed to capture Hexad types [10, 42], as well as entirely different methods such as co-creation workshops and qualitative co-design with the target audience [61], may be of value to both researchers and practitioners depending on the specific application context. However, in the area of questionnaire assessments, our study demonstrates the reliability and validity of the Hexad-12 and paves the way for its use in research projects on tailored gamification and personalized gamification design in practice.

5.1 Implications for Human-Computer Interaction Research and Practice

In general, our results show that Hexad-12 is not only considerably shorter but also has several advantages over the original Hexad-24 in terms of its psychometric properties and validation. Considering the significant uptake of Hexad within the CHI community (the initial Hexad scale paper is the most downloaded and second

most cited paper of the CHI PLAY conference series[2]), this has implications for gameful design researchers and practitioners:

5.1.1 Hexad-12 is a Valid New Instrument For Personalization of Interactive Systems, especially when Assessment Time is Limited or many other Measurement Instruments are Used. According to our data, the shorter assessment time of Hexad-12 does not seem to come at the expense of lower reliability or validity. In contrast, Hexad-12 has advantages over Hexad-24 in terms of its psychometric properties. Because it contains half as many items as the original Hexad questionnaire, Hexad-12 has the potential to reduce dropout rates and increase data quality in academic settings. It could also promote wider adoption of Hexad in the industry by enabling practitioners to integrate Hexad into rapid iteration cycles. In settings where interaction modalities are constrained, such as on mobile devices, Hexad-12 is a less cumbersome method for evaluating Hexad user types. For these reasons, it has great potential to facilitate gamification personalization compared to existing scales, which is an important and recent topic in gamification research [44].

In fact, Hexad-12 not only contributes to gamification research but also to the broader CHI community. As noted by Fischer, "a fundamental objective of human–computer interaction research [...] is to provide users with experiences fitting their specific background knowledge and objectives" [29]. With Hexad-12, we provide a concise instrument with good psychometric properties that directly contributes to better understanding and assessing user motivation, needs, and objectives in interactive systems. Thus, it can be used beyond gamification, to personalize and inform the design of any interactive system when motivation and engagement are of concern.

5.1.2 Our Results support the Theoretical Construct of the Six-Factor Hexad Model. Another important implication concerns the theoretical construct of Hexad. The result of the confirmatory factor analysis of Hexad-12 shows that the hypothesized model, consisting of six traits, fits the collected data well. Thus, our results provide support for the theoretical construct of the Hexad, especially when reducing the set of items to two per user type. In contrast to existing typologies such as Bartle [13] (proclaiming four player types) or BrainHex [57] (establishing seven different archetypes of players), which both have been shown to lack empirical validation [14, 19, 20, 83], the confirmatory model fit of our investigation indicates the existence of six types. This has implications for researchers and practitioners alike to make a more informed decision on which user or player typology to use when operationalizing user motivation or individual preferences in interactive systems.

6 LIMITATIONS AND FUTURE WORK

There are several limitations which should be considered when using the Hexad-12.

First, as mentioned in the description of our item selection, items f2 and f4 of the Hexad-24 Free Spirit scale presented a challenge to the development of the Hexad-12. Item f2 loaded on an entirely new factor of its own, which caused the other items of the Free Spirit scale to load on the Disruptor factor instead, making it difficult to apply our factor loadings selection criterion. By excluding f2 as a consequence, we omitted the theoretical facet of curiosity, which is included in Hexad-24 for the Free Spirit type. With reliability contribution as the second criterion, we chose f1 and f3 and against f4, although we thus dropped the theoretical facet of Free Spirit self-expression in the Hexad-24. Our results indicate that the items thus selected, f1 and f3, load adequately on their own factor in the Hexad-12, with appropriate scale reliability, convergent validity (but still in need of improvement), and discriminant validity, and still represent the Free Spirit of the Hexad-24 very well. Considering that previous validation studies have also indicated problems with the items f2 and f4 [47, 59, 84], it is important to further investigate the reasons for this. Potentially, the importance of self-expression and curiosity, albeit being related to autonomy, differs between people. Thus, there might be interpersonal differences in how the respective items of the Free-Spirit scale are rated, leading to reduced convergence. Another potential reason might be related to the Free-Spirit and the Disruptor sharing their underlying motivation, i.e., a relationship between creativity and disruption being part of becoming creative. Thus, further theoretical and empirical work is needed to determine whether self-expression and curiosity are important aspects of the Free Spirit type, and if so, how these aspects might be represented in new items that better capture all facets of this user type.

Second, we did not examine test-retest reliability. Given that only two items were used to measure each user type in the Hexad-12, a divergent response to a single item could potentially affect the rating of that user type more than it did for the Hexad-24. Because previous research has indicated that the Hexad user type changes over time [67, 70] and test-retest reliability analysis assumes that the concepts measured are stable over time, we have decided against including test-retest reliability analysis in the scope of this work. However, once more is known about what factors have an impact on the stability of Hexad user types, future work should consider these factors and examine the test-retest reliability of the Hexad-12.

Third, our decision to use a subset of the Hexad-24 was a methodological decision consistent with previous approaches to short-scale development [4, 41, 76]. We did not include mixed-methods approaches, such as qualitative interviews or workshops with subject matter experts or target audiences, in our development. Therefore, it would be a great avenue for further research to qualitatively test the validity and applicability of the Hexad-12 with experts in tailored gamification and benefit from their knowledge to further improve the Hexad-12 and its items.

Fourth, previous validation studies of the Hexad-24 in Turkish [2, 81], Dutch [59], German [47], Spanish [84] and Brazilian Portuguese [71] show that there is a great need for using the Hexad scale in languages other than English to improve its applicability in different countries and contexts. Because we mainly recruited native English speakers in our second study, to avoid problems due to language proficiency, our participants were mainly from the United Kingdom, Australia, Canada, the United States, New Zealand, and Ireland. However, this limits the transferability of the validity of the Hexad-12 to other countries, cultures, and languages, and calls for further research that adapts the Hexad-12 to other languages in order to enable widespread applicability.

Finally, it should be noted that the participants in our two studies were adults, with a mean age of $M = 29.00$ (first study) and $M = 31.70$ (second study). Previous studies have shown that personality

[2]https://bit.ly/3FBcAkm, last accessed March 10, 2023

assessments, such as the Big Five questionnaire and the Hexad types, are not appropriate for adolescents or children [59]. In addition, user typologies such as Hexad can also present challenges for elderly people with different experiences and needs than the target group for whom these archetypes were developed [23]. Therefore, we urge caution in using the Hexad-12, which may have similar problems to personalize gamification for children, adolescents, or the elderly, and call for further research to examine the validity of the Hexad-12 for these target audiences.

7 CONCLUSION

In this paper, we propose Hexad-12, a shortened version of the original Hexad scale with 24 items. First, we conducted an exploratory factor analysis and identified two items for each user type based on their factor (cross-)loadings, their contribution to scale reliability, and the extent to which they represent non-redundant information of the respective user type. We then examined the psychometric properties of Hexad-12 through confirmatory factor analysis. We found that Hexad-12 has a good model fit, discriminates all six user types well, and has acceptable reliability.

A comparison of Hexad-12 with Hexad-24 showed that Hexad-12 resolved the problems with Hexad-24 found in previous studies: the low to borderline model fit reported in the past was substantially improved, similar to discriminant validity. Although the reliability of some user types decreased in the Hexad-12 (only one user type fell below the threshold of $\alpha \geq 0.7$), the reliability of other user types that have been found to be problematic in previous research improved so that five of the six user types in the Hexad-12 have acceptable reliability (compared to three of the six subscales in the Hexad-24). Thus, this work provides a solid new instrument for an efficient assessment of Hexad user types, which can be considered advantageous because of its psychometric properties compared to the original Hexad scale.

REFERENCES

[1] Hina Akasaki, Shoko Suzuki, Kanako Nakajima, Koko Yamabe, Mizuki Sakamoto, Todorka Alexandrova, and Tatsuo Nakajima. 2016. One Size Does Not Fit All: Applying the Right Game Concepts for the Right Persons to Encourage Non-game Activities. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9735 (2016), 103–114. https://doi.org/10.1007/978-3-319-40397-7_11

[2] Özcan Erkan Akgün and Murat Topal. 2018. Adaptation of the Gamification User Types Hexad Scale into Turkish. *International Journal of Assessment Tools in Education* 5, 3 (2018), 389–402.

[3] Mousa Alavi, Denis C. Visentin, Deependra K. Thapa, Glenn E. Hunt, Roger Watson, and Michelle Cleary. 2020. Chi-square for model fit in confirmatory factor analysis. *Journal of Advanced Nursing* 76, 9 (2020), 2209–2211. https://doi.org/10.1111/jan.14399

[4] Catherine Alberola, Götz Walter, and Henning Brau. 2018. Creation of a Short Version of the User Experience Questionnaire UEQ. *i-com* 17, 1 (apr 2018), 57–64. https://doi.org/10.1515/icom-2017-0032

[5] Noora Aldenaini, Felwah Alqahtani, Rita Orji, and Srinivas Sampalli. 2020. Trends in Persuasive Technologies for Physical Activity and Sedentary Behavior: A Systematic Review. *Frontiers in Artificial Intelligence* 3 (2020), 7. https://doi.org/10.3389/frai.2020.00007

[6] Maximilian Altmeyer, Pascal Lessel, Subhashini Jantwal, Linda Muller, Florian Daiber, and Antonio Krüger. 2021. Potential and Effects of Personalizing Gameful Fitness Applications Using Behavior Change Intentions and Hexad User Types. *User Modeling and User-Adapted Interaction* 31, 1 (2021), 675–712. https://doi.org/10.1007/s11257-021-09288-6

[7] Maximilian Altmeyer, Pascal Lessel, Linda Muller, and Antonio Krüger. 2019. Combining Behavior Change Intentions and User Types to Select Suitable Gamification Elements for Persuasive Fitness Systems. In *Proceedings of the International Conference on Persuasive Technology (PERSUASIVE '19)*. Springer, Berlin, Germany, 337–349.

[8] Maximilian Altmeyer, Pascal Lessel, Marc Schubhan, and Antonio Krüger. 2019. Towards Predicting Hexad User Types from Smartphone Data. In *Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. ACM, New York, NY, USA, 315–322. https://doi.org/10.1145/3341215.3356266

[9] Maximilian Altmeyer, Marc Schubhan, Pascal Lessel, Linda Muller, and Antonio Krüger. 2020. Using Hexad User Types to Select Suitable Gamification Elements to Encourage Healthy Eating. In *Extended Abstracts Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '20)*. ACM, New York, NY, USA, 1–8. https://doi.org/10.1145/3334480.3383011

[10] Maximilian Altmeyer, Gustavo F. Tondello, Antonio Krüger, and Lennart E. Nacke. 2020. HexArcade: Predicting Hexad User Types By Using Gameful Applications. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, New York, NY, USA, 219–230. https://doi.org/10.1145/3410404.3414232

[11] Tapio Auvinen, Lasse Hakulinen, and Lauri Malmi. 2015. Increasing Students' Awareness of Their Behavior in Online Learning Environments with Visualizations and Achievement Badges. *IEEE Transactions on Learning Technologies* 8, 3 (2015), 261–273. https://doi.org/10.1109/TLT.2015.2441718

[12] Kirk Bansak, Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. 2018. The Number of Choice Tasks and Survey Satisficing in Conjoint Experiments. *Political Analysis* 26, 1 (2018), 112–119. https://doi.org/10.1017/pan.2017.40

[13] Richard Bartle. 1996. Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs. *Journal of MUD Research* 1, 1 (1996), 19.

[14] Chris Bateman, Rebecca Lowenhaupt, and Lennart E. Nacke. 2011. Player Typology in Theory and Practice. In *DiGRA Conference Proceedings (DiGRA '11)*. Digital Games Research Association, Utrecht, 1–24.

[15] María J. Blanca, Rafael Alarcón, Jaume Arnau, Roser Bono, and Rebecca Bendayan. 2017. Non-normal data: Is ANOVA still a valid option? *Psicothema* 29, 4 (2017), 552–557. https://doi.org/10.7334/psicothema2016.383

[16] Martin Böckle, Isabel Micheel, and Markus Bick. 2018. A Design Framework for Adaptive Gamification Applications. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS '18)*. Association for Information Systems, Honolulu, Hawaii, USA, 1227–1236.

[17] Martin Böckle, Jasminko Novak, and Markus Bick. 2017. Towards Adaptive Gamification: A Synthesis of Current Developments. In *Proceedings of the European Conference on Information Systems (ECIS '17)*. Association for Information Systems, Guimarães, Portugal, 158–174.

[18] Martin Böckle, Jasminko Novak, and Markus Bick. 2020. Exploring gamified persuasive system design for energy saving. *Journal of Enterprise Information Management* 33, 6 (jun 2020), 1337–1356. https://doi.org/10.1108/JEIM-02-2019-0032

[19] Marc Busch, Elke Mattheiss, Wolfgang Hochleitner, Christina Hochleitner, Michael Lankes, Peter Fröhlich, Rita Orji, and Manfred Tscheligi. 2016. Using Player Type Models for Personalized Game Design - An Empirical Investigation. *Interaction Design and Architectures* 28, 1 (2016), 145–163.

[20] Marc Busch, Elke Mattheiss, Rita Orji, Peter Fröhlich, Michael Lankes, and Manfred Tscheligi. 2016. Player Type Models - Towards Empirical Validation. In *Extended Abstracts Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, San Jose, California, USA, 1835–1841. https://doi.org/10.1145/2851581.2892399

[21] Gerry Chan, Ali Arya, Rita Orji, Zhao Zhao, and Anthony Whitehead. 2021. Personalizing Gameful Elements in Social Exergames: An Exploratory Study. In *The 16th International Conference on the Foundations of Digital Games (FDG) 2021*. ACM, New York, NY, USA, 1–6. https://doi.org/10.1145/3472538.3472578

[22] Jacob Cohen. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, New York, USA.

[23] Frederiek De Vette, Monique Tabak, Marit Dekker - van Weering, and Miriam Vollenbroek-Hutten. 2015. Engaging Elderly People in Telemedicine Through Gamification. *JMIR Serious Games* 3, 2 (2015), e9. https://doi.org/10.2196/games.4561

[24] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From Game Design Elements to Gamefulness: Defining Gamification. In *Proceedings of the International Academic MindTrek Conference (Mindtrek '11)*. ACM, Tampere, Finnland, 9–15. https://doi.org/10.1145/2181037.2181040

[25] Lisa Diamond, Gustavo F Tondello, Andrzej Marczewski, Lennart E Nacke, and Manfred Tscheligi. 2015. The HEXAD Gamification User Types Questionnaire : Background and Development Process. In *Workshop on Personalization in Serious and Persuasive Games and Gamified Interactions*. ACM, London, UK, 1–6.

[26] Brian S. Everitt and Graham Dunn. 2001. Exploratory Factor Analysis. In *Applied Multivariate Data Analysis*. John Wiley & Sons, Ltd,., West Sussex, United Kingdom, 271–290. https://doi.org/10.1002/9781118887486.ch12

[27] Leandre 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

[28] Andrew Farrell. 2010. Insufficient Discriminant Validity: A Comment on Bove, Pervan, Beatty, and Shiu (2009). *Journal of Business Research* 63, 3 (2010), 324–327.

[29] Gerhard Fischer. 2001. User modeling in human–computer interaction. *User modeling and user-adapted interaction* 11, 1 (2001), 65–86.

[30] 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.

[31] Victor Manuel García-Barrios, Felix Mödritscher, and Christian Gütl. 2005. Personalisation versus Adaptation? A User-centred Model Approach and its Application. In *Proceedings of the International Conference on Knowledge Management (I-KNOW)*. ACM, Graz, Austria, 120–127. https://doi.org/10.1.1.161.7376

[32] Aqeel Haider, Casper Harteveld, Daniel Johnson, Max V. Birk, Regan L. Mandryk, Magy Seif El-Nasr, Lennart E. Nacke, Kathrin Gerling, and Vero Vanden Abeele. 2022. MiniPXI: Development and Validation of an Eleven-Item Measure of the Player Experience Inventory. *Proc. ACM Hum.-Comput. Interact.* 6, CHI PLAY, Article 244 (oct 2022), 26 pages. https://doi.org/10.1145/3549507

[33] Lasse Hakulinen and Tapio Auvinen. 2014. The effect of gamification on students with different achievement goal orientations. In *Proceedings - 2014 International Conference on Teaching and Learning in Computing and Engineering, LATICE 2014*. IEEE, Kuching, Malaysia, 9–16. https://doi.org/10.1109/LaTiCE.2014.10

[34] Stuart Hallifax, Audrey Serna, Jean-charles Marty, Guillaume Lavoué, and Elise Lavoué. 2019. Factors to Consider for Tailored Gamification. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '19)*. ACM, Barcelona, Spain, 559–572.

[35] Juho Hamari and Jonna Koivisto. 2015. Why do people use gamification services? *International Journal of Information Management* 35, 4 (2015), 419–431. https://doi.org/10.1016/j.ijinfomgt.2015.04.006

[36] Juho Hamari, Jonna Koivisto, and Harri Sarsa. 2014. Does Gamification Work? - A Literature Review of Empirical Studies on Gamification. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS '14)*. Association for Information Systems, Honolulu, Hawaii, USA, 3025–3034. https://doi.org/10.1109/HICSS.2014.377

[37] Juho Hamari and Janne Tuunanen. 2014. Player Types: A Meta-synthesis. *Transactions of the Digital Games Research Association* 1, 2 (2014), 29–53. https://doi.org/10.26503/todigra.v1i2.13

[38] A. Regula Herzog and Jerald G. Bachman. 1981. Effects of Questionnaire Length on Response Quality. *Public Opinion Quarterly* 45, 4 (1981), 549–559. https://doi.org/10.1086/268687

[39] 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

[40] Joseph R. Keebler, William J. Shelstad, Dustin C. Smith, Barbara S. Chaparro, and Mikki H. Phan. 2020. Validation of the GUESS-18: A Short Version of the Game User Experience Satisfaction Scale (GUESS). *Journal of Usability Studies* 16, 1 (2020), 49–62.

[41] Joseph R Keebler Assoc, William J Shelstad, Dustin C Smith Google, Barbara S Chaparro, and Mikki H Phan Google. 2020. Validation of the GUESS-18: A Short Version of the Game User Experience Satisfaction Scale (GUESS). *Journal of Usability Studies* 16, 1 (2020), 49–62.

[42] Robbe Kimpen, Robin De Croon, Vero Vanden Abeele, and Katrien Verbert. 2021. Towards Predicting Hexad User Types from Mobile Banking Data. In *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play*. ACM, New York, NY, USA, 30–36. https://doi.org/10.1145/3450337.3483486

[43] Rex B. Kline. 2011. *Principles and Practice of Structural Equation Modeling*. Vol. 3. The Guilford Press, New York, NY, USA. 1689–1699 pages.

[44] Ana Carolina Tomé Klock, Isabela Gasparini, Marcelo Soares Pimenta, and Juho Hamari. 2020. Tailored Gamification: A Review of Literature. *International Journal of Human Computer Studies* 144 (2020), 1–22. https://doi.org/10.1016/j.ijhcs.2020.102495

[45] Dimosthenis Kotsopoulos, Cleopatra Bardaki, Stavros Lounis, and Katerina Pramatari. 2018. Employee Profiles and Preferences towards IoT-enabled Gamification for Energy Conservation. *International Journal of Serious Games* 5, 2 (2018), 65–85. https://doi.org/10.17083/ijsg.v5i2.225

[46] Jeanine Krath and Harald F.O. von Korflesch. 2021. Designing gamification and persuasive systems: a systematic literature review. In *5th International GamiFIN Conference*. CEUR Workshop Proceedings, online, 100–109.

[47] Jeanine Krath and Harald F. O. von Korflesch. 2021. Player Types and Game Element Preferences: Investigating the Relationship with the Gamification User Types HEXAD Scale. In *HCI in Games: Experience Design and Game Mechanics. HCII 2021. Lecture Notes in Computer Science, vol 12789*, X. Fang (Ed.). Springer Nature, online, 219–238. https://doi.org/10.1007/978-3-030-77277-2_18

[48] Élise Lavoué, Baptiste Monterrat, Michel Desmarais, and Sébastien George. 2018. Adaptive Gamification for Learning Environments. *IEEE Transactions on Learning Technologies* 12, 1 (2018), 16–28. https://doi.org/10.1109/TLT.2018.2823710

[49] C E López and C S Tucker. 2019. Implementing gamification in engineering bridge programs: A case study exploring the use of the Kahoot! application. In *2019 ASEE Zone I Conference & Workshop*. ASEE Peer, Niagara Falls, NY, 1–11.

[50] Christian E. Lopez and Conrad S. Tucker. 2019. The effects of player type on performance: A gamification case study. *Computers in Human Behavior* 91 (2019), 333–345. https://doi.org/10.1016/j.chb.2018.10.005

[51] Christian E Lopez and Conrad S Tucker. 2021. Adaptive Gamification and Its Impact on Performance. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Lecture Notes in Computer Science, Vol. 12789 LNCS)*, Xiaowen Fang (Ed.). Springer International Publishing, Cham, 327–341. https://doi.org/10.1007/978-3-030-77277-2_25

[52] Ana Manzano-León, Pablo Camacho-Lazarraga, Miguel A Guerrero-Puerta, Laura Guerrero-Puerta, Antonio Alias, Rubén Trigueros, and José M Aguilar-Parra. 2020. Adaptation and validation of the scale of types of users in gamification with the Spanish adolescent population. *International Journal of Environmental Research and Public Health* 17, 11 (2020), 4157.

[53] Andrzej Marczewski. 2015. *Even Ninja Monkeys Like to Play: Gamification, Game Thinking and Motivational Design*. CreateSpace Independent Publishing Platform, online.

[54] 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

[55] Alberto Mora, Gustavo F. Tondello, Laura Calvet, Carina González, Joan Arnedo-Moreno, and Lennart E. Nacke. 2019. The Quest for a Better Tailoring of Gameful Design: An Analysis of Player Type Preferences. In *Proceedings of the XX International Conference on Human Computer Interaction* (Donostia, Gipuzkoa, Spain) *(Interacción '19)*. Association for Computing Machinery, New York, NY, USA, Article 1, 8 pages. https://doi.org/10.1145/3335595.3335625

[56] Alberto Mora, Gustavo F. Tondello, Lennart E. Nacke, and Joan Arnedo-Moreno. 2018. Effect of Personalized Gameful Design on Student Engagement. In *Proceedings of the Global Engineering Education Conference (EDUCON '18)*. IEEE, Santa Cruz de Tenerife, Spain, 1925–1933. https://doi.org/10.1109/EDUCON.2018.8363471

[57] Lennart E. Nacke, Chris Bateman, and Regan L. Mandryk. 2014. BrainHex: A Neurobiological Gamer Typology Survey. *Entertainment Computing* 5, 1 (2014), 55–62. https://doi.org/10.1016/j.entcom.2013.06.002

[58] Lennart E. Nacke and Sebastian Deterding. 2017. The Maturing of Gamification Research. *Computers in Human Behavior* 71 (2017), 450–454. https://doi.org/10.1016/j.chb.2016.11.062

[59] Jeroen Ooge, Robin De Croon, Katrien Verbert, and Vero Vanden Abeele. 2020. Tailoring Gamification for Adolescents: a Validation Study of Big Five and Hexad in Dutch. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, New York, NY, USA, 206–218. https://doi.org/10.1145/3410404.3414267

[60] Rita Orji, Gustavo F. Tondello, and Lennart E. Nacke. 2018. Personalizing Persuasive Strategies in Gameful Systems to Gamification User Types. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, Montréal, Canada, 1–14. https://doi.org/10.1145/3173574.3174009

[61] Adam Palmquist, Ole Goethe, Jeanine Krath, Joacim Rosenlund, and Miralem Helmefalk. 2022. Design Implications for a Gamified Recycling House. In *HCI in Games. HCII 2022. Lecture Notes in Computer Science, vol 13334*. Springer, Cham, online, 289–305. https://doi.org/10.1007/978-3-031-05637-6_18

[62] Mario Passalacqua, Sylvain Sénécal, Marc Frédette, Lennart E. Nacke, Robert Pellerin, and Pierre-Majorique Léger. 2021. Should Gamification be Personalized? A Self-deterministic Approach. *AIS Transactions on Human-Computer Interaction* 13, 3 (sep 2021), 265–286. https://doi.org/10.17705/1thci.00150

[63] Daniel H. Pink. 2009. *Drive: The surprising truth about what motivates us*. Canongate Books, New York.

[64] Beatrice Rammstedt and Oliver P. John. 2007. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality* 41, 1 (2007), 203–212. https://doi.org/10.1016/j.jrp.2006.02.001

[65] Amon Rapp, Frank Hopfgartner, Juho Hamari, Conor Linehan, and Federica Cena. 2019. Strengthening Gamification Studies: Current Trends and Future Opportunities of Gamification Research. *International Journal of Human Computer Studies* 127 (2019), 1–6. https://doi.org/10.1016/j.ijhcs.2018.11.007

[66] Melanie Revilla and Carlos Ochoa. 2017. Ideal and Maximum Length for a Web Survey. *International Journal of Market Research* 59, 5 (2017), 557–565. https://doi.org/10.2501/IJMR-2017-039

[67] Inmaculada Rodríguez, Anna Puig, and Alex Rodríguez. 2021. We Are Not the Same Either Playing: A Proposal for Adaptive Gamification. In *Frontiers in Artificial Intelligence and Applications*, Mateu Villaret, Teresa Alsinet, Cèsar Fernández, and Aïda Valls (Eds.). Vol. 339. IOS Press BV, Amsterdam, Netherlands, 185–194. https://doi.org/10.3233/FAIA210133

[68] 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), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

[69] Ana Cláudia Guimarães Santos, Wilk Oliveira, Maximilian Altmeyer, Juho Hamari, and Seiji Isotani. 2022. Psychometric investigation of the gamification Hexad user types scale in Brazilian Portuguese. *Scientific Reports* 12, 1920 (2022), 1–11.

[70] Ana Claudia Guimaraes Santos, Wilk Oliveira, Juho Hamari, and Seiji Isotani. 2021. Do people's user types change over time? An exploratory study. In *5th International GamiFIN Conference*. CEUR Workshop Proceedings, online, 90–99. arXiv:2106.10148

[71] Ana Cláudia Guimarães Santos, Wilk Oliveira, Juho Hamari, Luiz Rodrigues, Armando M. Toda, Paula T. Palomino, and Seiji Isotani. 2021. The relationship between user types and gamification designs. *User Modeling and User-Adapted Interaction* 31, 5 (2021), 907–940. https://doi.org/10.1007/s11257-021-09300-z

[72] 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.

[73] Sofia Schöbel, Andreas Janson, Katharina Jahn, Bastian Kordyaka, Ozgur Turetken, Naza Djafarova, Mohammed Saqr, Dezhi Wu, Matthias Söllner, Martin Adam, Povl Heiberg Gad, Henrik Wesseloh, and Jan Marco Leimeister. 2020. A research agenda for the why, what, and how of gamification designs: Outcomes of an ecis 2019 panel. *Communications of the Association for Information Systems* 46 (2020), 706–721. https://doi.org/10.17705/1CAIS.04630

[74] Sofia Schöbel, Manuel Schmidt-Kraepelin, Andreas Janson, and Ali Sunyaev. 2021. Adaptive and Personalized Gamification Designs: Call for Action and Future Research. *AIS Transactions on Human-Computer Interaction* 13, 4 (2021), 479–494. https://doi.org/10.17705/1thci.00158

[75] Martin Schrepp. 2020. On the Usage of Cronbach's Alpha to Measure Reliability of UX Scales. *Journal of Usability Studies* 15, 4 (2020), 247–258.

[76] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2017. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence* 4, 6 (2017), 103. https://doi.org/10.9781/ijimai.2017.09.001

[77] Katie Seaborn and Deborah Fels. 2015. Gamification in Theory and Action: A Survey. *International Journal of Human-Computer Studies* 74 (2015), 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006

[78] Alissa Sherry and Robin K. Henson. 2005. Conducting and Interpreting Canonical Correlation Analysis in Personality Research: A User-Friendly Primer. *Journal of Personality Assessment* 3891 (2005), 37–48. https://doi.org/10.1207/s15327752jpa8401_09

[79] Noora Shrestha. 2021. Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics* 9, 1 (2021), 4–11. https://doi.org/10.12691/ajams-9-1-2

[80] John Sweller. 2010. Cognitive load theory: Recent theoretical advances. In *Cognitive Load Theory*, Jan L. Plass, Roxana Moreno, and Roland Brünken (Eds.). Cambridge University Press, Cambridge, UK, 29–47. https://doi.org/10.1017/CBO9780511844744.004

[81] Necati Taskin and Ebru Kiliç Çakmak. 2020. Adaptation of Modified Gamification User Types Scale into Turkish. *Contemporary Educational Technology* 12, 2 (2020), ep268.

[82] Bruce Thompson. 1984. *Canonical Correlation Analysis: Uses and Interpretation*. Sage, Thousand Oaks, CA, USA.

[83] Gustavo Tondello, Deltcho Valtchanov, Adrian Reetz, Rina R. Wehbe, Rita Orji, and Lennart E. Nacke. 2018. Towards a Trait Model of Video Game Preferences. *International Journal of Human–Computer Interaction* 34, 8 (2018), 732–748. https://doi.org/10.1080/10447318.2018.1461765

[84] Gustavo F. Tondello, Alberto Mora, Andrzej Marczewski, and Lennart E. Nacke. 2019. Empirical validation of the Gamification User Types Hexad scale in English and Spanish. *International Journal of Human Computer Studies* 127 (2019), 95–111. https://doi.org/10.1016/j.ijhcs.2018.10.002

[85] Gustavo F. Tondello, Alberto Mora, and Lennart E. Nacke. 2017. Elements of Gameful Design Emerging from User Preferences. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play (CHI PLAY '17)*. ACM, Amsterdam, The Netherlands, 129–142. https://doi.org/10.1145/3116595.3116627

[86] Gustavo F. Tondello, Rina R. Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E. Nacke. 2016. The Gamification User Types Hexad Scale. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*. ACM, New York, NY, USA, 229–243. https://doi.org/10.1145/2967934.2968082

[87] Margaret Verkuyl, Naza Djafarova, Paula Mastrilli, and Lynda Atack. 2022. Virtual gaming simulation: evaluating players' experiences. *Clinical Simulation in Nursing* 63 (2022), 16–22.

[88] Hao-Ting Wang, Jonathan Smallwood, Janaina Mourao-Miranda, Cedric Huchuan Xia, Theodore D. Satterthwaite, Danielle S. Bassett, and Danilo Bzdok. 2018. Finding the Needle in High-Dimensional Haystack: A Tutorial on Canonical Correlation Analysis. (2018). http://arxiv.org/abs/1812.02598

[89] Daniel Wessel, Christiane Attig, and Thomas Franke. 2019. ATI-S - An Ultra-Short Scale for Assessing Affinity for Technology Interaction in User Studies. In *Proceedings of Mensch und Computer 2019*. ACM, New York, NY, USA, 147–154. https://doi.org/10.1145/3340764.3340766

[90] Jamie Woodcock and Mark R Johnson. 2018. Gamification: What it is, and how to fight it. *The Sociological Review* 66, 3 (2018), 542–558.

[91] Adriana Zaiţ and PSPE Bertea. 2011. Methods for Testing Discriminant Validity. *Management & Marketing Journal* 9, 2 (2011), 217–224.

[92] Zhao Zhao, Ali Arya, Rita Orji, and Gerry Chan. 2020. Effects of a personalized fitness recommender system using gamification and continuous player modeling: System design and long-term validation study. *JMIR Serious Games* 8, 4 (2020), 1–27. https://doi.org/10.2196/19968

A FULL LIST OF ITEMS OF THE HEXAD-24

B RESULTS OF THE ONE-WAY ANOVA AS A PRE-TEST TO MERGING THE DATASETS

Table 15: Items and labels of the Hexad-24

Hexad type	Item	English Version (based on [84])	Included in Hexad-12
Philanthropist	p1	It makes me happy if I am able to help others	Yes
	p2	I like helping others to orient themselves in new situations	No
	p3	I like sharing my knowledge	No
	p4	The well-being of others is important to me	Yes
Socializer	s1	Interacting with others is important to me	No
	s2	I like being part of a team	Yes
	s3	It is important to me to feel like I am part of a community	No
	s4	I enjoy group activities	Yes
Achiever	a1	I like overcoming obstacles	No
	a2	I like mastering difficult tasks	Yes
	a3	It is important to me to continuously improve my skills	No
	a4	I enjoy emerging victorious out of difficult circumstances	Yes
Player	r1	I like competitions where a prize can be won	No
	r2	Rewards are a great way to motivate me	Yes
	r3	Return of investment is important to me	No
	r4	If the reward is sufficient, I will put in the effort	Yes
Free Spirit	f1	It is important to me to follow my own path	Yes
	f2	I often let curiosity guide me	No
	f3	Being independent is important to me	Yes
	f4	Opportunities for self-expression are important to me	No
Disruptor	d1	I like to provoke	No
	d2	I like to question the status quo	No
	d3	I see myself as a rebel	Yes
	d4	I dislike following rules	Yes

Table 16: Precondition tests, F-value and significance of the one-way ANOVA testing the effect of gender on Hexad types in both datasets

Hexad Type	Levene's	p	Shapiro-Wilk	p	F	p
[10]						
Philanthropist	.03	.97	.87	$< .001^{***}$	.25	.77
Socializer	.44	.65	.97	$< .01^{**}$	.43	.65
Achiever	1.12	.33	.81	$< .001^{***}$	.37	.69
Player	2.24	.11	.86	$< .001^{***}$	.25	.78
Free Spirit	.77	.47	.85	$< .001^{***}$	.05	.95
Disruptor	2.61	.07	.98	.07	.07	.93
[47]						
Philanthropist	.27	.77	.90	$< .001^{***}$	1.18	.31
Socializer	.08	.92	.96	$< .001^{***}$	2.78	.06
Achiever	2.06	.13	.91	$< .001^{***}$	3.21	$.04^{*}$
Player	2.22	.11	.96	$< .001^{***}$	1.25	.29
Free Spirit	1.49	.23	.95	$< .001^{***}$	.06	.95
Disruptor	1.30	.27	.99	$.03^{*}$	9.11	$< .001^{***}$