No Item is an Island Entire of Itself: A Statistical Analysis of Individual Player Difference Questionnaires

Nathan GJ Hughes, Paul Cairns

University of York, York, YO10 5GH

Abstract

There are many questionnaires to identify individual differences between digital game players. However, each is typically developed without reference to explicit conceptual underpinnings and its conceptual relationship to other questionnaires. This paper reports on a study where 11 such questionnaires were compiled into one 177-item survey. The analysis of responses ($N = 1683$) assesses individual questionnaires and compares across questionnaires through five analyses, including reliability analysis and both confirmatory and exploratory factor analyses. These showed only moderate support for the original questionnaire structures, and seemingly related sub-scales tended to measure unrelated concepts. An EFA of all 177 items found a 19-factor solution, suggesting considerable conceptual overlap. Overall, there are concerning shortcomings in the ability for these instruments to measure distinct player differences. A more robust theoretical basis is needed to address the value of such questionnaires in advancing player research.

Keywords:

Player Traits, Player Motivations, Player Preferences, Questionnaires

1. Introduction

Digital games, like other forms of entertainment media, contain a vast diversity of content. Such diversity has also led to an extensive and well-recognised diversity of players, and consequently both researchers and developers have sought to understand how differences in players impacts their

Email addresses: nathan.hughes@york.ac.uk (Nathan GJ Hughes),
paul.cairns@york.ac.uk (Paul Cairns)

Preprint submitted to Elsevier

March 28, 2022

preferences and behaviours within games. Such concern for differences between people is an important topic in Psychology, where these are generally referred to as individual differences.

Since the early days of games research there have been attempts to classify players into different types ([1]), to characterise differences in play behaviours. Later efforts have sought to go further and predict player behaviour, and what aspects of game design might influence it e.g., how personality traits affect behaviour in online games [2]. Methods for studying these individual differences vary (such as behavioural observations or interviews), but one of the most common methods for assessing individual difference is questionnaires. By using questionnaires, scores produced can help identify clusters of players, where individual differences in concepts such as motivations and disposition are expected to both affect and predict behaviours.

However, despite the common goal of player researchers, there has been a proliferation of different questionnaires all purporting to measure player traits, types, motivations, and preferences (all sub-concepts under the general idea of individual player differences). Indeed, there are over 23 currently available questionnaires that measure individual differences in players ([3]). Some questionnaires offer unique sub-scales not captured by others (such as Arousal by Sherry and colleagues[4]), but most show a large degree of semantic overlap (the concept of ‘Challenge’ appears in multiple scales, such as [5], [6], and [4]). This proliferation of questionnaires therefore presents a problem to the field. If all questionnaires are valid, then no single questionnaire captures all the relevant aspects of individual players differences — yet their overlap suggests there is also a degree of unnecessary repetition between different instruments. For the questionnaire approach to individual differences to be effective, the concepts used ought to be diverse enough to capture meaningful differences between players, but few enough to have wide applicability for reasonable groupings of large numbers of players. This suggests what is needed is a more systematic approach to comparing and synthesising all of these different questionnaires.

Despite this, to date there has been little comparison or distinction made between these different instruments. Previous work comparing questionnaires typically has done so by analysing the items and sub-scale names (e.g., [7]), however this does not indicate how questionnaires overlap statistically. Overlap between items, sub-scales, and whole questionnaires can be found empirically (for example via reliability and factor analyses), though currently this is rarely done in player individual differences. By contrast, the need for such

comparison is emerging in the area of player experience research where there have been several attempts to empirically compare different questionnaires aimed at measuring experience (e.g., [8], [9]). Without explicit empirical comparison, it is unclear which questionnaires measure these constructs in the most reliable and valid way. Each questionnaire claims to capture aspects of the individual differences present in players, but are they measuring the same underlying constructs? And if yes, are some scales better than others at doing so? These questions hint at a larger problem for the field: with so many questionnaires available, how much conceptual overlap is present? And what do we currently understand about these concepts?

This paper presents an empirical exploration of the overlap between existing individual player difference questionnaires. It is a statistical companion to a separate paper that discusses the common practices researchers use to report their questionnaire usage ([3]). For this reason, the current work references the companion paper frequently throughout, especially for theoretical considerations undertaken in this work.

For the purposes of gathering the necessary data, players were asked to complete 11 questionnaires that represent the scope of measurement currently available, to allow an analysis of their combined concepts. High overlap between questionnaires would indicate an underlying theoretical concern that new questionnaires created in the field are not adding extra nuance, and are instead re-observing the same concepts. This raises the question as to whether creating further questionnaires will overcome this problem. Before assessing their overlap, the general stability of the questionnaires is explored via reliability and factor analyses. This explores how good the current measures are at capturing what they intend to, providing a baseline of how likely current questionnaires are to capture the concepts explored in this work. Afterwards, exploratory analyses consider the overlaps between measurement tools, to assess what the field overall is currently able to measure. This is concluded with an exploratory factor analysis on all items studied (177 in total), which captures the overall trend in concepts studied in the field of individual player difference.

Therefore, this work has three aims:

1. To explore the structure of a sample of individual player difference questionnaires and evaluate their reliability and construct validity
2. To understand the extent of overlap between items and sub-scales between questionnaires via exploratory factor analysis

3. To use exploratory factor analysis on all questionnaire items to analyse the current focus of the field and identify gaps in research

2. Background

2.1. *The Convergence of Different Theories of Behaviour*

When looking at work broadly concerned with investigating individual differences in players, there are many theories used to motivate the work, with distinct framings to be considered. This has been highlighted in the past, such as Hamari [7] who found many different perspectives to player attitudes when considering player typologies. Overall, the dominant types of framing for such work are as follows:

- Player Traits
- Player Preferences
- Player Motivations, including Needs Satisfaction

Player traits can be considered aspects of people that are generally stable over time, which influence behaviours (usually built upon psychological trait theory). This implies they can be useful in predicting whether certain people will like/dislike a given game or aspect of a game. For example, players who identify as having a ‘Narrative Orientation’ show significant positive correlations to enjoying role-playing elements in games [5]. Building on this, preferences refer to a person’s attitude to a given stimulus, usually when asked to make a decision about it [10]. These may be less stable than traits, as several factors could influence preference for a game. For example, whilst personality traits may influence game genre preference [11], so too could gender [12]. Therefore, preferences are more flexible and build upon a number of factors, rather than a singular ‘trait’.

Finally, motivations refer to players seeking out or wishing to avoid a specific experience in a game. A common motivation theory in games research is Self Determination Theory (SDT [13]), that posits players pursue games/aspects of games that can satisfy a basic human need (autonomy, competence, and relatedness). Players are motivated to fulfil these needs, which in turn will drive their gaming choices. Therefore, this perspective argues whilst the underlying motivations may be the same between players, the level of satisfaction within each need will vary.

Despite these quite different terms, which serve to emphasise different ways of conceptualising individual player differences, the practical measurement of these concepts shows a high degree of convergence. Each theory typically employs the use of questionnaires to assess the thoughts of players, meaning whilst the underlying stance of the work may differ, the methodology used is the same. By reducing these abstract concepts into concrete items on a scale, a process of reification has occurred. Because of this, the distinct differences found when considering these concepts at the abstract level have been removed, and the questionnaires become what is actually being measured.

For example, preferring games that give a sense of freedom is one theoretical explanation for behaviour (such as Autonomy/Exploration in the GAMES questionnaire [14]), and being motivated to fulfil a need for freedom is another (such as the Autonomy sub-scale from the PENS questionnaire [13]). If players are asked a question in a questionnaire about this, the item tends to appear similar. ‘I like games that offer you a lot of options and choices’ is semantically comparable to ‘The game provides me with interesting options and choices’, despite the former being derived from preferences, and the latter from a motivation/needs satisfaction perspective.

Because of this, researchers using any of these questionnaires across theoretical domains could do so interchangeably, as the items have become comparable. They may even do so with little consideration for the theory, and care only for what the questionnaire claims to measure. There is evidence of this from recent work by Tyack and Mekler [15], that found many researchers using SDT in HCI games research gave little consideration or explanation to the theory that led to its construction. Of 110 papers referencing SDT, only 27.72% explained how the theory motivated the work. This suggests it is commonplace to only lightly consider the theory behind the construction of a questionnaire, and so these differences in theory may not always be considered by researchers.

Therefore, the current study considers any questionnaire that claims to measure how players differ from one another as comparable, even if they refer to the differing terms of preference, trait, or motivation. The term ‘individual player differences’ is used for the remainder of this article to refer to any questionnaire used in this study, to improve ease of reading. An important note is that the individual player difference questionnaires discussed in the current work are different from player experience questionnaires. This is because the interest is in what drives players to *seek out* such experiences,

and not what they *actually experience* when playing. This drive (be it a trait, a preference, or a motivation), leads them to seek it out, but is not related to the feeling once within the desired experience. As such, questionnaires relating to the experience of playing games were not included.

2.2. Previous Work on Comparing Questionnaires

Once created, many questionnaires go through some form of validation, typically carried out within the original publication. Sometimes a follow-up study that further validates the scale is published, such as the online multi-player motivation scale created by Yee[16] that was validated in a follow-up study 6 years later [17]. Other times scales are validated by external authors, such as Busch et al. [18] who validated BrainHex created by Nacke, Bateman, and Mandryk [19]. However, to date there is little work comparing scales against each other, where attempts are typically restricted to theoretical overlappings. For example, Hamari [7] compared the existing player trait models for similarities in what they measured, to conclude the overarching themes that have been studied. This meta-analytic approach was useful to assess the current state of the field and highlight areas that were understudied. It did not however empirically compare the scales, instead relying on the wording of items and scale names. So while scales may show commonalities and differences semantically, it is not known how this manifests empirically in responses to these questionnaires.

An empirical approach is needed to address this question, as comparing item responses to different questionnaires from the same sample can highlight these suspected overlaps and divergences. This empirical approach has been used to compare questionnaires in the wider research field of player experience. For example, Denisova, Nordin and Cairns [20] compared the Player Experience of Needs Satisfaction (PENS; [13]), Game Engagement Questionnaire (GEQ; [21]) and the Immersive Experience Questionnaire (IEQ; [22]) scales, to assess how these measures of player experiences converged. They found generally high reliability for the scales, as well as moderate correlations between questionnaires. This indicates each questionnaire can be reasonably used to measure player experiences, though the authors concluded a unified questionnaire would be beneficial for the field. Johnson, Gardner and Perry [8] conducted similar work by comparing PENS to the GEQ, with the aim of validating these questionnaires. They generally supported the factor structures of the questionnaires, but indicated the sub-scales Competence and Intuitive Controls of PENS could be combined into one factor. The GEQ

was found to have a 5-factor solution, in contrast to the proposed 7-factor. The structure of both questionnaires could also be improved with the removal of several items from each.

Therefore, we have adopted a similar empirical approach. In this work we reanalysed individual player difference questionnaires alongside each other, not only to assess the factor structures of the standalone instruments, but also to identify the relationships between what they (claim to) measure.

2.3. A Historical Overview of Individual Player Difference Questionnaires

Measuring individual player differences has been an interest to the field for many years, traced back to the seminal work of Bartle in 1996 ([1]) to distinguish players into types. Questionnaires designed to capture these differences began to populate in 2006, with the introduction of three popular scales from three separate disciplines; Games ([16]), Psychology ([13]), and Media ([4]). The existence of these disciplines, and the evolution of the field overall, is described in detail in the companion paper ([3]), but a brief summary is provided below.

As identified by the companion paper ([3]), there are currently 4 main disciplines that seek to understand individual player differences — Games, Media, Psychology, and a more recent edition, Education. Games research considers what is unique to games and the behaviour of its players, Media attempts to map media theory to the domain of games, Psychology aims to map psychological theory to games, and Education wishes to understand what is engaging about games to create better learning environments. These disciplines have created unique scales to measure individual player differences, with little communication between them. For example, the scale by Nacke et al., ([19]) is inspired by Yee ([16]), whilst the Gaming Motivation Scale ([23]) was inspired by Ryan et al., ([13]). These represent Games and Psychology disciplines respectively, that make little reference to one another.

This raises the question as to what influence originating discipline has on the construction of questionnaires. It is expected that scales created within the same discipline will show similarities in constructs measured, but do questionnaires created in separate disciplines also share overlaps? This is important to consider, as concepts measured within a multidisciplinary field such as individual player differences may be distinct from one another, yet appear similar on the surface. For example, is the concept of Challenge the same as measured by scales from Games research, as that from Psychology

scales? If it is unclear how distinct the concepts measured are, there is confusion to what is actually being measured between papers, as there is no common language between disciplines to communicate with. This makes research synthesis across disciplines complex, and prone to misunderstandings. Conversely, if there are concepts that appear distinct but are actually highly convergent, more concepts have been created than are necessary. Is the concept of Diversion from Media research (e.g., [4]) the same as Escapism from Games research (e.g., [16])?

To answer these questions, it is important to assess how questionnaires across the whole spectrum of disciplines within individual player differences measure concepts. By assessing the measurement tool common between all 4 disciplines, it is possible to highlight differences and similarities between them, to inform what is currently measured by the field overall. Comparing the convergence in concepts can therefore be done whilst accounting for theoretical differences present due to the unique disciplines.

What these questionnaires measure, and how this overlaps with other questionnaires, is therefore the focus of this study. To do so, an exploration of their current structure is required to assess a baseline of reliability and validity. For this reason, previous validation attempts of the questionnaires sampled are included in the following section. The current work uses the literature review of the companion paper to provide the sample of questionnaires to empirically study. Further, the influence of originating discipline is explored and commented on where appropriate. The following section details which questionnaires from the companion paper were chosen for this study.

2.4. Questionnaires Used

This section describes first how the questionnaires were chosen for this study, followed by sections explaining each questionnaire. Due to the high volume of questionnaires available, the literature review of [3] was used to collect the sample of individual player difference questionnaires used in the current study. Their review found 23 questionnaires that measured individual player differences, which would be impractical to administer to participants. Instead half of these were chosen, leading to 11 questionnaires studied for a total of 177 items. These 11 were chosen for the following reasons:

- They represent a mixture of old and new questionnaires, to counteract a bias in newer papers building on top of pre-existing work and as a result being naturally more robust

- They represent questionnaires that originate from all four disciplines of games, media, psychology, and education
- They represent questionnaires not restricted to a specific genre/play type such as online games. The goal of this research is to understand individual player differences generally across all types of players, and these would be too specific. Two exceptions to this are the inclusion of Yee and VKS, where the reasons for this are explained in their respective sections.

The following questionnaires are presented in the order in which they were published, to provide a sense of how the field has evolved over time. The labels given for each questionnaire (highlighted in bold for emphasis) are used following this section.

*2.4.1. Video Game Uses and Gratifications, **Sherry***

One of the first motivation scales was created by Sherry et al. [4], known as the Video Game Uses and Gratifications Scale. This questionnaire emerged from the field of Media Effects, which sought to explain why games were appealing and what players seek from them. It consists of 20 items that fall into 6 uses and gratifications — Competition, Challenge, Social Interaction, Diversion, Fantasy and Arousal. Research has used this questionnaire to study a wide range of variables, such as how play of *The Sims 2* varied by gender [24], and how consumption varies between traditional sports and esports [25]. Therefore, this questionnaire is included to represent the Media perspective to understand what players want from an experience and how this influences their game preference.

*2.4.2. Motivations for Play in Online Games, **Yee***

Yee [16] created the Motivations for Play in Online Games scale in the same year as Sherry, inspired by the player types discussed by Bartle [1]. Sampling a large sample size of over 3000 players, Yee found 3 overarching motivations with 10 sub-factors for play; Advancement, Mechanics, Competition, Socialising, Relationship, Teamwork, Discovery, Roleplaying, Customisation, and Escapism. These were revised in 2012 [17] and sorted into 3 overarching motivations. Research has used both version of the questionnaire, such as Kuss, Louws and Weirs [26] who used the original scale to measure how motivations to play relate to addictive play behaviours, and Billieux et al. [27] who sought to understand the specific motivations to play

World of Warcraft. Yee was chosen for this study because it represents one of the first attempts to measure player traits from a games research perspective. The questionnaire published in 2006 ([16], though not to be confused with [28]) was chosen over the revised 2012 version ([17]) because it has seen more usage in the field than the revised set ([3]), and has a wider range of different traits that may provide more nuanced insights.

2.4.3. *Player Experience of Needs Satisfaction, **PENS***

The Player Experience of Needs Satisfaction (PENS) questionnaire ([13]), also originally published in 2006, is widely used in player research. It is inspired by Self Determination Theory (SDT), which posits that motivation to perform actions can be summarised as a desire to fulfil a number of innate ‘needs’, which is done so in a self-directed way. These needs drive behaviour, and so can explain why people perform actions on a generalised, macro level. The PENS questionnaire therefore comes from the field of Psychology, which reflects its backing by a psychological theory of needs satisfaction. It originally consisted of 5 sub-scales from 21 items; Competence, Autonomy, Relatedness, Presence (though sometimes referred to as measuring Immersion [13]; for the purpose of this paper it is called Presence/Immersion), and Intuitive Controls. This theoretical backing has allowed researchers to extract what about games players are seeking from a play session. For example, those who showed a favouritism towards sport/simulation games tended to score low on the Autonomy and Presence/Immersion sub-scales [29]. PENS is arguably one of the most commonly used questionnaire in player research — recent work by Tyack and Mekler [15] showed how of 110 papers submitted to CHI and CHIPLAY conferences that referenced SDT, 40% used some or all of the PENS scale.

PENS has been validated independently by other researchers in the past (e.g., [29]), though the conclusions drawn from these studies are not always in agreement. Brühlmann and Schmid [9] found the factor structure of PENS to be consistent with the original study, but in 2 separate works by Johnson and Gardner ([29], [8]) the sub-scales Competence and Intuitive Controls required combining. This disagreement in factor structure warrants further investigation, and so will be investigated in this study.

However, PENS is not an individual player difference questionnaire per se, as it assesses underlying needs satisfaction of players. Tondello et al. [5] mentions in their discussion that PENS, being a general theory, cannot explain player preferences. This is because it is meant to assess the underlying

psychological needs of people, something that influences multiple activities in a persons life, rather than being game-specific. Yet the concept of ‘needs’ is similar to traits, in that players are motivated to fulfil some desire in their gaming. If a player has a trait or preference for something they experience within a game, and they have a need that needs to be satisfied, there is little semantic difference between them. Furthermore, there are examples of researchers using the PENS scale as if it is an individual player difference questionnaire (i.e., to understand player preferences; [3]), so including it is not outside current practice. By including PENS alongside these questionnaires, the semantic overlap can be assessed, to show either the items do indeed address the same concepts, or PENS (and by extension needs satisfaction) is in fact a separate construct.

Therefore, this study includes PENS in the analysis. However, as the PENS questionnaire is a commercial product, the items cannot be published in this paper. Instead, these are reported in relation to their presentation in the 2007 copy of the PENS documentation — for example, the first item in the Competence sub-scale is referred to as Competence 1.

2.4.4. *Gaming Motivation Scale, **GAMS***

Inspired by Self Determination Theory (SDT) and by extension the PENS scale, Lafrenière, Verner-Filion and Vallerand [23] developed the Gaming Motivation Scale in 2012. The questionnaire consists of 18 items across 6 sub-scales; Intrinsic Motivation, Integrated Regulation, Identified Regulation, Introjected Regulation, External Regulation, and Amotivation. Studies have used this scale to understand how consumer motivation relates to player engagement [30], how the experience of need frustration mediates the link between gaming motivation and problematic gaming [31], and why cancer patients are motivated to play games [32], as a few examples. The questionnaire has not been validated by others in its original form, though an adapted Italian version did show adequate validity and reliability [33]. Therefore this questionnaire is included to serve as a test of the original sub-scales, and to investigate how GAMS differs from PENS/SDT, the inspiration for the items.

2.4.5. *Gaming Attitudes, Motives, and Experiences Scale, **GAMES***

Though similar in name and in discipline to GAMS (as both come from Psychology), the GAMES scale by Hilgard, Engelhardt and Bartholow [14] is instead inspired from Sherry. It consists of 59 items in 9 sub-scales;

Story, Violent Catharsis, Violent Reward, Social Interaction, Escapism, Loss-Sensitivity, Customisation, Grinding, and Autonomy. As the sub-scale names may suggest, the authors were motivated to understand problematic behaviours such as violence, a topic too large and contentious to examine in this paper (see [34] for an overview). Uses of this scale include investigating how factors such as age and ethnicity affect gaming preferences [35], as well as how Native English speakers differ from non-native speakers in motivations to play [36]. There has been little validation work on this questionnaire, and so this paper will investigate its structure, as well as how the influence of Sherry has shaped the work in comparison to other questionnaires from the discipline of psychology.

2.4.6. *BrainHex*

Following on from Yee, the BrainHex scale was created by Nacke, Bateman and Mandryk [19], that aimed to provide a measurement of traits from a top-down approach. This was done by relating the theory behind it to neurobiological findings. It consists of 21 items in 7 sub-scales; Seeker, Survivor, Daredevil, Mastermind, Conqueror, Socialiser, and Achiever. An example of its usage is correlations of scores on BrainHex to attention as measured with eye-tracking data [37]. Whilst this questionnaire has been used in research, it has not been without critique. Validation studies have struggled to replicate the validity and reliability of the structure. Busch et al. [18] found that whilst they could confirm the structure of the Socialiser and Achiever factors, the other 5 were unclear and contained many cross-loaded items. Therefore, this study includes BrainHex to assess these structural concerns, as well as compare it to other questionnaires from Games research.

2.4.7. *Player Motivations in Massively Multiplayer Online Games, VKS*

Voulgari, Komis and Sampson [6] was similarly inspired by the work of Yee, but wished to situate this within the context of learning. It consists of 25 items in 5 sub-scales; Sociability, Competition, Immersion, Achievement, and Challenge. Whilst this questionnaire is not as widespread or actively in use, it provides a comparison to BrainHex — both were inspired by Yee, but originate from different disciplines, and so how these emerged differently is of interest. Furthermore, the VKS questionnaire was not validated by the original authors, and so this study aims to explore the structure.

2.4.8. *The Trojan Player Typology, **Trojan***

The questionnaire developed by Kahn et al. [38] is different in that, unlike the others mentioned so far, it draws inspiration across disciplines. Its items are built both from Yee and Sherry, and so combines the Games research literature with the research on Media. There are 15 items in 6 sub-scales; Socialiser, Completionist, Competitor, Escapist, Story-Driven, and Smarty-Pants. As the original sample used for this study was *League of Legends* players, studies using this questionnaire typically study related topics. For example, studies have looked into the motivation of esports players [39], and the motivations to play online collectable card games [40]. As a fusion of disciplines, the Trojan questionnaire is included in this study, to assess whether this joint influence affects the structure of the sub-scales and the concepts it purports to measure.

2.4.9. *Trait Model and Scale of Game Playing Preferences, **Tondello***

The questionnaire developed in 2019 by Tondello et al. [5] is the latest iteration in the chain of questionnaires from the Games research discipline. Inspired by both Yee and BrainHex, the questionnaire has 25 items and 5 sub-scales; Aesthetic, Narrative, Goal, Social, and Challenge. As this questionnaire has only been published recently it was yet to be used by other academics at the time of data collection, and by extension has not been validated by others. Therefore, Tondello is included in this study to explore the structure, and because it the latest iteration from the field of Games research. The concepts measured are therefore expected to be clearer.

2.4.10. *Five-Factor Inventory of Intrinsic Motivations to Gameplay, **IMG***

Similarly to Trojan, the IMG scale by Vahlo and Hamari [7] draws inspiration across three research fields; Games, Media, and Psychology. It consists of 15 items in 5 sub-scales; Relatedness, Competence, Autonomy, Fun, and Immersion. As another example of a recently published work there is little usage of it in the field, and so this study aims to explore its structure. This work also seeks to explore if it presents a clearer conceptualisation of individual player differences, by virtue of being newer and based on multiple disciplines.

2.4.11. *Video Game Pursuit Scale, **VGPu***

Finally, the questionnaire by Sanchez and Langer [41] is included in the analysis. Similarly to VKS, this questionnaire was built to understand mo-

Figure 1: Ancestry tree of the questionnaires assessing individual differences used in this study.

tivation in learning environments. It consists of 19 items in 4 sub-scales; Intentional Game Play, Generalised Game Self-Efficacy, Enjoyment of Games, and Prone to Game Immersion. Interestingly however, the study did not use any of the previously mentioned studies as a source of their inspiration for item generation. Therefore, whilst this questionnaire is the newest in the set, it is included in this study to assess how the items will perform in relation to the others, and see if there is overlap still.

2.5. Summary

As discussed above, many have found inspiration from previous questionnaires, highlighting how the field has constantly adapted and built upon itself. This ‘ancestry’ has influenced the development of the questions asked, and a diagram of these influences is presented in Figure 1.

Note that this diagram does not indicate *all* avenues of inspiration, and instead focuses solely on questionnaires referenced during the item creation process. For example, Yee is inspired by Bartle [1], but as the Bartle paper was not a questionnaire, it is not included in this diagram. Furthermore, Tondello uses SDT (the theory behind PENS) to explain the sub-scales cre-

ated, and GAMES discusses all three questionnaires published in 2006 in the background section. However, Tondello did not use or reference inspiration from the items of PENS in the creation process, and GAMES only used the items of Sherry. As these other questionnaires are not mentioned during item creation, their influence lines are not included, leading to a simplified influence diagram.

Furthermore, this diagram not only helps predict where similarities will occur between questionnaires, but whether these similarities are related to the disciplines that inspired them. All 4 research disciplines described by the companion paper ([3]) are present within this ancestry tree, allowing for the influences between them to be assessed; these are Games Research, Media Effects, Psychology, and Education. Yee, BrainHex, Tondello and Trojan fall under Games Research, as they primarily build on the literature of those wishing to understand game players. Sherry forms research from Media Effects, that build on the literature that explores how media can affect consumers (typically by transferring knowledge from TV media effects to games). PENS and GAMS form the research from Psychology, building on psychological theories of human behaviour to explain players. VGPU and VKS are derived from Education research, which aim to understand what about games is engaging so this can be used to create better learning environments. Interestingly, whilst VKS was inspired by Yee, VGPU shares no inspiration from any previous questionnaires studied here. Finally, IMG is unique in that it explicitly shares disciplines of influence, thereby potentially bridging the gaps between them.

Therefore, using this diagram we could predict there will be strong similarities between questionnaires that come from the same research disciplines. For example, Games Research questionnaires and their sub-scales should cluster together, as should the Media, Psychology, and Education questionnaires. This question is explored in Section V of the results, whilst exploring how items relate to one another. Further, because of the natural building on top of previous questionnaires, it may also be expected that newer questionnaires may perform better in terms of reliability and validity. For example, it may be that GAMES is a more robust questionnaire than Sherry, as it should have taken inspiration and improved on the issues found in the original. This is explored in Section I of the results, whilst discussing the reliability of these questionnaires.

Therefore, this study aims to perform a general assessment of the structure of each questionnaire to provide a baseline for comparison. Afterwards,

an exploration of how these questionnaires relate to one another is performed, making reference to the ancestry diagram and its hypothesised influence.

3. Method

3.1. Sub-Scale Selection

The previous section has explained the reasons for selecting each of the 11 questionnaires for this study. However, even by narrowing down the analysis to these specific questionnaires, there are still an excessive amount of items for participants to fill out. If all of the items were used from these questionnaires, there would be 283 items in 68 sub-scales. This would be a challenging survey to complete for many participants, which might lead to higher rates of attrition and fatigue. To reduce this, a number of sub-scales were removed from the sample and not placed into the survey. The decision was made to remove sub-scales instead of individual items, so that the structure of the sub-scales would not be altered. Omitting sub-scales is common practice in the field of individual player differences, as it is unlikely that new research requires all concepts to be measured in a specific research context ([3]). The types of sub-scales removed, and why this decision was made, is presented below.

3.1.1. Social Gaming

Social aspects of gaming are widely studied in the literature, and almost all the questionnaires sampled had at least one sub-scale related to how players feel towards other players. This makes sense, as in the meta-analysis of player types by Hamari [7] Sociability was one of the most common concepts in the literature. These could be assessing cooperative behaviour (e.g. Socialiser by BrainHex), competitive (e.g. Competition by Yee), or friendship formation (e.g. Social Interaction by GAMES).

This interest is understandable, as the social aspect of games is seen as fun and important to the gaming experience [42], sometimes the most important reason for players to engage with a game [43]. The social aspect does not even need to be purely cooperative, as socialising is considered fun even in competitive settings [44].

However, socialising is not present in all games. Indeed, most of the literature surrounding social play focuses specifically on Massively Multiplayer Online Games (MMOGs), such as *World of Warcraft*. Not all players may partake in these multiplayer games, and so would not be able to answer these

questions. Furthermore, many of the social sub-scales specifically ask about the relationships players form in these games. This is not related to individual player differences per se, as it asks them to reflect on the relationships they form within games. As this was not the focus of the questionnaires chosen, and as some players may exclusively play single-player games, we decided to remove any sub-scale that referenced the social aspects of games (even if only one or two items of the scale mentions other players). The removed sub-scales are as follows:

- Yee: Advancement, Competition, Socialising, Relationship, Teamwork
- Tondello: Social
- Trojan: Socialisers, Competitors
- VKS: Sociability, Competition
- GAMES: Social Interaction, Loss Aversion
- BrainHex: Conqueror, Socialiser
- Sherry: Competition, Social Interaction
- IMG: Relatedness
- VGpu: Intentional Pursuit of Game Play, Intimidation with Games
- PENS: Relatedness

Overall, this helped to make the questionnaire more appropriate for those that do not play multiplayer games, and reduced the number of overall items from 283 to 208.

3.1.2. Violence

Some questionnaires within the sample contain sub-scales that assess the extent to which players enjoy/engage with violent content, and why this may be. However, the topic of violence in video games is contentious. A brief overview will be given of the gaming violence literature to position the current work, but for further information see [34].

Generally speaking, the main concern of violent video games is that violent content exposure will lead to more violence/aggression in real life, especially amongst children. This so-called priming would mean those who play

violent games for an extended period of time would be more likely to commit violent acts in real life, as the behavioural response has been ‘primed’ (see [45] for an overview). However, many researchers dispute this link as being causal. Cross-sectional studies on longitudinal data show no short or long term predictive power of violent games on behaviour [46], and fMRI data shows no link between excessive use of violent video games and a decrease in empathy responses [47]. The influential power of violence in video games is an ongoing debate, and one that does not appear to be resolved in the near future.

Another consideration is that players may be aware of the research surrounding violence and video games, due to widespread media coverage around the topic. Players may take offence to any item asking them about this connection in their own play, and so their inclusion could provide unnecessary discomfort.

Therefore, unlike the other sub-scales relating to individual player differences, violence is uniquely contentious to both researchers and players. To avoid this debate, we decided to remove any sub-scale referring to violence. In the questionnaires selected for analysis, only GAMES did so in 2 sub-scales; Violence Catharsis, and Violent Reward. Removing these scales reduced the number of items from 208 to 195.

3.1.3. Non-Individual Difference Sub-Scales

The final sub-scales removed for analysis were those concerned with concepts outside of individual player differences. For example, Intuitive Controls from PENS is about the experience of the controls of a game, and not about a quality the player possesses. Therefore, this sub-scale was removed.

Another sub-scale removed from the analysis was Amotivation from GAMES, as this assesses players who feel they no longer feel motivated to play games. For example, ‘I used to have good reasons, but now I am asking myself if I should continue’ asks players to reflect on why they don’t wish to play anymore. This would be counter-intuitive to add to the item pool, as the goal was to assess players who had motives to play.

The final sub-scales removed were from VGPU: Intentional Pursuit of Game Play, and Intimidation with Games. Similarly to the above reasons, these sub-scales assessed concepts outside of individual player differences. The item ‘I deliberately seek out video games to play’ from the Intentional Pursuit of Game Play sub-scale asks players about their gaming habits generally rather than what they enjoy about games specifically. Further, the item

‘Video games are intimidating to me’ from the Intimidation with Games sub-scale assesses if players are reluctant to play due to a fear of games, which is a perception of games rather than a reason for engaging with gaming aspects. Neither of these sub-scales therefore assess individual player differences, and so were removed from the analysis.

With their removal, the item pool decreased from 195 to the final selection of 177.

3.1.4. Summary

Overall, these changes helped reduced the item list from 283 to 177, and the number of sub-scales from 68 to 45. This is still a large item set, but less so than the original, and so more manageable for participants. Table 1 displays the questionnaires used, and the sub-scales selected for analysis.

Table 1: A table of the questionnaires used, and the sub-scales used in this study.

Questionnaire	Sub-Scales Used
Tondello [5]	Aesthetic, Narrative, Goal, Challenge
Yee [16]	Mechanics, Discovery, Role-Playing, Customisation, Escapism
Trojan [38]	Completionist, Escapist, Story Driven, Smarty Pants
VKS [6]	Immersion, Achievement, Challenge
BrainHex [19]	Seeker, Survivor, Mastermind, Daredevil, Achiever
GAMES [14]	Story, Escapism, Customisation, Grinding/Completion, Autonomy/Exploration
PENS [13]	Competence, Autonomy, Presence/Immersion
Sherry [4]	Challenge, Diversion, Fantasy, Arousal
GAMS [23]	Intrinsic Motivation, Integrated Regulation, Identified Regulation, Introjected Regulation, External Regulation
IMG [48]	Competence, Immersion, Fun, Autonomy
VGPu [41]	Generalised Game Self-Efficacy, Enjoyment of Games, Prone to Game Immersion

3.2. Item Preparation

The 177 items were compiled into one Qualtrics questionnaire, and the order of items was randomised to avoid fatigue in later answers skewing a particular questionnaire. Due to the mix of 7-point and 5-point Likert scales present in the original scales, questions were refactored to all be 7-points for consistency. This refactoring is valid, as it does not significantly influence how participants respond (as explained in [49]), and is a common practice in the field of individual player differences ([3]).

A further inconsistency between questionnaires was the wording of items, which may introduce unnecessary processing load on participants. For example, most items refer to ‘games’ generally, except for VKS and PENS which refer to a specific game. To standardise this, items were reworded to read in the same style; agree/disagree statements that refer to games generally (to relate to player’s thoughts on games as a whole). This was done whilst being mindful to not change the meaning of the items. As an example, the first item in the Mechanics scale of Yee is worded as ‘How interested are you in the precise numbers and percentages underlying the game mechanics?’. This item was rewritten to be, ‘I am interested in the precise numbers and percentages underlying the game mechanics’.

Overall, rewording was done with minimal adjustments (see <https://osf.io/95a4c/> for a comparison), where stems were changed to be consistent for an agree/disagree Likert scale, and items rephrased to reflect on games generally rather than on a specific gaming experience. This allowed for a coherent flow in items which increased the ease of reading for participants. Consequently, this reduced the risk of fatigue and attrition, as well as less confusion which could bias results.

This does mean the items used are not identical to those in the original papers. However, it is common practice to reword items in such a way — 39% of papers using individual player difference questionnaires explicitly reworded items, where only 6% of papers made it explicitly clear that items had not been reworded and 55% not providing information ([3]). This rewording varies from simple changes for clarity (e.g., specifying the game in the items; [50]), to fully rewriting items (e.g., [51]). Small degrees of rewording do not always affect the statistics of questionnaires ([49]), though it is advised to be done as little as possible to maintain validity. Therefore, whilst rewording items in this study reduces the ability to infer the reliability and validity of the original scales, as this is a common practice it does inform on how questionnaires perform in the ways they are used. This makes the findings of this work comparable to previous uses of questionnaires.

Overall, the wording was changed for 49 out of 177 items. These are from Yee, VKS, BrainHex, PENS and GAMS. A comparison of the original items to the reworded items used in this study can be found at <https://osf.io/95a4c/>.

3.3. Procedure

Ethical approval was first gained from the Computer Science department of [Blinded for Review]. The questionnaire was distributed through websites

such as *Reddit* (posted on the subreddits *r/SampleSize* and *r/TrueGaming*), *Facebook* (participant survey groups) and *Twitter*. When the link was clicked, participants were briefed on what they were expected to do, with a specific focus on mentioning the length of the survey. This was to encourage them to decide if it would be too much work up front, to reduce attrition. To thank participants for their time, upon completion of the questionnaire participants were invited to enter a prize draw to win one of 20 Steam or Amazon vouchers worth £20 each.

After consent was given, participants were asked to fill in information about their gaming habits (time spent playing games and average weekly play time), as well as their age and gender. Participants were then asked to complete the questionnaire items, with an attention question at the halfway point ('Please tick the box "Second" before continuing to the next set of questions.'). After completing all items, participants could opt to provide an email address to take part in the prize draw. All analysis was completed using R version 4.1.0 ([52]).

The data collected in this study is available for download, along with further analysis conducted not covered in the following result sections. This information can be found at <https://osf.io/95a4c/>.

4. Results

4.1. Data Cleaning

To prepare data for analysis, a number of cleaning steps were performed. Table 2 illustrates this process, along with the number of participants left after each step was completed. An initial 3000 participants responded, and after cleaning, 1683 responses remained. Exclusion criteria included those that did not complete the questionnaire, insincere responses (answering strongly disagree to all items for example), failing the control question, and answering under 18 for participant age. Specifically for the last group, a number of participants (9 in total) answered no to the age question, thereby ending the survey. However, they then proceeded to complete the questionnaire by answering yes within a minute of being taken to the end of the survey. Because of this, they were removed from the analysis as it was unclear how old they truly were, thereby breaking the ethical concern of involving minors.

Table 2: Data cleaning steps undertaken to achieve the sample used.

Removal Steps Taken	Participants Remaining
Starting sample	3004
Did not complete survey	1757
Did not pass control question	1748
Answered under 18	1693
Fake answers (e.g. all 1s)	1684
Answered under 18, then over 18	1683
Final sample	1683

4.2. Demographics

The gender of the sample is as follows: 220 females (13.07%), 1353 males (80.39%), 35 nonbinary (2.08%), 67 preferred not to answer/gave no response (3.98%), and 8 gave insincere responses (such as objects or profanities; 0.48%). The length of gaming habits are as follows: 2 had played for less than a year (0.12%), 18 for 1-5 years (1.07%), 106 for 5-10 years (6.30%), 318 for 10-15 years (18.89%), 474 for 15-19 years (28.16%), and 765 had played for 20+ years (45.45%). In relation to their per week hourly gaming, 12 played for less than an hour (0.71%), 57 for 1-2 hours (3.39%), 148 for 3-4 hours (8.79%), 146 for 4-5 hours (8.67%), 181 for 5-6 hours (10.75%), and 1339 for 6+ hours (67.68%). On average, participants took 29.37 minutes to complete the survey (median = 18.05). Overall, the sample was heavily skewed towards being well-experienced male gamers that played for multiple hours a week.

4.3. PART I: Reliability

The aim of this section is to assess the reliability of the questionnaires. This is required to provide a baseline of how likely it is the questionnaire can measure the concepts they intend to. Firstly, this section explains what is meant by ‘good’ reliability, and how it has been considered in the current work.

Reliability is the degree to which any measure provides the same results when applied to the same situation. In the context of personality questionnaires, this typically means statistical reliability or internal consistency; the

degree to which the different items of a scale give a consistent indication of the overall value of the scale. Typically, this is measured using Cronbach's Alpha. A strength of calculating Alpha is the analysis can reflect on individual items in a sub-scale, as it highlights how the Alpha changes if an item is dropped. This helps to understand the importance of each item for achieving reliability. For example, some items may be unrelated to the other items, and so their inclusion would reduce the overall reliability. By removing such items, the overall reliability can be improved.

However, calculating Alpha scores requires that the items and scale behave according to Classical Test Theory [49]. This is not always the case, and in particular not for some of our sub-scales which showed significant skew. We therefore also use the more robust Omega measure [53], to help in interpreting statistical reliability. Alpha and Omega scores typically agree, especially when using a 95% Confidence Interval (CI). Indeed, in all cases except 2 Alpha and Omega are broadly in agreement to within 0.10 of each other. These exceptions are Integrated Regulation from GAMS, and Competence from IMG. In both cases, Omega is higher than Alpha, which could be due to skew in the distributions of individual items in the scales. There is evidence of this, where a full analysis of these scales can be found at <https://osf.io/95a4c/>. This highlights the need to consider the individual distributions of items within scales, as these can help explain reliability scores.

When to consider a score an indication of 'good' reliability is dependent on the context the sub-scale is to be used in. Nunally and Bernstein [54] suggest a 0.70 cutoff as the minimum for when a scale becomes reliable, but for use as a 'basic research tool' this ideally should be higher than 0.80. This should be increased to above 0.90 for clinical settings, as concluding diagnostic categories based on a sub-scale requires more stringent reliability. Not all researchers agree on this, and instead argue any sub-scale above 0.90 is too high, because this more likely reflects upon the items being too similar to one another, which would make their inclusion redundant [55]. When items are too similar to one another, this can lead the scale to be a 'bloated specific' — asking participants the same question slightly reworded each time can create an unintentional sub-factor where these items cluster [56]. This creates two problems. Firstly, if there are too many of these similar items in a sub-scale, this can artificially inflate the reliability, as the strong correlations between these items will skew the overall reliability. Secondly, these items in a sense change the nature of the sub-scale to be about the item, and not the concept it is supposed to measure. This is a problem more of validity, as

the sub-scale is no longer measuring what it is predicted to. Relatedly, high reliability can arise when there are a small number of items in a sub-scale. Catell [56] goes on to explain how it is unlikely a sub-scale with a handful of items (for example, less than 3) can accurately capture something as complex as human behaviour.

Therefore, for the purpose of this paper any sub-scale above 0.70 is considered moderately reliable, while those above 0.80 indicate strong reliability. Further, any sub-scale with a low number of items is considered to likely be an inaccurate measurement of the concept it is associated with. Table 3 shows the alpha and omega scores for each scale.

Out of the 45 sub-scales, 28 were reliable over 0.70 on either the Alpha or Omega tests, and 14 above 0.80. This does not necessarily mean the remaining 18 sub-scales (approximately 40%) are not still useful, as each item within the sub-scale may load together well. They will however require closer inspection to confirm this. Only two questionnaires had all sub-scales returned with reliability over 0.70; Tondello and GAMES. This indicates these may be the most reliable of the questionnaires, though further testing is required.

Due to the large amount of questionnaires and sub-scales analysed, it is not possible to explore in detail each reliability test for each sub-scale. Therefore, a selection of sub-scales has been chosen that highlight the range of outcomes; each section contains an example of a highly reliable questionnaire, a weakly reliable questionnaire, as well as an ‘averagely’ reliable questionnaire. These show the general trends in the reliability of the sub-scales, that gives an overview of these questionnaires. For a full analysis of the reliability of all questionnaires, go to <https://osf.io/95a4c/>.

4.3.1. Highly Reliable Questionnaire; Tondello

As shown in Table 3, each sub-scale in this questionnaire has an overall reliability of over 0.73, indicating each item is likely to be relevant to its sub-scale. These tests for the Tondello items can be found in Table 4. From this it can be seen that all the sub-scales apart from Narrative can be improved by removing an item (as the overall scale reliability increases if these items are dropped).

For example, removing ‘I like to customise how my character looks in a game’ from Aesthetic improves the reliability from 0.73 to 0.75. This makes sense, as the other items appear to specifically discuss the gameworld and not customisation. Removing ‘I like finishing quests’ from Goal improves the

Table 3: Reliability tests for all questionnaire scales (95% CIs in brackets). Scales in bold highlight scores over 0.70 on either Cronbach’s alpha (α) or omega (ω) tests.

Questionnaire	Scale	Alpha	Omega
Yee	Mechanics	0.66 (0.63-0.69)	0.66 (0.63-0.69)
	Roleplaying	0.73 (0.72-0.75)	0.79 (0.77-0.81)
	Discovery	0.69 (0.66-0.71)	0.69 (0.66-0.72)
	Customisation	0.76 (0.74-0.78)	0.76 (0.74-0.78)
	Escapism	0.61 (0.58-0.64)	0.68 (0.65-0.71)
Tondello	Aesthetic	0.73 (0.70-0.75)	0.72 (0.69-0.75)
	Narrative	0.84 (0.82-0.85)	0.84 (0.82-0.86)
	Goal	0.81 (0.80-0.82)	0.84 (0.83-0.85)
	Challenge	0.87 (0.86-0.88)	0.88 (0.87-0.89)
Trojan	Completionist	0.62 (0.59-0.65)	0.66 (0.62-0.69)
	Escapist	0.79 (0.77-0.81)	0.79 (0.76-0.82)
	Story Driven	0.77 (0.75-0.79)	-
	Smarty Pants	0.74 (0.71-0.76)	0.74
VKS	Immersion	0.66 (0.63-0.68)	0.67 (0.64-0.70)
	Achievement	0.73 (0.71-0.75)	0.77 (0.75-0.79)
	Challenge	0.53 (0.49-0.56)	0.58 (0.54-0.62)
GAMES	Story	0.88 (0.87-0.89)	0.88 (0.87-0.89)
	Escape	0.86 (0.85-0.87)	0.88 (0.87-0.89)
	Customisation	0.87 (0.86-0.88)	0.88 (0.87-0.89)
	Grinding/Completion	0.81 (0.79-0.82)	0.81 (0.80-0.83)
	Autonomy/Exploration	0.75 (0.73-0.77)	0.76 (0.74-0.78)
PENS	Competence	0.71 (0.69-0.74)	0.72 (0.69-0.75)
	Autonomy	0.65 (0.62-0.68)	0.66 (0.62-0.70)
	Immersion	0.87 (0.86-0.88)	0.88 (0.87-0.89)
BrainHex	Seeker	0.63 (0.59-0.66)	0.64 (0.60-0.68)
	Survivor	0.45 (0.40-0.49)	0.45 (0.41-0.50)
	Mastermind	0.63 (0.60-0.66)	0.64 (0.60-0.68)
	Daredevil	0.55 (0.52-0.59)	0.56 (0.52-0.60)
	Achiever	0.87 (0.86-0.88)	0.88 (0.87-0.89)
Sherry	Challenge	0.60 (0.57-0.63)	0.60 (0.57-0.64)
	Diversion	0.88 (0.87-0.90)	-
	Fantasy	0.83 (0.81-0.84)	0.83 (0.81-0.84)
	Arousal	0.68 (0.65-0.70)	0.68 (0.66-0.71)
GAMS	Intrinsic	0.50 (0.46-0.55)	0.51 (0.46-0.55)
	Integrated	0.58 (0.55-0.62)	0.70 (0.65-0.75)
	Identified	0.71 (0.68-0.73)	0.75 (0.72-0.77)
	Introjected	0.77 (0.75-0.79)	0.78 (0.76-0.80)
	External	0.58 (0.54-0.61)	0.58 (0.54-0.62)
IMG	Competence	0.57 (0.53-0.60)	0.68 (0.62-0.74)
	Immersion	0.72 (0.70-0.74)	0.73 (0.70-0.75)
	Fun	0.62 (0.59-0.65)	0.62 (0.58-0.67)
	Autonomy	0.80 (0.79-0.82)	0.80 (0.79-0.83)
VGPU	Generalised Game Self-Efficacy	0.83 (0.82-0.84)	0.85 (0.83-0.86)
	Enjoyment of Games	0.87 (0.86-0.88)	0.86 (0.84-0.89)
	Prone to Immersion	0.64 (0.62-0.67)	0.68 (0.65-0.71)

Table 4: The Cronbach’s alpha scores for each item in the Tondello questionnaire, as well as the overall sub-scale reliability. Bolded sub-scale names indicates overall reliability above 0.70. Bolded alpha scores indicate scores that would improve the overall reliability if dropped.

Scale	Item	Alpha if dropped	Overall Alpha
Aesthetic	I like games with detailed worlds or universes to explore	0.66	0.73 (0.70-0.75)
	I like games which make me feel like I am actually in a different place	0.66	
	I often feel in awe with the landscapes or other game imagery	0.67	
	I like to customize how my character looks in a game	0.75	
	I like to spend some time exploring the game world	0.66	
Narrative	I feel like storytelling often gets in the way of actually playing games	0.81	0.84 (0.82-0.85)
	Story is not important to me when I play games	0.79	
	I like games that pull me in with their story	0.80	
	I enjoy complex narratives in a game	0.81	
	I usually skip the story portions or the cutscenes when I am playing	0.81	
Goal	I like finishing quests	0.83	0.81 (0.80-0.82)
	I feel stressed if I do not complete all the tasks in a game	0.79	
	I usually do not care if I do not complete all optional parts of a game	0.75	
	I like completing games 100%	0.74	
	I like to complete all the tasks and objectives in a game	0.73	
Challenge	I like it when games challenge me	0.85	0.87 (0.86-0.88)
	I like it when progression in a game demands skill	0.85	
	I enjoy highly difficult challenges in games	0.81	
	I like it when goals are hard to achieve in games	0.83	
	I usually play games at the highest difficulty setting	0.89	

sub-scale from 0.81 to 0.83, and as it is the only item not to use the word ‘complete’, this may be why. Another reason is that ‘quests’ is a vague example of a task, whereas the other items are more specific. Finally, removing ‘I usually play games at the highest difficulty setting’ from Challenge improves the sub-scale from 0.87 to 0.89, a small increase for a question that appears semantically similar on the surface.

However, if the items of the Challenge sub-scale are looked at closer, the final item does not show the same distribution of responses to the others, shown in Figure 2. In fact, the final item has a bimodal distribution, in that players either agreed they played on the highest difficulty, or not at all. This highlights the importance of looking closer at items to see if unusual distributions are at play, and also justifies its removal from the sub-scale¹.

Overall, even though Tondello is one of the most reliable questionnaires analysed, it can still be improved by removing certain unreliable items. With the improvements made, this questionnaire shows promise of being a useful

¹Indeed, the other scales in this questionnaire do not show this difference in distributions; see the full analysis at <https://osf.io/95a4c/> for more information

Figure 2: Individual histograms of each item from the Tondello Challenge sub-scale.

way to measure player traits.

4.3.2. *Averagely Reliable Questionnaire; Yee*

As an example of a questionnaire which shows both high reliability in some sub-scales but low reliability in others, Yee has been selected. 2 out of 5 of the sub-scales have Alpha scores over 0.70 (Roleplaying and Customisation), with Discovery just under at 0.69, as shown above in Table 3. Table 5 shows the reliability scores for each individual item.

Table 5: The Cronbach’s alpha scores for each item in the Yee questionnaire, as well as the overall scale reliability. Bolded scale names indicates overall reliability above 0.70. Bolded alpha scores indicate scores that would improve the overall reliability if dropped.

Scale	Item	Alpha if dropped	Overall Alpha
Mechanics	I often use a character builder or a template to plan out my character’s advancement at an early level	0.67	0.66 (0.63-0.69)
	It is important that my character is as optimised as possible for their profession/role	0.58	
	It is important for me to know as much about the game mechanics and rules as possible	0.54	
	I am interested in the precise numbers and percentages underlying the game mechanics	0.58	
Roleplaying	I often make up stories and histories for my characters	0.61	0.73 (0.72-0.75)
	I enjoy being immersed in a fantasy world	0.74	
	I often role-play my character	0.57	
	I enjoy trying out new roles and personalities with my characters	0.71	
Discovery	I enjoy finding quests, NPCs or locations that most people do not know about	0.63	0.69 (0.66-0.71)
	I enjoy exploring every map or zone in the world	0.57	
	I enjoy exploring the world just for the sake of exploring it	0.57	
	I enjoy collecting distinctive objects or clothing that have no functional value in games	0.72	
Customisation	It’s important that my character’s armour/outfit matches in colour and style	0.68	0.76 (0.74-0.78)
	It’s important that my character looks different from other characters	0.69	
	I spend a lot of time customising my character during character creation	0.65	
Escapism	It’s important that games allow me to escape from the real world	0.38	0.61 (0.58-0.64)
	I often play to relax from the day’s work	0.69	
	I often play so you I avoid thinking about some of my real-life problems or worries	0.32	

Only Customisation does not have any items highlighted for removal —

coincidentally, this is the most reliable scale of the questionnaire. Mechanics, Roleplaying, Discovery and Escapism are all highlighted for the removal of one item each. When looking at the items this is understandable, as they share little semantic similarity to the others. For example, ‘I often use a character builder or a template to plan out my character’s advancement at an early level’ from Mechanics relates to strategic thinking rather than to making the character unique. ‘I enjoy being immersed in a fantasy world’ from Roleplaying refers to the immersion in the game world rather than embodying a character, and ‘I enjoy collecting distinctive objects or clothing that have no functional value in games’ from Discovery relates more closely to collecting than simply finding the objects. The item ‘I often play to relax from the day’s work’ from Escapism shows the largest increase in reliability with its removal (from 0.61 to 0.69), which is understandable as it is less to do with escaping and more about relaxing. However, in practice removing this item would create a 2-item sub-scale, which would make it less valid or accurate for measuring Escapism overall.

In summary, Mechanics, Roleplaying and Discovery can be improved by removing unreliable items, however removing the item from Escapism could make the sub-scale less valid. This would make it less useful as a way to measure Escapism. Therefore, Yee is a questionnaire that has room for improvement to increase reliability, but some improvements would call into question the validity of the sub-scales. This mixed collection of reliable and unreliable elements is common across many of the questionnaires analysed, with more information found at <https://osf.io/95a4c/>.

4.3.3. *Weakly Reliable Questionnaire; BrainHex*

Table 3 showed that only the Achiever sub-scale had an Alpha or Omega score over 0.70, making the questionnaire mostly unreliable. To investigate this further, a closer look at how each item performs in its respective sub-scale is shown in Table 6.

From this it can be seen that only one sub-scale has an item that can be removed to increase reliability. Removing ‘I enjoy hanging from a high ledge in a game’ from Daredevil increases the sub-scale reliability from 0.55 to 0.58, but even then this is still a very low reliability. Furthermore, removing this item would only leave 2 items in the sub-scale, making it more likely to be a bloated specific, a larger validity problem. The items in the reliable sub-scale Achiever do load well with each other, making it reliable and could be used reliably to measure this specific player preference. Overall, BrainHex is a

Table 6: The Cronbach’s alpha scores for each item in the BrainHex questionnaire, as well as the overall scale reliability. Bolded scale names indicates overall reliability above 0.70. Bolded alpha scores indicate scores that would improve the overall reliability if dropped.

Scale	Item	Alpha if dropped	Overall Alpha
Seeker	I enjoy exploring to see what I can find	0.44	0.63 (0.59-0.66)
	I often wonder what’s behind a locked door	0.62	
	I enjoy looking around just to enjoy the scenery	0.52	
Survivor	I enjoy frantically escaping from a terrifying foe	0.37	0.45 (0.40-0.49)
	I often feel scared, terrified or disturbed in games	0.32	
	I feel relief when I escape to a safe area	0.36	
Mastermind	I enjoy working out what to do on my own	0.50	0.63 (0.60-0.66)
	I enjoy devising a promising strategy when deciding what to try next	0.58	
	I enjoy working out how to crack a challenging puzzle	0.52	
Daredevil	I enjoy being in control at high speed in a game	0.38	0.55 (0.52-0.59)
	I enjoy responding quickly to an exciting situation in a game	0.41	
	I enjoy hanging from a high ledge in a game	0.58	
Achiever	I enjoy picking up every single collectible in an area	0.78	0.87 (0.86-0.88)
	I enjoy finding what I need to complete a collection	0.86	
	I enjoy getting 100% (completing everything in a game)	0.80	

questionnaire with deep reliability flaws and could not be recommended for usage in player research.

4.3.4. Discussion

Overall, a number of improvements can be made to the sub-scales to increase their reliability, by removing a number of unreliable items. A full summary of the reliability of each scale (and how each individual item performs) can be found at <https://osf.io/95a4c/>. These changes are summarised in Table 7.

22 sub-scales have items removed, leading to 15 2-item sub-scales (4 of which were not created from removing items, but existed beforehand). If these items are removed, the result is 29 sub-scales that have alphas over 0.70. Removing the 2-item sub-scales leaves 20 that could be reliably used in player research.

Overall, Tondello and GAMES were the most reliable, though even they would benefit from certain items being removed. This does make sense, as these sub-scales have built upon previous work and seemingly have been improved because of this. Tondello performs better than both Yee and Brain-Hex, the two scales that influenced its creation, whilst GAMES is more reliable than Sherry. Trojan and Sherry also had high reliability but had very short sub-scales (2-4 items), making these unlikely to be concepts but rather

Table 7: A summary of the reliability analyses of all sub-scales. Alpha scores above 0.70 are **bolded**.

Questionnaire	Sub-Scale	Item dropped?	Improved Reliability	Two-Item Sub-Scale
Yee	Mechanics	✓	0.67	-
	Roleplaying	✓	0.74	-
	Discovery	✓	0.72	-
	Customisation	-	0.76	-
Tondello	Escapism	✓	0.69	✓
	Aesthetic	✓	0.75	-
	Narrative	-	0.84	-
	Goal	✓	0.83	-
Trojan	Challenge	✓	0.89	-
	Completionist	-	0.62	-
	Escapist	-	0.79	✓
	StoryDriven	-	0.77	✓
VKS	SmartyPants	-	0.74	✓
	Immersion	-	0.66	-
	Achievement	✓	0.76	✓
	Challenge	✓	0.57	✓
GAMES	Story	-	0.88	-
	Escapism	✓	0.88	-
	Customisation	✓	0.88	-
	Grinding/Completion	✓	0.82	-
PENS	Autonomy/Exploration	-	0.75	-
	Competence	-	0.71	-
	Autonomy	-	0.65	-
	Presence/Immersion	✓	0.88	-
BrainHex	Seeker	-	0.63	-
	Survivor	-	0.45	-
	Mastermind	-	0.63	-
	Daredevil	✓	0.58	✓
Sherry	Achiever	-	0.87	-
	Challenge	✓	0.61	-
	Diversion	-	0.88	✓
	Fantasy	-	0.83	-
GAMS	Arousal	-	0.68	-
	IntrinsicMotivation	-	0.50	-
	IntegratedRegulation	✓	0.70	✓
	IdentifiedRegulation	✓	0.75	✓
IMG	IntrojectedRegulation	✓	0.79	✓
	ExternalRegulation	✓	0.60	✓
	Competence	✓	0.64	✓
	Immersion	-	0.72	-
VGPU	Fun	✓	0.78	✓
	Auto-my	-	0.80	-
	GeneralisedGameSelfEfficacy	-	0.83	-
	EnjoymentOfGames	-	0.87	-
	ProneToGameImmersion	✓	0.67	✓

singular questions reworded a number of times. Trojan, by fusing together two earlier questionnaires, has an improved reliability. However, it is let down in its conception, as the sub-scales are too short and so are likely to be invalid for use. Sherry is the most reliable of the earliest scales, but also is let down by the use of short sub-scales. Therefore, these 4 make up the most reliable of the questionnaires surveyed, but still contain problems that would need addressing to improve their reliability.

Considering the ancestry diagram presented in Figure1, there is a general trend in newer questionnaires performing better. From the Games field, BrainHex is more reliable than Yee, and Tondello is more reliable than BrainHex. Similarly, VKS is slightly more reliable than Yee, and GAMES shows a similar reliability to Sherry. However, other scales predicted to perform better do not share this trend. GAMS is less reliable than PENS despite being newer, and despite IMG being a new combination of 3 research disciplines, Table 3 shows only two of its sub-scales are reliable above 0.70. This is less than PENS and Yee, two of the scales that it was inspired by, and performs approximately as well as Sherry. These questionnaires are not completely comparable, as they do not all measure the same constructs. However, it can be seen that overall the trend is for little improvement in reliability. On a more granular analysis, there are 3 sub-scales that IMG shares with PENS; Autonomy, Competence, and Immersion. Only Autonomy is more reliable in IMG than PENS, indicating it has not improved on the original questionnaire. Therefore, whilst new questionnaires may build on previous work, this does not always mean that the resulting questionnaire will be more reliable.

Overall, this section highlights how the reliability of many questionnaires are questionable, and so also brings into question their usefulness. If these scales cannot reliably measure concepts, then any conclusions based on answers to these items may not be accurate. This has widespread repercussions for the field, and highlights the need to reevaluate the questionnaires currently in use.

However, high reliability is not an indicator of conceptual coherence. Reliability can be falsely high if the scale is not unidimensional (which it ought to be). Therefore, to assess the structure of the sub-scales, factor analyses are needed, which are performed in the next section.

4.4. PART II: Confirmatory Factor Analysis (CFA)

The aim of this section is to assess how well each item fits onto its proposed factor. This is achieved by a Confirmatory Factor Analysis (CFA),

a technique that evaluates how well a hypothesised model ‘fits’ a data set. In this case, the proposed factor structures of the questionnaires are tested against the data collected, and an evaluation of this fit is made. Typically when reporting a CFA, certain measures of fit would also be reported on, such as Chi-Squared (χ^2) or Root Mean Square Error of Approximation (RMSEA). However, these measures are useful specifically when comparing between models. This is not the aim of this section, which is instead to simply assess the fit of the proposed models to the data collected. For this reason, measures of fit are not reported.

To calculate the CFAs, the *mirt* package in R was used [57]. As there were more than 3 factors for each questionnaire, the Metropolis-Hastings Robbins-Monro (MHRM) method was selected [58]. The output is assessed, and compared to the reliability measures calculated earlier.

Similarly to reliability, there is no definitive cutoff for when a loading becomes sufficient, yet each item is expected to reasonably measure its underlying concept. A loading above 0.30-0.40 is considered adequate, but when used in questionnaire creation researchers tend to seek higher [59]. For example, Matsunaga [60] suggests that, whilst 0.40 may be the lowest loading considered appropriate, a more conservative cut-off is 0.70. Others suggest items above 0.60 indicate strong fit [61]. For the purpose of this work, items above 0.60 are considered adequate fit, though items below this cut-off are commented upon when appropriate.

Similarly to Part I, a selection of questionnaires are presented in detail that summarise the range of model fits found from the analysis. The full analyses can be found at <https://osf.io/95a4c/>.

4.4.1. *Good Model Fit; GAMES*

GAMES is an example of a factor structure well supported by the data collected, which was also found to be highly reliable. Most items load highly onto their factors, and each sub-scale consists of numerous varied items, reducing the chance of bloated specifics. The CFA is shown in Table 8, along with the reliability of each item for comparison.

The proposed factors of GAMES fits the data well, with all items loading onto their respective sub-scales above 0.40, with 25 out of the 33 loading above 0.60. 2 of the items flagged earlier as unreliable do not load as strongly as the other items in the sub-scales, further suggesting these be removed. However, a few items in Story do not load as strongly as expected. Instead they hang together around 0.40, but were not flagged as unreliable. This

Table 8: The loadings for each item of GAMES, from a CFA. Loadings above 0.60 are in bold. Alpha items in bold indicates items flagged as unreliable in Part I.

Sub-Scale	Item	Story	Escapism	Customisation	Grinding/ Completion	Autonomy/ Exploration	Alpha if Dropped	Overall Sub- Scale Alpha
Story	Some of my favourite stories are in video games	0.685	0	0	0	0	0.87	0.88
	I really do my best to put myself into the main character's shoes	0.613	0	0	0	0	0.87	
	I'm interested in learning the lore and history of video game worlds	0.669	0	0	0	0	0.87	
	I'm excited to find out what happens next in the story	0.835	0	0	0	0	0.86	
	In video games, it's hard for me to identify with my character	0.477	0	0	0	0	0.88	
	It's hard for me to play a game if I can't relate to my character	0.412	0	0	0	0	0.88	
	I mostly play video games for their stories	0.774	0	0	0	0	0.86	
	Video game stories aren't important to me	0.802	0	0	0	0	0.86	
	I feel emotionally attached to the characters in my favourite games	0.694	0	0	0	0	0.87	
	I love to learn about the backstories of the character in video games	0.804	0	0	0	0	0.86	
	The feeling of immersion is important to me	0.593	0	0	0	0	0.87	
	Stories in video games just get in the way	0.755	0	0	0	0	0.87	
	I play video games because they let me do things I can't do in real life	0	0.694	0	0	0	0.84	
	I play video games to keep my mind off my problems	0	0.875	0	0	0	0.83	
Escapism	I play video games because it allows me to escape real life	0	0.925	0	0	0	0.81	0.86
	I like to do something that I could not normally do in real life through a video game	0	0.649	0	0	0	0.85	
	Games calm me down when I'm feeling nervous	0	0.488	0	0	0	0.88	
	Video games allow me to escape from the problems associated with everyday life	0	0.949	0	0	0	0.81	
	I really like to customise my character's outfit	0	0	0.903	0	0	0.82	
Customisation	I like making things in video games, like houses or outfits	0	0	0.689	0	0	0.88	0.87
	I like to personalise and customise my character	0	0	0.917	0	0	0.83	
	I'll put considerable time into designing my character's appearance (e.g. clothes, face)	0	0	0.912	0	0	0.81	
Grinding/ Completion	I like taking the time to pick up every single collectible item in games	0	0	0	0.817	0	0.76	0.81
	I'm excited to unlock achievements or earn trophies in games	0	0	0	0.635	0	0.79	
	I will often level up my characters until they reach the level cap	0	0	0	0.563	0	0.79	
	I rarely complete collections of in-game items	0	0	0	0.843	0	0.76	
	I don't mind grinding for an hour or two to get an item I want	0	0	0	0.421	0	0.82	
	I'll play a game until I get 100% on it, completing everything in games	0	0	0	0.859	0	0.75	
Autonomy/ Exploration	I like games that offer you a lot of options and choices	0	0	0	0	0.644	0.71	0.75
	I like games that offer different ways to get to the next level or area	0	0	0	0	0.591	0.72	
	I like having a choice of several different places or levels to try	0	0	0	0	0.587	0.72	
	I like games that do not put a lot of constraints on the player	0	0	0	0	0.736	0.72	
	I prefer games that allow me to play however I want	0	0	0	0	0.869	0.66	

suggests there may be more than one factor in this scale — a question explored later in Part III: Emergent Subfactors. Overall the proposed factor structure of GAMES is good, confirming the structure of the questionnaire.

4.4.2. Average Model Fit; IMG

The CFA for IMG reflects the majority of the CFAs for the data set, demonstrating good to average model fit. Table 9 shows this, along with the reliability of individual items. IMG consists of small sub-scales where

Table 9: The loadings for each item of IMG, from a CFA. Loadings above 0.60 are in bold. Alpha items in bold indicates items flagged as unreliable in Part I.

Sub-Scale	Item	Competence	Immersion	Fun	Autonomy	Alpha if Dropped	Overall Sub- Scale Alpha
Competence	I play because of the challenge	0.621	0	0	0	0.41	0.57
	I play to master my skills and to win myself	0.867	0	0	0	0.24	
	I play to make progress and to achieve objectives	0.374	0	0	0	0.64	
Immersion	I play because game events bring about emotions	0	0.620	0	0	0.70	0.72
	I play because I want to identify with the game characters	0	0.760	0	0	0.60	
	I play because I want to be part of the gameworld and its events	0	0.804	0	0	0.59	
Fun	I play because games are enjoyable	0	0	0.922	0	0.44	0.62
	I play because games are entertaining	0	0	0.890	0	0.41	
	I play because playing games is relaxing	0	0	0.462	0	0.78	
Autonomy	I play because in games I can make meaningful choices	0	0	0	0.824	0.73	0.80
	I play because in games I can make my own decisions	0	0	0	0.804	0.74	
	I play because in games I can make a difference with my actions	0	0	0	0.819	0.73	

each item loads together well, but with a few unreliable items flagged earlier contributing to low loadings. Indeed, only 2 items are below 0.60, which

were also flagged as unreliable in the above section. This is reflective of many questionnaires in the sample, that show good model fit once unreliable items are removed, but also highlights a trend in factors containing a small number of items. If so much as one item is deemed unreliable, the sub-scale is only left with 2 items, making it much more likely a bloated specific rather than able to assess the proposed concept. This means that, whilst the sub-scale may be reliable, from a validity standpoint it is difficult to defend.

Overall, IMG has a good structure that can explain most aspects of the data, but the small construction of each sub-scale decreases their validity.

4.4.3. Poor Model Fit; GAMS

Overall the GAMS questionnaire is an example of a poor model fit. The CFA for GAMS is shown in Table 10. The large amount of unreliable items

Table 10: The loadings for each item of GAMS, from a CFA. Loadings above 0.60 are in bold. Alpha items in bold indicates items flagged as unreliable in Part I.

Sub-Scale	Item	Intrinsic	Integrated	Identified	Introjected	External	Alpha if Dropped	Overall Sub-Scale Alpha
Intrinsic Motivation	I enjoy the feeling of efficacy I experience when I play	0.501	0	0	0	0	0.43	0.50
	I find pleasure in trying/experiencing new game options (e.g., classes, characters, teams, races, equipment)	0.482	0	0	0	0	0.47	
	I find it stimulating to play	0.767	0	0	0	0	0.33	
Integrated Regulation	I think the games I play align with my personal values	0	0.312	0	0	0	0.70	0.58
	I think games are an integral part of my life	0	0.769	0	0	0	0.40	
	I think games are an extension of me	0	0.849	0	0	0	0.25	
Identified Regulation	I think games are a good way to develop important aspects of myself	0	0	0.899	0	0	0.46	0.71
	Games have personal significance to me	0	0	0.511	0	0	0.75	
	I think are a good way to develop social and intellectual abilities that are useful to me	0	0	0.766	0	0	0.55	
Introjected Regulation	I feel that I must play regularly	0	0	0	0.648	0	0.79	0.77
	If I don't play I would feel bad about myself	0	0	0	0.897	0	0.62	
	I must play to feel good about myself	0	0	0	0.828	0	0.67	
External Regulation	I like the prestige of being a good player	0	0	0	0	0.461	0.60	0.58
	I like to acquire powerful and rare items and virtual currency or to unlock hidden/restricted elements of games	0	0	0	0	0.783	0.41	
	I like to gain in-game awards and trophies or character/avatar's levels and experiences points	0	0	0	0	0.668	0.44	

flagged earlier already highlighted concerns for the structure of the factors, as only Intrinsic Motivation did not have any unreliable items. Even then, Intrinsic Motivation had an Alpha score of 0.50, which is considered very low and is reflected in the factor loading of the sub-scale; only the item 'I find it stimulating to play' loaded above 0.60. The remaining sub-scales contain unreliable items flagged before, where these items loaded considerably lower than the other items in the scale. Removing them would improve both the reliability of all sub-scales, and would also allow each item to load above 0.60. However, doing so would create 2-item sub-scales, reducing the validity of the sub-scales to capture their intended concepts.

Therefore, GAMS reflects a questionnaire with a poor model fit, where many items across multiple sub-scales are shown to have low reliability and associated weaker loadings. These findings make it difficult to defend its use in player research without considerable changes being made.

4.4.4. Discussion

Overall, the CFAs confirm many of the unreliable items flagged earlier also fit poorly onto their respective sub-scales, furthering the need to remove them. Tondello and GAMES therefore seem the most reliable and structurally sound questionnaires from the sample, whereby all sub-scales show good fit to the data, and could reliably be used in research (though some items may still need removing to increase reliability). The remaining questionnaires still contain sub-scales that could be reliably and validly used in player research, though many only contain a small number of items each. Multiple sub-scales suffer from being short, and made shorter when unreliable items are removed, decreasing their validity.

A summary of all the sub-scales can be found in Table 11, which shows the percentage of items that load above 0.60 in each. This provides an overview of the structure of the questionnaires, and a full analysis can be found at <https://osf.io/95a4c/>. The majority of sub-scales have more than 50% of items load above 0.60 (41 out of 45). Indeed, 17 sub-scales had all items above 0.6, indicated strong structures. Further, 16 sub-scales have more than half of the items above 0.80, an even more stringent measure. However, it should also be noted that many of these strong structures come from sub-scales with a low number of items — 7 sub-scales had all items above 0.80, but 4 of these come from 2-item sub-scales. Therefore, the CFA structures indicate most sub-scales have an adequate structure, but those at the highest end tend to be small and likely invalid for measurement of player traits.

In the majority of cases, items that showed poor loading onto their sub-scales were also flagged as unreliable. There were however a few exceptions, such as a number of items in Story from GAMES, which showed weaker loading than the others but were not deemed unreliable. This suggests the potential for sub-factors, where there was also evidence of this in Immersion by VKS. To see if this is true, the next section explores the likelihood of sub-factors present within each of the sub-scales.

4.5. Part III: Emergent Sub-Factors

The aim of this section is to further explore the structure of some of the questionnaire sub-scales. This is because some items did not load as predicted when a CFA was run. For example, Story from GAMES showed good reliability for all items, but some items loaded together below 0.60. It is possible these items form a distinct sub-factor, which would affect the structure of the sub-scale if not taken into account.

Table 11: The percentage of items that load above 0.60 and 0.80 for each sub-scale. Names in bold indicate sub-scales where 100% of items load above 0.60.

Questionnaire	Sub-scale	Items above 0.60	Items above 0.80
Yee	Mechanics	50% (2/4)	25% (1/4)
	Roleplaying	50% (2/4)	50% (2/4)
	Discovery	75% (3/4)	25% (1/4)
	Customisation	100% (3/3)	0% (0/3)
	Escapism	66% (2/3)	33% (1/3)
Tondello	Aesthetic	80% (4/5)	20% (1/5)
	Narrative	100% (5/5)	60% (3/5)
	Goal	80% (4/5)	40% (2/5)
	Challenge	100% (5/5)	80% (4/5)
Trojan	Completionist	33% (1/3)	33% (1/3)
	Escapist	100% (2/2)	100% (2/2)
	Story Driven	100% (2/2)	100% (2/2)
	Smarty Pants	100% (2/2)	100% (2/2)
VKS	Immersion	43% (3/7)	0% (0/7)
	Achievement	66% (2/3)	33% (1/3)
	Challenge	66% (2/3)	0% (0/3)
GAMES	Story	75% (9/12)	25% (3/12)
	Escapism	83% (5/6)	50% (3/6)
	Customisation	100% (4/4)	75% (3/4)
	Grinding/Completion	66% (4/6)	50% (3/6)
	Autonomy/Exploration	60% (3/5)	20% (1/5)
PENS	Competence	100% (3/3)	33% (1/3)
	Autonomy	100% (3/3)	0% (0/3)
	Presence/Immersion	89% (8/9)	33% (3/9)
BrainHex	Seeker	66% (2/3)	33% (1/3)
	Survivor	0% (0/3)	0% (0/3)
	Mastermind	100% (3/3)	0% (0/3)
	Daredevil	66% (2/3)	0% (0/3)
	Achiever	100% (3/3)	100% (3/3)
Sherry	Challenge	50% (2/4)	0% (0/4)
	Diversion	100% (2/2)	100% (2/2)
	Fantasy	100% (4/4)	50% (2/4)
	Arousal	75% (3/4)	0% (0/4)
GAMS	Intrinsic Motivation	33% (1/3)	0% (0/3)
	Integrated Regulation	66% (2/3)	33% (1/3)
	Identified Regulation	66% (2/3)	33% (1/3)
	Introjected Regulation	100% (3/3)	66% (2/3)
	External Regulation	66% (2/3)	0% (0/3)
IMG	Competence	66% (2/3)	33% (1/3)
	Immersion	100% (3/3)	33% (1/3)
	Fun	66% (2/3)	66% (2/3)
	Autonomy	100% (3/3)	100% (3/3)
VGPU	Generalised Game Self Efficacy	83% (5/6)	33% (2/6)
	Enjoyment of Games	100% (4/4)	100% (4/4)
	Prone to Game Immersion	66% (2/3)	33% (1/3)

To assess if this is true, a parallel analysis was run on each questionnaire sub-scale. Parallel analysis compares the eigenvalues of a data set to those of a randomly generated data set of the same size, and outputs the number of factors to keep [62]. If a sub-scale is truly 1 factor, the parallel analysis should reflect this and the number of components would be 1, which was calculated using the psych package in R [63]. Any sub-scale that has more than 1 component is investigated further, by running an Exploratory Factor Analysis (EFA). EFA is a technique to assess the underlying relationships between variables, when there are no prior assumptions of how data should be structured [64]. This was achieved using the mirt package in R [57], with an oblique rotation (oblimin) to allow for correlations between factors.

Whilst typically factors are associated with Exploratory Factor Analysis and components with Principal Component Analysis, for the purposes of this analysis and those in following sections, components are used to establish the number of factors to extract for analysis. There is a difference between components and factors mathematically, but for practical use their difference is in fact minimal as they perform similarly ([65]). Furthermore, the suggested components provided a clearer number than suggested by factors overall, as sometimes a factor would return as 0. Finally, each EFA conducted was also accompanied with factor models computed either side of the suggested component (e.g., if 3 components were suggested, 2 and 4 factor solutions were also computed), and the difference between component and factor numbers were consistently minimal, leading to the choice of components over factors for extraction for ease.

As Table 12 shows, 3 of the sub-scales do not indicate a 1-component solution; Story from GAMES, Immersion from VKS, and Presence/Immersion from PENS. Therefore, there is a need to investigate these sub-scales more closely. To do so, a 1-factor solution is run and compared to the solution suggested by the parallel analysis. Each parallel analysis suggested 2-factor solutions, but this may not reflect the most accurate description of the data, due to the somewhat arbitrary cutoff imposed for a purely exploratory technique [66]. Therefore, a 3-factor solution is also calculated for each and compared to. Once a solution has been chosen as the most accurate, a reliability analysis is run on the newly proposed factors, and correlations between factors are considered.

Table 12: Parallel Analyses for each sub-scale analysed. Components higher than 1 are bolded.

Questionnaire	Sub-Scale	Factors	Components
Yee	Mechanics	2	1
	Roleplaying	2	1
	Discovery	1	1
	Customisation	1	1
	Escapism	1	1
Tondello	Aesthetic	2	1
	Narrative	2	1
	Goal	2	1
	Challenge	2	1
Trojan	Completionist	1	1
	Escapist	1	1
	Story Driven	1	1
	Smarty Pants	1	1
VKS	Immersion	3	2
	Achievement	1	1
	Challenge	0	1
GAMES	Story	4	2
	Escapism	3	1
	Customisation	3	1
	Grinding/Completion	2	1
	Autonomy/Exploration	2	1
PENS	Competence	1	1
	Autonomy	1	1
	Presence/Immersion	3	2
BrainHex	Seeker	1	1
	Survivor	0	1
	Mastermind	1	1
	Daredevil	0	1
	Achiever	1	1
Sherry	Challenge	3	1
	Diversion	1	1
	Fantasy	2	1
	Arousal	2	1
GAMS	Intrinsic	0	1
	Integrated	1	1
	Identified	1	1
	Introjected	1	1
	External	1	1
IMG	Competence	1	1
	Immersion	1	1
	Fun	1	1
	Autonomy	1	1
VGpu	Generalised Self Efficacy	3	1
	Enjoyment	2	1
	Prone To Immersion	1	1

Table 13: EFA of the Story scale from GAMES. Loadings above 0.30 are bolded.

Item	1 Factor	2 Factor		3 Factor		
	F1	F1	F2	F1	F2	F3
Video game stories aren't important to me	0.669	-0.939	0.078	-0.968	0.020	0.061
Stories in video games just get in the way	0.776	-0.913	0.090	-0.829	0.069	-0.088
I mostly play video games for their stories	0.837	-0.729	-0.112	-0.711	-0.151	-0.011
I'm excited to find out what happens next in the story	0.759	-0.655	-0.246	-0.498	-0.137	-0.291
Some of my favourite stories are in video games	0.687	-0.507	-0.236	-0.410	-0.164	-0.186
I really do my best to put myself into the main character's shoes	0.615	0.059	-0.835	0.060	-0.802	-0.080
In video games, it's hard for me to identify with my character	0.805	0.031	-0.631	-0.016	-0.645	0.031
It's hard for me to play a game if I can't relate to my character	0.481	0.004	-0.515	-0.096	-0.602	0.155
The feeling of immersion is important to me	0.804	-0.161	-0.538	-0.172	-0.492	-0.055
I feel emotionally attached to the characters in my favourite games	0.413	-0.246	-0.564	-0.108	-0.403	-0.334
I love to learn about the backstories of the character in video games	0.697	-0.476	-0.412	-0.023	-0.030	-0.903
I'm interested in learning the lore and history of video game worlds	0.592	-0.417	-0.315	-0.021	0.023	-0.784

4.5.1. Story from GAMES

As mentioned previously, Story from GAMES has a large range of loaded items (from 0.412 to 0.835), where none are flagged as unreliable. This implies there may be more than one factor, supported by the parallel analysis that suggests a 2-factor solution. To investigate the likelihood of this, a 1-factor solution EFA was conducted, alongside 2 and 3-factor solutions. These solutions were compared to see which was most descriptive of the data. This is shown in Table 13.

The loading of each item in the 1-factor solution are all adequate, with none below 0.40. There is however still a large variance, similarly to the solution provided in the previous CFA. The 2-factor solution pulls apart 2 distinct groups of items, but with a strong cross-loading for the final 2 items. In comparison, the 3-factor solution pulls apart 3 groups with only one cross-loaded item, 'I feel emotionally attached to the characters in my favourite games'. This suggests the best description of the data is the 3-factor model, with the cross-loaded item removed to improve clarity (as the loading is still far lower than the other items in either factor).

Looking at these 3 factors, the following names are proposed: Story, Character Identification, and Lore Learning². To further assess the likelihood of this 3-factor solution, Cronbach's alpha was calculated for these new factors, shown in Table 14. The overall reliability of the new Story factor is strong at 0.85, as well as the reliability of Lore Learning at 0.80. Lore Learning however only has 2 items, so is not likely to be useful as its own sub-scale, but may indicate a distinct concept that requires future elaboration. The

²or 'Lorening' if you will

reliability of Character Identification is lower at is 0.69, although no items are flagged as unreliable.

Table 14: Reliability of the proposed subfactors of the Story scale from GAMES.

Proposed Scale	Item	Alpha if dropped	Overall Sub-Scale
Story	Video game stories aren't important to me	0.79	0.85 (0.84 - 0.86)
	Stories in video games just get in the way	0.81	
	I mostly play video games for their stories	0.82	
	I'm excited to find out what happens next in the story	0.81	
	Some of my favourite stories are in video games	0.84	
Identification	I really do my best to put myself into the main character's shoes	0.54	0.69 (0.66 - 0.71)
	In video games, it's hard for me to identify with my character	0.64	
	It's hard for me to play a game if I can't relate to my character	0.67	
	The feeling of immersion is important to me	0.63	
Lore Learning	I love to learn about the backstories of the character in video games	0.67	0.80 (0.78 - 0.82)
	I'm interested in learning the lore and history of video game worlds	0.66	

Correlations between the proposed factors were calculated, and are presented in Table 15.

There are moderate to strong correlations between the 3 factors, indicating the sub-factors within this sub-scale are distinct but likely related. Therefore, whilst Story is presented as a single factor, this research suggests it may contain 2-3 distinct sub-factors. The 3rd factor is smaller than the others, indicating it is not currently a fully elaborated concept, but it is distinct enough to be considered on its own.

4.5.2. Immersion from VKS

The parallel analysis suggested 2 components for the Immersion sub-scale. The exploratory 1-factor solution indicated a large range of loaded items (from 0.332 to 0.768), with only 3 items above 0.60. This further indicates the potential for more factors, so 2 and 3-factor solutions were then calculated to investigate this further. These solutions are shown in Table 16.

From this it can be seen that no solution fully captures the data. The lack of highly loaded items in the 1-factor solution makes it unlikely there

Table 15: Correlations between the proposed Story sub-factors.

	Story	Identification	Lore Learning
Story	-		
Identification	0.55	-	
Lore Learning	0.69	0.59	-

Table 16: EFA of the Immersion scale from VKS. Loadings above 0.30 are bolded.

Item	1 Factor	2 Factor		3 Factor		
	F1	F1	F2	F1	F2	F3
I like taking up quests or missions in games	0.503	-0.594	-0.118	-0.724	-0.037	0.103
I enjoy immersing in the world of games	0.768	-0.831	-0.037	-0.714	0.034	-0.152
The satisfaction I feel when I accomplish a task successfully in games is stronger than the satisfaction I feel when I accomplish a task successfully in real life	0.353	0.011	0.787	0.015	0.792	0.007
Games are an important activity in my life	0.489	-0.334	0.354	-0.289	0.357	-0.089
I enjoy the feeling of escape from real life through games	0.706	-0.592	0.171	0.019	0.046	-0.807
I enjoy transforming to somebody else in games	0.630	-0.601	0.030	-0.064	-0.067	-0.681
I enjoy learning in games new skills that may be useful in real life	0.332	-0.290	0.088	-0.299	0.130	0.007

is a unidimensional overarching factor, meaning the items may be unrelated to one another. The item ‘I enjoy learning in games new skills that may be useful in real life’ does not strongly load onto any factor, with the item ‘Games are an important activity in my life’ showing a similar trend (though does slightly cross-load on some factors). This makes sense, as these items are asking about wider impacts of gaming on the player, rather than immersion. Discarding these 2 items provides a clearer factor structure, with the item ‘The satisfaction I feel when I accomplish a task successfully in games is stronger than the satisfaction I feel when I accomplish a task successfully in real life’ loading on its own factor in both the 2 and 3-factor solutions. The correlations between these factors is shown in Table 17.

Table 17: Correlations between the proposed Immersion sub-factors.

	F1	F2	F3
F1	-		
F2	0.28	-	
F3	-0.60	-0.47	-

There are a range of correlations between the 3 factors, though some are calculated as negative. This is most likely an artefact of the factor analysis, where the direction of the 1st factor has ‘flipped’. This can be seen in Table 16, where the 1-item sub-factor is positively loaded, but the other 2 factors are negative. Because of this, the correlation suggests negative correlations, but this is unlikely to be true. For the purpose of this interpretation, the sign is ignored and negative correlations are treated as positive.

The 1-item factor does show weak to moderate correlation with the other 2 factors (0.28 and 0.47 respectively), but it is not as strong as the correlation between them (0.60). Therefore, discarding this 1-item factor would leave

4 items in total across 2 factors (shown in the 3-factor solution). However, this leaves 2 items in each, and as the correlation between these factors is moderate to high, it may be best to combine these items into a singular 4-item sub-scale. Put another way, this sub-scale could be reduced in size from the original to improve the structure of the factor.

To test this, Cronbach’s alpha was computed on this new 4-item sub-scale, and is shown in Table 18.

Table 18: Reliability of the proposed new Immersion sub-scale from VKS. The item in bold would improve the overall reliability if dropped.

Proposed Sub-scale	Item	Alpha if dropped	Overall Sub-scale
New Immersion	I enjoy transforming to somebody else in games	0.57	0.67 (0.65 - 0.70)
	I enjoy the feeling of escape from real life through games	0.58	
	I enjoy immersing in the world of games	0.57	
	I like taking up quests or missions in games	0.68	

The new immersion factor does not reach over the 0.70 threshold, even with dropping the item ‘I like taking up quests or missions in games’ to increase it to 0.68. This means that this sub-scale, even after considerable alterations, has questionable reliability and structure.

4.5.3. Presence/Immersion from PENS

The parallel analysis suggested a 2-factor solution, so a 1-factor solution was computed to see if the sub-scale is likely to contain sub-factors. Apart from one item, all items loaded above 0.60, making it likely if there are sub-factors, these will be highly related. 2 and 3-factor solutions were then computed to explore the most descriptive model of the data. This is shown in Table 19.

The one item to load below 0.60 in the 1-factor solution is understandable, as the item relates to the feeling of pride, which does not seem related to immersion. The 2-factor solution appears the most descriptive of the data, especially if discarding the pride item mentioned above (which loads strongly as a 1-item factor in the 3-factor solution). Doing so would leave one strongly cross-loaded item, which appears to be a bridge between the two sub-factors of immersion. To clarify the structure, it may be necessary to also remove this item from the sub-scale. This leaves 2 factors, that could be named Emotional Immersion and Transportation Immersion.

The reliability of these factors was then tested using Cronbach’s alpha, shown in Table 20. The alpha for both scales is strong, with both over

Table 19: EFA of the Presence/Immersion sub-scale from PENS. Loadings above 0.30 are highlighted in bold.

Item	1 Factor	2 Factor		3 Factor		
	F1	F1	F2	F1	F2	F3
Item 1	0.691	-0.802	0.010	-0.843	0.024	-0.017
Item 7	0.636	-0.846	0.109	-0.817	0.076	-0.008
Item 2	0.721	-0.704	-0.105	-0.721	-0.095	0.008
Item 9	0.763	-0.760	-0.104	-0.658	-0.139	0.130
Item 3	0.446	-0.396	-0.098	-0.002	0.006	0.960
Item 5	0.803	-0.330	-0.535	-0.319	-0.552	-0.015
Item 8	0.827	0.036	-0.939	0.050	-0.959	-0.005
Item 6	0.804	0.001	-0.869	-0.019	-0.847	0.004
Item 4	0.765	-0.032	-0.795	-0.021	-0.789	0.031

0.80. This supports the plausibility of splitting this sub-scale in to 2 sub-factors. Furthermore, the correlation between factors in the 2-factor solution is 0.67, indicating they are highly related (as predicted by the strong 1-factor solution). This means the 2 factors could be considered unidimensional when combined, meaning the original structure is still feasible.

Overall, the sub-scale appears to measure Immersion but has 2 distinct sub-factors, with 2 items dropped for being unrelated/strongly cross-loaded. Being aware of these sub-factors would help to improve the sub-scale’s validity, and usefulness in explaining presence/immersion.

4.6. Part IV: Semantically Similar Factors

The aim of this section is to explore whether sub-scales that share similar names can be considered interchangeable measurements. It is common when a new questionnaire is created to compare the sub-scales to those that exist in previous works, as this helps to place the work in the literature and supports the authors’ assertion they have captured a new concept. For example, Trojan compares its Competitors sub-scale to the Competition sub-scale of Yee, as well as the similarities in the Socialisers and Socialising sub-scales. Trojan then goes on to show that its Smarty Pants scale is a concept not measured previously, highlighting the uniqueness of this new questionnaire.

However, this comparison implies that these sub-scales are similar in ways

Table 20: Reliability of the proposed subfactors of the PENS Presence/Immersion sub-scale.

Proposed Sub-scale	Item	Alpha if dropped	Overall Sub-scale
Emotional	Item 1	0.74	0.81 (0.80 - 0.83)
	Item 7	0.78	
	Item 2	0.77	
	Item 9	0.77	
Transportation	Item 8	0.76	0.85 (0.83 - 0.86)
	Item 6	0.78	
	Item 4	0.82	

more so than their namesake. Is it enough to say that, because Competitors from Trojan is named similar to Competition from Yee, they share any commonalities in what they measure? The answer to this question has implications if it is both true or false. If true, it may be possible to combine items from similar sub-scales into 1 larger sub-scale, rather than having multiple separate smaller sub-scales. This would overcome the problems highlighted earlier, where sub-scales are so small they become unreliable, and are more likely to measure a bloated specific rather than a player trait. If false, the way in which questionnaires are discussed in the future may need to change, as it would not be enough to look for semantic similarities. Instead, statistical testing would be necessary to conclude the true convergence between questionnaire sub-scales.

This section therefore aims to answer this question. To do so, we conducted a series of comparative assessments of similarly named sub-scales. There are many examples of this found within the questionnaires studied, and the following concepts were selected for analysis: Escapism (Yee, Trojan and GAMES), Challenge (Tondello, VKS and Sherry), and Customisation (Yee and GAMES), as these were examples of sub-scales with the exact same name. Building upon previous findings in this study, a further comparison was made. As Part I highlighted, many items from the sub-scales were deemed unreliable, and so potentially unrelated to the concept being measured. These may alter the relationship between sub-scales unfairly, and so we ran each assessment step with two sets of data; the original items from the sub-scales, and an ‘altered’ set with the unreliable items removed.

Differences in findings are discussed when appropriate.

Firstly, correlations were run between sub-scales to indicate the likelihood they are measuring similar concepts. Note that, as this work is purely exploratory, no significance testing is run on these correlations. Correlations above 0.30 are considered weak, and correlations above 0.60 are considered moderate ([67]). Secondly, the items from all the sub-scales were combined and a parallel analysis was run. Similarly to Part III, it is expected the number of components will be 1, if the sub-scales all measure the same concept. Following on from this, a 1-factor EFA was run to assess the structure, and if this is deemed a poor model, further solutions are explored.

4.6.1. *Escapism*

Firstly, the Escapism sub-scales from Yee, Trojan and GAMES were analysed. The first two originate from Games research, whilst GAMES is from Psychology. To see how related these are, a Spearman’s correlation was run on both the original and altered sub-scales (those with items flagged as unreliable in Part I removed). The correlations are shown in Table 21.

Table 21: Spearman’s correlations for the Escapism sub-scales, both original and altered.

Original				Altered			
	Trojan	GAMES	Yee		Trojan	GAMES	Yee
Trojan	-	0.74	0.55	Trojan	-	0.74	0.55
GAMES		-	0.83	GAMES		-	0.85
Yee			-	Yee			-

There are strong correlations between the sub-scales, though the correlation between Yee and Trojan is notably weaker. This suggests there is some relation between the sub-scales, suggesting they may measure similar concepts. However, a parallel analysis suggested 3 components for the original sub-scales and 2 components for the altered. Therefore, 1, 2 and 3-factor solutions were run on both the original and altered sub-scales to find the best description of the data. The 1-factor solutions are shown in Table 22.

The 1-factor solutions show all items load above 0.70, except for 2 items that were flagged as unreliable earlier. This suggests the items form an overarching theme, but there is still a chance sub-factors exist, due to the results of the parallel analysis. 2 and 3-factor solutions were therefore also

Table 22: A 1-factor EFA for the Escapism sub-scales, both original items and altered.

Questionnaire	Item	Original	Altered
GAMES	Video games allow me to escape from the problems associated with everyday life	0.938	0.933
GAMES	I play video games because it allows me to escape real life	0.934	0.941
Yee	I often play so you I avoid thinking about some of my real-life problems or worries	0.866	0.860
GAMES	I play video games to keep my mind off my problems	0.856	0.852
Yee	It's important that games allow me to escape from the real world	0.802	0.804
Trojan	Video games allow me to pretend I am someone/somewhere else	0.750	0.752
GAMES	I play video games because they let me do things I can't do in real life	0.748	0.751
Trojan	I like to do things in games which I cannot do in real life	0.736	0.738
GAMES	I like to do something that I could not normally do in real life through a video game	0.711	0.711
GAMES	Games calm me down when I'm feeling nervous	0.502	-
Yee	I often play to relax from the day's work	0.399	-

conducted. These are shown in Table 23 for both the original and altered scales.

Table 23: 2 and 3-factor solutions for the original and altered Escapism sub-scales. Loadings above 0.30 are highlighted in bold.

Questionnaire	Item	Original						Altered					
		2 Factor		3 Factor				2 Factor		3 Factor			
		F1	F2	F1	F2	F3		F1	F2	F1	F2	F3	
Yee	I often play so you I avoid thinking about some of my real-life problems or worries	0.975	0.076	0.994	0.009	-0.052		-0.975	0.076	-0.975	0.043	-0.025	
GAMES	I play video games to keep my mind off my problems	0.962	0.087	0.849	0.053	0.113		-0.960	0.082	-0.978	0.035	-0.065	
GAMES	Video games allow me to escape from the problems associated with everyday life	0.936	-0.050	0.785	-0.070	0.163		-0.931	-0.056	-0.902	-0.068	0.033	
GAMES	I play video games because it allows me to escape real life	0.737	-0.256	0.226	-0.082	0.719		-0.736	-0.260	-0.597	-0.155	0.332	
Yee	It's important that games allow me to escape from the real world	0.562	-0.293	0.032	-0.101	0.743		-0.559	-0.299	-0.439	-0.293	0.200	
GAMES	Games calm me down when I'm feeling nervous	0.413	-0.131	0.524	-0.223	-0.188		-	-	-	-	-	
Yee	I often play to relax from the day's work	0.233	-0.223	0.284	-0.267	-0.075		-	-	-	-	-	
Trojan	Video games allow me to pretend I am someone/somewhere else	0.219	-0.669	0.030	-0.586	0.284		-0.218	-0.671	-0.125	-0.588	0.231	
GAMES	I play video games because they let me do things I can't do in real life	0.067	-0.867	-0.010	-0.818	0.135		-0.066	-0.870	-0.033	-0.834	0.089	
GAMES	I like to do something that I could not normally do in real life through a video game	-0.030	-0.947	0.031	-0.946	-0.053		0.030	-0.946	0.003	-0.965	-0.058	
Trojan	I like to do things in games which I cannot do in real life	-0.045	-0.998	-0.012	-1.000	-0.024		0.046	-0.998	0.029	-1.017	-0.043	

The 2-factor solution for the original sub-scales shows a clear split, with one item not loading onto either factor. The altered 2-factor solution shows a similar trend, with the unloaded item already removed for unreliability. The 3-factor solution is similar, with a clear separation of 3 factors and the unreliable item loading onto no factor, however the altered 3-factor solution created a cross-loaded item, 'I play video games because it allows me to escape real life'. It also contains a loading below -1.00, a rare occurrence but one that can happen when highly correlated factors are present in an oblique rotation (and therefore the measurement is based on regression coefficients, not correlations [68]).

Therefore both solutions are descriptive, but the 2 factor is likely more useful as the separation of factors is cleaner (especially in the altered version). This would mean the sub-scale contains 2 sub-factors — assessing the items in each, we suggest the names escaping *from* the real world, and escaping *into* a game.

Overall, this means Escapism is present across disciplines, and could be considered a single factor with 2 sub-factors. This means using any one of these sub-scales to measure ‘Escapism’ is likely valid. However, this distinction between the 2 sub-factors is lost by using one of the sub-scales instead of combining them, as the items are split across all 3. For example, GAMES contains items from both of the sub-factors, meaning the sub-scale could measure Escapism as a whole, but not be able to measure the sub-factors. This can only be done by using all 3 sub-scales together.

4.6.2. Challenge

The Challenge sub-scales in this analysis are from Tondello, VKS and Sherry, presenting three separate originating disciplines (Games, Education, and Media respectively). The correlations between each sub-scale are shown in Table 24.

Table 24: Spearman’s correlations for the Challenge sub-scales, both original and altered.

Original				Altered			
	Tondello	Sherry	VKS		Tondello	Sherry	VKS
Tondello	-	0.32	0.21	Tondello	-	0.36	0.11
Sherry		-	0.30	Sherry		-	0.24
VKS			-	VKS			-

Unexpectedly, the sub-scales do not strongly correlate — the highest correlation is 0.36 between Tondello and Sherry for the altered sub-scales, which is still only a small correlation (especially in comparison to Escapism mentioned above). If these sub-scales are measuring the same concept of challenge, these correlations should be higher. Furthermore, the parallel analysis suggested 3 components for both the original and 3 altered sub-scales, indicating there may be factors that do not overall form a single factor. Therefore, 1 and 3-factor solutions were run for both sub-scale variations. Table 25 shows the 1-factor solutions.

The original sub-scales show a large range of loaded items, between a very low 0.136 to a high 0.926. Even with the unreliable items removed the altered sub-scales show a similar trend, with a low of 0.129 and a high of 0.922. This suggests these sub-scales are not measuring the same overarching concept, but may instead be a mixture of different concepts. To investigate this further, the 3-factor solutions were calculated, shown in Table 26.

Table 25: A 1-factor EFA for the Challenge sub-scales, both original items and altered. Loadings above 0.30 are highlighted in bold.

Questionnaire	Item	Original	Altered
Tondello	I enjoy highly difficult challenges in games	0.926	0.922
Tondello	I like it when goals are hard to achieve in games	0.893	0.897
Tondello	I like it when games challenge me	0.869	0.871
Tondello	I like it when progression in a game demands skill	0.825	0.832
Tondello	I usually play games at the highest difficulty setting	0.715	-
VKS	I try to think of new ways to react to ingame situations	0.443	-
Sherry	I feel proud when I master an aspect of a game	0.433	0.431
Sherry	I enjoy finding new and creative ways to work through video games	0.409	0.403
Sherry	I play until I complete a level or win a game	0.285	-
Sherry	I find it very rewarding to get to the next level	0.277	0.280
VKS	Before taking up a quest, I analyse it and plan my tactics	0.191	0.183
VKS	Before I begin a task or quest, I try to find out more about it	0.136	0.129

Table 26: 3-factor EFAs for the Challenge sub-scales, both original items and altered. Loadings above 0.30 are highlighted in bold.

Questionnaire	Item	Original			Altered		
		F1	F2	F3	F1	F2	F3
Sherry	I find it very rewarding to get to the next level	-0.716	0.085	-0.008	0.073	0.752	-0.022
	I enjoy finding new and creative ways to work through video games	-0.588	-0.086	-0.076	-0.130	0.507	-0.074
	I feel proud when I master an aspect of a game	-0.523	-0.258	-0.011	-0.272	0.527	-0.001
	I play until I complete a level or win a game	-0.449	-0.059	-0.036	-	-	-
Tondello	I enjoy highly difficult challenges in games	0.037	-0.968	0.021	-0.963	-0.054	0.015
	I like it when goals are hard to achieve in games	-0.001	-0.899	-0.006	-0.917	-0.042	-0.020
	I like it when games challenge me	-0.071	-0.834	-0.015	-0.851	0.036	-0.017
	I like it when progression in a game demands skill	-0.199	-0.735	0.023	-0.757	0.169	0.008
	I usually play games at the highest difficulty setting	0.132	-0.788	-0.044	-	-	-
VKS	Before taking up a quest, I analyse it and plan my tactics	0.055	-0.034	-0.758	-0.044	-0.059	-0.740
	Before I begin a task or quest, I try to find out more about it	-0.139	0.088	-0.596	0.068	0.112	-0.625
	I try to think of new ways to react to ingame situations	-0.303	-0.249	-0.190	-	-	-

Interestingly, the EFA shows the 3 factors suggested are exactly the original scales for both versions (with a weak loading of a VKS item onto Sherry in the original). This means each Challenge sub-scale is measuring a different concept, that shows little relation to the others. It is highly unlikely that each sub-scale can be used interchangeably, or indeed semantically compared to. This is perhaps unsurprising, as each sub-scale came from different originating disciplines, meaning the conceptualisation of Challenge may be unique for each. Therefore, the concept of ‘Challenge’ should be treated with more caution in the future.

4.6.3. Customisation

The two customisation sub-scales come from Yee and GAMES, where only one item from GAMES was flagged as unreliable. The sub-scales correlated at 0.80 on a Spearman’s correlation for both the original and altered sub-

scales, which are strong and supports the notion the two are related. The parallel analyses also supported this, as they suggested only 1 component. To complete this investigation a 1-factor EFA was run, and is presented in Table 27.

Table 27: A 1-factor EFA for the Customisation sub-scales, both original items and altered.

Questionnaire	Item	Original	Altered
GAMES	I'll put considerable time into designing my character's appearance (e.g. clothes, face)	0.967	0.968
Yee	I spend a lot of time customising my character during character creation	0.926	0.927
GAMES	I like to personalise and customise my character	0.873	0.868
GAMES	I really like to customise my character's outfit	0.865	0.863
Yee	It's important that my character's armour/outfit matches in colour and style	0.681	0.679
GAMES	I like making things in video games, like houses or outfits	0.677	-
Yee	It's important that my character looks different from other characters	0.676	0.676

The lowest loading is 0.676, supporting a strong 1-factor solution. Therefore, it is highly likely these sub-scales are measuring the same concept, and could be used interchangeably.

4.6.4. Discussion

This section has sought to understand if semantically similar sub-scales can be considered the same. If true, it may be possible to combine sub-scales together to create stronger structures that are more reliable. If false, how questionnaires are compared to one another may need to be done with a more nuanced approach than only considering their names.

In this analysis, Customisation was the only concept that appears to be one single factor. Escapism did also show evidence of 1 factor, but also contained 2 distinct sub-factors; escaping *from* and escaping *to*. Challenge in comparison has 3 separate factors, each one unrelated to the other, where each factor represents the original sub-scale the items came from. This means sub-scales cannot be considered the same based on name alone. Sub-factors that have not been considered in the past may be present, which may only be visible when pooling items across questionnaires. Future work should compare measurements in a similar way to truly assess convergence, rather than relying on the names of sub-scales alone.

However, sub-scales with different names may still show overlaps with other sub-scales from different questionnaires. Further, items that are unreliable in some sub-scales may load better with other items in different sub-scales. A logical next step would be to assess how the sub-scales load together in an Exploratory Factor Analysis using mean scores as data. This

was completed, and is available to read in the OSF supplementary materials (<https://osf.io/95a4c/>). However, as it was possible to run an EFA on all 177 items, and its findings were similar to that of Section V, this comparison of sub-scales is not included here. Therefore, the next section allows items to move freely into factors by computing an EFA using all 177 individual items.

4.7. *PART V: Item Comparisons*

This final section aims to assess how all 177 items across the 11 questionnaires compare to one another. A parallel analysis was computed on the 177 items, and an Exploratory Factor Analysis (EFA) was run on the proposed factor solution. By doing so, factors are allowed to form across sub-scales, indicating where there are overlaps between questionnaires and sub-scales. These factors are then assessed in terms of reliability, and correlations between factors are considered.

A parallel analysis suggested a 19-factor solution from the 177 items, and so a 19-factor EFA was run. The output is presented in Table 28 and Table 29, split due to size. Some items cross-loaded across factors (15 in total), such as the item ‘I enjoy exploring every map or zone in the world’ from the Discovery sub-scale of Yee, which loaded onto factors 3 and 5. This is understandable, as the item contains aspects of both factors (each explained later in this section), being both about Completionism (‘every map or zone’) and Exploration (‘enjoy exploring’). Due to the high volume of items it was deemed best to remove these items, as this allows for a clearer structure to be seen. Further, some items did not load onto any factor, and so are also absent from this table (20 items). For example, the item ‘I feel proud when I master an aspect of a game’ from the Challenge scale of Sherry. Overall, 142 items were left across 19 factors.

The following factor names are proposed, ordered the same as the factor analysis. These will be discussed in depth in the following discussion section:

1. Transportation Immersion	9. I Like Stories
2. Competence	10. Play My Way
3. Completionism	11. Games as Life
4. Escape From	12. Forward Thinking
5. Exploration	13. Roleplaying
6. I Like Games	14. Escape To
7. Customisation	15. Challenge
8. Smarten	16. Strategic Thinking

Table 28: The first 8 factors for the 177 item EFA, with bolded items over 0.60. Any item below 0.30 is removed, as well as any remaining items that cross-loaded above 0.30.

Factor	Item	F1	F2	F3	F4	F5	F6	F7	F8
Transportation Immersion	The feeling of immersion is important to me	-0.646							
	Immersion PENS 8	-0.598							
	I like games which make me feel like I am actually in a different place	-0.575							
	Immersion PENS 6	-0.571							
	I enjoy immersing in the world of games	-0.561							
	I enjoy being immersed in a fantasy world	-0.490							
Competence	I play because I want to be part of the gameworld and its events	-0.309							
	Competence PENS 3	-0.886							
	I have good video game skills	-0.863							
	I am good at video games, compared to others	-0.794							
	I am confident playing video games	-0.628							
	I can keep up with a video game that moves quickly	-0.615							
	I have a lot of experience with playing video games	-0.598							
	Competence PENS 2	-0.508							
	Based on my knowledge of previous video games, I can easily see through the rules of a game	-0.425							
	Competence PENS 1	-0.416							
Completionism	I like the prestige of being a good player	-0.367							
	I enjoy getting 100% (completing everything in a game)	-0.966							
	I like completing games 100%	-0.964							
	I'll play a game until I get 100% on it, completing everything in the game	-0.852							
	I like to complete all the tasks and objectives in a game	-0.851							
	I enjoy picking up every single collectible in an area	-0.825							
	I rarely complete collections of in-game items	-0.778							
	I like taking the time to pick up every single collectible item in games	-0.714							
	I usually do not care if I do not complete all optional parts of a game	-0.698							
	I enjoy finding what I need to complete a collection	-0.692							
	I'm excited to unlock achievements or earn trophies in games	-0.573							
	I like to gain in-game awards and trophies or character/avatar's levels and experiences points	-0.508							
Escape From	I like to try everything that is possible to do in a game	-0.433							
	I will often level up my characters until they reach the level cap	-0.339							
	I often play to avoid thinking about some of my real-life problems or worries	-0.925							
	Video games allow me to escape from the problems associated with everyday life	-0.920							
	I play video games to keep my mind off my problems	-0.911							
	I play video games because it allows me to escape real life	-0.756							
	I enjoy the feeling of escape from real life through games	-0.612							
	It's important that games allow me to escape from the real world	-0.579							
Exploration	Games calm me down when I'm feeling nervous	-0.388							
	I enjoy exploring the world just for the sake of exploring it	-0.797							
	I enjoy exploring to see what I can find	-0.734							
	I like to spend some time exploring the game world	-0.727							
	I enjoy looking around just to enjoy the scenery	-0.599							
	I often wonder what's behind a locked door	-0.354							
I Like Games	I enjoy finding quests, NPCs or locations that most people do not know about	-0.337							
	I play because games are enjoyable						0.754		
	I like playing video games						0.709		
	I play because games are entertaining						0.699		
	I enjoy playing video games						0.671		
	I think video games are entertaining						0.510		
	Video games are fun						0.471		
	I play video games because they excite me						0.422		
	I find it stimulating to play						0.403		
	I play because playing games is relaxing						0.375		
	I often play to relax from the day's work						0.373		
	I like finishing quests						0.365		
	Autonomy PENS 1						0.314		
	I like to customize how my character looks in a game						-0.946		
Customisation	I'll put considerable time into designing my character's appearance (e.g. clothes, face)						-0.917		
	I like to personalise and customise my character						-0.894		
	I really like to customise my character's outfit						-0.894		
	I spend a lot of time customising my character during character creation						-0.871		
	It's important that my character's armour/outfit matches in colour and style						-0.682		
	It's important that my character looks different from other characters						-0.629		
	I like making things in video games, like houses or outfits						-0.599		
	I enjoy collecting distinctive objects or clothing that have no functional value in games						-0.494		
Smarten	I play games to enhance my intellectual abilities							0.796	
	Games make me smarter							0.755	
	I think are a good way to develop social and intellectual abilities that are useful to me							0.747	
	I think games are a good way to develop important aspects of myself							0.693	
	I enjoy learning in games new skills that may be useful in real life							0.584	

Table 29: Factors 9 through to 19 for the 177 item EFA, with bolded items over 0.60. Any item below 0.30 is removed, as well as any remaining items that cross-loaded above 0.30. Strategic Thinking has no items in this table, as all items cross-loaded above 0.30

Factor	Item	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19
I Like Stories	Video game stories aren't important to me	-0.896										
	I like stories in a game	-0.875										
	Stories in video games just get in the way	-0.865										
	Story is not important to me when I play games	-0.846										
	I like games that pull me in with their story	-0.756										
	I mostly play video games for their stories	-0.733										
	I feel like storytelling often gets in the way of actually playing games	-0.724										
	I enjoy complex narratives in a game	-0.711										
	I usually skip the story portions or the cutscenes when I am playing	-0.691										
	I'm excited to find out what happens next in the story	-0.615										
	I like to the feeling of being part of a story	-0.494										
	I love to learn about the backstories of the character in video games	-0.478										
	Some of my favorite stories are in video games	-0.444										
Play My Way	I'm interested in learning the lore and history of video game worlds	-0.402										
	I prefer games that allow me to play however I want	0.694										
	I like games that offer different ways to get to the next level or area	0.625										
	I like games that do not put a lot of constraints on the player	0.561										
	I like games that offer different ways to get to the next level or area	0.494										
	I like games that offer you a lot of options and choices	0.492										
	Autonomy PENS 3	0.334										
Games as Life	Autonomy PENS 2	0.302										
	I think games are an integral part of my life	-0.727										
	Games are an important activity in my life	-0.679										
	If I don't play I would feel bad about myself	-0.551										
	I feel that I must play regularly	-0.513										
	I think games are an extension of me	-0.492										
	I must play to feel good about myself	-0.465										
	The satisfaction I feel when I accomplish a task successfully in games is stronger than the satisfaction I feel when I accomplish a task successfully in real life	-0.359										
	Games have personal significance to me	-0.352										
	It is important for me to know as much about the game mechanics and rules as possible	0.691										
Forward Thinking	I am interested in the precise numbers and percentages underlying the game mechanics	0.652										
	I like to figure out how game work inside and out	0.510										
	It is important that my character is as optimized as possible for their profession/role	0.491										
	I often use a character builder or a template to plan out my character's advancement at an early level	0.392										
Roleplaying	I often role-play my character	0.610										
	I often make up stories and histories for my characters	0.562										
	I really do my best to put myself into the main character's shoes	0.445										
	I enjoy the excitement of assuming an alter ego in a game	0.376										
	I often role-play my character	0.350										
Escape To	Immersion PENS 5	0.316										
	I like to do things in games which I cannot do in real life	-0.954										
	I like to do something that I could not normally do in real life through a video game	-0.921										
	I like to do something that I could not normally do in real life through a video game	-0.915										
	I play video games because they let me do things I can't do in real life	-0.846										
	I play video games because they let me do things I can't do in real life	-0.806										
	Video games allow me to pretend I am someone/somewhere else	-0.463										
	Video games allow me to pretend I am someone/somewhere else	-0.367										
	I play because in games I can make my own decisions	-0.351										
	I play because in games I can make a difference with my actions	-0.328										
Challenge	I enjoy highly difficult challenges in games	-0.903										
	I like it when goals are hard to achieve in games	-0.867										
	I like it when games challenge me	-0.792										
	I play because of the challenge	-0.740										
	I like it when progression in a game demands skill	-0.727										
	I usually play games at the highest difficulty setting	-0.605										
	I enjoy frantically escaping from a terrifying foe	-0.408										
	I play to master my skills and to win myself	-0.336										
	I enjoy working out how to crack a challenging puzzle	-0.336										
	I enjoy responding quickly to an exciting situation in a game	-0.325										
Strategic Thinking*												
Emotional Engagement	Immersion PENS 1									0.700		
	Immersion PENS 7									0.681		
	Immersion PENS 9									0.638		
	Immersion PENS 2									0.594		
	I feel emotionally attached to the characters in my favourite games									0.572		
	I play video games because they stimulate my emotions									0.571		
	I play because game events bring about emotions									0.516		
Progression	In video games, it's hard for me to identify with my character									0.317		
	I play to make progress and to achieve objectives										0.313	
	A failure in games makes me feel stressed											-0.705
	A failure in games makes me feel disappointed											-0.617
	In games, very often I feel stressed and pressured											-0.543
	I feel relief when I escape to a safe area											-0.397
	Immersion PENS 3											-0.382

17. Emotional Engagement
18. Progression

19. Emotional Reaction

Next, the correlations between factors were considered, shown in Table 30. In total, there were 12 correlations above 0.30 or below -.30. This indicates some factors share weak to moderate relationships. Interestingly, no factors load above 0.60, suggesting that overall the factors appear to represent distinct concepts.

Table 30: A Correlation matrix between the 19 factors. Correlations above 0.30 are in bold.

	Transportation Immersion	Competence	Completionism	Escape From	Exploration	I Like Games	Customisation	Smarten	I Like Stories	Play My Way	Games as Life	Forward Thinking	Roleplaying	Escape To	Challenge	Strategic Thinking	Emotional Engagement	Progression	Emotional Reaction
Transportation Immersion	1																		
Competence	0.041	1																	
Completionism	0.011	0.057	1																
Escape From	0.213	-0.054	0.088	1															
Exploration	0.277	0.028	0.245	-0.007	1														
I Like Games	-0.174	-0.339	-0.118	-0.051	-0.182	1													
Customisation	0.18	-0.028	0.11	0.121	0.229	-0.015	1												
Smarten	-0.163	-0.289	-0.122	-0.087	-0.077	0.127	-0.052	1											
I Like Stories	0.258	-0.02	0.036	0.006	0.197	-0.248	0.138	-0.082	1										
Play My Way	-0.186	-0.179	-0.036	-0.145	-0.324	0.182	-0.218	0.164	-0.089	1									
Games as Life	0.094	0.206	0.059	0.306	-0.023	-0.132	0.071	-0.287	0.017	0.027	1								
Forward Thinking	0.014	-0.304	-0.265	-0.031	-0.105	0.052	-0.047	0.203	0.071	0.201	-0.093	1							
Roleplaying	-0.289	0.039	-0.02	-0.098	-0.123	-0.066	-0.225	0.14	-0.273	0.174	-0.012	0.031	1						
Escape To	0.374	0.069	0.033	0.487	0.131	-0.21	0.204	-0.208	0.091	-0.327	0.081	-0.014	-0.239	1					
Challenge	-0.01	0.539	0.042	-0.086	0.102	-0.157	-0.077	-0.208	-0.144	-0.119	0.096	-0.339	0.06	-0.015	1				
Strategic Thinking	0.025	0.152	-0.04	-0.071	0.116	-0.175	0.031	0.008	0.03	-0.167	0.01	-0.09	0.06	0.017	0.232	1			
Emotional Engagement	-0.334	-0.071	0.042	-0.143	-0.119	0.141	-0.144	0.318	-0.363	-0.097	-0.212	-0.055	0.284	-0.298	-0.007	-0.002	1		
Progression	-0.117	-0.064	-0.202	-0.101	-0.057	0.106	-0.106	0.194	-0.055	0.124	-0.013	0.116	0.077	-0.119	-0.095	-0.022	0.107	1	
Emotional Reaction	0.064	-0.035	0.165	0.268	0.057	-0.061	0.069	-0.1	0.033	-0.04	0.233	-0.191	0.008	0.117	0.066	0.068	-0.115	-0.251	1

Many of the higher correlations relate to Emotional Engagement, as well as Competence, Escape From and Escape To. These relations are explained in their respective discussion sections. Each factor is now discussed in turn, to explore what it informs about the field of player traits.

4.7.1. Discussion

This analysis presents a clear picture of how items from a variety of questionnaires can load together. Many of the factors analysed for being semantically similar have again separated here, such as the creation of 2 Escapism sub-scales. There are more items in these categories than were originally analysed however — for example, Fantasy from Sherry also has formed the items in the Escape To sub-scale, but not the Escape From sub-scale.

Overall, the EFA has found 19 factors that summarise what the questionnaires studied have found in combination. Next, each is discussed in turn in terms of what its composition tells us about the concept, what is already known about it in the literature, and what it tells us about the field as a whole.

4.7.2. *Transportation Immersion*

This factor contains 7 items, where only 1 loads above 0.60. It has formed from a variety of sub-scales across questionnaires, and refers to players enjoying feeling as if they are ‘in’ the gameworld. Only items from 2 of the 4 scales named ‘Immersion’ have loaded here, suggesting the concept of immersion is more nuanced than is currently measured by the questionnaires in use. No items from VGPU loaded here, and only 2 from PENS, 1 from VKS, and 1 from IMG are found here, indicating an overlap in the measurement of Immersion across Psychology and Games research.

This factor is a specific type of immersion, discussed in detail by Jennet et al [22]. The authors found whilst Immersion can be seen as a unified concept, it has distinct sub-components. One of these is Transportation, which relates to the extent to which players feel they are ‘in’ a game rather than the real world. This is similar to the concept of Presence, where players feel they exist within the game world [69]. The fact this factor correlated with the Escape To factor therefore makes sense, as by entering into a gameworld players may feel more connected to it.

It is therefore unsurprising that this aspect of Immersion loaded distinctly by itself, and further supports the idea that Immersion has sub-components. The fact it is the only Immersion factor is interesting however, implying this is the most coherent concept currently measured by these questionnaires. Overall, this factor highlights that players enjoy the feeling of Immersion in games, though the way it is currently measured is specific to the Immersion felt via transportation into a game world.

4.7.3. *Competence*

This factor contains 10 items, where 5 items loaded over 0.60. These items address players feeling good about their gaming skills, and by extension think they are good at playing them. Competence has been studied widely in the literature, as a player’s confidence is seen as key to their enjoyment. For example, players enjoy a gaming experience more if they also feel competent in their playing abilities [13]. Because of this, Competence has been measured by a few separate questionnaires, such as PENS, IMG, and VGPU (in the form of the Generalised Game Self Efficacy sub-scale).

Interestingly, each item on this factor above 0.60 came only from PENS and VGPU, and no item from the IMG Competence sub-scale loaded onto this factor. This is especially interesting as IMG was inspired by PENS, so it was expected they would share similarities in loading. 2 of the IMG

Competence items instead load into the Challenge factor, and the item ‘I play to make progress and to achieve objectives’ loaded onto its own single-item factor. Therefore, the IMG Competence sub-scale does not measure the same concept of Competence as PENS. The PENS Competence sub-scale instead shows more similarity to the VGPU sub-scale Generalised Game Self Efficacy.

This factor has a medium correlation with the Challenge factor at 0.54, which is the strongest correlation found in the current analysis. This is understandable, as players who wish to feel good about their skills also may seek an environment in which to push the boundary of them. This in turn may give them higher feelings of competence. Overall, this factor highlights the similarities between the PENS and VGPU Competence sub-scales, and captures the concept that players enjoy games because they wish to feel good about their skills.

4.7.4. Completionism

This factor contains 13 items, where 9 loaded over 0.60. These items seem to assess whether players seek to engage with all the tasks and achievements available in games. Specifically, the language around ‘100%’ is common here, implying it is not simply a desire to engage with a lot of content, but to interact with everything a game offers. The items mostly come from the following sub-scales: Achiever from BrainHex, Grinding/Completion from GAMES, and Goal from Tondello. In fact, all items from these sub-scales loaded into this factor, except 1 from GAMES and 2 from Tondello. The item ‘I feel stressed if I do not complete all the tasks in a game’ from Tondello did load weakly onto this factor, but also weakly onto the ‘Emotional Reaction’ factor, and so was dropped. The other Tondello item ‘I like finishing quests’ instead loaded onto the ‘I Like Games’ factor. The item ‘I don’t mind grinding for an hour or two to get an item I want’ from GAMES did not load above 0.30 in the 19-factor solution.

Interestingly, the names of each of these original sub-scales would not imply they are so closely related — Achiever, Grinding/Completion, and Goal, sound like similar yet distinct concepts. This shows it can be difficult to semantically compare sub-scales, as the naming conventions researchers use can obscure similarities. Overall, the fact these sub-scales load together is understandable, as the items they contain all assess the concept of seeking to complete objectives. This implies that, although they are named differently, they are in fact addressing the same underlying concept.

The level of commitment a player has to the game content has been discussed in the past, and is mostly associated with a need to achieve. Specifically, players are motivated by the rewards given by a game, such as achievements [70]. Therefore, this Completionism factor has been created from multiple sub-scales, and shows that players can be motivated to play games to achieve many of the objectives available.

4.7.5. Escape From

There are 7 items in the Escape From scale, with 5 loading above 0.60. This factor specifically discusses using games to leave some aspect of real life behind, and therefore is about escaping from something. All these items originate from Escapism sub-scales, except ‘I enjoy the feeling of escape from real life through games’ from the Immersion sub-scale of VKS. This is not surprising however, as the item does contain the phrase ‘escape from real life’ within it. 4 items from Escapism by GAMES loaded here, with the other 2 loading onto the ‘Escape To’ factor. One item is missing from the Escapism sub-scale by Yee, which is ‘I often play to relax from the day’s work’, that loaded onto the ‘I Like Games’ factor. This makes sense, as it is unrelated to escaping and is instead about relaxing, a weakly related form of escaping.

This factor showed a medium correlation with the factor Escape To of 0.49, the other Escapism factor identified in this analysis. The relationship here is unsurprising, as when the Escapism sub-scales were discussed in Part IV there were strong correlations found. It also makes conceptual sense, as to wish to escape from something and escape into a game involves similar motivations. However, they were distinct enough to form separate concepts, highlighting the need to consider them as related but unique types of escapism. There was also a weak correlation with the ‘Games as Life’ factor at 0.31, again highlighting that players may use games to escape because this makes them feel better about themselves. Interestingly, the ‘Escape To’ factor showed little correlation to ‘Games as Life’, at 0.08.

Using games as an escape has been noted in the literature before, as games can help people to alleviate problems in their own life. Players sometimes use games to cope with real-life stressful events, or to help understand their own feelings towards a topic [71]. Therefore, this factor demonstrates that player may wish to use games as an escape from real life, potentially as it helps to improve mood.

4.7.6. *Exploration*

This factor contains 6 items, where 3 loaded above 0.60. It was created from 3 sub-scales; Discovery by Yee, Seeker by BrainHex, and Aesthetic by Tondello. Further, all items from Seeker load here, and half of Discovery. This factor relates to players exploring game environments, and finding enjoyment from doing so. Interestingly, no items from Autonomy/Exploration by GAMES loaded here, despite the name. This is because all 5 items loaded into Play My Way factor. When considering the items, this is not surprising — they appear to assess control over actions (the ‘Autonomy’ of Autonomy/Exploration), and have little to do with exploring the game world. It does however again highlight the need to carefully assess items within a sub-scale, as the name may not accurately capture what the items are measuring.

Exploration is considered an important behaviour in certain games. Open world games are especially known for this ability, as they provide large spaces for players to discover ([72]). Indeed, the definition of these games seems inherently tied to this ability; Szymanczyk et al. [73] defines them as “the use of large open worlds with a distinct gameplay focus on player exploration and discovery”.

Overall, this factor shows players wish to explore environments within games. However, as this factor only exists by combining sub-scales, it is currently difficult to assess exploration as a trait in players. No one sub-scale is currently apt to capture this, meaning further work is needed to understand the concept of Exploration.

4.7.7. *I Like Games*

This factor contains 12 items, where 4 items loaded over 0.60. The wording of many items relates to enjoying games, and shows a high degree of overlap. This indicates the structure may be bloated, where many similar items have been repeated without providing extra nuance. The items with lower loadings provide more interesting variance, such as ‘I play video games because they excite me’ (Arousal, Sherry) that loaded at 0.422. This item shows that the enjoyment of games may be related to the feeling of excitement. Interestingly, there are also items that relate to relaxing, such as ‘I play because playing games is relaxing’ (Fun, IMG) that loaded at 0.375. It is possible the highly loaded items with very similar wordings may be affecting this factor, but it is also possible players enjoy games in a variety of ways, with different intentions.

Overall, this factor is not surprising, as players do indeed enjoy games.

This topic has been widely studied in the past, as games are, after all, a form of entertainment [74]. Therefore, this factor does not add any new knowledge to the field, but reaffirms players can be motivated to play games because they are fun. It also highlights many similarities in item wording across various questionnaires.

4.7.8. Customisation

This factor contains 9 items, where all but two load above 0.60. All items from the original Yee and GAMES Customisation sub-scales loaded here, which was expected due to the similar findings in Part IV. Apart from the last 2 items, this factor exclusively refers to customising the player character, in a variety of ways. The items mention both the time spent customising, as well as what they choose to customise (e.g. outfit, appearance). Research has focused on player customisation in games previously, as they are a number of motivations for players to do so [75]. The ability to customise a game character is positively related to the experience of flow [76], as well as the enjoyment of a game [77].

Therefore this factor is unsurprising, and reflects a player's desire to control how they appear in the games they play.

4.7.9. Smarten

This factor contains 5 items, where all but one loaded above 0.60. All items from Smarty Pants by Trojan loaded here, and 2 of the 3 items from Identified Regulation by GAMS. A single item from the VKS Immersion sub-scale also loaded here. The factor seems to mostly reflect improving skills and learning beyond the game environment, something that has been studied in the literature before. This is because whether games can transfer skills to other domains, or be used to enhance skills, has been an interesting topic to many. For example, due to the socialising present in multiplayer online games such as *World of Warcraft*, it is possible for players to use these environments to learn key social skills, such as cooperation and group dynamics [78]. There is also evidence that those who play action games show improved visual skills outside of games [79].

Therefore, this factor shows a strong cluster of items that assess how players view games as places to improve themselves via learning and developing skills.

4.7.10. *I Like Stories*

This factor consists of 14 items, where 10 loaded above 0.60. These items primarily come from Story by GAMES and Narrative by Tondello. Indeed, all of the Narrative sub-scale loaded here, as well as 7 from GAMES' original 11, and both items of Story Driven by Trojan. However, the factor appears bloated by the concept of appreciating stories, without much nuance to any specific aspect of game narratives. This may indicate the field of individual player differences has keenly focused on this concept at a general level, resulting in numerous sub-scales being created with little depth.

The concept of appreciating game narratives is not new, and has been discussed at length in the literature. Adding narratives to previously non-narrative games can increase enjoyment and presence [80], and players typically enjoy complex, branching narratives more so than 'linear' ones [81]. Therefore, this factor does not add new understanding to individual player differences, but does highlight a notable focus on asking players about their feelings about stories.

4.7.11. *Play My Way*

This factor contains 7 items, with two loaded above 0.60. The entirety of the Autonomy/Exploration of GAMES sub-scale can be found here, along with two Autonomy items from PENS. Interestingly, no items from the IMG Autonomy sub-scale load here, despite it being based on PENS. Paired with the fact most of this sub-scale loaded weakly (5 out of 7 item below 0.60), this raises the question how clearly 'autonomy' is measured by the current questionnaires.

Autonomy was expected to be a more cohesive concept from this analysis, as it has been discussed extensively in the literature. Indeed, 3 questionnaires studied here contained sub-scales related to Autonomy, highlighting its popularity as a measurable concept. Self Determination Theory (SDT) for example states Autonomy is a key to the enjoyment of games, being something they are particularly good at providing [13]. Autonomy, alongside a sense of control, is shown to be a good predictor of game enjoyment when customising characters, more so than the attachment a player may have to a character [82].

It is therefore considered an important concept, but was relatively weak in this analysis — especially in comparison to other concepts that are equally discussed in the literature, such as Challenge. Furthermore, this factor did not show a strong correlation to the Customisation factor ($r = -0.218$), despite

the association found by Kim et al. [82]. This suggests the concept of Autonomy, when applied to individual player differences, may need further theoretical work.

Overall, the items in this factor inform on how players wish to control how they play games, though does not seem particularly related to exploration as named by GAMES. More work may be needed to further explore how this concept is measured.

4.7.12. Games as Life

This factor contains 8 items, where 2 loaded above 0.60. The items relate to players finding games an important aspect of their lives, and something they actively seek out. Many of the items came from a variety of the GAMS sub-scales, as well as 2 items from the Immersion sub-scale by VKS. It has been known that games are meaningful to players; they hold importance both in a person's culture and their identity (e.g. [83]). Therefore, this factor highlights that games are important to players, and is a motivation to play them.

4.7.13. Forward Thinking

This factor contains 5 items, where 2 loaded above 0.60. Most of this factor originates from Mechanics by Yee, where all the original items load here, and a single item from Completionist by Trojan. This factor appears to discuss players 'meta-gaming' (a term coined by Garfield [84]), where they consider their play tactics by bringing in other experiences and strategies, sometimes before the playing has truly started. This concept of strategising has been discussed in the past; Donaldson [85] explores how the ability to 'metagame' can help expert players in *League of Legends*, a competitive multiplayer game. Players who score highly on this concept look beyond the game to bring in information that will give them an 'edge', and so it makes sense this factor weakly correlated with both 'Competence' and 'Challenge', due to overlaps in wishing to achieve and to perform optimally.

Therefore, this factor highlights how players may wish to think more tactically about their play, using the mechanics of a game to achieve the best outcome.

4.7.14. Roleplaying

This factor consists of 6 items, where only 1 item loaded above 0.60. Three out of the four Roleplaying sub-scale items from Yee are found here, along

with three items from three separate sub-scales. The Yee item ‘I enjoy being immersed in a fantasy world’ instead loaded onto Transportation Immersion, which is understandable.

Roleplaying and embodying characters has been studied extensively in the past. Players typically form attachments by roleplaying a character, as they can embody their actions (e.g. [86]). Roleplaying can also allow players to reflect on their own identity [87], by experimenting with different aspects of themselves. However, this sub-scale consists of mainly low-loaded items, suggesting more work is needed to expand this concept to increase its measurability.

4.7.15. Escape To

The Escape To factor contains 9 items, with 5 loaded above 0.60. The fact these items load together is unsurprising, as many are repeated items — this is because many questionnaires were inspired from Sherry, and so used the same items in their sub-scales. Many of the items found here come from Escapism sub-scales. For example, all items of Fantasy by Sherry loaded here, except for ‘I enjoy the excitement of assuming an alter ego in a game’ that instead loaded onto Roleplaying (an understandable loading), as well as both items from Escapist by Trojan. Other items came from sub-scales such as 2 of the 3 items of Autonomy by IMG (the final item ‘I play because in games I can make meaningful choices’ loaded onto no other factors).

Escape To was created from a variety of sub-scales, in contrast to Escape From which was formed solely from 2 Escapism sub-scales (and 1 item from Immersion by VKS). This makes it a less defined concept in the literature at current, potentially because engaging with a game and the reason for doing so are not always distinct in questions asked of players. A player could enjoy embodying a character, and therefore be asked a question from an Immersion sub-scale such as the one from VKS. However, if they are doing so because they wish to escape into a game, this reason has been overlooked and so has been treated as only a question of Immersion.

Overall, Escape To indicates players wish to engage with activities within games that do not occur in real life, and so implies they wish to experience something different and/or unique. This could be done in a variety of ways, such as taking on the role of a character, or simply being a part of a fantasy game world that provides a new experience dissimilar to real life. However, the distinction between the types of Escapism found in this study is not deeply explored in the literature to date. The two are typically discussed

in close proximity to one another, such as Lazarro [88] who explains players may wish to play games to escape the real world, as well as experience new things they could not do otherwise do so (e.g., flying). Despite this, the conceptual difference is noticeable; whilst escaping *from* something implies a negative association with real life, this is not inherent for escape *to*. Players may wish to experience new things purely for the enjoyment it brings, and so ‘escape’ the real world in search of them. This may explain why Roleplaying and Transportation Immersion factors correlated stronger with Escape To than Escape From. The distinction in factors found here indicates a need to look more closely at the separation of these concepts, to understand how and when they differ.

Therefore, there appears to be two distinct instances of Escapism within individual player differences, that highlights how people may wish to escape from a negative experience in their real life, or experience something new and different within a game. This distinction may help explain how games affect players differently, as this different motivation may account for some of the variance found in Escapism-focused studies.

4.7.16. *Challenge*

This factor contains 10 items, where 6 loaded above 0.60. The majority of this factor comes from all 5 items of the Challenge sub-scale by Tondello, with 2 items from the Competence sub-scale by IMG, and the final 3 from a variety of BrainHex sub-scales. Notably, no other sub-scales named Challenge loaded onto this factor, a point previously discussed in Part IV. 2 of the items from Challenge by Sherry cross-loaded onto other factors, whilst the other 2 loaded onto no factors above 0.30. This is similar for the Challenge sub-scale from VKS, where two items cross-loaded elsewhere and the final item loaded nowhere. This further highlights how ‘Challenge’ as it is currently studied is not the same concept, and will require further defining.

Challenge has been studied widely from a number of perspectives. An important distinction to make here is the difference between enjoying Challenge in games, and the experience of a Challenging game. The factor found here relates to the trait of enjoying Challenge, rather than the outcomes associated with experiencing Challenges in games (as explored by [89]). There has been work that looks at trait Challenge, mostly associated with flow theory [90], which states players can achieve a state of flow when the game difficulty matches their current skill level. Because of this, players with higher skill will need more challenges to stay in flow, and enjoy the game. Research in this

area has shown this to be true, as higher skilled players experienced greater flow with more difficult game settings, whereas lower skilled did not [91].

Therefore, this concept of Challenge refers to players enjoying games that put their skill level to the test, and overcoming this adversity. This factor also highlights that the concept of Challenge must be considered more closely, as sub-scales measuring ‘Challenge’ are not interchangeable.

4.7.17. Strategic Thinking

This factor contains 3 items, however all cross-loaded above 0.30 with other factors. These items are two from the BrainHex Mastermind sub-scale (‘I enjoy working out what to do on my own’ and ‘I enjoy devising a promising strategy when deciding what to try next’), and one from the Challenge sub-scale of Sherry (‘I enjoy finding new and creative ways to work through video games’).

The factors these items cross-loaded with is understandable, as they all relate to players strategising in some way. Strategic Thinking items cross-loaded with Play My Way, Forward Thinking, and Challenge. This makes the factor small and not a complete concept, though strong enough to exist above 0.30, unlike many of the items that did not load onto any factor.

Strategic Thinking appears to be a sub-factor/related concept of players analysing their behaviour and formulating tactics to overcome game challenges. It will however need further investigation to understand how distinct it is.

4.7.18. Emotional Engagement

This factor contains 8 items, where 3 loaded above 0.60. Many of these items originate from Presence/Immersion of PENS, though each item specifically relates to players feeling as though the games they play have engaged their emotions. An interesting note is that this is not the same as *reacting* emotionally to in-game situations, a factor covered later in this discussion. Instead these items are specific to the feeling of emotional attachment, and how the game presents experiences that allow these feelings. Indeed, this factor showed correlations to other factors such as ‘I Like Stories’ and ‘Transportation Immersion’. Again, this is understandable — players are emotionally engaged by taking on the role of a character and being ‘transported’ into the game world.

Emotional Engagement has some links to Presence in the literature. It is known that games can provide players with a number of emotional ex-

periences, from a range of sources, such as the Story, the Artwork, or the Game Environment [92]. These emotional experiences can help keep players engaged, often by giving players a character to embody [93].

Overall, Emotional Engagement highlights how players seek to feel engaged in the games they play in an emotional sense.

4.7.19. Progression

This factor contains a single item; ‘I play to make progress and to achieve objectives’. It comes from Competence by IMG, and refers to players wishing to progress in games. The loading is weak, but for a single factor to load onto itself is impressive, highlighting how distinct the concept of ‘progression’ is to provide a unique variance in the model.

There has been discussion in the literature around progression and why it is a specific concept. Juul [94] discusses the difference between games of Emergence and games of Progression, explaining that Progression games are typically a series of challenges for a player to complete. This progression helps to keep players motivated, as it presents them with a series of tasks, and a sense of achievement when completed. The enjoyment of completing tasks in sequence appears separate to simply achieving objectives overall, which is covered in the Completionism factor.

Therefore, this factor highlights how Progression is a distinct concept not specifically measured at current in the questionnaires sampled. It is perhaps in need of further study before this is possible, as only 1 item out of 177 captured this concept.

4.7.20. Emotional Reaction

The final factor has 5 items, where 2 loaded above 0.60. These items came from Achievement by VKS, along with one item from Survivor by BrainHex and 1 Presence/Immersion by PENS. These items refer to general emotional reactions, such as pride, relief, and the feeling of adrenaline. This implies the original VKS sub-scale is fairly unique, as all its items loaded together and not many other items from other sub-scales loaded with it. However, ‘Achievement’ is perhaps not the best name for the factor, as these items did not load with other similarly named sub-scales (such as Achiever from BrainHex, or Goal from Tondello).

This factor is similar to the Emotional Engagement factor, but it is not related to being engaged with the game. Instead, this factor represents the feelings a game elicits, the emotional consequences of playing, which can

be both positive or negative (e.g., stressed vs relieved). Whilst there was a correlation between these factors ($r = 0.115$) this is weak, furthering the distinction between these factors.

Reacting to in-game scenarios emotionally has been studied in the past, as it is known that games can cause similar emotional reactions as though they are happening in real-life. Usually, this is related to the moral decision-making literature, which shows people react the same to moral dilemmas in games as they do in real-life [95]. These emotional reactions may influence the experience of a game, similarly to how being emotionally engaged can influence the experience.

Therefore, this factor highlights how reacting to emotional events is not entirely the same as engaging with them, as the process of reacting may imply a reflection process.

4.7.21. Summary

The above 19 factors highlight what concepts have been studied in the field of individual player differences, many of which have been discussed at length in the literature before. To see them here provides further evidence to their existence — such as how players enjoy the narrative of games, and enjoy customising their characters. However, there are also examples of concepts that have been understudied, or have not been discussed with enough nuance. For example, there are potentially two types of Escapism, which at current cannot accurately be assessed with the questionnaires in use. Another example is Progression, which is a distinct enough concept to load by itself, but has not been extensively studied or measured.

This analysis also highlights that, just because a specific sub-scale is unreliable does not mean its items need to be wasted. Out of the 177 items, only 20 did not load onto any items, and so many were salvaged and able to provide important insights. This is because many items provided more insight loading onto different concepts than the ones originally suggested, and by allowing each item to move independently led to a clearer and stronger structure. From this, the field of individual player differences can look at this analysis to see what concepts are currently studied, understudied, and in need of further nuance.

5. General Discussion

This work has sought to empirically assess the individual player difference questionnaires currently available, to reflect on the concepts currently measured by the field. With an ever-increasing number of questionnaires available for researchers to use, it has become increasingly difficult to understand which are the best suited for assessing unique concepts. Whilst it is possible to compare questionnaires and their sub-scales at face value (such as the work of [7]), little empirical testing has been done to confirm that these semantic comparisons do indeed reflect the statistical nature of the concepts. Further, despite individual player differences being an interdisciplinary field, there has been little focus on the influence of originating discipline on the development of concepts measured by questionnaires.

To explore these questions, 11 questionnaires were selected that provided an overview of individual player differences measurement, resulting in an analysis of 177 items from 45 sub-scales. The analysis was carried out in two parts; firstly, tests were conducted on the structure of the individual questionnaires. This work found at best weak to moderate reliability for all questionnaires involved, with two exceptions: Tondello [5] and GAMES [14]. Even scales with good reliability could be improved, indicating many questionnaires need to be reconsidered — if they cannot reliably measure individual player differences, any work founded upon them is open to false conclusions. Whilst there was indication of newer questionnaires improving on older versions (e.g., Tondello improves upon BrainHex), this was not commonplace. Typically, new questionnaires may ‘solve’ older issues, but also tend to create equally unstable factor structures that do not fully capture the concepts identified in the 177-item EFA.

Secondly, how questionnaires and their sub-scales compare to one another was assessed, via a series of exploratory correlations and factor analyses. A mixture of convergence and divergence was identified: some factors with semantic similarity do not always share statistical similarity, as many sub-scales with similar names were found to be weakly correlated and would load onto separate factors in an EFA. Notably, the concept of ‘Challenge’ varies greatly between questionnaires, whilst the concept of Escapism appears to be two distinct but related concepts; ‘Escaping From’ and ‘Escaping To’. An EFA run on the 177 items indicated a 19-factor solution, that summarises what concepts within individual player differences can currently be measured from the questionnaires in use. Items from all 4 originating disciplines freely loaded

together, indicating strong conceptual overlap between disciplines. One notable exception is again the concept of Challenge, which appears unique and distinct for each discipline.

These findings highlight that, in the case of questionnaires, more is not necessarily better. It is therefore important to stress here the aim of this analysis was not to create another questionnaire. This would be counter-intuitive; this paper exists because there are too many questionnaires already in existence, and creating another would not solve this. Instead, the goal was to explore what this combination of questionnaires currently measures, to identify convergences in items and concepts. This allows a reflection on the field as a whole, driven by how the field measures concepts. To discourage the EFA created here being used as a questionnaire, the current analysis purposefully did not include a Confirmatory Factor Analysis (CFA) or reliability analysis of the factor solution.

Overall, whilst there is high overlap in sub-scales and concepts, no one current questionnaire is apt to measure any of them. If there are indeed upwards of 19 factors of individual player differences (and almost certainly more not covered in this analysis), this highlights a deeper problem with the use of such scales for research. If all 19 factors are considered to vary in each individual player, the common ternary split of players into high, indifferent and low on each factor would lead to 3^{19} (approximately 1.2bn) distinct partitions of players into ‘types’. Even the simpler (and arguably) naive binary split would lead to approximately 0.5M distinct player categories. This would likely make any exploratory approach to individual player differences futile, and open to spurious significant findings. Certainly, an atheoretical approach seems likely either to fail or find false positives in a morass of disinterested significance tests. Therefore, to move the field forward, a reconsideration of what should be measured and why is required.

5.0.1. Future Work

A different approach is needed for measuring concepts. New questionnaires created should reflect upon the entire field, to identify what it will supposedly capture that others before it could not. This is beyond taking scales at face value, but evaluating the items within them, as well as conducting more empirical comparisons such as in this work. This will help reduce overlaps in scales, and help identify what is truly needed that could not be measured already. In doing so, a stronger focus on theoretical justifications for measurement is required, to ensure the concepts measured are valid.

Further, closer care should be taken when creating questionnaires. Few questionnaires employed basic item creation techniques, such as reversed items or generally having many items per scale. Adhering to these techniques can improve structure of the factor model; it is perhaps not by chance that the most reliable questionnaires are also those that included reversed items, and many items per sub-scale (especially noticeable in GAMES). By making these changes, less questionnaires need be created, and those that are will be to a higher standard.

5.1. Limitations

There are a number of limitations with this work, beyond those previously discussed in earlier sections. Firstly, this work uses 11 questionnaires, though there are certainly more available and in use in the literature (at least 23 as identified by the companion paper; [3]). These 11 were chosen to give an overview of the field, though it may be that those not included are more structurally sound and/or assess distinct concepts. As these are not included, they cannot be discussed.

Secondly, to reduce the number of items, all sub-scales relating to social interactions with other players were removed. This was a large number of items, and social behaviours within games is a widely studied area. As not all games involve others, these sub-scales were removed. Therefore, no motivation to play influenced by social factors has been analysed, so there is no conclusion to be made of the structure of these concepts. Future research could apply the same methodology here to these social sub-scales, to assess their structure in a similar way.

Thirdly, a number of research decisions were made to run such a study, including the rewording of certain items and the standardisation of Likert points. Whilst this is common practice both in the field ([3]) and in other comparison works ([20]), it does mean a typical comparison of reliability and validity cannot be gained from this analysis. Instead, this paper explored the general structure of the questionnaires, as a prelude to exploring the underlying concepts they measure. Future work could replicate this work without taking such decisions, however the practicality of such a task would likely result in different decisions being taken that equally affects comparability.

Other future work building upon this study could move beyond the current practice of only comparing sub-scales semantically to one another. Instead, anytime a new questionnaire is created, it should be compared empirically to other similar questionnaires, to identify any overlaps. This has

been done infrequently in the past (e.g., [20]), but there should be a move to make this a more common occurrence. This will help extend the field in its understanding of how each new questionnaire is capturing a distinct and important aspect of player preference, and not circling the same content in a slightly reworded way.

6. Conclusion

With such a large array of individual player difference questionnaires available to researchers, it has become difficult to know which questionnaire is most fit for purpose, if any. A lack of empirical comparison has led to the creation of evermore questionnaires, with little evidence any meaningfully build upon previous work. To overcome this problem, this work has sought to understand how the questionnaires currently used to measure individual player differences perform, both internally and in comparison to one another. By exploring similarities/differences between questionnaires, this work has also provided a reflection of what the field currently assesses overall.

40% of the sub-scales analysed were found to have weak reliability, with only 2 questionnaires containing a full set of reliable sub-scales. Many showed indications of unreliable items, and CFAs suggested many questionnaire structures were only adequate at best. These findings call into question any study using these measures to assess players, as this general trend of weak reliability may lead to false interpretations or conceptually unsound conclusions. When comparing across questionnaires, sub-scales with semantically similar names can share little to no empirical overlap, making it all but impossible to choose between them from a cursory glance. The final analysis, compiling 177 items into a 19-factor solution, provided a deep insight into how the field has measured individual player differences so far, and highlighted nuances that may need further consideration.

The sheer volume of factors, and the variability a player could therefore have for each, calls for a revaluation of how players are assessed. Without a theoretical underpinning for why certain individual player differences should correlate with behaviours/other personality measures, there is a high chance for false positive results, and the creation of even more questionnaires. Simply put, before moving forward the field must consider what are its aims for research, and how questionnaires will help answer these questions. Without a reconsideration of these goals, the field will likely continue to circle around the same concepts, with no hope to move forward. This is not to say

questionnaires are not a worthy methodology, capable of providing valid and meaningful insights. But there must, of course, be *method* to the madness.

7. Acknowledgements

This work was supported by [Blinded for Review]

References

- [1] R. Bartle, Hearts, clubs, diamonds, spades: Players who suit muds, *Journal of MUD research* 1 (1996) 19.
- [2] J. Park, Y. Song, C.-I. Teng, Exploring the links between personality traits and motivations to play online games, *Cyberpsychology, Behavior, and Social Networking* 14 (2011) 747–751.
- [3] N. Hughes, P. Cairns, Growing together: An analysis of measurement transparency across 15 years of player motivation questionnaires (2021).
- [4] J. L. Sherry, K. Lucas, B. S. Greenberg, K. Lachlan, Video game uses and gratifications as predictors of use and game preference, *Playing video games: Motives, responses, and consequences* 24 (2006) 213–224.
- [5] G. F. Tondello, K. Arrambide, G. Ribeiro, A. J.-I. Cen, L. E. Nacke, “i don’t fit into a single type”: A trait model and scale of game playing preferences, in: *IFIP Conference on Human-Computer Interaction*, Springer, 2019, pp. 375–395.
- [6] I. Voulgari, V. Komis, D. G. Sampson, Player motivations in massively multiplayer online games, in: *2014 IEEE 14th International Conference on Advanced Learning Technologies*, IEEE, 2014, pp. 238–239.
- [7] J. Hamari, J. Tuunanen, Player types: A meta-synthesis, *Transactions of the Digital Games Research Association—Vol. 1, No. 2* (2014) 29.
- [8] D. Johnson, M. J. Gardner, R. Perry, Validation of two game experience scales: the player experience of need satisfaction (pens) and game experience questionnaire (geq), *International Journal of Human-Computer Studies* 118 (2018) 38–46.

- [9] F. Brühlmann, G.-M. Schmid, How to measure the game experience? analysis of the factor structure of two questionnaires, in: Proceedings of the 33rd annual acm conference extended abstracts on human factors in computing systems, 2015, pp. 1181–1186.
- [10] P. Slovic, The construction of preference., *American psychologist* 50 (1995) 364.
- [11] N. Peever, D. Johnson, J. Gardner, Personality & video game genre preferences, in: Proceedings of the 8th australasian conference on interactive entertainment: Playing the system, 2012, pp. 1–3.
- [12] T. Hartmann, C. Klimmt, Gender and computer games: Exploring females’ dislikes, *Journal of Computer-Mediated Communication* 11 (2006) 910–931.
- [13] R. M. Ryan, C. S. Rigby, A. Przybylski, The motivational pull of video games: A self-determination theory approach, *Motivation and Emotion* 30 (2006) 344–360.
- [14] J. Hilgard, C. R. Engelhardt, B. D. Bartholow, Individual differences in motives, preferences, and pathology in video games: the gaming attitudes, motives, and experiences scales (games), *Frontiers in psychology* 4 (2013) 608.
- [15] A. Tyack, E. Mekler, Self-determination theory in hci games research: Current uses and open questions, in: CHI 2020 - Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, ACM, 2020. doi:10.1145/3313831.3376723.
- [16] N. Yee, Motivations for play in online games, *CyberPsychology & Behavior* 9 (2006) 772–775.
- [17] N. Yee, N. Ducheneaut, L. Nelson, Online gaming motivations scale: development and validation, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, 2012, pp. 2803–2806.
- [18] M. Busch, E. Mattheiss, R. Orji, P. Fröhlich, M. Lankes, M. Tscheligi, Player type models: Towards empirical validation, in: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, 2016, pp. 1835–1841.

- [19] L. E. Nacke, C. Bateman, R. L. Mandryk, Brainhex: A neurobiological gamer typology survey, *Entertainment Computing* 5 (2014) 55–62.
- [20] A. Denisova, A. I. Nordin, P. Cairns, The convergence of player experience questionnaires, in: *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, ACM, 2016, pp. 33–37.
- [21] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, J. N. Pidruzny, The development of the game engagement questionnaire: A measure of engagement in video game-playing, *Journal of experimental social psychology* 45 (2009) 624–634.
- [22] C. Jennett, A. L. Cox, P. Cairns, S. Dhoparee, A. Epps, T. Tijs, A. Walton, Measuring and defining the experience of immersion in games, *International journal of human-computer studies* 66 (2008) 641–661.
- [23] M.-A. K. Lafrenière, J. Verner-Filion, R. J. Vallerand, Development and validation of the gaming motivation scale (gams), *Personality and Individual Differences* 53 (2012) 827–831.
- [24] J. Jansz, C. Avis, M. Vosmeer, Playing the sims2: An exploration of gender differences in players’ motivations and patterns of play, *New Media & Society* 12 (2010) 235–251.
- [25] D. Lee, L. J. Schoenstedt, Comparison of esports and traditional sports consumption motives., *ICHPER-SD Journal Of Research* 6 (2011) 39–44.
- [26] D. J. Kuss, J. Louws, R. W. Wiers, Online gaming addiction? motives predict addictive play behavior in massively multiplayer online role-playing games, *Cyberpsychology, Behavior, and Social Networking* 15 (2012) 480–485.
- [27] J. Billieux, M. Van der Linden, S. Achab, Y. Khazaal, L. Paraskevopoulos, D. Zullino, G. Thorens, Why do you play world of warcraft? an in-depth exploration of self-reported motivations to play online and in-game behaviours in the virtual world of azeroth, *Computers in Human Behavior* 29 (2013) 103–109.

- [28] N. Yee, The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments, *Presence: Teleoperators and virtual environments* 15 (2006) 309–329.
- [29] D. Johnson, J. Gardner, Personality, motivation and video games, in: *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*, 2010, pp. 276–279.
- [30] J. Banyte, A. Gadeikiene, The effect of consumer motivation to play games on video game-playing engagement, *Procedia economics and finance* 26 (2015) 505–514.
- [31] D. J. Mills, M. Milyavskaya, N. L. Heath, J. L. Derevensky, Gaming motivation and problematic video gaming: The role of needs frustration, *European Journal of Social Psychology* 48 (2018) 551–559.
- [32] M. L. G. Comello, D. B. Francis, L. H. Marshall, D. R. Puglia, Cancer survivors who play recreational computer games: Motivations for playing and associations with beneficial psychological outcomes, *Games for health journal* 5 (2016) 286–292.
- [33] S. Peracchia, F. Presaghi, G. Curcio, Pathologic use of video games and motivation: Can the gaming motivation scale (gams) predict depression and trait anxiety?, *International journal of environmental research and public health* 16 (2019) 1008.
- [34] C. J. Ferguson, The good, the bad and the ugly: A meta-analytic review of positive and negative effects of violent video games, *Psychiatric quarterly* 78 (2007) 309–316.
- [35] D. Loffredo, A. Tavakkoli, Analyzing motives, preferences, and experiences in video game play, in: *The 20th World Multi-Conference on Systemics, Cybernetics and Informatics (WMSCI 2016)*, 2016, pp. 45–49.
- [36] S. Tekofsky, P. Miller, P. Spronck, K. Slavin, The effect of gender, native english speaking, and age on game genre preference and gaming motivations, in: *International Conference on Intelligent Technologies for Interactive Entertainment*, Springer, 2016, pp. 178–183.

- [37] V. Moreira, M. L. Okimoto, The relation of attention between player profiles: A study on the eye-tracking and profile brainhex, in: International Conference on Applied Human Factors and Ergonomics, Springer, 2018, pp. 425–434.
- [38] A. S. Kahn, C. Shen, L. Lu, R. A. Ratan, S. Coary, J. Hou, J. Meng, J. Osborn, D. Williams, The trojan player typology: A cross-genre, cross-cultural, behaviorally validated scale of video game play motivations, *Computers in Human Behavior* 49 (2015) 354–361.
- [39] S. García-Lanzo, A. Chamarro, Basic psychological needs, passion and motivations in amateur and semi-professional esports players, *Aloma: revista de psicologia, ciències de l’educació i de l’esport Blanquerna* 36 (2018) 59–68.
- [40] S. Turkay, S. Adinolf, Understanding online collectible card game players’ motivations: a survey study with two games, in: Proceedings of the 30th Australian Conference on Computer-Human Interaction, 2018, pp. 501–505.
- [41] D. R. Sanchez, M. Langer, Video game pursuit (vgpu) scale development: Designing and validating a scale with implications for game-based learning and assessment, *Simulation & Gaming* (2019) 1–32.
- [42] M. Griffiths, Z. Hussain, S. M. Grüsser, R. Thalemann, H. Cole, M. N. Davies, D. Chappell, Social interactions in online gaming, *International Journal of Game-Based Learning (IJGBL)* 1 (2011) 20–36.
- [43] C. Klimmt, H. Schmid, J. Orthmann, Exploring the enjoyment of playing browser games, *CyberPsychology & Behavior* 12 (2009) 231–234.
- [44] L. K. Kaye, J. Bryce, Putting the fun factor into gaming: The influence of social contexts on the experiences of playing videogames, *International Journal of Internet Science* 7 (2012) 24–38.
- [45] D. Zendle, P. Cairns, D. Kudenko, No priming in video games, *Computers in Human Behavior* 78 (2018) 113–125.
- [46] C. J. Ferguson, A. Garza, J. Jerabeck, R. Ramos, M. Galindo, Not worth the fuss after all? cross-sectional and prospective data on violent video game influences on aggression, visuospatial cognition and mathematics

- ability in a sample of youth, *Journal of youth and adolescence* 42 (2013) 109–122.
- [47] G. R. Szycik, B. Mohammadi, M. Hake, J. Kneer, A. Samii, T. F. Münte, B. T. Te Wildt, Excessive users of violent video games do not show emotional desensitization: an fmri study, *Brain imaging and behavior* 11 (2017) 736–743.
 - [48] J. Vahlo, J. Hamari, Five-factor inventory of intrinsic motivations to gameplay (img), in: *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019, pp. 2476–2485.
 - [49] P. Cairns, *Doing better statistics in human-computer interaction*, Cambridge University Press, 2019.
 - [50] J. J. Davies, T. J. Hemingway, Guitar hero or zero?: Fantasy, self-esteem, and deficient self-regulation in rhythm-based music video games, *Journal of Media Psychology* (2014).
 - [51] B. Monterrat, É. Lavoué, S. George, Adaptation of gaming features for motivating learners, *Simulation & Gaming* 48 (2017) 625–656.
 - [52] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. URL: <https://www.R-project.org/>.
 - [53] T. J. Dunn, T. Baguley, V. Brunsden, From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation, *British journal of psychology* 105 (2014) 399–412.
 - [54] J. C. Nunally, I. Bernstein, *Psychometric theory*. ed, NY: McGraw-Hill (1978).
 - [55] D. L. Streiner, Starting at the beginning: an introduction to coefficient alpha and internal consistency, *Journal of personality assessment* 80 (2003) 99–103.
 - [56] R. Cattell, *The scientific use of factor analysis in behavioral and life sciences*, Springer Science & Business Media, 2012.

- [57] R. P. Chalmers, mirt: A multidimensional item response theory package for the R environment, *Journal of Statistical Software* 48 (2012) 1–29. doi:10.18637/jss.v048.i06.
- [58] L. Cai, Metropolis-hastings robbins-monro algorithm for confirmatory item factor analysis, *Journal of Educational and Behavioral Statistics* 35 (2010) 307–335.
- [59] T. A. Brown, *Confirmatory factor analysis for applied research*, Guilford publications, 2015.
- [60] M. Matsunaga, How to factor-analyze your data right: Do’s, don’ts, and how-to’s., *International journal of psychological research* 3 (2010) 97–110.
- [61] P. Cabrera-Nguyen, Author guidelines for reporting scale development and validation results in the journal of the society for social work and research, *Journal of the Society for Social Work and Research* 1 (2010) 99–103.
- [62] J. L. Horn, A rationale and test for the number of factors in factor analysis, *Psychometrika* 30 (1965) 179–185.
- [63] W. Revelle, *psych: Procedures for Psychological, Psychometric, and Personality Research*, Northwestern University, Evanston, Illinois, 2018. URL: <https://CRAN.R-project.org/package=psych>, r package version 1.8.12.
- [64] F. J. Floyd, K. F. Widaman, Factor analysis in the development and refinement of clinical assessment instruments., *Psychological assessment* 7 (1995) 286.
- [65] P. Kline, *A psychometrics primer*, London: Free Association, 2000.
- [66] L. R. Fabrigar, D. T. Wegener, R. C. MacCallum, E. J. Strahan, Evaluating the use of exploratory factor analysis in psychological research., *Psychological methods* 4 (1999) 272.
- [67] H. Akoglu, User’s guide to correlation coefficients, *Turkish journal of emergency medicine* 18 (2018) 91–93.

- [68] K. G. Jöreskog, How large can a standardized coefficient be (1999).
- [69] R. Tamborini, P. Skalski, The role of presence in the experience of electronic games, *Playing video games: Motives, responses, and consequences* (2006) 225–240.
- [70] C. Cruz, M. D. Hanus, J. Fox, The need to achieve: Players’ perceptions and uses of extrinsic meta-game reward systems for video game consoles, *Computers in Human Behavior* 71 (2017) 516–524.
- [71] I. Iacovides, E. D. Mekler, The role of gaming during difficult life experiences, in: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–12.
- [72] N. G. Hughes, P. Cairns, Opening the world of contextually-specific player experiences, *Entertainment Computing* 37 (2021) 100401.
- [73] O. Szymanczyk, P. Dickinson, T. Duckett, et al., From individual characters to large crowds: Augmenting the believability of open-world games through exploring social emotion in pedestrian groups (2011).
- [74] P. Vorderer, J. Bryant, K. M. Pieper, R. Weber, *Playing video games as entertainment*. (2006) 1–7.
- [75] H. Lin, H. Wang, Avatar creation in virtual worlds: Behaviors and motivations, *Computers in Human Behavior* 34 (2014) 213–218.
- [76] A. R. B. Soutter, M. Hitchens, The relationship between character identification and flow state within video games, *Computers in human behavior* 55 (2016) 1030–1038.
- [77] S. Treppe, L. Reinecke, Avatar creation and video game enjoyment, *Journal of Media Psychology* (2010).
- [78] N. Ducheneaut, R. J. Moore, More than just ‘xp’: learning social skills in massively multiplayer online games, *Interactive Technology and Smart Education* (2005).
- [79] C. S. Green, D. Bavelier, Action video game modifies visual selective attention, *Nature* 423 (2003) 534–537.

- [80] E. F. Schneider, A. Lang, M. Shin, S. D. Bradley, Death with a story: How story impacts emotional, motivational, and physiological responses to first-person shooter video games, *Human communication research* 30 (2004) 361–375.
- [81] C. Moser, X. Fang, Narrative structure and player experience in role-playing games, *International Journal of Human-Computer Interaction* 31 (2015) 146–156.
- [82] K. Kim, M. G. Schmierbach, M.-Y. Chung, J. D. Fraustino, F. Dardis, L. Ahern, et al., Is it a sense of autonomy, control, or attachment? exploring the effects of in-game customization on game enjoyment, *Computers in Human Behavior* 48 (2015) 695–705.
- [83] D. Muriel, G. Crawford, *Video games as culture: considering the role and importance of video games in contemporary society*, Routledge, 2018.
- [84] R. Garfield, *Metagames, Horsemen of the apocalypse: Essays on role-playing* (2000) 14–21.
- [85] S. Donaldson, Mechanics and metagame: Exploring binary expertise in league of legends, *Games and Culture* 12 (2017) 426–444.
- [86] M. L. Lewis, R. Weber, N. D. Bowman, “they may be pixels, but they’re my pixels:” developing a metric of character attachment in role-playing video games, *CyberPsychology & Behavior* 11 (2008) 515–518.
- [87] S. L. Bowman, *The functions of role-playing games: How participants create community, solve problems and explore identity*, McFarland, 2010.
- [88] N. Lazarro, Why we play games: Four keys to more emotion without story, in: *Game developer’s conference*, San Jose, 2004.
- [89] A. Denisova, P. Cairns, C. Guckelsberger, D. Zendle, Measuring perceived challenge in digital games: Development & validation of the challenge originating from recent gameplay interaction scale (corgis), *International Journal of Human-Computer Studies* 137 (2020) 102383.

- [90] M. Csikszentmihalyi, I. S. Csikszentmihalyi, Optimal experience: Psychological studies of flow in consciousness, Cambridge university press, 1992.
- [91] S.-A. A. Jin, “toward integrative models of flow”: Effects of performance, skill, challenge, playfulness, and presence on flow in video games, *Journal of Broadcasting & Electronic Media* 56 (2012) 169–186.
- [92] J. Frome, Eight ways videogames generate emotion., in: *Digra conference*, 2007.
- [93] M. D. Dickey, Engaging by design: How engagement strategies in popular computer and video games can inform instructional design, *Educational technology research and development* 53 (2005) 67–83.
- [94] J. Juul, The open and the closed: Games of emergence and games of progression., in: *CGDC Conf.*, 2002.
- [95] A. J. Weaver, N. Lewis, Mirrored morality: An exploration of moral choice in video games, *Cyberpsychology, Behavior, and Social Networking* 15 (2012) 610–614.