Online Gaming Motivations Scale:
Development and Validation

Nick Yee, Nicolas Ducheneaut, Les Nelson
Palo Alto Research Center
3333 Coyote Hill Road, Palo Alto, CA
[nyee, nicolas, lnelson]@parc.com

ABSTRACT
Understanding gaming motivations is important given the growing trend of incorporating game-based mechanisms in non-gaming applications. In this paper, we describe the development and validation of an online gaming motivations scale based on a 3-factor model. Data from 2,071 US participants and 645 Hong Kong and Taiwan participants is used to provide a cross-cultural validation of the developed scale. Analysis of actual in-game behavioral metrics is also provided to demonstrate predictive validity of the scale.

Author Keywords
Online games; player motivations; taxonomy; scale development; scale validation; cross-cultural.

ACM Classification Keywords
H5.m. [Information interfaces and presentation (e.g., HCI)]: Miscellaneous;

INTRODUCTION
Research in gameplay motivations across multiple fields has shown repeatedly that it is myopic to treat gamers as a monolithic group [1,4,10]: different people play games for very different reasons. Therefore, having a validated motivations taxonomy and a robust measure of those motivations would provide a crucial theoretical and methodological bridge between players and in-game behaviors and outcomes. This taxonomy and measurement tool could, for example, help in examining the links between demographics, motivation, engagement, retention, and learning or behavioral outcome.

Understanding player motivations is an important topic for CHI because gaming-related mechanisms are being implemented in many non-gaming applications, such as: location tracking (e.g., foursquare), news aggregation (e.g., Google News), and exercise monitoring (e.g., fitbit). Others have made the bolder claim that enterprises will have to design gaming mechanisms into everyday business to fully engage the gamer generation [6], In fact, Gartner has predicted that by 2014, more than 70% of Global 2000 organizations will have at least one gamified application [2]. Thus, understanding gamer psychology, and specifically their motivations for playing, is valuable even outside the context of games.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

Many player and motivation taxonomies have been proposed, but most of these were not developed using statistical methods and do not provide a means for quantitative assessment (see for instance [4]). One exception in the domain of online gaming motivations is Yee's motivation taxonomy [10], based on a factor analytic examination and restructuring of Bartle's player taxonomy of MUD players [1]. Yee's research identified 10 motivations that fall into 3 higher-level categories related to: achievement, social, and immersion motivations. Given that current gamification efforts draw inspiration from online games (including both Facebook games and role-playing games like World of Warcraft), an existing taxonomy grounded in online gaming is a good starting point for our current research effort.

While Yee's research provided both a taxonomy and survey instrument for assessment, it suffers from 3 weaknesses:

1. Although Yee's taxonomy identified 3 higher-level motivation factors, his 39-item survey instrument assesses the underlying 10 components without providing a direct means of assessing the 3 high-level factors. Given that these factors parsimoniously capture many different motivations, it is important to construct and validate a shorter survey instrument that can directly assess them.
2. Yee's taxonomy was derived using an English-speaking participant sample, and it is unclear whether the factor analysis results would be consistent in other cultures. A replication in a non-Western culture would provide much-needed evidence of the applicability of the motivation taxonomy to other cultures.
3. And finally, Yee's research does not provide data on predictive validity--how well the self-report survey measures correlate with actual in-game behaviors. After all, just because a survey scale has high internal reliability doesn't mean it actually measures anything meaningful [8]. Data showing meaningful correlations with in-game behavioral metrics would provide evidence for predictive validity.

The current study attempts to resolve these 3 weaknesses by: 1) constructing and validating a survey measure of the 3 high-level motivation factors via factor analysis, 2) replicating the factor structure in a non-Western culture, and 3) examining the predictive validity of the scale in terms of correlations with in-game behavioral metrics.

STUDY ONE - SCALE DEVELOPMENT

We first describe how the revised scale was constructed and the initial validation of the scale using exploratory factor analysis.

Scale Construction

First, strong inventory items in each of the 10 components from Yee's 39-item scale were selected via factor loading. Then, in several intermediate pilot surveys (each with N > 300), we tested variants of these inventory items and iteratively selected items that had high factor loadings. In creating variants, we attempted to create shorter, more direct inventory items. For example, the item "How important is it to you that your character is as optimized as possible for their profession or role?" was revised to "Optimizing your character as much as possible". We also avoided items that were semantically similar. Finally, we streamlined the response options. Thus, while the original inventory items varied in using "how much", "how often", "how interested", and "how important" question stems, the revised items are all framed using "how important are these gameplay elements when you play online games". Thus, the response options are identical for all inventory items--a 5-point Likert scale ranging from 1 (Not Important At All) to 5 (Extremely Important). In the final iteration of this process, we selected 12 items (4 for each factor) as the final inventory set to be tested in this study (see Table 1).

Participants

To collect data for scale validation, we recruited *World of Warcraft* (WoW) players in the US to participate in an online survey version of the scale by posting announcements in high traffic websites and forums dedicated to WoW. Altogether, 2,071 WoW players participated in the survey. Of the participants, 1,358 were male, 709 were female, with an average age of 29.95 (SD = 9.20).

Exploratory Factor Analysis

To examine the underlying factor structure of the revised scale, we conducted an exploratory factor analysis (EFA)--a statistical procedure that examines the covariances among a set of variables to identify latent factors.

Kaiser's Meyer-Olkin measure of sampling adequacy was .78, while Bartlett's Test of Sphericity was significant at $p < .001$. These two measures indicate that the data set was appropriate for factor analysis.

We conducted an EFA in SPSS 18 using principal factors analysis. Given that most psychometric factors are mildly correlated, we used an oblique rotation. Three factors emerged with eigenvalues greater than 1. Examination of the scree plot also suggested a solution with 3 factors. Together, these 3 factors accounted for 59.3% of the overall variance. In Table 1, we report the pattern matrix.

Item	Soc.	Imm.	Ach.
Chatting with other players	**0.73**		
Being part of a guild	**0.70**		
Grouping with other players	**0.67**		
Keeping in touch with your friends	**0.60**	0.11	
Learning about stories and lore of the world		**0.78**	0.12
Feeling immersed in the world		**0.73**	0.11
Exploring the world just for the sake of exploring it		**0.60**	-0.14
Creating a background story and history for your character		**0.54**	-0.13
Becoming powerful			**0.73**
Acquiring rare items		0.13	**0.68**
Optimizing your character as much as possible		-0.12	**0.67**
Competing with other players			**0.55**
% of Variance	23.9%	21.4%	13.9%
Cronbach's α	.77	.75	.74

Table 1. Factor loadings. For ease of reference, we sort the items using their factor loadings with the primary loadings bolded, and exclude any factor loadings less than .10.

The pattern matrix shows that the scale items loaded onto 3 factors that correspond to Immersion, Social, and Achievement motivations respectively. The factor loadings are high; most are over .60. None of the cross-loadings are higher than .2. Finally, the scale items within each factor all have Cronbach α's above .70. Together, these findings suggest that the revised scale has a good factor structure and good internal reliability. In terms of inter-factor correlations, the Social factor was correlated with the Immersion and Achievement factors at .09 and .24 respectively. The Immersion factor was correlated with the Achievement factor at -.21.

Factor	Gender Means (SD)	Gender Differences	Age Corr.
Achievement	Male = .14 (.83) Female = -.27 (.91)	t = 10.08 $p < .001$	r = -.27 $p < .001$
Social	Male = -.05 (.88) Female = .09 (.90)	t = -3.43 $p < .005$	r = -.13 $p < .001$
Immersion	Male = -.07 (.89) Female = .13 (.87)	t = -4.70 $p < .001$	r = -.01 p = .64

Table 2. Gender and age differences.

For completeness, we report the gender and age differences related to the motivation factors in Table 2. We note that these differences are consistent with those previously reported [10].

STUDY TWO - CROSS-CULTURAL VALIDATION

To validate the scale cross-culturally, we recruited participants from Hong Kong (HK) and Taiwan (TW). These two non-Western cultures were selected because they share the same server pools (in Taiwan) and share a written language (traditional Chinese script).

Scale Preparation

To localize the scale, a bilingual translator first translated the scale into Chinese (traditional script). We then piloted the translated instrument with several WoW players in the two locations for idiomatic fluency and made several revisions. Finally, a different translator back-translated the scale into English to ensure the two English scales were comparable.

Participants

To collect data for the study, we recruited WoW players from HK and TW to participate in the online survey version of the scale by posting announcements in high traffic websites and forums dedicated to WoW. Altogether, 645 players participated in the study--406 from TW, 239 from HK. 514 participants were male, 128 were female. The average age was 23.59 (SD = 5.16).

Confirmatory Factor Analysis

To verify that the 3-factor structure replicates in a non-Western culture, we conducted a confirmatory factor analysis (CFA)--a statistical procedure that compares the fit of the data with a factor model specified by the researcher, or in essence, the opposite of the EFA procedure.

We conducted a CFA with Amos 18. In the specified CFA model, the three motivation factors were included as latent variables, each hypothesized to have a direct effect on the 4 corresponding measured scale items. Unique measurement errors were hypothesized to have a direct effect on each measured scale item. Finally, the 3 latent factors were hypothesized to co-vary with each other.

Maximum likelihood estimation was used for the analysis. The chi-square statistic of the model was significant (χ^2 [51] = 140.88, $p < .001$) as is expected from the large sample size. In these scenarios, goodness of fit indices provide a more meaningful assessment. We report here the common set of indices recommended: CFI = .95, SRMR = .04, and RMSEA = .05. All these values fall within the range of current recommendations for good model fit [3]. Thus, data from HK and TW were consistent with the 3-factor structure identified in the US data. More importantly, this implies that the factor structure of motivations we have identified is not simply an artifact of Western culture, and has some level of cross-cultural applicability.

STUDY THREE - PREDICTIVE VALIDITY

To examine predictive validity of the scale, we collected behavioral metrics within WoW and examined the correlations between the self-report and behavioral data.

Behavioral Metrics in WoW

The context of WoW and the availability of in-game metrics via the Armory have been described in the CHI literature [9]. Of particular relevance to our research goal is the existence of hundreds of in-game goals known as *achievements*. For example, there are achievements for exploring every zone within a continent, and achievements for killing a dungeon boss within a short amount of time. More importantly, these achievements are grouped into 6 categories and an aggregate achievement score is provided for each. These behavioral categories are automatically calculated and compiled by the game and require no additional coding on our part. These achievement categories map on to very different aspects of gameplay within WoW:

- Quests: Individual, easy, goal-based missions.
- Exploration: Systematic geographical exploration.
- PvP: Competitive, player-vs-player activity.
- Dungeons/Raids: Team-based collaboration with large rewards.
- Professions: Non-combat crafting skills.
- World Events: Thematic, seasonal, story-based events.

Our first hypothesis is that each of the motivation factors has a significant impact across this set of in-game behavioral categories.

To examine whether the relationships between the motivation factors and in-game behaviors are aligned with theory, we leveraged the behavioral expectations of the motivation factors from Yee's original work [10]. We hypothesized that the Achievement factor would correlate positively with Dungeons/Raids and PvP (related to Yee's Advancement and Competition components respectively). We hypothesized that the Social factor would correlate positively with Dungeons/Raids, suggested by Yee's Teamwork component. And we hypothesized that the Immersion factor would correlate positively with Exploration, as suggested by Yee's Discovery component.

Data Collection

From the data sets already collected, we randomly sampled 500 participants from the US region, as well as 500 participants from the HK and TW region for behavioral data collection. In the online survey, participants were asked to list their active characters. On average, each participant had 2.79 active characters (SD = 1.51). We used an XML scraper to collect data from the Armory for each character over a 6-month period and used the most recent character snapshots for the following analysis.

To generate a player-level metric across characters, we calculated the score ratio for each achievement category (i.e., = category achievement score/total achievement score) across all of a player's characters. This allowed us to avoid confounding achievement scores with character levels.

Results

We calculated the factor score for each motivation factor and conducted a multivariate regression using these factor scores to predict the in-game achievement category ratios, controlling for gender and age. The multivariate tests were significant for the Achievement ($F = 19.95$, $p < .001$), Social ($F = 11.61$, $p < .001$), and Immersion ($F = 3.38$, $p < .01$) factors. Thus, it is clear that the self-report data has a significant relationship with the set of in-game behavioral metrics. The individual regression coefficients are shown in Table 2. Our hypothesized correlations are all supported by the data. In addition, we note that the Achievement factor is strongly negatively correlated with Professions, Exploration, and Quests. In hindsight, it makes sense that activities that do not lead directly to functional rewards are not appealing to Achievement-oriented players. The same is true of the negative correlation between Quests and the Social factor. Given that quests are now designed to be completed alone, it is not surprising in hindsight that Social-oriented players find quests less appealing.

Category	ACH	SOC	IMM	Adj R^2	p
Quests	**-.21**	**-.16**	**.10**	.13	< .001
Exploration	**-.21**	**-.09**	**.09**	.17	< .001
PvP	**.14**	-.05	**-.07**	.18	< .001
Dungeons	**.28**	**.18**	**-.10**	.30	< .001
Professions	**-.19**	**-.14**	.06	.25	< .001
World Event	**-.11**	.03	.05	.11	< .001

Table 3. Standardized beta coefficients and significance of regression models. Bolded coefficients are p < .01.

CONCLUSION

Data provided in this paper show that online gaming motivations can be parsimoniously captured using a 3-factor model. Our assessment tool for this model was validated in both a Western and non-Western culture using rigorous statistical methods. And finally, self-report data using this measure is significantly correlated with actual in-game behavioral metrics. Together, this data demonstrates the robust validity and reliability of the developed scale.

On the other hand, there are several weaknesses that should be mentioned. First, we only collected data from players of one online game. However, the concordance between the current findings and the earlier work (using a broader online gaming sample) provides moderate assurance that these factors generalize more broadly than just WoW. Nevertheless, future research should attempt to validate this scale with players of additional online games. Secondly, the behavioral correlations with the Immersion factor were not particularly strong. Future studies should examine whether this is an artifact of WoW lacking good in-game variables relating to Immersion, or whether the subjective nature of the Immersion items are poorly captured by behavioral metrics in general. Finally, validating the scale in other cultures would help us understand how universal the 3-factor structure is.

As we mentioned in the introduction, game-related mechanisms have been predicted to slowly permeate many application categories. We believe that the development of a robust and well-validated scale is an important foundation in understanding why players play, what mechanisms may work better for different demographic segments, and variances in learning or behavioral outcomes.

Beyond this usage, leveraging the scale to understand a player's profile before they start using a game (or game-like) system opens the possibility of tailoring their experience to better match their motivations – and such personalization has been a topic of interest at CHI for many years [5,7]. In fact, our in-game data shows that this process could even be automated to infer player motivations directly from their in-game behaviors, raising the possibility of dynamically tailoring a player's experience over time and without explicitly asking them for information. Such personalization, dynamic or otherwise, would most likely increase the efficiency of many of the game-like systems starting to appear at CHI and elsewhere.

ACKNOWLEDGEMENTS

This research was sponsored by the Air Force Research Laboratory.

REFERENCES

1. Bartle, R. Hearts, Clubs, Diamonds, Spades: Players Who Suit MUDs. *The Journal of Virtual Environments 1*, (1996).
2. Gartner. Gartner Newsroom Press Release. 2011. http://www.gartner.com/it/page.jsp?id=1629214.
3. Hu, L. and Bentler, P. Cutoff criteria for fit indexes in covariance structure analysis: Coventional criteria versus new alternatives. *Structural Equation Modeling 6*, (1999), 1-55.
4. Lazarro, N. Why We Play Games: Four Keys to More Emotion Without Story. *GDC*, (2004).
5. Mackay, W. Triggers and Barriers to Customizing Software. *Proceedings of SIGCHI 1*, (1991), 153-160.
6. Reeves, B. and Reed, L. *Total Engagement: Using Games and Virtual Worlds to Change the Way People Work and Businesses Compete*. Harvard Business School Press, 2009.
7. Riecken, D. Personalized Views of Personalization. *Communications of the ACM 43*, (2000), 26-28.
8. Slater, M. How colorful was your day? Why questionnaires cannot assess presence in virtual environments. *Presence: Teleoperators and Virtual Environments 13*, (2004), 484-493.
9. Yee, N., Ducheneaut, N., Nelson, L., and Likarish, P. Introverted Elves and Conscientious Gnomes. The Expression of Personality in World of Warcraft. *Proceedings of CHI 1*, (2011), 753-762.
10. Yee, N. Motivations for Play in Online Games. *Journal of CyberPsychology and Behavior 9*, (2007), 772-775.