Towards a Player Age Model

Shoshannah Tekofsky  
Tilburg University  
PO Box 90153, 5000 LE  
Tilburg, The Netherlands

Pieter Spronck  
Tilburg University  
PO Box 90153, 5000 LE  
Tilburg, The Netherlands

Martijn Goudbeek  
Tilburg University  
PO Box 90153, 5000 LE  
Tilburg, The Netherlands

Jan Broersen  
Utrecht University  
PO Box 80089, 3508 TB  
Utrecht, The Netherlands

Abstract

This paper proposes a Player Age (PA) model with the potential to be generalized to many different games. The model offers insight into the relationship between age and play style. Game developers can use the PA model to gain a better understanding of their target audience, and to optimize adaptive game features (i.e., AI, targeted marketing). In order to become generically applicable, the PA model is based on the literature on life-span developments in physiology and psychology. The PA model states that player age is a linear function of four factors: Speed of Play (+), Performance (-), Preference (+/-), and Time Played (+/-). The model is validated on a data set from Battlefield 3 (FPS). It explains 33.7% of the variance in age (range: 12-65 years) with a standard error of 6.743. To determine the generic quality of the PA model, future work will validate it on games of other genres.

In past decades video games were an activity set aside almost exclusively for the youngest generation. Yet over the years, many of the people that grew up playing games, continued playing in their adult years. The result is that the gamer demographic has strongly diversified in terms of age (?). The Entertainment Software Association reports that the average age of the American gamer is 30, and that 68% of game players are 18 years or older.¹

The age diversification of the gamer demographic is relevant to game development. To reach the largest possible audience, a game must appeal to widely differing age groups. However, age influences both how we can play and how we want to play. The changes are in part due to a number of physiological and psychological developments related to aging. Three major examples of such developments are a decline in cognitive performance, a shift to a more conscientious personality, and a decrease in achievement-based gaming motivations (See Theory Section).

We propose a model of the relationship between age and play style: the Player Age (PA) model. The purpose of the model is two-fold. First, the model can assist game developers in reaching their target audience by offering an increased understanding of the relationship between age and play style. By adapting the design of a game towards the play style of a certain age group, the game may become more appealing to the target audience. Secondly, the model can be used to dynamically adapt the game to the age of the player. Such adaptations can be used for targeted marketing or innovative game mechanics.

The PA model is inspired by our earlier work on the relationship between age and play style (?). We found that around half the variance in age can be explained by play style in Battlefield 3. The current paper takes our research on age a step further. Here we suggest a model of age (PA model) based on the literature from psychology and cognitive science. Our aim is for the model to become applicable to all games (generic). In the current paper, the model is validated on a large sample of Battlefield 3 players. We consider it future work to determine the generic quality of the model by validating the model on representative games from different genres.

Player Age Model

The PA model has two prominent characteristics. First, the model is a linear equation. Previous research among World of Warcraft and Battlefield 3 players has shown that a linear model performs well as an approximation of the relationship between age and play style (?). Secondly, the factors of the PA model are theory-based instead of computationally derived. The model is intended to become generically applicable to video games of all genres. Computational techniques cannot distinguish generic play style features across different video games. Therefore, potential generic play style features are distinguished using the literature on the physiological and psychological effects of aging.

Theory

The PA model explains the relationship between the age A in years of a given player i and his play style with the equation:

\[ A_i = S_i + P_{ei} + P_{ri} + T_i \]

Play style is described in terms of four main factors: Speed of Play (S), Performance (Pe), Preference (Pr), and Time Played (T). The factors are based on the physiological and psychological effects of aging. The effects are at least in part caused by life-span developments in cognitive performance, motivation, and personality (?). Each of the factors consists of one or more subfactors depending on the...
The four factors are defined and derived as follows.

**Speed of Play** (*S*) The frequency of game actions which fulfill two criteria: 1) the game actions contribute to the player’s progress toward the goal of the game; and 2) the game actions are (nearly) always available to the player. Example: ‘Actions per Minute’ in RTS games.

Speed of Play declines with age for two reasons. First, age negatively correlates to cognitive performance on attentional tasks (?). Attention mediates cognitive speed which in turn mediates Speed of Play. Secondly, age is positively correlated to Conscientiousness (?; ?). Conscientiousness is a personality dimension that is characterized by careful, dutiful, organized, and meticulous behavior. An increase in Conscientiousness would therefore lead to a decrease in Speed of Play (?).

The two criteria for game actions that contribute to Speed of Play serve to (1) exclude behavioral ticks or erratic behavior, and (2) ensure that Speed of Play is measured continuously over the entire play time.

**Performance** (*P*) The scores and/or ranking attained on game performance measures. Example: ‘Experience per Minute’ in RPGs.

Performance declines with age for two reasons. First, age is negatively correlated with performance on various components of spatial tasks (?), such as spatial pattern completion (?), and spatial memory (?). Spatial skills are relevant for efficient navigation of a game world. Secondly, age is negatively correlated with learning and memory in general (?). Both learning and memory are crucial in mastering game mechanics and completing tasks in video games.

**Preference** (*P*) Proportional time spent on a single game play choice among a range of equivalent game play choices. Example: ‘Time in Car X per Total Play Time’ in Racing games.

Preference may increase or decrease with age depending on the game play choice in question. The increases and decreases are due to two reasons. First, gaming motivation shifts with age. ? (?; ?) found that gaming motivations cluster into three categories: Achievement, Social, and Immersion. All three motivations decrease significantly with age. Achievement motivation decreases the most with an effect size around -0.3. Personality shifts with age, while correlating significantly with play style. Research (?; ?) has shown that Extraversion and Openness decrease with age, while Agreeableness and Conscientiousness increase. It has been shown that such differences in personality correlate with differences in play style for *NeverWinter Nights* and *Battlefield 3* players (?; ?; ?).

**Time Played** (*T*) Time spent playing the game. Example: ‘Time Logged In per Month’ in MMORPGs.

Time Played may increase or decrease with age depending on the game in question. Ideally, the increase or decrease is in line with the target audience: if the game is targeted at younger players, Time Played decreases with age; if the game is targeted at older players, Time Played increases with age. Naturally, age also increases while one is playing a game. However, we consider that increase negligible due to scale difference: Age is measured in years, while Time Played falls in the range of dozens to hundreds of hours.

Time Played has a unique role in the PA model as it does not describe how someone plays, but how much someone plays. The factor is included in the model for two reasons. First, Time Played improves Speed of Play and Performance by improving cognitive performance. ? (2) demonstrated multiple cognitive performance improvements due to video game training, such as improvements in spatial cognition and attention. ? (2) specifically explored the cognitive effect of video game training on older adults. They found that improvements in game performance were accompanied by improvements in various cognitive functions, including memory. ? (?) found that Time Played correlates positively or negatively to age depending on the game in question.

**Application**

During game development, the PA model can be applied to estimate the age of the player. When applying the model, the four factors serve to inform the selection of play style variables to be tracked for the estimation of age. The four factors themselves are not calculated at any point. Calculating the exact scores for the factors is an unnecessary interim step when estimating player age for a particular game. In contrast, the interim step is crucial in validation (which we will do for *Battlefield 3* below). The application of the model progresses in three steps.

First, domain knowledge of the game is used to select or construct play style variables that track the four factors of the PA model. If it is unclear if a variable belongs to a certain factor, it will be included. If the variable does not contribute to the model, it will be removed in the third step. All play style variables for the PA model are ratios of two or more absolute values. A play style variable is only informative if it is made relative to some criteria. For instance, the absolute number of kills is not informative. However, the number of kills per *death* describes survivability, while the number of kills per *time unit* describes kill efficiency.

Secondly, the age data and selected play style variables for a representative player sample are tracked. Beta tests or online games offer the most expedient means to gain a representative player sample that includes age data.

Thirdly, multiple linear regression (MLR) is performed with age as the outcome variable, and the play style variables as the predictor variables. The MLR yields the coefficients of the PA model, as well as the variance explained ($R^2$) by the model. Variance is defined as the average error between the mean and the collected data (see (?) for more statistical background.) The MLR will remove variables that do not contribute significantly to the model. The age of players outside the sample can now be estimated by a summation of the selected play style variables multiplied by the coefficients calculated by the MLR.

According to the PA model, the three steps outlined above lead to an estimation of age. However, the coefficients of the play style variables in the model cannot be easily interpreted. If an interpretation of the coefficients is desirable, then the play style variables need to be filtered on collinearity before...
being entered in the MLR. When constructing the PA model, there is a trade-off between the variance explained by the model and a meaningful interpretation of the model coefficients. If variance explained is a priority, then all available play style variables are included in the model. If a meaningful interpretation of the model coefficients is a priority, then play style variables with high collinearity are excluded from the model.

Validation: Battlefield 3
To determine the strength of the PA model as a model of the relationship between age and play style, it must be validated on a number of data sets that represent a varying range of video game genres. In this paper we present the first validation of the PA model on a popular FPS: Battlefield 3.

Data Collection
The Battlefield 3 data was collected as part of an investigation into the link between play style and personality (?). Participants were recruited from Battlefield 3 community platforms through a promotional campaign dubbed ‘PsyOps’. All data was automatically collected and stored via a dedicated website. Data collection took place over a period of six weeks. During this time, participants could visit the website to submit their data. The data form contained six fields: age, player name, gaming platform, 100-item IPIP questionnaire, country of residence, and credits. Participants were asked to give permission for anonymous use of their game statistics, which were then automatically retrieved from a public database. Player name was used as the key for game statistics retrieval. It is a unique identifier of a player account in Battlefield 3. Therefore, it was used to ensure that all participants were unique individuals. The credits field was a tick box where participants indicated if they wished to have their player name listed on the credits page of the final research report. After submitting all their data, participants were forwarded to a page with their results on the personality test and an overview of what the personality dimensions entail.

To determine the participant’s play style, 826 game statistics were gathered. Domain knowledge was employed to combine and process the game statistics to reflect play style more accurately. The result was that 60 play style variables were defined. While these variables only reflect behaviors that every player can show at any time in the game, it does not follow that every behavior a player can exhibit is exhibited by each player. If a player never engages in a certain behavior, then he will show a missing value for the relevant play style variable. We chose to enter zero for missing values on play style variables. The rationale for this choice is that lower values on play style variables generally indicate less skill with the relevant behavior. We consider it plausible that (barring a few exceptions) a player who never exhibits a certain behavior, has little to no skill with that behavior. Therefore, it follows that zero is a representative value for play style variables with missing values.

Data Description
The final data set contained data from 13,376 participants. During the data collection phase, the game statistics database was restructured to accommodate an upcoming expansion of the game. The restructuring process shifted the format of the collected data so only the first 9,367 submissions were usable. To ensure data integrity, 31 participants who indicated an age below 12 or above 65 were excluded from the sample. An additional 6 participants were removed from the remaining sample because they had a total play time of 0 seconds. The resulting effective sample size was 9,330.

The sample was highly heterogeneous in terms of gaming platform, personality scores, country of residence, and willingness to be mentioned in the credits of the research. For the sake of brevity, only age and play style data will be reviewed in detail in this paper.

Age. The distribution of age in the sample can be seen in Figure ???. The distribution is a skewed normal distribution with an average of 25, and spread over all mature ages. From the age of 17 to 18 there is a noticeable dip followed by a spike. Battlefield 3 is a game rated 18+ in most countries. It is likely that some participants that were 17 years old reported their age as 18 due to the age threshold for the game.

Play Style. Across the 60 play style variables, 1,263 participants exhibited one or more missing values. Missing values were substituted with zero as outlined above. The play style variables show two key characteristics of the sample. First, the sample is biased toward more experienced and skilled players, with performance variables showing means above those of the Battlefield 3 populace. The bias is likely caused by the method of data collection. Participants were sought out on Battlefield 3 community platforms. Community platforms are more popular with expert players. Therefore, the Battlefield 3 data set is likely to contain more expert players than the overall Battlefield 3 populace. Secondly, the distributions of the play style variables are likely to be equal to those in the Battlefield 3 populace. Most variables are

http://bf3stats.com/

http://ipip.org/
Creating the PA Model for Battlefield 3

To validate the PA model on the Battlefield 3 data set, the factors of the model must be defined explicitly. We constructed the model in five steps:

1. Four sets of play style variables were selected to represent each of the four factors.
2. Principal Component Analysis (PCA) with varimax rotation was used on each individual set of variables (provided it consisted of more than one variable).
3. The components returned by the PCA were evaluated for inclusion in the PA model as a subfactor. A component was included when it met two criteria: (1) it had an Eigen value greater than 1.0, and (2) it formed a semantically coherent component consisting of two or more component variables with an absolute component score > 0.5.
4. PCA was repeated on each of the sets of component variables that met the criteria for inclusion as a subfactor.
5. Subfactor scores were calculated for each participant in the sample by applying the component scores of the subfactors from step 4.

The PCA used in steps 2 and 4 returns a number of components that maximally explains the variance in a set of variables. The varimax rotation ensures that each component variable loads maximally onto one component. The components the PCA returns are the subfactors of a factor that stand candidate for inclusion in the PA model (step 3). A component is considered semantically coherent when its component variables reflect a logically coherent element of play style for the game in question.

The PA model for Battlefield 3 was created using the five steps outlined above. The following four sections describe the results of steps 1 through 3 for each factor. Almost all component variables discussed are ratios where the value without brackets is divided by the value within brackets. The combined result of steps 4 and 5 for each factor is shown in Table 1. It reflects the complete PA model for Battlefield 3.

Speed of Play Table 1 shows the selected variables for Speed of Play (step 1 & 2). Component 1 is the only component that meets both criteria for inclusion as a subfactor (step 3). Kill Assists (Time), Kills (Time), and Suppression Assists (Time) form a complete partition of all offensive actions a player performs per time unit. Components 2 and 3 are not semantically coherent.

Performance Table 2 shows the selected variables for Performance (step 1 & 2). Component 1 is the only component that meets both criteria for inclusion as a subfactor (step 3). Grenade Hits (Grenade Shot) is excluded from the subfactor resulting from component 1 due to the large gap in component scores between this variable and the remaining component variables of component 1. Component 2 consists of only one component variable. On closer examination, Deaths (Time) does not meet the criteria for the Performance factor. In Battlefield 3 it is possible to never die by remaining in certain safe zones. Remaining in safe zones is not a measure of Performance. Therefore, component 2 is not included in the PA model.

Preference Table 3 shows the selected variables for Preference (step 1 & 2). Both components 1 and 2 meet the criteria for inclusion as subfactors (step 3). Additional subfactors were created for each of the four classes in the game: Assault, Support, Engineer, and Recon. In Battlefield 3 every player must select exactly one class to play with at all times. The values of the four classes add up to exactly 1.0 for every player per definition of the subfactors. Therefore, it is redundant to include all four classes in the model. Of the four classes, the Assault class has the highest correlations with the subfactors of Speed of Play, Performance, and Time Played. It follows that the Preference subfactor Assault was excluded from the model.

Time Played The Battlefield 3 data set contained only one variable measuring Time Played: Total Play Time. There was no data on relative time spent playing the game. More-
over, the data collection took place over a period of six weeks. Therefore, the Time Played factor is imprecisely instantiated in the current model.

The Model Table 2 displays the PA model for Battlefield 3. The generic PA model is a linear equation \( A_i = S_i + Pe_i + Pr_i + T_i \). Therefore, the linear equation for the PA model of Battlefield 3 is of the form: \( A_i = (a \times \text{Attack Speed}_i) + (b \times \text{Kills & Score}_i) + (c_1 \times \text{Aircraft}_i + c_2 \times \text{Armor}_i + c_3 \times \text{Support}_i + c_4 \times \text{Engineer}_i + c_5 \times \text{Recon}_i) + (d \times \text{Play Time}_i) \).

Testing the PA Model

MLR was performed to determine the coefficients \((a, b, c_1, c_2, c_3, c_4, c_5, d)\) of the PA model of Battlefield 3. Age was entered as the outcome variable, and the eight subfactors were entered as the predictor variables.

Table 2 displays the beta, \( t \), and \( p \) values returned by the MLR. The beta values are the standardized coefficients of the (sub) factors in the linear equation of the PA model. The model explains 33.7\% of the variance in age (Adjusted \( R^2 \)) with \( F[9330] = 594.294, p < 0.001 \), and a standard error of 6.743. The direction of the beta coefficients is in line with the theory of the PA model.

The performance of the PA model can be put in perspective by comparing it to the results of MLR on the full set of 60 play style variables (brute force, see Table 3.) The PA model combines high variance explained (\( R^2 \)) with a generic quality that gives it the potential to be applied to a wide range of games. The brute force model yields a higher \( R^2 \) than the PA model, because it makes use of all available data. However, the PA model offers an understanding of the underlying factors mediating the relationship between age and play style. This understanding can be used to better appeal to target audiences, as well as to anticipate in advance what data needs to be collected to estimate age.

The accuracy of the age estimate of the PA model can be put in perspective by comparing it to the accuracy of an age estimate obtained by an optimal sampling of the age distribution (base line). Figure 2 shows the results of the comparison. The accuracy of the brute force model indicates the upper bound of the accuracy gain achievable by a model based on play style. Age estimations are evaluated per interval size. The base line percentage of correct age estimations is set to the zero point. The PA model and brute force model are plotted to show relative difference in age estimation accuracy compared to the base line. Accuracy of the age estimate is expressed in terms of percentage of the sample which is given an accurate age estimation. An accurate estimation of age is an estimation that falls within an age range centered on the actual age of the participant. The PA model improves on the base line from an interval size of 10 onward. The magnitude of the improvement is small.

Discussion

This section will discuss two major considerations relevant to the PA model. We elaborate on how the model can be validated further, and discuss potential candidate factors.

Firstly, to determine if the PA model is indeed a generic model, it must be tested on a wide range of games. In this paper we have validated the model on a major First Person Shooter title. Future validations will gain most from apply-
<table>
<thead>
<tr>
<th>Factor</th>
<th>Subfactor</th>
<th>A. C.*</th>
<th>Component Variable**</th>
<th>C. S.***</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>Attack Speed</td>
<td>-0.315</td>
<td>Kill Assist (Time)</td>
<td>0.933</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Kills (Time)</td>
<td>0.881</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Suppression Assists (Time)</td>
<td>0.831</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Score (Time)</td>
<td>0.899</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Deaths (Kill)</td>
<td>-0.840</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MVP123 Ribbons (Round)</td>
<td>0.836</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ELO Rating</td>
<td>0.829</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Hits (Shot)</td>
<td>0.801</td>
</tr>
<tr>
<td>Pe</td>
<td>Kills &amp; Score</td>
<td>-0.291</td>
<td>Scout Helicopter Time (Time)</td>
<td>0.867</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Jet Time (Time)</td>
<td>0.846</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Assault Helicopter Time (Time)</td>
<td>0.795</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Main Battle Tank Time (Time)</td>
<td>0.934</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Infantry Fighting Vehicle Time (Time)</td>
<td>0.934</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Support Time (Time)</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Engineer Time (Time)</td>
<td>1.000</td>
</tr>
<tr>
<td>Pr</td>
<td>Armor</td>
<td>0.178</td>
<td>Recon Time (Time)</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Support</td>
<td>0.156</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Engineer</td>
<td>0.182</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Recon</td>
<td>-0.196</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>Play Time</td>
<td>0.204</td>
<td>Total Play Time</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Table 4: Final Factor Definitions of the PA Model (BF3)

* Effect size of correlation of subfactor with age. Significant at $\alpha < 0.01$.
** Play style variables are made relative to the variable indicating in brackets.
*** Component Score of the play style variable.

Secondly, the PA model currently consists of four factors. Some might argue for the inclusion of additional candidate factors. We conjecture that at least one candidate factor exists: Secondary Objectives (SO). The SO factor measures the progress players achieve on secondary objectives proportional to their progress on the primary objective. The factor was not included in the current model because the authors know of no physiological or psychological effects of aging that could serve as a basis for the SO factor. However, direct research of gaming behavior has shown that children are significantly less interested in secondary objectives than adults (7). By extension, we would have liked to test the hypothesis that younger players are less interested in secondary objectives than older players in the Battlefield 3 data set. However, the hypothesis could not be tested as the data set lacks variables tracking secondary objectives. Therefore, the SO factor remains not validated, but holds promise as a candidate factor for the PA model.

Currently we are not considering other candidate factors apart from the SO factor. The existing four factors encompass nearly all relevant subfactors such as ‘patience’ (Speed of Play/Time Played), ‘strategic insight’ (Performance), and ‘Competitiveness’ (Preference). The possible depth of subfactors was not fully illustrated with the Battlefield 3 data set as the game is rather straightforward. Speed of Play and Performance could not be measured in more than one way each. Many other games will offer multiple subfactors for Speed of Play and Performance. Additionally, we do not exclude the possibility of a nonlinear relationship between age and play style, but decided to focus on a linear model.

**Conclusion**

We have proposed a Player Age (PA) model with the potential to be generalized to many different games. The model is based on the physiological and psychological developments of aging. It states that life-span developments in cognitive performance, motivation, and personality have a linear relationship with play style in terms of Speed of Play, Performance, Preference, and Time Played. Validating the PA model on a large sample of Battlefield 3 data, it explained 33.7% of the variance in age over an age range of 12-65 (years). The model benefits game developers by offering a deeper understanding of the relationship between age and play style. This understanding can be used to better reach target audiences as well as design game mechanics that play into the estimated age of the player.

**Acknowledgments**

The authors would like to thank Gwen Ferdinandus, Armon Toubman, Jay Abdeljalil, Finn Görges, and Andreas Lundgren for their technical and artistic contributions to PsyOps.

**References**


van Lankveld, G.; Spronck, P.; van der Herik, J.; and Arntz, A. Games as personality profiling tools. In 2011 IEEE Conference on Computational Intelligence and Games (CIG’11) (ed. Mike Preuss).

