# Assessing Type Agreeability in the Unified Model of Personality and Play Styles

Alexander Brooke
*Department of Computing and Mathematics*
*Manchester Metropolitan University*
Manchester, England
a.brooke@mmu.ac.uk

Matthew Crossley
*Department of Computing and Mathematics*
*Manchester Metropolitan University*
Manchester, England
m.crossley@mmu.ac.uk

Huw Lloyd
*Department of Computing and Mathematics*
*Manchester Metropolitan University*
Manchester, England
huw.lloyd@mmu.ac.uk

Stuart Cunningham
*School of Computer and Engineering Sciences*
*University of Chester*
Chester, England
s.cunningham@chester.ac.uk

***Abstract*—Classifying players into well defined groups can be useful when designing games and gamified systems, with many models relating to player or personality 'type'. The Unified Model of Personality and Play Styles groups together many player and personality taxonomies, but whilst similarities have been noted in previous work, the overlap between models has not been analysed ahead of its use. This study provides evidence both for and against aspects of the Unified Model, with model agreeability assessed through comparison of participant classifications. Results show that representations of types related by the Unified Model do correlate significantly greater than types unrelated by the model, but do so with only weak-to-moderate correlation coefficients. Ranking classifications leads to results better mapping to the Unified Model, but also reduces the overall strength of correlations between types. The Unified Model is therefore considered fit for purpose as an explanatory tool, but without additional study should be used with caution in further use cases.**

***Index Terms*—games, player typology, personality, play style**

## I. Introduction

Various models describing personality and play styles exist to provide perspectives from multiple disciplines. Such models are then utilised in relation to gaming, with the design of both game content and mechanics often making use of theory describing player 'types'. Gamified systems that provide users with a level of game-adjacent interactivity, from virtual learning environments to fast-food marketing campaigns, then make further use of such models, extending the use of player types beyond stereotypical gaming environments. User research then provides further applications for both player and personality types, with the underlying psychological implications of these models often studied in relation to player or user experience, whether via behaviour or affect.

As one example, Tondello [1] presents five player orientations, that describe an individual's preference for different game elements and play styles. This is created in response to statistical analysis of results collected from previous studies into player motivation [2], and is validated through the use of a self-report questionnaire. Deviating from previous examples, gamification and design efforts outside of the sphere of academia are commonly rooted in older typologies, such as Bartle's taxonomy of players [3], despite their relative lack of analysis.

Grouping models in this vein together, Stewart [4] describes "Personality and Play Styles: A Unified Model", in which a series of existing player and personality models are combined under the assertion that they share many of the same ideologies. This is then used as a key reference in multiple academic studies [5–11], often as a means to extend findings to a greater set of applications. This is done despite both the lack of validation for many of the models grouped by Stewart, and the lack of validation for the groupings themselves. Previous work assessing the validity of the model describes both promising and concerning points made by Stewart, but is again done without analysis of collected data [12, 13].

It is with the Unified Model's use in previous research in mind, that this work seeks to understand whether constructed measures of the theoretical models put forward by Stewart [4] show any inherent agreeability. With the predominant use-case of the Unified Model being justification of model similarities, likened models are tested against each other, with validation of constructed measures and comparison to external validated tests being a potential focus for future work.

## II. The Unified Model of Personality and Play Styles

Before discussing previous work comparing existing player and personality models, we discuss the Unified Model of Personality and Play Styles itself, looking at the various typologies and taxonomies that Stewart [4] compiles. To aid clarity, we define four terms:

1) Model or Unified Model - The Unified Model of Personality and Play Styles [4].

2) model(s) - The personality and play style models that the Unified Model aggregates. Additional sets of descriptors put forward by Stewart [4] are also defined as models for brevity.
3) Type(s) - Individual classifications within each model, these may be personality traits, play styles or descriptors.
4) Archetype(s) - The four core personality and play style categories that the Unified Model proposes. Each archetype relates to one type from each model.

### A. Four Archetypes

The Unified Model posits that there are four core personality archetypes, upon which many personality models and player typologies have been built, despite their changes in perspective. These four archetypes are (named as they are by Bartle [3]):

A) Killer - Killers enjoy adrenaline seeking, thrills, chaos and impacting others for their own amusement.
B) Achiever - Achievers enjoy achievement, collecting, completion, speed running and competition.
C) Explorer - Explorers enjoy exploration of game worlds: physical space, narratives, and the use and limitations of game mechanics.
D) Socialiser - Socialisers enjoy social interaction and forming meaningful connections, activities are secondary to the experience with others.

Stewart [4] likens each archetype to types found in theories and frameworks describing players [3], play [14], fun [15], temperament [16], organisational cultures [17] and more, across his Unified Model. Additional descriptors relating to player personification, motivation, problem solving attitude, overall goal and exemplary character class are then given to each archetype to provide additional context.

### B. Notes of Concern

Despite the Unified Model's appearances in previous work, it should be noted that various points of concern arise from its design. Other than its lack of validation in any form, models accumulated by Stewart often do no fit the intent of their original creation, or have been shown to lack validity on multiple occasions. Keirsey's temperaments [16] for example, are used as a core focus of Stewart's model, and yet are built on Myers' framework [18] which has been shown to inaccurately describe personality in previous work [19]. Multiple models are then expanded or condensed to fit Stewart's four archetype schema [20–22]. Bartle's taxonomy of players [3] is used in its four archetype form, despite later extension of the model by Bartle [23], and the MDA framework is included despite it being a model of design components [22]. Later research into player traits and motivations then describes qualities that do not easily map to the Unified Model, whilst providing a greater level of empirical validation [24, 25].

### C. A Complete Model

A summary of the complete Unified Model, put forward by Stewart [4], can be seen in Table I, with the addition of the Holland 'themes', as included in an earlier version of the Model [26] and the remainder of this study.

TABLE I
THE UNIFIED MODEL OF PERSONALITY AND PLAY STYLES

| Model | *Type A* | *Type B* | *Type C* | *Type D* |
|---|---|---|---|---|
| KEIRSEY [16] | Artisan | Guardian | Rational | Idealist |
| BARTLE [3] | Killer | Achiever | Explorer | Socialiser |
| CAILLOIS [14] | Ilinx | Agôn | Mimesis | Alea |
| LAZZARO [15] | Serious Fun | Hard Fun | Easy Fun | People Fun |
| GNS+ [4, 21], | Experientialist | Gamist | Simulationist | Narrativist |
| MDA+ [4, 22] | Kinetics | Mechanics | Dynamics | Aesthetics |
| HANDY [17] | Power | Role | Task | People |
| GALLUP [27] | Impacting | Striving | Thinking | Relating |
| COVEY [28] | Power | Security | Wisdom | Guidance |
| HOLLAND [29] | Things | Data | Ideas | People |
| MOTIVATION [4] | Power | Security | Knowledge | Identity |
| PROBLEM SOLVING [4] | Performance | Persistence | Perception | Persuasion |
| OVERALL GOAL [4] | Do | Have | Know | Become |
| PERSONIFICATION [4] | Hands | Heart | Head | Spirit |
| CHARACTER [4] | Thief | Warrior | Wizard | Cleric |

## III. RELATED WORK

Work by Hamari and Tuunanen [12] breaks down many of the core motivations put forward in previous work on player taxonomies, and determines key concepts found across the literature in an attempt to gain insight to the same end as the Unified Model, understanding the connection between existing typologies. Their work discusses the implications of Stewart's grouping of typologies into one universal Model, with some combinations deemed "strange and unfitting".

Work making use of the Unified Model by Al-Taei [10] showed poor results when attempting to predict Bartle type classifications calculated using a mapping from the Myers-Briggs Type Indicator [18], as would be suggested possible by the Unified Model. Poor results in this direct translation between types describes a failure of the model, although this is never reported as a finding of the work.

Testing the overlap between existing personality and player types unrelated by the Unified Model has however shown some previous success, with elements from the Big 5 personality model [30] showing relation to elements from the DGD1 [20] in previous work [31]. Similarly to the majority of models included in the Unified Model, the DGD1 itself does not have a validated measure, with participants in previous work categorising themselves based on author-constructed classification descriptions. Despite the potential misrepresentation or under-representation of each type due to this, statistically significant differences were found between how participants from each type responded to measures of conscientiousness and openness.

Francis et al., [32] determine the relationship between the Keirsey Temperament Sorter (KTS) [16] and the Short-Form Revised Eysenck Personality Questionnaire (EPQ) [33]. Relationships between the types evaluated by each survey were determined and described using Pearson's correlation coefficient, with multiple significant correlations found between measured personality types. Consideration of this data highlighted potential implications for the division of participants in each model.

With similar aims to this study, Hughes and Cairns [34] describe the comparison and validation of a series of questionnaires created to categorise players by their traits. Analysis was conducted following the collection of participant responses to each questionnaire, with comparisons drawn between models using Spearman's rank correlation tests. Validation of the surveys utilised provided key insights into the structure and stability of each test, whilst the comparison of models provided understanding applicable to the field at large.

This study seeks to take initial steps towards the level of analysis performed by Hughes and Cairns. A method similar to that seen in previous work on models without existing measures is employed [31], with a hope to understanding the face-value usability of the Unified Model and it's constituent parts.

## IV. Testing the Model

With an understanding of the aims of the Unified Model and each of the models it proposes to accumulate, a series of research questions were posed to test the agreeability between type and model evaluations-

RQ1) Do personality and play style types correlate more strongly with types within the same archetype as described by the Unified Model than with types in other archetypes?

RQ2) Which individual model types are most strongly correlated within the Unified Model?

RQ3) Which models have the highest average correlation between their types from within the Unified Model, using this as a measure of model agreeability?

To answer these questions, participants were recruited using convenience sampling via social media, conducted following ethical approval from Manchester Metropolitan University. Participants were provided with a questionnaire that classified them under each of the frameworks described in the Model, using constructed measures where required, similarly to the successful work of McMahon et al. [31]. Analysis of the relationship between responses therefore tests the Unified Model for its general usability in the field, with further work required to create validated self-report tests for each of the theoretical models in question, should there be further need of the Model beyond consideration of previous work making use of it.

### A. Survey Design

Requiring participant classifications under each model in the Unified Model, this study made use of applicable existing self-report methods where available, with additional survey questions created in the absence of such methods. Survey items were evaluated by the research team, with an understanding that self-categorisation based on representations of existing models has shown success in previous work [31]. Likert scales were employed to allow for a level of intensity attributable to each classification as also shown previously [31]. The finalised questionnaire totalled 205 items and was made publicly available[1].

*1) Demographic and Background Questions:* Demographic information was collected from participants to provide context to the findings presented in this study.

*2) Keirsey, BrainHex and Bartle:* Participants were provided with links to existing surveys where applicable, with results collected on completion. The NERIS Type Explorer [35] (via the 16Personalities.com website) was selected as a widely used implementation of Myers' [18] framework with previous validation [36]. From this, participant classifications under Keirsey's temperaments were calculated. BrainHex [2] survey results were then requested as an updated analogue for results for the DGD1 [20] model, although this was later removed due to a technical error on the BrainHex website. Results were then requested for the Bartle Test of Gamer Psychology [37].

*3) Constructed Measures:* Survey questions were then constructed to evaluate participants in models that do not have existing, or freely available, measures. These were formed from a series of statements describing the participant, such as "I enjoy games that give me challenges." specifically relating to each of the types in each model. These statements were rated by participants using a five-point Likert scale, allowing for responses ranging from "Strongly Disagree" to "Strongly Agree" (or similar). Whilst the survey items used in this study have not been psychometrically validated, findings related to the scales used act to give initial understanding to the Unified Model. Further work utilising the Unified Model would do well to compare results to validated tests such as those described in Hughes and Cairn [34], although this is beyond the scope of this study.

*4) Additional Descriptors:* The process described previously for models without existing measures was then applied to the additional descriptors created by Stewart [4].

### B. Survey Limitations

Findings from this study should be considered with respect to the following limitations.

*1) Model Representation:* A prominent limiting factor of data collected in this study comes from the absence of existing measures for each of the models in the Unified Model. Representing personality and play style models in a format suitable for self-report should therefore be seen as an inherent limitation of any research into this subject, with previous work making use of the Unified Model despite its reliance on models that have not been validated. Models were also presented in English, limiting recruitment to only English speaking participants.

*2) Survey Fatigue:* Due to the large number of models tested, response credibility was improved through the use of Instructed Response Items (IRIs) following advice given in previous work [38], combatting the effects of "survey fatigue". Two IRIs were used to assess whether participants had lost focus over the course of the survey.

[1]Questionnaire available at: https://forms.office.com/r/hXmLv2cbri

*3) Participant Count:* 125 participants were surveyed as a part of this study, with data from 15 discarded inline with the previously described engagement requirements. Further understanding may be taken from future studies with larger participant counts, focused on specific demographics or a smaller subset of the models encompassed by the Unified Model.

### C. Classifying Participants

A series of methods for classifying participants were tested to understand how best to utilise the collected data.

*1) Base Classifications:* 'Base classifications' were calculated from the responses collected, with results then normalised appropriately. Keirsey temperament classifications were calculated using participant's NERIS Type Explorer results, converting them to a one-hot encoded classification vector.

*2) Ranked Classifications:* A second classification vector was created for each participant by ranking base classification types within each model. Equal type values were given an averaged, shared rank.

### D. Statistical Analysis

Analysis was performed on the classification data acquired, with $\alpha = 0.01$ used as a threshold for statistical significance in all appropriate tests. A correlation matrix was created for both methods of classifying participants, representing each pair of types using the Pearson's correlation coefficient, or the Spearman's rank correlation coefficient as appropriate. $p$ values were then calculated in a further matrix per classification method. This was used to collect results in response to RQ2. An unpaired t-Test was conducted to assess the overall fit of the archetypes put forward by the Unified Model, testing for a significant difference between the populations of *same-archetype* and *different-archetype* type correlation coefficients. This was then repeated on a per-model type basis, responding to RQ1. Further to this, correlation coefficients related to each pair of models in the Unified Model were averaged in answer to RQ3.

## V. Results and Discussion

A total of 125 survey responses were collected through Microsoft Forms. Of these, the majority of participants responded incorrectly to the first instructed response item, leading to it being removed from the study (the instructed response being considered too hidden within the question). 15 participants were therefore removed because they responded incorrectly to the second IRI. BrainHex results were only collected for the first 50 respondents, after which a technical fault in the externally hosted implementation of the survey led to it being removed from the study. Results relating to BrainHex will therefore also not feature in the further discussion of this dataset.

Results from the analysis of data collected from the remaining 110 participants (60 male, 37 female, 11 other, 2 unknown, approximate age range 18-64) will be provided in this section.

### A. Type-to-Type Relationships

When looking at individual type-to-type relationships, this study shows that the Unified Model is often correct in its assertions over which types relate to each other, with many of the most strongly correlated pairs relating to the same archetype. Tables II and III provide a summary of the most strongly correlated types from various subsets of the data, using Pearson's and Spearman's rank correlations respectively. All values provided are statistically significant, except for negative correlations between types from the same archetype, of which only the strongest Spearman's rank correlation is statistically significant. The most strongly correlated pairings across the dataset are shown to be types from the same archetype, a success of the Model. The category "Strongest Positive Correlations from the Same Archetype" has been omitted from each table due to its overlap with the "Strongest Correlations Between Any Types" category due to this.

TABLE II
Most Highly Correlated Pairs of Types Under Various Categories Using Pearson's Correlation

| Ranking Category | Model Pairing | $r$ | $p$ |
|---|---|---|---|
| Strongest Correlations Between Any Types | (C) Lazzaro x (C) Caillois | 0.702 | $1.40 \times 10^{-17}$ |
| | (A) Covey x (A) Handy | 0.694 | $4.29 \times 10^{-17}$ |
| | (D) Holland x (D) Gallup | 0.642 | $4.00 \times 10^{-14}$ |
| | (A) Goal x (A) MDA+ | 0.631 | $1.46 \times 10^{-13}$ |
| | (A) Goal x (A) Caillois | 0.587 | $1.58 \times 10^{-11}$ |
| Strongest Positive Correlations from Different Archetypes | (D) Handy x (A) Gallup | 0.573 | $6.31 \times 10^{-11}$ |
| | (D) GNS+ x (C) GNS+ | 0.544 | $7.91 \times 10^{-10}$ |
| | (B) Goal x (A) Motivations | 0.543 | $9.08 \times 10^{-10}$ |
| Strongest Negative Correlations from the Same Archetype | (B) Personification x (B) Problems | -0.173 | 0.071 |
| | (B) Personification x (B) Holland | -0.158 | 0.098 |
| | (A) Holland x (A) Lazzaro | -0.155 | 0.106 |
| Strongest Negative Correlations from Different Archetypes | (D) Keirsey x (C) Keirsey | -0.672 | $9.83 \times 10^{-16}$ |
| | (C) Bartle x (B) Bartle | -0.523 | $4.43 \times 10^{-9}$ |
| | (D) Bartle x (A) Bartle | -0.496 | $3.65 \times 10^{-8}$ |

TABLE III
Most Highly Correlated Pairs of Types Under Various Categories Using Spearman's Rank Correlation

| Ranking Category | Model Pairing | $r$ | $p$ |
|---|---|---|---|
| Strongest Correlations Between Any Types | (A) Covey x (A) Handy | 0.561 | $1.87 \times 10^{-10}$ |
| | (D) Holland x (D) Gallup | 0.538 | $1.34 \times 10^{-9}$ |
| | (C) Lazzaro x (C) Caillois | 0.538 | $1.39 \times 10^{-9}$ |
| | (A) Motivations x (A) Covey | 0.527 | $3.47 \times 10^{-9}$ |
| | (D) Holland x (D) Handy | 0.501 | $2.56 \times 10^{-8}$ |
| Strongest Positive Correlations from Different Archetypes | (D) Motivations x (C) Caillois | 0.455 | $5.84 \times 10^{-7}$ |
| | (D) MDA+ x (B) Personification | 0.414 | $6.94 \times 10^{-6}$ |
| | (B) Bartle x (A) Holland | 0.378 | $4.59 \times 10^{-5}$ |
| Strongest Negative Correlations from the Same Archetype | (B) Personification x (B) Holland | -0.281 | 0.003 |
| | (C) Personification x (C) Caillois | -0.223 | 0.019 |
| | (C) Personification x (C) Lazzaro | -0.221 | 0.020 |
| Strongest Negative Correlations from Different Archetypes | (D) Goal x (A) Handy | -0.381 | $4.03 \times 10^{-5}$ |
| | (C) Caillois x (A) Covey | -0.369 | $7.18 \times 10^{-5}$ |
| | (D) Gallup x (C) Keirsey | -0.367 | $8.06 \times 10^{-5}$ |

In response to RQ2, the strongest correlations between any types are all: positive, between types from the same archetype, and statistically significant below $\alpha = 0.01$, with a preference for Lazzaro's Easy Fun and Caillois' Mimesis showing the strongest correlation. Similarly, when looking at only the strongest negative correlations, it can be seen that all describe statistically significant negative relationships between types from dissimilar archetypes, as the Unified Model predicts.

Tables II and III also describe a series of datapoints that mark failures in the Unified Model's assertions. Multiple statistically significant positive correlations are found between types from different archetypes, using both Pearson's correlation and Spearman's rank correlation coefficients, in opposition to what would be assumed using the Unified Model. Further apparent failures of the Unified Model can be seen in negative Spearman's rank correlation coefficients between types attributed to the same archetype. Of these pairings however, only the most strongly correlated is statistically significant.

### B. Archetype-to-Archetype Relationships

Whilst description of the strongest correlations gives some understanding of the individual successes and failures of the Unified Model, more generalised results provide a clearer overview of the Model's overall performance. Figures 1 and 2 display the probability distribution of correlations between types from the same and different archetypes using Pearson's correlation and Spearman's rank correlation respectively. Ranked classification reduces the overall strength of correlations, with the mean of distributions in Figure 2 marginally lower than those in Figure 1. Despite this, and the overall weakness in mean correlation, the difference between the 'Same' and 'Different' distributions is statistically significant for both methods. This study has therefore shown a small but significant difference in how types related by the Unified Model relate to each other, over types that the Model does not relate to each other, responding to RQ1.

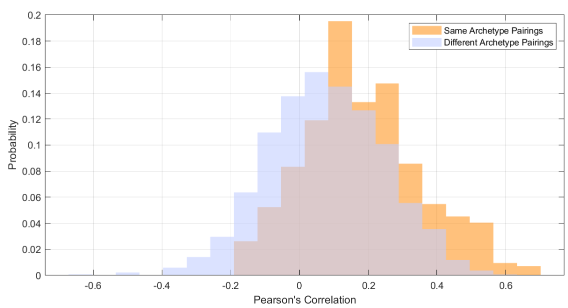


Fig. 1. Probability distribution of Pearson's correlation coefficients between types in the same and different archetypes. The 'Same' type distribution has a mean of 0.185 and a standard deviation of 0.172. The 'Different' type distribution has a mean of 0.08 and a standard deviation of 0.170. The two distributions are significantly different, with a $p$ value of $1.7 \times 10^{-28}$. Probability distributions are given to allow for the disparity in sample sizes between the two populations.

Whilst same-archetype relationships have been shown to outperform different-archetype relationships overall, comparisons between each pair of archetypes provides further detail. Table IV details the count of significant positive correlations between types attributed to each combination of archetypes (discounting comparisons that would pair each type with itself).

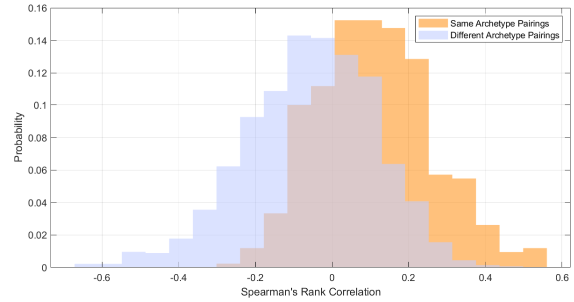


Fig. 2. Probability distribution of Spearman's rank correlations between types in the same and different archetypes. The 'Same' type distribution has a mean of 0.110 and a standard deviation of 0.149. The 'Different' type distribution has a mean of -0.056 and a standard deviation of 0.169. The two distributions are significantly different, with a $p$ value of $5.0 \times 10^{-67}$. Probability distributions are given to allow for the disparity in sample sizes between the two populations.

TABLE IV
COUNT (AND PERCENTAGE) OF POSITIVE CORRELATIONS SIGNIFICANT AT $\alpha = 0.01$

| **Significant Positive Pearson's Correlations** | | | | |
|---|---|---|---|---|
| | A | B | C | D |
| A | 68 (32.4%) | 51 (22.7%) | 26 (11.6%) | 51 (22.7%) |
| B | 51 (22.7%) | 44 (21.0%) | 33 (14.7%) | 28 (12.4%) |
| C | 26 (11.6%) | 33 (14.7%) | 80 (38.1%) | 39 (17.3%) |
| D | 51 (22.7%) | 28 (12.4%) | 39 (17.3%) | 90 (42.9%) |
| **Significant Positive Spearman's Rank Correlations** | | | | |
| | A | B | C | D |
| A | 28 (13.3%) | 13 (5.8%) | 1 (0.4%) | 2 (0.9%) |
| B | 13 (5.8%) | 20 (9.5%) | 6 (2.7%) | 5 (2.2%) |
| C | 1 (0.4%) | 6 (2.7%) | 32 (15.2%) | 6 (2.7%) |
| D | 2 (0.9%) | 5 (2.2%) | 6 (2.7%) | 64 (30.5%) |

Table IV provides perspective of the successes of archetype D, whilst also framing the failures of the Model overall. Whilst archetype D provides the greatest number of significantly paired types, these still only account for 42.9% of the possible archetype D to archetype D comparisons when using Pearson's correlation. When using Spearman's rank this value drops to only 30.5%, leaving over two thirds of type pairings insignificantly correlated. Significant Pearson's correlations are then found between more archetype A-B pairings than B-B pairings, although this is removed by ranking classifications.

### C. Model Relationships

Further examples of archetype B underperforming can be seen in Table V, in which the results of unpaired t-Tests between the correlations of each model type with types from other models attributed to the *same* and *different* archetypes are displayed. Results therefore describe the probability of each type being equally related to all archetypes, in opposition to the assertions of the Unified Model. The Bonferroni correction was applied to the significance threshold of $\alpha = 0.01$, accounting for the 120 tests performed per classification type.

Results are therefore considered significant in Table V at $\alpha = 8.33 \times 10^{-5}$.

TABLE V
UNPAIRED T-TEST RESULTS COMPARING EACH TYPE WITH TYPES FROM THE SAME AND DIFFERENT ARCHETYPES

| **Base Classification** | | | | |
|---|---|---|---|---|
| Model | A | B | C | D |
| Keirsey | 0.027 | 0.011 | 0.001 | **0.000*** |
| Bartle | 0.009 | 0.015 | **0.000*** | **0.000*** |
| Caillois | 0.027 | 0.299 | 0.000 | 0.497 |
| Lazzaro | 0.529 | 0.430 | 0.009 | 0.001 |
| GNS+ | 0.348 | 0.065 | 0.071 | 0.551 |
| MDA+ | 0.001 | 0.070 | 0.160 | 0.012 |
| Handy | 0.002 | 0.354 | 0.517 | 0.002 |
| Gallup | 0.588 | 0.378 | **0.000*** | **0.000*** |
| Covey | 0.002 | 0.802 | 0.051 | 0.069 |
| Holland Themes | 0.202 | 0.726 | 0.004 | **0.000*** |
| Motivations | 0.001 | 0.265 | **0.000*** | 0.010 |
| Problem Solving | 0.005 | 0.113 | 0.000 | 0.003 |
| Goal | 0.006 | 0.087 | 0.001 | 0.000 |
| Personification | 0.949 | 0.011 | 0.092 | 0.023 |
| Class | 0.040 | 0.194 | 0.000 | 0.399 |
| **Ranked Classification** | | | | |
| Model | A | B | C | D |
| Keirsey | 0.039 | 0.001 | 0.006 | **0.000*** |
| Bartle | 0.003 | 0.001 | **0.000*** | **0.000*** |
| Caillois | 0.004 | 0.025 | 0.001 | 0.378 |
| Lazzaro | 0.831 | 0.013 | 0.002 | 0.005 |
| GNS+ | 0.018 | **0.000*** | 0.035 | 0.116 |
| MDA+ | **0.000*** | 0.010 | 0.004 | **0.000*** |
| Handy | **0.000*** | 0.003 | 0.009 | **0.000*** |
| Gallup | 0.157 | 0.001 | **0.000*** | **0.000*** |
| Covey | **0.000*** | 0.001 | 0.002 | 0.002 |
| Holland Themes | 0.105 | 0.042 | **0.000*** | **0.000*** |
| Motivations | 0.000 | 0.121 | **0.000*** | 0.000 |
| Problem Solving | 0.002 | 0.008 | **0.000*** | 0.005 |
| Goal | **0.000*** | 0.000 | **0.000*** | **0.000*** |
| Personification | 0.085 | 0.425 | 0.162 | 0.025 |
| Class | 0.035 | 0.195 | 0.000 | 0.232 |

* **Significant** at $\alpha = 8.33 \times 10^{-5}$
Corrected for 120 tests using the Bonferroni correction

*1) Model-to-Archetype Relationships:* Few models perform well when comparing their relationships with types in the same and different archetypes. Using Pearson's correlation, only four models contain significant results when applying Bonferroni's correction, with only Bartle and Gallup containing multiple significant results. Spearman's rank correlation does however lead to a greater number of significant results per model on average, with a rise from four to eleven models containing at least one type showing a significant difference in its relation to the archetypes. The most successful model under this metric is Stewart's Goal descriptors [4], with significant results for archetype A, C and D. The success of Stewart's Goal descriptors is not unexpected, with each acting as a summary of the overall goal, wants and needs of the types in each model. This success may therefore describe an underlying relationship between each type and its archetype, even if the types themselves share a less distinct relationship. In contrast, using the base significance threshold of $\alpha = 0.01$ provides a clear perspective of Stewart's Personification descriptors [4] as the least successful model, with no significant results obtained using either correlation method. Whilst the summary of participant's overall goals are captured successfully by Stewart's Goal descriptors [4], the abstraction away from this (describing archetypes through the personification of Heart, Head, Hands and Spirit) is seemingly too subjective an approach to aptly map responses between models.

*2) Archetype-to-Model Relationships:* Considering the archetypes themselves, Table V displays the relative success of archetypes C and D, in relating to types that are significantly more correlated with each other than with types from the other archetypes. The difference between same-archetype and different-archetype correlation coefficients is significant in a number of archetype C and D types, although this is shown to be decreased when comparing Pearson's correlation coefficients. Using the base significance threshold of $\alpha = 0.01$ then highlights the further poor performance of archetype B, with no significant results using Pearson's correlation, in comparison to the relative success of archetypes A, C and D, which gain a majority of significant results.

*3) Limitations:* When drawing conclusions from Table V, the overall poor results described by Table IV should not be forgotten. Table V demonstrates how responses differed between types from the same and different archetypes for each model, but does not describe direct links between types from the same archetypes. A type could, for example, be completely uncorrelated with other types from the same archetype, but be negatively correlated with types from differing archetypes. This would still represent a small success of the Unified Model, although the usability of the Model in such cases is greatly diminished. Table V provides further detail, describing models with few significant differences in how each of their types correlates with types from the same and different archetypes. Most notably, these are the Personification and Character Class descriptors, with no significant results using either correlation method. Whilst Caillois' and Lazzaro's models also provide no significant results, the continued failure of the Personification and Character Class descriptors upon removal of the Bonferroni correction implies greater inability to divide players amongst the archetypes.

*4) Model-to-Model Relationships:* Failures of the Personification and Character Class descriptors are explained further when calculating the count of positive significantly correlated types for each pair of models. When using Pearson's correlation, many models show multiple links to other models. Despite this, the only model pairings to feature significant correlations between all four pairs of types are:

1) Gallup and Handy
2) Gallup and Covey
3) Motivations and Covey

This is more than likely presented by the related nature of the Gallup, Handy and Covey models, with Stewart seemingly basing the Motivation descriptors heavily on these. Types from the Gallup model specifically provide 30 significant positive Pearson's correlation coefficients, out of a possible 52, a weak majority, despite this being the most successful model.

Whilst Spearman's rank correlations describe fewer significant links between models, the Personality and Character Class descriptors provide noticeably fewer, providing little evidence for being related to the other models at all. Both do show relation to the Motivations descriptor also put forward by Stewart [4], but fail to show significant correlation with the models to which it was previously associated.

The models featuring the two highest counts of significant Spearman's rank correlations with types from other models are the Overall Goal descriptor, and the MDA+ framework, with 15 and 13 respectively. The Overall Goal descriptor scoring well provides some evidence that the links made between models may exist even if the models themselves are not directly attributable to each other, as suggested by Table V. The successes of the MDA+ framework may also be explained via the same means. With the original framework describing game design as a series of abstract layers, and therefore not describing player behaviour at all. The MDA+ framework posited by Stewart [4] is therefore far from the intentions of the original literature, instead forming an amalgamation of other models and theory put forward by Stewart in an attempt to include it. Due to this, the MDA+ framework may act similarly to the Overall Goal descriptor, in that it may describe the overall themes of the Unified Model better than any of its other constituent models could.

*5) A Summary of Model Relationships:* In answer to RQ3, the two models with the highest average Pearson's correlation between their respective types are the Overall Goal descriptors and the Gallup model, with all four pairings significantly correlated, and an average correlation of 0.426. Using Spearman's rank correlation, this pairing is beaten only by Overall Goal and MDA+, with three significant pairings and an average correlation of 0.323. Explanation of the potential overlap in descriptors and extended models has already been discussed, but the pairing of Overall Goal and Gallup's model is unexplained by this. Instead, responses to each correlate despite their change in perspective, providing some evidence for the underlying link between types described by the Unified Model.

Considering only existing models commonly related to games, the most highly correlated models using Pearson's correlation are GNS+ [4, 21] and Caillois' types of play [14], with an average coefficient of 0.401. The most highly correlated game-related models using Spearman's rank correlation are Lazzaro's 4 Keys 2 Fun [15] and Bartle's taxonomy [3], with an average coefficient of 0.255.

Findings of this study are summarised in Figures 3 and 4, in which the correlation between types in each model and the related types from each other model are summarised. The median correlation from all tested pairings for each model is positive, and describe a general tendency towards positive, but weak, overall fit with the Unified Model. Ranking classifications has been shown to decrease the overlap between types unrelated by the Model, but doing so also reduces the overall strength of correlations between models. The Personification and Character Class descriptors perform especially poorly in this respect, and share very few significant correlations with any other models when ranked.

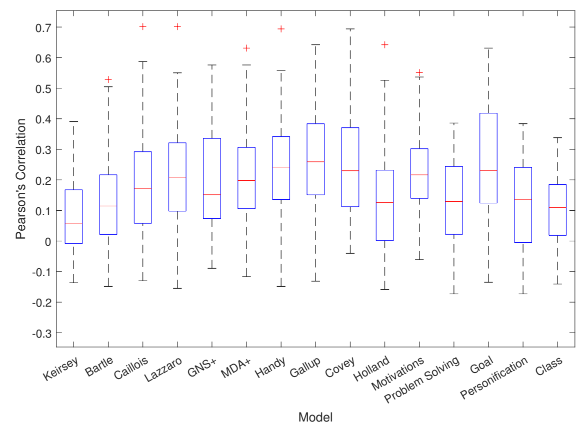


Fig. 3. Pearson's correlations between types in each model with types attributed to the same archetypes from each other model.

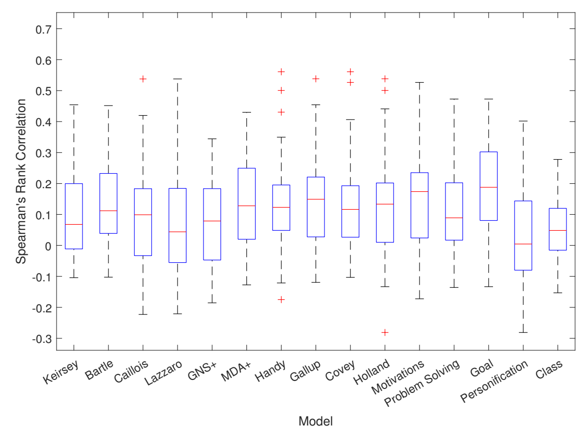


Fig. 4. Spearman's rank correlations between types in each model with types attributed to the same archetypes from each other model.

## VI. Conclusion

This study has provided evidence both for and against various aspects of Stewart's Unified Model of personality and play styles [4]. In response to RQ1, this study provides evidence that there is a link between many of the personality and play style models combined by the Model, with model types often correlating significantly more with types within their archetype than with those from dissimilar archetypes. The Model is however not directly usable to make accurate translations from one model to the next, whether due to poor representation, or the changes in perspective posed by each model's original literature. Types generally correlate significantly, but very weakly. The models sharing the strongest correlation are described in response to RQ3, with successes from Stewart's Overall Goal descriptors [4], Gallup's personality model [27] and the adapted MDA framework [4, 22] seen across multiple methods. The encompassing nature of Stewart's Overall Goal descriptors aligns with the nature of Gallup's model, whilst the MDA+ model correlates despite its misuse in the Unified Model. Despite the overall weakness of correlations observed, the strongest correlations are shown to all align with the assertions of the Unified Model, with responses to RQ2 describing the most strongly correlated individual types. Under multiple methods, these are shown to be Lazzaro's Easy Fun [15] and Caillois' Mimesis [14], and Covey [28] and Handy's [17] Power types. Both pairs describe successes of the Unified Model, with each pairing relating types from the same archetypes.

Whilst multiple significant results have been described, the overall inability of the Model to be used as a predictor across models is clear. Previous work making use of the Unified Model as justification of methodology should therefore be considered with the findings of this study in-mind, with the relatedness of models largely lying in theory alone. Future work considering the use of player type models should not rely on the Unified Model to provide methodological backing.

## References


[1] G. Tondello, K. Arrambide, G. Ribeiro, A. Cen, and L. Nacke, "“i don't fit into a single type": A trait model and scale of game playing preferences," 09 2019.

[2] L. E. Nacke, C. Bateman, and R. L. Mandryk, "Brainhex: A neurobiological gamer typology survey," *Entertainment Computing*, vol. 5, no. 1, pp. 55–62, 2014.

[3] R. Bartle, "Hearts, clubs, diamonds, spades: Players who suit muds," *Journal of MUD research*, vol. 1, no. 1, p. 19, 1996.

[4] B. Stewart, "Gamasutra - personality and play styles: A unified model," 2011, accessed Feb. 11, 2024. [Online]. Available: www.gamedeveloper.com/design/personality-and-play-styles-a-unified-model

[5] L. F. Bicalho, A. Baffa, and B. Feijó, "A game analytics model to identify player profiles in singleplayer games," in *2019 18th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*. Rio de Janeiro, Brazil: Institute of Electrical and Electronics Engineers, 2019, pp. 11–20.

[6] B. Cowley and D. Charles, "Behavlets: a method for practical player modelling using psychology-based player traits and domain specific features," *User Modeling and User-Adapted Interaction*, vol. 26, pp. 257–306, 2016.

[7] T. van Dam and S. Bakkes, "The ace2 model: Refining bartle's player taxonomy for creation play," 2019.

[8] L. De Simone, D. Gadia, D. Maggiorini, and L. A. Ripamonti, "Design of a recommender system for video games based on in-game player profiling and activities," in *Proceedings of the 14th Biannual Conference of the Italian SIGCHI Chapter*, ser. CHItaly '21. New York, NY, USA: Association for Computing Machinery, 2021.

[9] J. Pirker, C. Guetl, and J. Löffler, "Ptd: Player type design to foster engaging and playful learning experiences," in *Teaching and Learning in a Digital World*. Springer International Publishing, 01 2018, pp. 487–498.

[10] A. Al-Taei, "Automated classification of game players among the participant profiles in massive open online courses," Master's thesis, Çankaya University, Ankara, Turkey, 2015.

[11] C. Cruz, M. D. Hanus, and J. Fox, "The need to achieve: Players' perceptions and uses of extrinsic meta-game reward systems for video game consoles," *Computers in Human Behavior*, vol. 71, pp. 516–524, 6 2017.

[12] J. Hamari and J. Tuunanen, "Player types: A meta-synthesis," *Transactions of the Digital Games Research Association*, vol. 1, pp. 29–53, 3 2014.

[13] N. Carneiro, D. Miranda, G. Pereira, G. Mendonça, and T. Darin, "A systematic mapping on player's profiles: Motivations, behavior, and personality characteristics," *Journal on Interactive Systems*, vol. 13, pp. 257–273, 10 2022.

[14] R. Caillois, *Man, Play, and Games. Translated From the French by Meyer Barash*. Champaign, Illinois, USA: University of Illinois press, 1961.

[15] N. Lazzaro, "Why we play games: Four keys to more emotion without story," 2004.

[16] D. Keirsey and M. Bates, "Please understand me: An essay on temperament styles. del mar," 1978.

[17] C. B. Handy, *Gods of management: The changing work of organizations*. Oxford University Press, USA, 1995.

[18] I. B. Myers, *The Myers-Briggs Type Indicator: Manual (1962)*. Consulting Psychologists Press, 1962.

[19] D. J. Pittenger, "The utility of the myers-briggs type indicator," *Review of educational research*, vol. 63, pp. 467–488, 1993.

[20] C. Bateman and R. Boon, *21st Century Game Design (Game Development Series)*. Rockland, MA, USA: Charles River Media, Inc., 2005.

[21] R. Edwards, "Narrativism: Story now," 2003.

[22] R. Hunicke, M. Leblanc, and R. Zubek, "Mda: A formal approach to game design and game research," *AAAI Workshop - Technical Report*, vol. 1, 4 2004.

[23] R. A. Bartle, *Designing virtual worlds*. New Riders, 2004.

[24] G. Tondello, D. Valtchanov, A. Reetz, R. Wehbe, R. Orji, and L. Nacke, "Towards a trait model of video game preferences," *International Journal of Human-Computer Interaction*, pp. 1–17, 04 2018.

[25] N. Yee, N. Ducheneaut, and L. Nelson, "Online gaming motivations scale: development and validation," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 2803–2806.

[26] B. Stewart, "Styles of play - the full chart," 2005, accessed Feb. 11, 2024. [Online]. Available: flatfingers-theory.blogspot.com/2005/01/styles-of-play-full-chart.html

[27] G. Inc. (2023) Gallup personality test. Gallup. Accessed Feb. 11, 2024. [Online]. Available: www.gallup.com/cliftonstrengths

[28] S. R. Covey, "The 7 habits of highly successful people," *New York: Fireside*, vol. 1, pp. 183–190, 1989.

[29] J. L. Holland, *Making vocational choices: A theory of vocational personalities and work environments*. Psychological Assessment Resources, 1997.

[30] R. R. McCrae and P. T. Costa, "Reinterpreting the myers-briggs type indicator from the perspective of the five-factor model of personality," *Journal of personality*, vol. 57, pp. 17–40, 1989.

[31] N. McMahon, P. Wyeth, and D. Johnson, "Personality and player types in fallout new vegas," in *Proceedings of the 4th International Conference on Fun and Games*. ACM, 9 2012, pp. 113–116.

[32] L. J. Francis, C. L. Craig, and M. Robbins, "The relationship between the keirsey temperament sorter and the short-form revised eysenck personality questionnaire," *Journal of Individual Differences*, vol. 29, no. 2, pp. 116–120, 2008.

[33] S. B. Eysenck, H. J. Eysenck, and P. Barrett, "A revised version of the psychoticism scale," *Personality and individual differences*, vol. 6, no. 1, pp. 21–29, 1985.

[34] N. Hughes and P. Cairns, "Player trait questionnaires: An (in)validation study," 06 2020.

[35] N. A. Limited, "16personalities," 2011, accessed Feb. 11, 2024. [Online]. Available: www.16personalities.com

[36] K. Makwana and G. Dave, "Confirmatory factor analysis of neris type explorer® scale – a tool for personality assessment," *International Journal of Management*, vol. 11, pp. 257–265, 9 2020.

[37] E. Andreasen and B. Downey, "The bartle test of gamer psychology," 2001, accessed Feb. 11, 2024. [Online]. Available: www.andreasen.org/bartle

[38] T. Gummer, J. Roßmann, and H. Silber, "Using instructed response items as attention checks in web surveys: Properties and implementation," *Sociological Methods & Research*, vol. 50, no. 1, pp. 238–264, 2021.