

# **Fusing Quantitative and Qualitative Methods in Virtual Worlds Behavioral Research**

**Carl Symborski, Gary M. Jackson, Meg Barton,  
Geoffrey Cranmer, Byron Raines, Mary Magee Quinn**

SAIC

4001 N Fairfax Dr.

Arlington, VA 22203

+1 703-558-2958

[carl.w.symborski@saic.com](mailto:carl.w.symborski@saic.com), [marguerite.r.barton@saic.com](mailto:marguerite.r.barton@saic.com),

[geoffrey.cranmer@saic.com](mailto:geoffrey.cranmer@saic.com), [byron.a.raines@saic.com](mailto:byron.a.raines@saic.com)

**Celia Pearce**

Georgia Institute of Technology

85 Fifth Street NW

Atlanta, GA 30308

+1 310-866-8014

[celia.pearce@lmc.gatech.edu](mailto:celia.pearce@lmc.gatech.edu)

## **ABSTRACT**

In this study, Science Applications International Corporation (SAIC) and Georgia Institute of Technology (GT) developed a quantitative-qualitative mixed methods research technique to investigate the extent to which real world characteristics of Massively Multiplayer Online Role-Playing Game (MMORPG) players can be predicted based on the characteristics and behavior of their avatars. SAIC used three primary assessment instruments to quantitatively rate videos of participant gameplay sessions, while GT produced detailed qualitative descriptions of avatar activities and behavior. Automated textual analysis was then used to identify conceptual themes across all of the descriptions produced by the qualitative team. Using the themes generated by the automated textual analysis in combination with the quantitative variables, we were able to demonstrate the efficacy of the hybrid method for the prediction of real world characteristics from avatar characteristics and behavior.

## **Keywords**

Mixed methods, behavioral studies, online games, MMORPGs, virtual worlds

## **INTRODUCTION**

Quantitative and qualitative behavioral and social science methods and analyses are often conducted separately. Each method yields its own unique results and serves different purposes. Mixed methods behavioral studies of virtual worlds offer the possibility of a richer understanding of the context of the data collected, and have the potential to strengthen findings through a combined perspective (Johnson and Onwuegbuzie 2004); furthermore, they have been used successfully in the past on virtual worlds research

**Proceedings of DiGRA 2013: DeFragging Game Studies.**

© 2013 Authors & Digital Games Research Association DiGRA. Personal and educational classroom use of this paper is allowed, commercial use requires specific permission from the author.

(Feldon and Kafai 2008). However, combining quantitative and qualitative research methods can also pose unique challenges for researchers. To explore this interdisciplinary approach, Science Applications International Corporation (SAIC) and Georgia Institute of Technology (GT) collaborated to develop a mixed methods virtual worlds research approach, involving the use of text analysis tools to extract quantitative data from qualitative thick description text. The purpose of this research study was to determine whether or not it is possible to predict a person's real world characteristics from the characteristics and behaviors of his or her avatar in an MMORPG, using variables generated by the quantitative-qualitative approach.

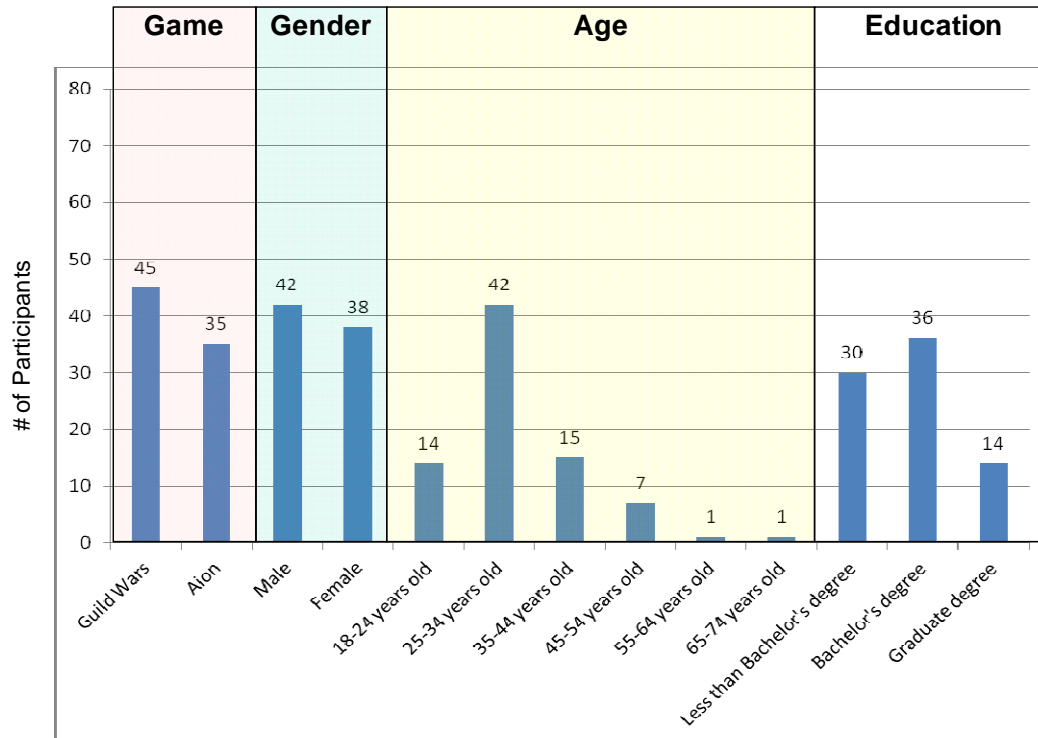
To investigate the prediction of real world characteristics using avatar characteristics and behavior, an innovative protocol was followed using videos recorded by participants during their regular play activities: (1) SAIC conducted quantitative rating of the video to determine avatar characteristics, behavior, and personality using standardized assessment forms; (2) GT conducted qualitative observations of the video generating thick description notes typically used in ethnographic participant observation (Geertz 1973); and, (3) automated text analysis was conducted to extract themes from the qualitative thick description notes, which could be used to supplement the quantitative data. The statistical technique of discriminant analysis (DA) was then conducted to determine whether or not it was possible to produce predictive models using the collected data.

While automated text analysis is common in virtual worlds research, to our knowledge this is the first time that it has been applied to qualitative thick description notes to convert such notes to quantitative and predictive data. In this study, the interdisciplinary team used the quantitative and qualitative variables gathered to generate statistical models for the prediction of real world gender, age, extraversion level, submissive ideology, and aggressive ideology.

## **METHOD**

### **Participants**

Participants were recruited via Facebook/Google advertisements, flyers, and online game forums. All participants were required to be at least 18 years of age and to have a minimum of 50 hours of experience with either *Guild Wars*<sup>®</sup> (ArenaNet 2005) or *Aion*<sup>®</sup> (NCSOFT 2009), two popular MMORPGs. A total of 80 participants completed the study. Recruitment occurred primarily within the Washington DC region, USA.

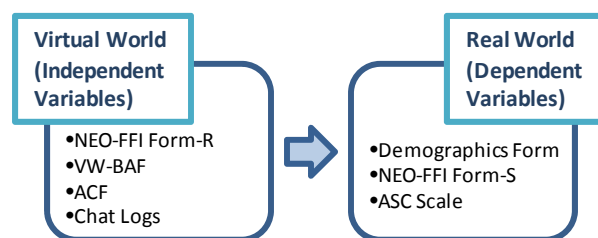


**Figure 1:** Participant demographics.

More *Guild Wars*<sup>®</sup> (45, or 56%) than *Aion*<sup>®</sup> players (35, or 44%) and slightly more males (42, or 52%) than females (38, or 48%) participated in the study. As might be expected, the majority of the participants fell into the young adult age ranges, with a dwindling number of participants divided across the middle- to older-adult ranges. A slight majority of participants held a Bachelor's degree (36, or 45%), while 30 participants (37%) had less than a Bachelor's degree and 14 participants (18%) held a graduate degree.

## Instruments

An important facet of the research design was the distinction between the dependent variables (DVs) to be predicted, or the selected real world (*RW*) *characteristics* – gender, age, extraversion, submissive ideology, and aggressive ideology – and the independent variables (IVs) to be used as predictors for the selected DVs, or the *characteristics and behavior of the players' avatars*. Several instruments were used to capture these data (see Figure 2).



**Figure 2:** Sources of real world DVs and virtual world IVs.

### *Data Collection Instruments: Real World Characteristics*

The DVs to be predicted, or the RW characteristics of the participants, were measured using a basic demographics form, the NEO Five-Factor Inventory (NEO-FFI) personality assessment form, and the Aggression-Submission-Conventionalism Scale (ASC scale).

The **Demographics Form** was used to collect detailed demographic information on the participants, including gender, age, and highest educational level achieved.

The **NEO Five-Factor Inventory (NEO-FFI)** is a standardized and validated personality assessment instrument constructed from the five-factor model of personality. The instrument has been validated to provide an accurate assessment of the so-called “Big Five” traits of neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness. The NEO-FFI has two forms: the NEO-FFI Form-S (NEO-S), a self-report form, and the NEO-FFI Form-R (NEO-R), an observer rating form; the two forms are identical, except that the NEO-S is written from a first person perspective and the NEO-R is written from a third person perspective (Costa and McCrae 1992). To establish ground truth on the participants’ personalities, each individual completed the NEO-S.

Each participant’s ideology was measured using the **Aggression-Submission-Conventionalism Scale (ASC scale)**. The ASC scale includes three subscale domains, as the name implies - submissive, aggressive, and conventional - and has demonstrated high predictive validity (Dunwoody and Funke n.d.).

### *Data Processing Instruments: Avatar Characteristics and Behavior*

Multiple mechanisms for collecting IV data on the characteristics and behaviors of the participants’ avatars were employed. The NEO-R was used to gather information on avatar personality, the Virtual Worlds Behavior Analysis Form (VW-BAF) and Avatar Characteristics Form (ACF) were developed to collect data on avatar behaviors and characteristics, chat logs were analyzed for linguistic data, and qualitative data were recorded. All forms were designed to be generalizable across all MMORPGs.

While the NEO-S was completed by the participants, the NEO-R (observer rating form) was used to assess avatar personality. The NEO-R was completed by trained assessors observing the participant’s avatar, facilitated by a “bridge” developed to resolve inconsistencies or conflicts of any NEO-R item that caused rating difficulty within the virtual world (VW), considering that the original form had been developed to rate humans in the RW.

The **Virtual Worlds Behavior Analysis Form (VW-BAF)** was created to measure avatar activities and behavior in MMORPGs. The form had two pages. The first allowed a rater to tally occurrences of behaviors on a minute-by-minute basis (e.g., approaches toward other avatars, verbal initiations, instances of verbal humor, battles completed). The second page allowed a rater to record presence or absence of major activity categories, such as “quest activity” or “social activity,” during the interval. The second page of the form also measured how much time an avatar spent in a group versus solo, and whether the mode was Player vs. Player or Player vs. Environment.

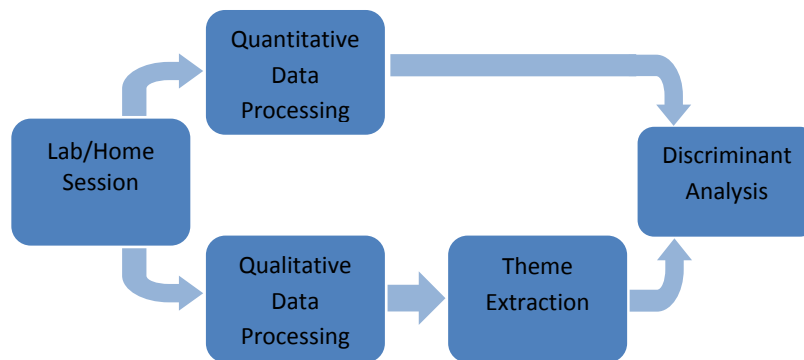
The **Avatar Characteristics Form (ACF)** was developed to capture characteristics of avatars in the VW, such as gender, appearance, character class, and combat role. It also captured information about play style; for example, whether an avatar belonged to a guild, was a social player, or was a strategic player.

The research team also transcribed participant **chat logs**. Care was taken only to record chat by the target participant/avatar, not initiations or responses by others. These chat logs were run through a parser, which extracted variables such as number of words and average word length.

The remaining IVs were generated through **qualitative observation** and converted into a form suitable for statistical analysis and prediction using automated textual analysis. This process will be described below in the sub-section titled “Extraction of Quantitative Variables from Qualitative Data.”

## Design and Procedure

The basic research design included five primary phases which will be described here: the initial laboratory session with the participant to establish ground truth, the processing of quantitative and qualitative data, the conversion of the qualitative data into quantitative variables that could be combined with the original quantitative data, and the development of predictive statistical models using discriminant analysis. Figure 3 depicts the research design, from the time a recruited participant arrived at the laboratory through analysis of the data.



**Figure 3:** Research design.

### *Laboratory/Home Session*

Upon arriving at the laboratory, each participant viewed a standard introductory video describing the study and completed the informed consent process, if he/she desired to participate. The participant then completed the demographics form, the NEO-S, and the ASC scale (described above). These collected data would form the basis for ground truth on the participant’s RW characteristics.

At the participant’s leisure after returning home, the participant recorded a one-hour gameplay session of the selected game using a screen capture program. Participants were requested to choose one avatar and to use only that avatar for the entire one-hour home session and to play the game just as they normally would.

Once the home session was completed, the participant saved the recorded session on a USB flash drive and mailed it back to the laboratory. Participants were asked to complete the home session within two weeks of the laboratory session. Once the flash drive was received, participants were compensated for their contribution to the study.

### *Quantitative Data Processing*

Once a participant's USB flash drive was received at the laboratory, trained assessors rated that session using the NEO-R, the VW-BAF, and the ACF. Separately, chat was transcribed. A minimum of 80% inter-rater agreement was obtained for all rating forms.

For completion of the NEO-R, raters observed the entire hour of the recorded session and then provided avatar ratings using a developed "bridge" to help answer specific items (mentioned above).

The VW-BAF rating occurred across three 10-minute sessions, which were standardized in terms of start times across the full one-hour session to ensure a representative sample of avatar behavior. Observations were recorded once a minute. As described previously, the VW-BAF had two pages: the first was a series of counts for how many times particular behaviors occurred each minute, and the second recorded whether or not certain behaviors occurred within each minute. Both kinds of variable were summed over each 10-minute sample, and then averaged across the three 10-minute samples for each participant before being used for analysis. Hence, each VW-BAF variable either corresponds to how many times a behavior occurred in an average minute, or to how many minutes out of 10 included that behavior on average.

At the same time that the VW-BAF ratings were recorded, the ACF was completed. The ACF items consisted of presence or absence variables; for example, presence or absence of a costume or of a ranged damage-per-second (DPS) combat role.

Once all of the quantitative data had been gathered, it was integrated into a data array in preparation for analysis.

### *Qualitative Data Processing*

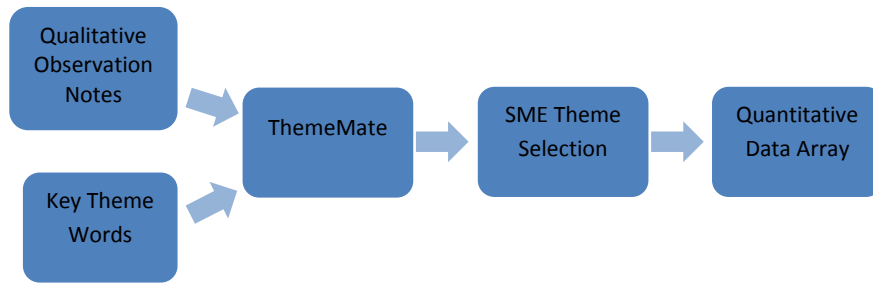
The qualitative team protocol consisted of observing the recorded session and creating a detailed textual description of participant behavior. Researchers conducted and recorded observations using a note-taking technique known as "thick description" (Geertz 1973) to write descriptive text of the participant gameplay sessions. The thick description note-taking technique is significantly different from traditional coding in that it provides a detailed *interpretive, contextualized* narrative account of events, relying on specified domain knowledge by the authors. Details of avatar appearance, including the use of costumes and whether armor items were set to visible or invisible, were noted. Player activities were noted in detail, including specific behaviors pertaining to modes of navigation, weapons, skills and skill combinations used, as well as combat style and interaction style with teammates.

### *Theme Extraction of Quantitative Variables from Qualitative Data*

The extraction of quantitative variables from the qualitative thick description notes is at the crux of the mixed methods approach taken in this study. By processing thick description notes to generate quantitative variables, the data generated by both qualitative and quantitative methods could be combined to produce statistically predictive models of players' RW characteristics.

In order to accomplish this, the text from the observational notes was processed with ThemeMate, an Automated Behavior Analysis application (Jackson 2012). ThemeMate is a statistically-based application used to extract "themes" from bulk text. The ThemeMate application produces a data array of all extracted themes. Figure 4 depicts the

methodology used to convert qualitative thick description notes into quantitative variables, useful as predictors in predictive statistical analysis.



**Figure 4:** Methodology used to convert qualitative notes into quantitative variables.

A list of key theme words was chosen based on subject matter expertise of *Guild Wars*<sup>®</sup> and *Aion*<sup>®</sup> and familiarity with the thick description notes. The ThemeMate software used these key words to perform a directed theme search of all the thick description notes and generated 253 conceptual theme constructs ranked by importance. From these 253 themes, 20 were picked, with consultation from subject matter experts (SMEs), as potential predictors for subsequent statistical modeling. A theme was eliminated from the larger set if it a) was already represented in the existing quantitative data, b) had the same meaning or indicated the same result as another theme as assessed by SMEs, c) was spurious as a result of standard note formatting conventions (e.g., “the,” “he,” etc.), or d) was representative of standard/common avatar behavior (e.g., wearing armor, carrying a weapon) and therefore unlikely to distinguish players from one another. From the remaining list, the 20 most important themes by ThemeMate ranking were selected for use. The data array for the 20 selected themes was added to the quantitative data array and made available for statistical analysis.

The importance of subject matter expertise in vetting these themes cannot be overemphasized. Without some domain expertise in the genre and specific games being studied, the automated analysis generated a good deal of “noise” that could have easily been misinterpreted.

#### ***Statistical Technique: Discriminant Analysis***

Discriminant analysis (DA) was used to generate predictive models for participants’ RW characteristics based on their avatars’ characteristics and behavior. The purpose of DA is to predict group membership, based on RW DVs such as gender or extraversion, from a linear combination of VW IVs, such as avatar gender or type of armor worn. DA begins with a data set containing many cases (participants), where both the values of the IVs and the group membership (DV) are known. The end result is an equation or set of equations that predict group membership for new cases where only the values of the IVs are known (Stockburger).

Specifically, DA using *backward* stepwise reduction was conducted. This form of stepwise reduction begins with a given set of variables and reduces the set by eliminating the variables that are associated with the DV to a lesser degree than remaining variables (Brace et al. 2009). In this way, only the variables that are most predictive of the DV

remain as part of the predictive model when the analysis is complete. As an additional quality control, leaving-one-out cross-validation was used for all stepwise reduction DA runs. This form of validation accuracy is the process by which a model is trained on all cases but one and tested on the one case that was withheld. The process repeats until all cases have been withheld and tested blindly, eliminating any chance of predicting a case based on extracted knowledge of that case.

## RESULTS AND DISCUSSION

### The Discriminant Function

As described above, DA generates a predictive model in the form of a linear combination of independent, or predictor, variables. There is also a constant term for each equation, which is used as the linear offset in the discriminant functions. The general form of the discriminant function is as follows:

$$DV \text{ Group } 0 = (b_1 \times x_1) + (b_2 \times x_2) + \cdots + (b_n \times x_n) + c$$

$$DV \text{ Group } 1 = (b_1 \times x_1) + (b_2 \times x_2) + \cdots + (b_n \times x_n) + c$$

Where:  $b_n$  = the Fisher Coefficient (or weight) for that variable,  
 $x_n$  = the value of the independent variable, and  
 $c$  = the value of the constant.

Once the values of the variables are substituted in the above equations for the names of the variables, whichever of the two equations evaluates to the greater number will be the prediction for that case, or participant. In other words, whichever of the two Fisher's discriminant functions produces a higher value "wins," and the participant will be predicted as a member of the corresponding category.

### Accuracy Metrics

Several accuracy metrics are reported for each of the predictive models. Overall accuracy is the number of cases correctly classified divided by the total number of cases. Precision and recall are also presented to provide information about the accuracy of the models.

### Overall Results

Table 1 presents the accuracy results of models combining quantitative and qualitative variables predicting RW characteristics from VW observations. Individual results sections that follow present each model including a description of how the DV was defined, an overview of the accuracy of the model, and a brief discussion of the IVs relevant to the prediction of the target DV.



	<b>Overall Accuracy</b>	<b>Precision</b>	<b>Recall</b>
Gender	84%	85%	83%
Approximate Age	66%	71%	63%
Extraversion Level	73%	68%	71%
Submissive Ideology	71%	71%	79%
Aggressive Ideology	73%	53%	75%

**Table 1:** Overall accuracy for predicting RW characteristics from VW observations.

## Gender

### *Definition of the DV*

In the case of gender, defining the DV was straightforward: males were assigned as one group, and females were assigned as the other group. Forty-two (42) participants were male and 38 participants were female.

### *Accuracy of Gender Model*

The gender model achieved 84% overall accuracy (67 participants correctly classified; see Table 2). Precision and recall were nearly balanced at 85% and 83%, respectively.

<b>Overall Accuracy</b>	<b>Precision</b>	<b>Recall</b>
84% (67)	85%	83%

**Table 2:** Accuracy of gender model.

### *Discriminant Function for Gender*

As described above, in determining which group a new case should be classified into, the values for each of the IVs would be plugged into each of the following two equations. Whichever equation yields the largest value represents the group that the new case would be classified into. The following are the equations for gender:

*Male*

$$= (4.319 \times MaleAV) + (3.643 \times T58Heals) + (3.061 \times HairAccNA) - 2.789$$

*Female*

$$= (0.645 \times MaleAV) + (0.699 \times T58Heals) + (0.160 \times HairAccNA) - 0.746$$

Table 3 presents descriptions of the predictor variables in the above equations.

Variable	Description of Variable
MaleAV	Avatar is male
T58Heals	Avatar heals other avatars
HairAccNA	Hair Accessories cannot be observed, likely because head is covered (e.g., by a helmet or costume)

**Table 3:** Description of IVs relevant to prediction of gender.

Through interpretation of the discriminant functions above, the models can be described loosely in simple English as follows:

If the avatar is male, heals others, and/or has covered hair, then it is likely that the player's RW gender is male.

Otherwise, it is likely that the RW gender is female.

**NOTE:** These English sentences are not to be substituted for the above equations in making predictions; they are only intended to aid the reader in understanding the statistical models.

### *Discussion of IVs Relevant to the Prediction of Gender*

#### *Avatar Gender (ACF)*

It is well-known that avatar gender is closely related to RW gender (Yee n.d.); therefore, it is not surprising that avatar gender surfaced as a predictor variable in this analysis. In our study worlds, the appearance of a male avatar strongly predicted a RW gender of male. Female avatars required the support of additional IVs to distinguish between female avatars operated by RW females and female avatars operated by RW males. It should also be noted that while “gender-bending” is common among male players in MMORPGs (Yee 2003; MacCallum-Stewart 2008), our findings replicate other research findings that it is rare among women (Yee 2003): only three participants in our pool were females playing male avatars.

#### *Avatar Heals Others (Theme)*

This theme, extracted from the qualitative notes, was identified as present if the notes indicated that an avatar was healing other avatars during the observation period. It is commonly assumed that females gravitate toward healing roles in MMORPGs, since women are considered to be more nurturing and supportive by nature (Bergstrom et al. 2012). Our research suggests the opposite; the presence of this theme was a predictor of RW male gender. Because male avatar gender is such a strong predictor of RW male gender, the remaining variables in the discriminant function above primarily indicate the remaining RW males who were “gender-bending.” Yee et al. conducted a study that offers an explanation as to why the theme of healing others in combination with the presence of a female avatar might predict RW males: they discovered that “[male] players

enact this stereotype [of women as healers] when gender-bending” (2011, 776). Hence, a male player creating an avatar for a healing role might tend to choose a female.

#### *Covered Hair (ACF)*

The variable “Hair Accessories N/A” indicated if an avatar’s hair accessories could not be observed because the avatar’s hair was covered, due to a helmet or costume that occluded the hair (e.g., a hood). The model indicated that RW male players playing female characters in the sample were less likely to expose their hair and hair accessories than RW female players were. While there does not appear to be a ready explanation for this phenomenon in the literature, the research team postulated that RW male players operating female avatars might be less concerned with displaying their hair than RW female players, instead being more focused on gameplay activities, while RW female players may have more interest in ensuring that their avatars have a feminine appearance, complete with displayed flowing locks.

## **Age**

#### *Definition of the DV*

The DV of age was divided into two groups: under the age of 30 (younger), and 30 and over (older). This is consistent with demographic research (Yee 2006) and qualitative observations indicating that MMORPG players tend to be younger than social game players. Furthermore, the age group of 18-29 is one that is frequently cited in research studies, particularly in voting-related and medical contexts, and thus seemed a logical breakdown to use. Using this division, 43 participants were age 30 or over and 37 participants were under age 30.

#### *Accuracy of Age Model*

With 66% overall accuracy, the age model correctly categorized 53 participants (see Table 4). Precision and recall were 71% and 63%, respectively.

Overall Accuracy	Precision	Recall
66% (53)	71%	63%

**Table 4:** Accuracy for predicting age.

#### *Discriminant Function for Age*

The following are the two equations for age:

*30 or Over*

$$= (0.999 \times BAF35) + (0.984 \times Mage) + (2.029 \times T44exploring) - 1.658$$

*Under 30*

$$= (0.426 \times BAF35) + (2.100 \times Mage) + (2.768 \times T44exploring) - 2.076$$

Table 5 presents descriptions of the variables used.

Variable	Description of Variable
BAF35	Avatar does not move for the full 60 seconds
Mage	Avatar class is Mage
T44exploring	Avatar actively traverses the environment

**Table 5:** Description of IVs relevant to prediction of age.

In simple English, this model becomes:

If the avatar spends more time stationary, is not a Mage, and/or does not actively traverse the environment, then it is likely that the player's age is 30 or over.  
Otherwise, it is likely that the player's age is under 30.

### *Discussion of IVs Relevant to Age*

#### *Avatar does not move for the full 60 seconds (VW-BAF)*

This item measured how many full minutes an avatar spent stationary over the observation period. In the study sample, an avatar that spent more time stationary was more likely to be older (age 30 or over). It is common knowledge that younger people tend to be more active than older people, and perhaps more prone to bouts of fidgeting and nervous activity. In the virtual world, this may translate into moving frequently, walking in circles, or jumping repeatedly, and an avatar that is frequently in motion is unlikely to be stationary for a full minute at a time. Conversely, an avatar operated by an older individual is more likely to spend time comfortably still.

#### *Mage (ACF)*

The discriminant function indicated that those who played a Mage class avatar were *less* likely to be age 30 or over. Mages are generally ranged damage dealers with powerful attacks, capable of quickly bringing down opponents. In the opinion of the research team, ranged damage-per-second (DPS) tends to be one of the least complex and most self-sufficient of the major combat roles. Mages may be attractive to younger players, who are more interested in fire-power and in quickly becoming proficient in the game, whereas older players may be more interested in more strategic, cooperative roles such as healing or tanking.

#### *Avatar actively traverses the environment (Theme)*

This theme was identified as present in the qualitative notes if the avatar was actively moving through the environment throughout the observation period. In the study sample, an avatar that did *not* actively traverse the environment was more likely to be age 30 or older. This indicator is synergistic with the previous movement-related variable from the VW-BAF. The model indicated that an avatar who spends a great deal of time moving through the game environment is likely to be a younger player, actively fighting her/his way across the landscape. An older player may be less driven, but have a more balanced play experience doing inventory management or other activities along with battling.

## **Extraversion Level**

### *Definition of the DV*

The NEO scoring system groups individuals into very high, high, average, low, and very low categories of extraversion (Costa and McCrae 1992). To split the sample into groups, those with high or very high extraversion were placed in the high extraversion group, and

those with very low, low, or average extraversion were placed in the low extraversion group. For this model, 35 participants were considered to have high extraversion and 45 participants were considered to have low extraversion.

#### *Accuracy of Extraversion Level Model*

For the extraversion level model, 73% overall accuracy was obtained, with 58 participants correctly classified (see Table 6). Precision and recall were fairly comparable, at 68% and 71%, respectively.

Overall Accuracy	Precision	Recall
73% (58)	68%	71%

**Table 6:** Accuracy for predicting extraversion level.

#### *Discriminant Function for Extraversion Level*

The following are the two equations for extraversion level:

##### *High extraversion*

$$= (0.967 \times T76backtrack) + (0.641 \times T78teamheal) + (0.191 \times BAF18) - 0.973$$

##### *Low extraversion*

$$= (2.701 \times T76backtrack) + (1.859 \times T78teamheal) + (0.615 \times BAF18) - 1.770$$

Table 7 presents descriptions of the variables used.

Variable	Description of Variable
T76backtrack	Avatar turns around or goes back
T78teamheal	Avatar heals party members
BAF18	# of occasions avatar follows a command issued by another group member

**Table 7:** Description of IVs relevant to prediction of extraversion level.

Translated into simple English, this model becomes:

If the avatar does not backtrack, does not heal party members, and/or does not follow commands issued by others, then it is likely that the individual has high extraversion.

Otherwise, it is likely that the individual has low extraversion.

#### *Discussion of IVs Relevant to Extraversion Level*

##### *Avatar turns around or goes back (Theme)*

This theme was identified as present if the qualitative notes indicated that an avatar turned around and backtracked or returned to a previously visited person or place during the observation period. In the study sample, an avatar that did *not* turn around or go back was more likely to be operated by a highly extraverted individual. It may be that extraverted players exhibit confidence in the navigation decisions they make during game play by backtracking less, pressing forward more.

#### *Avatar heals party members (Theme)*

This second theme was identified as present if the qualitative notes indicated that an avatar was involved with healing other members of his/her group during the observation period. In the study sample, *not* being involved with healing other party members helped predict high extraversion in the RW. In MMORPGs, the group healer is generally a background figure, healing party members from a distance while preserving his/her own health for the sake of the group's success, whereas tanking and melee DPS players tend to "lead the charge" into battle and otherwise direct the group on attack strategy (Bergstrom et al. 2012). Therefore, it might make sense that a highly extraverted, leader-like individual would be drawn to tanking/melee DPS roles rather than healing roles.

#### *Number of occasions avatar follows a command issued by another group member (VW-BAF)*

This item represented the extent to which an avatar was observed following others' commands during the session. In the study sample, an avatar who did *not* spend time following others' commands was more likely to be highly extraverted in the RW. In keeping with the theme above, it makes sense that the extraverted, who frequently emerge as leaders in groups (Judge et al. 2002), may spend more time issuing commands than following commands issued by others.

## **Submissive Ideology**

### *Definition of the DV*

An item from the ASC scale was used to measure a submissive ideology:

ASC 2. Our leaders know what is best for us (Dunwoody and Funke n.d.).

This item was selected on the basis that it was the most representative item from the ASC authoritarian submissive subscale. The idea that DV groups could be formed using a specific item from an assessment scale was considered to be particularly interesting; by achieving accuracy in the prediction of submissive ideology, it is possible to demonstrate the prediction of an individual's response to a specific statement. Those participants who responded neutral, agree or strongly agree to the item were considered submissive; those who responded disagree or strongly disagree were considered not submissive. Forty-three (43) participants had a submissive ideology and 37 participants did not have a submissive ideology.

### *Accuracy of Submissive Ideology Model*

For this model, 57 participants were correctly predicted, for 71% overall accuracy (see Table 8). Precision and recall were 71% and 79%, respectively.

Overall Accuracy	Precision	Recall
71% (57)	71%	79%

**Table 8:** Accuracy for predicting submissive ideology.

### *Discriminant Function for Submissive Ideology Model*

The following are the two equations for submissive ideology:

#### *Submissive ideology*

$$= (1.935 \times NEOR11) + (2.837 \times NEOR24) + (2.219 \times T44exploring) - 5.697$$

#### *No submissive ideology*

$$= (2.705 \times NEOR11) + (2.291 \times NEOR24) + (2.968 \times T44exploring) - 6.212$$

Table 9 presents descriptions of the variables used.

Variable	Description of Variable
NEOR11	When he's under a great deal of stress, sometimes he feels like he's going to pieces.*
NEOR24	He tends to be cynical and skeptical of others' intentions.* [reverse scored]
T44exploring	Avatar actively traverses the environment

**Table 9:** Description of IVs relevant to prediction of submissive ideology.

Stated in simple English, this model becomes:

If the avatar does not go to pieces under stress, is not cynical and skeptical of others' intentions, and/or does not actively traverse the environment, then it is likely that the individual has a submissive ideology.

Otherwise, it is likely that the player does not have a submissive ideology.

### *Discussion of IVs Relevant to Submissive Ideology*

*“When he's under a great deal of stress, sometimes he feels like he's going to pieces.”\* (NEO-R)*

This personality assessment item is the rater's evaluation of whether or not the avatar “does not go to pieces under stress.” As specified on the developed NEO Bridge, a rater agreed with this item if the avatar, in the presence of a stressor (e.g., a battle, an argument), requested attention or assistance by demanding heals or resurrects or by rapidly issuing repeated commands, and/or responded emotionally to the situation. A rater disagreed with this item if the avatar, in the presence of a stressor, successfully retreated from, won, or returned to the battle, and/or responded calmly to or ignored aggressive social situations (e.g., being harassed or insulted by other players). In the study sample, *not* going to pieces under stress helped predict a RW submissive ideology. The research team postulated that a submissive person might not be particularly unsettled when in a stressful situation, relying on confidence that the designers of the game or leaders in group play “know what is best for [them]” (see “Definition of the DV” section above).

*“He tends to be cynical and skeptical of others' intentions.”\* (NEO-R)*

This NEO-R item was the rater's evaluation of whether or not the avatar was “cynical and skeptical of others' intentions.” A rater agreed with this item if the avatar voiced doubt; complained about the game's programming or other players; or exerted influence over others' (avatar or NPC) behavior, particularly by planting flags or issuing commands. A rater disagreed with this item if the avatar did not exert influence over others' (avatar or NPC) behavior with commands, and/or did not express complaints or doubts about others. Instead, the avatar might have followed others or complimented other avatars or the game. In the study sample, *not* being cynical and skeptical of others' intentions

helped predict a submissive ideology. It makes sense that a submissive person would not be particularly cynical and skeptical of others' intentions, particularly since the ASC scale item that was used to define this DV -- "Our leaders know what is best for us" (Dunwoody and Funke n.d.) -- is clearly devoid of cynicism or skepticism.

#### *Active player traversing the environment (Theme)*

This theme also appeared in the age model, and is defined in the section "Discussion of IVs Relevant to Age." In the study sample, players who did *not* actively traverse the environment were more often those with a submissive ideology. This suggests that players with a submissive ideology may seek a more balanced play experience, perhaps conforming to game design expectations which include activities beyond focused seek-and-destroy missions. In contrast, a player with an aggressive ideology might be more likely to spend time traversing the environment, on a constant mission to fight enemies along the way.

## **Aggressive Ideology**

### *Definition of the DV*

For aggressive ideology, the most representative item from the ASC authoritarian aggressive subscale was selected:

ASC 18. Strong punishments are necessary in order to send a message (Dunwoody and Funke n.d.).

In order to be considered to have an aggressive ideology, the individual had to agree or strongly agree with the statement. Those who responded neutral, disagree, or strongly disagree were considered to be non-aggressive. Twenty-four (24) participants had an aggressive ideology and 56 participants did not have an aggressive ideology.

### *Accuracy of Aggressive Ideology Model*

As shown in Table 10, the overall accuracy for the aggressive ideology model was 73% (58 participants correctly classified). Though precision was low at 53%, recall was high at 75%.

<b>Overall Accuracy</b>	<b>Precision</b>	<b>Recall</b>
73% (58)	53%	75%

**Table 10:** Accuracy for predicting aggressive ideology.

### *Discriminant Function for Aggressive Ideology Model*

The following are the two equations for aggressive ideology:

#### *Aggressive ideology*

$$= (1.158 \times Strategizes) + (2.599 \times T87melee) + (3.026 \times T18quest) - 2.178$$



*No aggressive ideology*

$$= (2.607 \times \text{Strategizes}) + (1.133 \times \text{T87melee}) + (2.115 \times \text{T18quest}) - 1.822$$

Table 11 presents descriptions of the variables used.

Variable	Description of Variable
Strategizes	Player has a strategic play style
T87melee	Avatar inflicts melee damage
T18quest	Player checks the quest log

**Table 11:** Description of IVs relevant to prediction of aggressive ideology.

In simple English, this model becomes:

If the avatar does not strategize, inflicts melee damage in combat, and/or checks the quest log, then it is likely that the individual has an aggressive ideology.  
Otherwise, it is likely that the player does not have an aggressive ideology.

### *Discussion of IVs Relevant to Aggressive Ideology*

#### *Strategizes (ACF)*

This item captured the general play style of an avatar over the observation period. An avatar would be coded as one who strategized if the avatar displayed strategic behavior, demonstrating that the player has given some thought to how best to fight enemies in the game. Some examples of strategic behavior include pulling, a behavior in which the avatar lures a small group of enemies into safe territory to fight, giving the player a competitive advantage; and using a gimmick build, which is a character build that exploits game mechanics to render the avatar maximally efficient. In the study sample, *not* strategizing helped predict an aggressive ideology. One might expect an aggressive individual not to be particularly strategic, especially if he/she is constantly rushing into battle without taking time to consider strategy.

#### *Avatar inflicts melee damage (Theme)*

This theme was identified as present if the qualitative notes indicated that an avatar dealt melee damage during battles. In the study sample, inflicting melee damage predicted an aggressive ideology. Avatars involved in melee combat are in close proximity to enemies under attack; as such, this style of “in your face” play is arguably the most physically aggressive in MMORPGs. Therefore, it makes sense that an aggressive individual would gravitate toward melee combat.

#### *Player checks the quest log (Theme)*

This theme was identified as present if the qualitative notes indicated that an avatar accessed the quest log during the observation period. Each avatar has a quest log that tracks active quests, progress made toward completing quests, and next step instructions specific to each quest. In the study sample, checking the quest log helped predict an aggressive ideology. Attention to active quests suggests a goal-orientated play style. This suggests that players with an aggressive ideology might be more focused on advancement, rather than other activities such as socializing.

## **CONCLUSIONS**

The aim of this study was to combine quantitative and qualitative research methods in a mixed methods approach to develop a deeper understanding of the relationship between

virtual world avatar behavior and real world characteristics. To that end, the research team generated statistical models for five RW characteristics from VW observations. The average overall accuracy across all five models – 73% – suggests that it is indeed possible to develop predictive models for RW characteristics from observations of avatar characteristics and behavior. Interestingly, though, all of the models require the input of several variables in order to generate predictions, some of which are more easily explained from a face validity perspective than others. This suggests that the association between an individual's RW and VW characteristics is not intuitively clear cut: a female avatar is not necessarily being operated by a RW female, and an avatar that appears extraverted is not necessarily representing a person who is extraverted in the RW.

An additional finding related to the absence of chat variables from any of the predictive models was noted. The research team, curious as to why none of the chat variables had proved significant predictors of any of the DVs, performed a qualitative investigation of the chat data. Two conclusions were reached: first, in the sample, many participants did not chat at all, and those who did spoke very little, with most conversation content limited to current game activities. Second, qualitative analysis combined with demographic data allowed us to determine that the use of all lowercase letters and absence of punctuation was a universal behavior pattern that transcended real world demographic categories such as age.

It is true that there were several limitations in this study. First, the small sample size ( $N = 80$ ) may be an impediment to the statistical power of the model development. The requirement to bring each participant into the laboratory, while it enhanced certainty in the validity of the ground truth data, limited recruitment – both geographically and in the sense that only MMORPG players willing to come to the laboratory participated. Furthermore, the observation of only one hour of game-play video by the quantitative and qualitative rating teams was less than ideal; however, due to time and staffing constraints associated with the rating process, it would have been very difficult within the scope of this study to process more participant data. Future research should seek to further validate the predictive models developed, with more participants and by rating the participants over multiple hour-long observation sessions.

That being said, we believe that this study presents several major contributions to the gaming-related research base. For one, our approach included unique methods of virtual world observation, including the use of the NEO-R, bridged for use in the VW, to record avatar personality ratings; the development of a VW behavior assessment instrument (VW-BAF) that standardized ratings of avatar behaviors and allowed for reliable behavioral measures across all avatars; and the consistent recording of observable avatar characteristics via the ACF. The developed NEO bridge provides guidance for how to utilize the NEO-R instrument, originally designed for use with people in RW proximal settings, to rate avatars in a VW by direct observation. To our knowledge, the use of the NEO-R instrument and this form of RW-to-VW rating bridge is new in VW assessment.

The development of a new mixed methods approach for studying avatar behavior in VWs is also a valuable contribution to the research community. This technique allowed researchers to extract IVs that held promise as potential predictors with semi-automated assistance, without having to manually analyze an entire corpus of qualitative thick-description notes. The results of the interdisciplinary hybrid study reveal that automated extraction of conceptual themes across thick description notes, guided by subject matter expertise, can be combined with quantitative measures to predict RW characteristics. The

importance of the contribution of the SMEs should not be underemphasized. By fusing quantitative and qualitative approaches under the guidance of SMEs, we gain a better understanding of the relative contribution of each and how they may be used together to predict RW characteristics from VW observations. In the future, the mixed methods approach described in this work may be generalized to research in other contexts.

## ACKNOWLEDGMENTS

We express our thanks to Kathleen Wipf, Jasmine Pettiford, Nic Watson, Hank Whitson, and Patrick Coursey for their work and contributions on this project.

This work was supported by the Air Force Research Laboratory (AFRL). The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of AFRL or the U.S. Government.

\*Reproduced by special permission of the Publisher, Psychological Assessment Resources, Inc., 16204 North Florida Avenue, Lutz, Florida 33549, from the NEO Personality Inventory-Revised by Paul T. Costa Jr., PhD and Robert R. McCrae, PhD, Copyright 1978, 1985, 1989, 1991, 1992 by Psychological Assessment Resources, Inc. (PAR). Further reproduction is prohibited without permission of PAR.

## BIBLIOGRAPHY

- ArenaNet. (2005). *Guild Wars*. [PC Computer, Online Game], NCSOFT, North America: played September, 2011-April, 2012.
- K. Bergstrom, J. Jennifer, and S. de Castell, "What's 'Choice' Got to Do With It? Avatar Selection Differences between Novice and Expert Players of World of Warcraft and Rift," In *Proceedings of the International Conference on the Foundations of Digital Games*, Raleigh, NC: 2012, ACM Press, pp. 97-104.
- N. Brace, R. Kemp, and R. Snelgar, *SPSS for Psychologists (4th ed.)*, New York: NY: Routledge, 2009.
- P. T. Costa and R. R. McCrae, *NEO PI-R Professional Manual*, Odessa, FL: Psychological Assessment Resources, Inc., 1992.
- P. Dunwoody and F. Funke, "Testing three three-factor authoritarianism scales," *Journal of Social and Political Psychology*, submitted for publication.
- D. Feldon and Y. Kafai, "Mixed Methods for Mixed Reality: Understanding Users' Avatar Activities in Virtual Worlds," *Educational Technology Research and Development*, vol. 56, 2008, pp. 575-593.
- C. Geertz, *The Interpretation of Cultures*, New York, NY: Basic Books, 1973.
- G. M. Jackson, *Predicting Malicious Behavior: Tools and Techniques for Ensuring Global Security*, Hoboken, NJ: John Wiley & Sons, 2012.
- R. B. Johnson and A. J. Onwuegbuzie, "Mixed Methods Research: A Research Paradigm Whose Time Has Come," *Educational Researcher*, vol. 33, no. 7, 2004, pp. 14-26.
- T. A. Judge, J. E. Bono, R. Illies, and M. W. Gerhardt, "Personality and Leadership: A Qualitative and Quantitative Review," *Journal of Applied Psychology*, vol. 87, 2002, pp. 765-780.
- NCSOFT. (2009). *Aion*. [PC Computer, Online Game], NA NCSOFT, North America: played September, 2011-April, 2012.
- E. MacCallum-Stewart, "Real Boys Carry Girly Epics: Normalizing Gender Bending in Online Games," *Eludamos*, vol. 2, 2008, pp. 27-40.

- D. W. Stockburger, "Discriminant Function Analysis," *Multivariate Statistics: Concepts, Models, and Applications*, Accessed October 7, 2012,  
<http://www.psychstat.missouristate.edu/multibook/mlt03.htm>
- N. Yee, "Gender-bending," *The Daedalus Project: Psychology of MMORPGs*. Accessed August 2, 2012, <http://www.nickyee.com/eqt/genderbend.html#5>
- N. Yee, "The Demographics of Gender-bending," *The Daedalus Project: Psychology of MMORPGs*, September 3, 2003, Accessed August 2, 2012,  
<http://www.nickyee.com/daedalus/archives/000551.php?page=1>
- N. Yee, "The Demographics, Motivations and Derived Experiences of Users of Massively Multi-user Online Graphical Environments," *PRESENCE: Teleoperators and Virtual Environments*, vol. 15, 2006, pp. 309-329.
- N. Yee, N. Ducheneaut, M. Yao, and L. Nelson, "Do Men Heal More When in Drag? Conflicting Identity Cues Between User and Avatar," In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (CHI '11)*, New York, NY: 2011, ACM Press, pp. 773-776.