

Inferring Player Rating from Performance Data in Massively Multiplayer Online Role-Playing Games (MMORPGs)

Kyong Jin Shim, Muhammad Aurangzeb Ahmad, Nishith Pathak, Jaideep Srivastava

Department of Computer Science and Engineering

University of Minnesota

Minneapolis, Minnesota 55455, USA

{kjshim, mahmad, npathak, srivasta}@cs.umn.edu

Abstract—This paper examines online player performance in EverQuest II, a popular massively multiplayer online role-playing game (MMORPG) developed by Sony Online Entertainment. The study uses the game's player performance data to devise performance metrics for online players. We report three major findings. First, we show that the game's point-scaling system overestimates performances of lower level players and underestimates performances of higher level players. We present a novel point-scaling system based on the game's player performance data that addresses the underestimation and overestimation problems. Second, we present a highly accurate predictive model for player performance as a function of past behavior. Third, we show that playing in groups impacts individual performance and that player-level characteristics alone are insufficient in explaining an individual's performance, which calls for a different set of performance metrics methods.

I. INTRODUCTION

A. Motivation

Massively Multiplayer Online Role-Playing Games (MMORPGs) are personal computer or console-based digital games where thousands of players can simultaneously sign on to the same online, persistent virtual world to interact and collaborate with each other through their in-game characters. In recent years, researchers have taken notice that virtual environments such as EverQuest II serve as a major mechanism for socialization [25]. In particular, educational research has found virtual environments to be a sound venue for studying learning, collaboration, social participation, literacy in online space, and learning trajectory at the individual level as well as at the group level. Online communities and virtual worlds alike frequent journals and conference proceedings in the field of Learning Sciences. Learning takes place beyond classroom doors, and virtual worlds have allowed researchers to study learning in naturally occurring contexts [11]. A more recent study [24] sets out to examine the discourse of MMORPG gaming wherein the primary emphasis of research lies in understanding individual-level participation, social and material practices, literacy, community membership, and individual learning trajectory in MMORPGs. The present research is concerned with learning in virtual environments and examines online

player performance in EverQuest II, a popular massively multiplayer online role-playing game (MMORPG) developed by Sony Online Entertainment. The study uses the game's player performance data to devise performance metrics for online players.

B. Performance Metrics

In Operations Research, performance has been studied extensively for decades and has resulted in the development of quantitative measures to optimize assembly line production, customer satisfaction, and employee retention. One area of interest in Operations Research is manufacturing strategy, defined in terms of capabilities and resources, and how they are linked to manufacturing performance. Assembly line balancing problems are an example and have been the subject of rigorous research for decades. Although most of the existing literature is focused on single-product assembly lines, recent work in this area has explored the measurement and optimization of performance in team-oriented assembly lines. The main objective of assembly line balancing problem is to maximize efficiency through minimization of idle time [5], [7], [9]. Along the same line, prior literature exists that discusses more complex systems. Mixed model assembly lines are one such example, and it assembles several models of the same product on the same line [1], [2], [4], [10].

In much the same way manufacturing performance is measured in terms of performance, quality and inventory [3], we define performance in EverQuest II as a function of productivity and quality. We define productivity as a measure of how many tasks a given player completes and how many points he/she gains as a result of completing the task(s) in a given time duration. In Operations Research, quality is discussed in terms of defects. In a similar manner, we define quality as success ratio, which is a measure of how successful a given player is at completing a task. In the game, there are multiple types of tasks a player can perform. Monster kills and quests are two prominent types of task in EverQuest II. In the case of monster kills, success ratio is formulated as (number of successful attempts) / (number of successful attempts + number of unsuccessful attempts).

C. Impact of Group Formation on Performance

Manufacturing plants over the years have adopted the formation of work teams as a practice [12], [18], [21], [22]. Many companies have adopted team approaches to produce high quality products and services which would lead to improved customer satisfaction [8]. Additionally, huge cost savings coupled with quality improvements have been reported in numerous studies [13]–[17], [19]. A more recent study conducted empirical studies on the impact of team formations at workplaces on manufacturing performance over an extended period of time [23].

As is the case in manufacturing, team formation is a common occurrence in many MMORPGs [29]–[31]. The games are designed to encourage social interactions in such a way that certain quests must be done as a group. Not only that, certain quests require players each with a different set of skills. In order to successfully complete a given quest, the team members must collaborate and rely on one another. In EverQuest II's monster kills, a player can choose to group with and collaborate with one or more players in killing monsters. Such grouping behaviors are often observed in the case of killing difficult or vicious monsters. Also, novice players can team up with more advanced players to get familiarized with the game via the game's mentorship system.

The present study investigates the impact of team formation on individual players' performances: Is teaming up necessarily better or worse for individual players? Does the answer to this question vary depending on participating players' levels?

II. EVERQUEST II GAME MECHANICS

A. Monster Kills and Point System in EverQuest II

EverQuest II is rich in types of task players can perform with monster kills being one of the most popular. In monster kills, each monster has a level and a tier. The two values indicate the difficulty of a monster. The higher the two values, the more difficult or challenging it is for a given player to kill the monster. The monster level increase is not a monotonic function (i.e., monster level 17 is not necessarily difficult than monster level 16 because difficulty is an aggregate function of monster levels and tiers). In successfully killing the monster, a player obtains points. The amount of points assigned is minimally dependent upon three factors: 1) monster's level, 2) monster's tier, and 3) player's level. Table I shows performance data from killing a baby dune cobra. This example shows two different baby cobras: one having level 13 and tier 5 and the other having level 15 and tier 5. Two players of levels 16 and 19, respectively, performed the first task and obtained scores of 52 and 12. In performing the same task, the player with a lower level obtains more points. The same trend is shown in the second example where three players performed the same task, and the player with the lowest level obtains the highest points amongst the three. These examples illustrate how EverQuest II rewards adjusted points based on task difficulty and player skill level.

Monster	M-Level	M-Tier	Player Level	Points
Baby dune cobra	13	5	16	52
Baby dune cobra	13	5	19	12
Baby dune cobra	15	5	13	141
Baby dune cobra	15	5	21	27
Baby dune cobra	15	5	22	12

TABLE I
MONSTER LEVEL AND TIER

B. Point-Scaling System in EverQuest II

In EverQuest II, there is a concept of Ding Points, which is the amount of points one needs to obtain in order to move from one level to the next higher level [32]. For instance, to move from Level 2 to Level 3, one needs to obtain 1,000 points whereas 20,000 points are required to move from Level 73 to 74. The amount of ding points increases as one advances to the next level. As players gain more experience with the game and advance to higher levels, the types of task they can perform increase and the task difficulty also increases. The higher the task difficulty, the higher the potential point gain. Does this increase in point gain scale well with the increase in task difficulty?

Numerous online and offline posts and articles within gaming circles report that the relative player rating systems and point-scaling systems are not perfect, often causing distress among advanced players who assert that the systems over-reward low level players and under-reward high level players. Sony's EverQuest II is no exception in that it adopts a relative point-scaling system which requires that higher level players accumulate relatively more points than lower level players in order to advance to the next level [32].

C. Performance-Based Point-Scaling System

An experimental study is conducted to evaluate the existing Ding Points-based Point-Scaling System. The main objective of any point-scaling system should be to raise or lower expectations in terms of performance based on the player's level so that advancing from Level i to Level $i + 1$ carries the same amount of difficulty throughout the different levels after factoring in player skill. One way to verify that this condition is currently being met in the game is to measure the average time spent by players to advance from Level i to Level $i + 1$. The reasoning behind the measurement of time spent working on tasks is that time spent is generally proportional to task difficulty. If players at Level i are spending substantially more time than what is expected in the entire distribution from Level i to Level $i + 1$, it is an indication that the game's ding points-based point-scaling system is imposing expectations too high, and that it is not fair for players at Level i .

III. PERFORMANCE METRICS

A. Productivity as a Performance Measure

Upon completing a task, a player obtains a certain amount of points, which we call Experience (XP) points. Table II shows

	Task	Task	Task	Task	Total
Player A (Level 25)	1 (50)	2 (200)	3 (120)	4 (200)	570 XP
Time spent	3 min	8 min	5 min	10 min	26 mins
Player B (Level 25)	5 (150)	6 (200)	7 (220)		570 XP
Time spent	6 min	9 min	11 min		26 mins
Player C (Level 25)	5 (150)	6 (200)	7 (220)		570 XP
Time spent	4 min	8 min	10 min		22 mins
Player D (Level 25)	10 (50)	11 (40)	12 (40)	13 (50)	180 XP
Time spent	3 min	1.25 min	1 min	2 min	9.25 mins

TABLE II
TASK PERFORMANCE AND EXPERIENCE POINTS

an example of performance data of four players A, B, C, and D.

Case 1 - Differing Number of Tasks, Same XP Points, Same Time Durations

Let us take performance data of Player A and B. Both players gained 570 XP points in 26 minutes. However, Player A completed four tasks to achieve this score whereas Player B completed only three tasks to achieve the same.

Case 2 - Same Number of Tasks, Same XP Points, Differing Time Durations

Let us take performance data of Player B and C. Both players gained 570 XP points by completing three tasks. However, Player B took less time than Player C to achieve this score.

Case 3 - Same Number of Tasks, Differing XP Points, Differing Time Durations

Let us take performance data of Player A and D. Both players completed four tasks. However, Player A achieved a much higher XP score than Player D. Additionally, Player A took more time than Player D achieve higher XP score.

Given the above use cases, the most reasonable measure of productivity in the game is XP points gain.

The reasoning behind leveraging only XP points, not number of tasks, is that as shown in Use Case 1, the two players took different paths (one choosing three relatively difficult tasks and the other choosing one easy task and three relatively difficult tasks) but they both achieved the same XP points at the end of the day, in 26 minutes. For Player B, the 26 minute duration might have been more intensive than that of Player A because he faced more difficult monsters but one can make a similar argument by saying that Player A kept himself busy by completing one more task than Player B.

Apart from the number of tasks, the two primary reasonings behind using XP points are 1) XP points reflect task difficulty and 2) tasks with higher difficulty levels take more time. Hence, coupled with time measure (time taken to complete a set of tasks), XP points gain can provide a good measure of a player's productivity. The performance of Player K as a function of Productivity (Performance Metric 1) is defined as the following:

$$PerformanceMetric1_k = \frac{\sum_{i=1}^N XP_i}{\sum_{j=1}^M ST_j}$$

where

XP = Experience points

N = Total number of tasks completed by Player K

ST = Session time

M = Total number of sessions during which Player K completed tasks

B. Quality as a Performance Measure

In Operations Research, quality is discussed in terms of defects. In a similar manner, in EverQuest II, quality is defined as success ratio, which is a measure of how successful a given player is at completing a task. Success ratio is formulated as (number of successful attempts) / (number of successful attempts + number of unsuccessful attempts). Success ratio is specific to a task and performing of that task, and therefore, the aforementioned Performance formula can be adjusted to account for task-specific quality. Hence, the performance of Player K as a function of productivity and quality (Performance Metric 2) is as follows:

$$PerformanceMetric2_k = \frac{\sum_{i=1}^N \frac{XP_i + (XP_i \times Q_i)}{2}}{\sum_{j=1}^M ST_j}$$

where

XP = Experience points

N = Total number of tasks completed by Player K

ST = Session time

M = Total number of sessions during which Player K completed tasks

Q = Quality or success ratio associated with completing Task i

IV. EXPERIMENTS AND RESULTS

A. Dataset

The study uses one month worth of performance data from March 1, 2006 to March 31, 2006. The dataset contains over 36 million player-to-task records where over 4 million of them are monster kills related tasks. The dataset contains 24,571 distinct players across player levels 1 through 70. For the evaluation of ding points-based point-scaling system, the study uses eight months worth of data ranging from March, 2006 to August, 2006.

B. Evaluation of Ding Points-Based Point-Scaling System

An experimental study is conducted to evaluate the existing Point-Scaling System. The game's ding points are indicative of player level difficulty or how much effort is needed to move from Level i to Level $i + 1$. Another source of player level difficulty is the game's performance data. From the performance data, we extract session time, which is indicative

of how much effort is actually being spent to move from Level i to Level $i + 1$. We compare this against the ding points-based point-scaling system to see how well the ding points-based point-scaling system reflects the actual player performance.

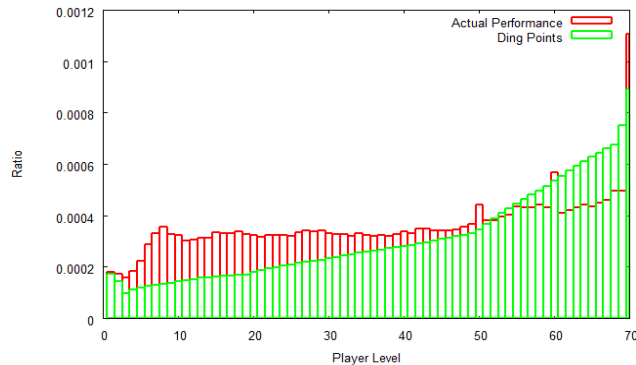


Fig. 1. Point-Scaling System versus Actual Time Spent

Figure 1 is meant to compare ding point system and the actual performance of the players. The x-axis shows the player levels while the y-axis is a ratio. The green bars show the ratios between ding points required to move from level i to level $i + 1$ while the red bars show the ratios between amounts of time spent in moving from level i to level $i + 1$ divided by the maximum time spent. From this figure, it can be observed that up until Level 49, the actual time spent by players performing tasks is more than expected. This could potentially mean that the tasks performed by players up until Level 49 were more challenging than expected as time spent increases with an increasing level of task difficulty. Between Levels 50 and 55, the actual time spent is well in accordance with what is expected. Beyond Level 55 up until Level 68, the actual time spent is well below what is expected. This could potentially mean that the tasks performed were not challenging enough as time spent decreases with a decreasing level of task difficulty. Game developers can use the above two pieces of information to do the following.

First, they could lower standards/expectations for players at levels below 49 by decreasing the amount of ding points required to move up to the next level. Players will need to either complete less number of tasks or complete less challenging tasks. Alternatively, the score adjustment formula can reduce the penalty imposed on lower level player.

Secondly, they could raise the standards/expectations for players at Levels 55 through 68 by increasing the amount of ding points required to move up to the next level. Players will respond by trying more challenging tasks as increasing challenge level positively correlates with increasing experience points necessary to move up to the next level. Alternatively, players will need to complete more tasks if they opt for not trying more challenging tasks as completing more tasks will result in experience points gain necessary to move up to the

next level. Another way for the game developers to respond to this finding is that in EverQuest II, experience points are adjusted based on the player level and perhaps, the score adjustment formula as discussed in Section II-A can reduce the reward amount imposed on higher level player.

Analyzing the game's performance data reveals information valuable in understanding at what rate players advance through completing tasks of different difficulty levels. The results from this analysis can be used to establish a measure of how difficult it is to move from Level i to Level $i + 1$. This measure would be similar to the existing Ding Points-based Point-Scaling System, however, as the study reveals, this is not completely reflective of actual player performance.

This observation has practical consequences for game development since it is known that one of the main attractions for game players in any game is the challenge associated with the game [27]. Understanding the balance between keeping players relatively busy/challenged versus bombarding them with difficult tasks is valuable in devising training routines for soldiers and novices.

C. Impact of Group Formation on Individual Performance

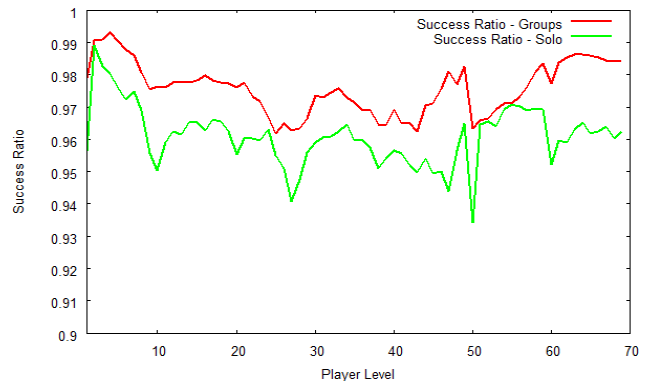


Fig. 2. Success Ratio - Playing Solo versus Playing in Groups

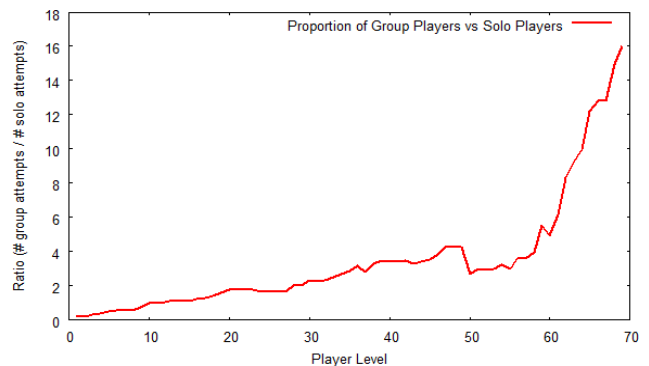


Fig. 3. Proportion of Group Players versus Solo Players

To evaluate the impact of group formation on individual players' performances, we computed each player's quality

measure only in the context of monster kills. We then aggregated over all players at each level, and the resulting Figure 2 shows that in most levels, individual players' success ratios are higher when they played in groups. Figure 3 shows a trend that as the player level increases, the proportion of players playing in groups increases. Higher level players tend to fight more difficult monsters which presents the necessity to group with other players in order to successfully kill the monsters. Additionally, players reaching higher levels are more inclined to join guilds or raid groups. Moreover, a player's higher level status attracts other players to group with him.

D. Evaluation of Performance Metrics

In this set of experiments we evaluate the first Performance Metric. Player performance as a function of productivity and/or quality reveals a given player's rate at which he or she advances through player levels by completing tasks of different difficulty levels and/or by grouping with other players. Given the player's past performance, it is possible to predict his or her future performance.

At each player level (Level i), we select N players and compute their Performance Metric 1 scores. The game's existing ding points-based point-scaling system dictates that there is a fixed amount of points to be gained between any Level i and Level $i + 1$. Given a player's Performance Metric 1 score and the fixed amount of points between Level $i + 1$ and Level $i + 2$, we can compute the total session time (play time) and this becomes our prediction as to how fast this player will advance to the next higher level in the future.

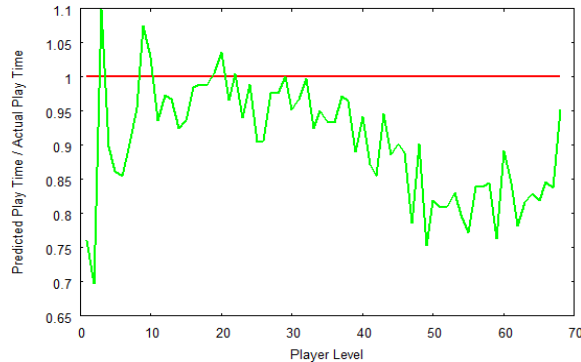


Fig. 4. Ratio of Predicted Play Time and Actual Play Time by Player Level

We compare the predicted play time against the actual play time. We take the ratio between the two and observe at each player level what is the margin of prediction error is. Figure 4 shows that for Levels 2 and 3, our method underestimated the actual play time. Between Levels 4 and 48, the margin of prediction error stays well within 18% boundary. Beyond Level 48, the margin of error increases and players' performances become less predictable. For higher level players, our method tends to underestimate the actual play time.

Coupling this finding with Figure 2 and Figure 3 reveals an interesting pattern. As the player level increases, group

formation becomes a more common occurrence. And playing in groups leads to higher success ratio at the individual player's level. There is a tradeoff between playing solo versus playing in groups. From timing perspective, playing solo allows a given player to advance faster than it would if he were to play in groups potentially due to process loss or coordination overhead incurred by having to get multiple players together. From the perspective of successful task completion and success ratio, playing in groups serves as an advantage in that the chance of getting a given task done is higher for a given individual player in this setting.

Analyzing historical performance data of a player is expected to yield information valuable in understanding his learning trajectory over time. To further evaluate Performance Metric 1, we conducted an experiment. In one use case, we used the immediate past performance data (from Level $i + 1$ to $i + 2$) as a predictor for future performance (from Level $i + 2$ to $i + 3$). In the other use case, we used a more distant past performance data (from Level i to $i + 1$) as a predictor for future performance (from Level $i + 2$ to $i + 3$).

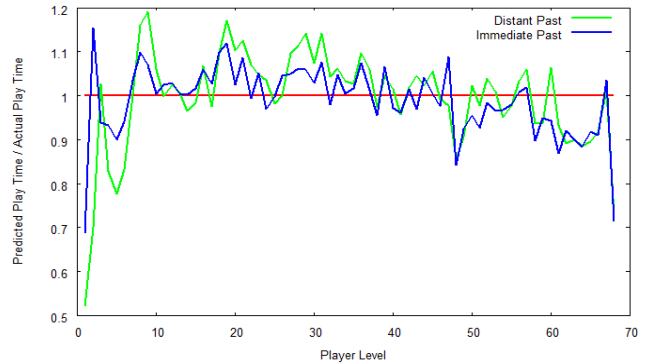


Fig. 5. Predicted Play Time versus Actual Play Time (Distant past as a predictor of future performance)

Figure 5 shows that the margin of prediction error is larger when we use a more distant past performance data as predictor for future performance. It is evident from the results that the quality of performance data as a predictor for future performance decays with an increasing distance on the player level scale. Our finding indicates that in order to incorporate more distant performance data into the proposed Performance Metrics method, some sort of a weight assignment or decay function must be applied in such a way that the most weight is given to the most immediate past.

Our findings indicate that our proposed Performance Metric 1 is suitable for predicting individual players' future performances in absence of impact of group formation. These findings call for more thorough and comprehensive studies on different types of group formations (homogeneous, heterogeneous, social interactions amongst the group members, etc.) and their impact on individual players' performances.

We conducted a similar experiment to evaluate Performance Metric 2. In this experiment, we used only monster kills for analysis because these tasks are readily amenable to analysis

in terms of success and failure. Failure in this context would mean the death in the game, while failure cannot be readily described for most other tasks. The results were not significantly different from the Performance Metric 1 evaluation results. In monster kills, taking quality into consideration for performance metric does not lead to better predictions. A planned addition to the current research is to expand this analysis onto other more complex types of tasks (i.e. quests) and evaluate whether quality does or does not play a role in predicting individual performance in those types of tasks.

V. CONCLUSION

The present study analyzes EverQuest II's player performance data to devise individual player performance metrics. First, our analysis reveals that the game's existing ding points-based point-scaling system is in general well in accordance with the actual player performance observed in the game's historical performance data. It also reveals that the level of granularity that the performance data offers can potentially lead to fine tuning of the existing point-scaling system. Secondly, the proposed performance metrics define performance as a function of productivity and/or quality. Our findings demonstrate that a given player's past performance can be used as a predictor for his future performance. Our findings also indicate that the proposed performance metrics yield less than optimal predictions about individual players' future performances in higher levels where group formation increasingly becomes a common occurrence. Additionally, the study reveals that in a certain type of task (i.e. monster kills), the quality aspect of individual performance plays an insignificant role in predicting a player's future performance. Future directions include studying different ways of defining quality in all types of task in the game to devise more generalizable individual player performance metrics, conducting a more thorough and comprehensive analysis on the impact of group compositions and social interactions on individual player performance, investigating individual learning trajectory over a larger time span, and developing group performance metrics.

ACKNOWLEDGMENT

The research reported herein was supported by the National Science Foundation via award number IIS-0729421, and the Army Research Institute via award number W91WAW-08-C-0106. The data used for this research was provided by the SONY corporation. We gratefully acknowledge all our sponsors. The findings presented do not in any way represent, either directly or through implication, the policies of these organizations. Acknowledgements are also due to the members of the University of Minnesota, Northwestern University, University of Southern California, and University of Illinois.

REFERENCES

- [1] Thomopoulos, N. T. (1970). Mixed model line balancing with smoothed station assignments. *Mgmt. Sci.* 16, 593-603.
- [2] Macaskill, J. L. C. (1972). Production-line balances for mixed-model lines. *Mgmt. Sci.* 19, 423-434.
- [3] Kaplan, R. (Oct 1983). Measuring Manufacturing Performance: A New Challenge for Managerial Accounting Research, *The Accounting Review*, Vol. 58, No. 4., pp. 686-705.
- [4] Chakravarty, A. K. & Shtub, A. (1985). Balancing mixed model lines with in-process inventories. *Mgmt. Sci.* 31, 9.
- [5] Baybars, I. (1986). A survey of exact algorithms for the simple assembly line balancing problem. *Mgmt. Sci.* 32, 909-932.
- [6] Manufacturing Studies Board. (1986). Human resources practices for implementing advanced manufacturing technology. Washington, DC: National Academy Press.
- [7] Gosh, S. & Gagnon, R. J. (1989). A comprehensive literature review and analysis of the design, balancing and scheduling of assembly systems. *Int. J. Prod. Res.* 27, 637-670.
- [8] Boyett, J. H. & Conn, H. P. (1991). *Workplace 2000: The revolution reshaping American business*. New York: Penguin.
- [9] Dar-El, E. M. (1991). Design of assembly systems "DAS", ICPR. 1991, Workshop 1, 18-23 August 1991, Hefei, China.
- [10] Fremery, F. (1991). Model-mix balancing: more flexibility to improve the general results. In: *Production Research: Approaching the 21st. century*, London.
- [11] Lave, J. & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
- [12] Safizadeh, M. H. (1991). The case of workgroups in manufacturing operations. *California Management Review*, 33(4): 61-82.
- [13] Hitchcock, N. A. (1993). Employee participation paves way to product perfection. *Modern Material Handling*, 48(6): 44-45.
- [14] Tjosvold, D. (1991). *Team organization: An enduring competitive advantage*. New York: Wiley.
- [15] Schilder, J. (1992). Work teams boost productivity. *Personnel Journal*, 71: 67-71.
- [16] Katzenbach, J. R. & Smith, D. K. (1993). *The wisdom of teams: Creating the high-performance organizations*. New York: HarperCollins.
- [17] Sheinkman, J. (1993). Amalgamated clothing and textile workers union. *Quality Progress*, 26(9): 57-58.
- [18] Business Week. (1994). The team player: no more "same-ol'-same-ol'." October 17, 95-96.
- [19] Gupta, Y.P. & Ash, D. (1994). Excellence at Rohm and Haas Kentucky: A case study of work team introduction in manufacturing. *Production and Operations Management*, 3(3): 186-200.
- [20] New York Times. (1994). A new labor design at Levi Strauss. October 13: C1-C2.
- [21] Osterman, P. (1994). How common is workplace transformation and who adopts it? *Industrial and Labor Relations Review*, 17: 173-188.
- [22] Ichniowski, C., Shaw, K., & Prennushi, G. (1994). The impact of human resource management practices on productivity. Working paper 014, Columbia University, Carnegie Mellon University, and World Bank.
- [23] Banker, R. D., Field, J. M., Schroeder, R. G., & Sinha, K. K. (Aug., 1996). Impact of Work Teams on Manufacturing Performance: A Longitudinal Field Study. *The Academy of Management Journal*, Vol. 39, No. 4, pp. 867-890.
- [24] Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method*. New York: Routledge.
- [25] Jewels (2002, December 9). But in the end, they're still nothing more than video games. *Jive Magazine*.
- [26] Gee, J. P. (2003). *What videogames have to teach us about learning and literacy*. New York: Palgrave Macmillan.
- [27] Gao, Y. (2004). Appeal of online computer games: A user perspective. *The Electronic Library*, 22, 74-78.
- [28] Williams, D. (2005). Bridging the methodological divide in game research. *Simulation & Gaming*, 36(4), 447-463.
- [29] Williams, D. (2006). Groups and goblins: The social and civic impact an online game. *Journal of Broadcasting and Electronic Media*, 50(4), 651-670.
- [30] Williams, D., Ducheneaut, N., Xiong, L., Zhang, Y., Yee, N., & Nickell, E. (2006). From tree house to barracks: The social life of guilds in World of Warcraft. *Games & Culture*, 1(4), 338-361.
- [31] Williams, D., Caplan, S., & Xiong, L. (2007). Can you hear me now? The impact of voice in on online gaming community. *Human Communication Research*, 33(4), 427-449.
- [32] Prima Development. *Everquest II: Official Strategy Guide (Prima Official Game Guides)*.