THE STONE AGE IS BACK
HCI Effects on Recommender Systems

Yuval Dan-Gur
Graduate School of Management, University of Haifa, Mt. Carmel, Haifa, Israel

Keywords: Recommender systems, HCI, Friends group, Social IS.

Abstract: We addressed HCI and social aspects of recommender systems by studying the uncharted domain of the advising group and the user's control over it. We conducted a longitudinal field study in which, for two years, our research tool, QSIA (which means QUESTION in Hebrew language), was free for use on the web and was adopted by various institutions and classes of heterogeneous learning domains. QSIA enables the user to be involved in the formation of the advising group. The user was free to choose advising group for each recommendation sought, while the default choice is the common 'neighbors group'. QSIA yielded high internal validity of acceptance and rejection ratios due to the immediate "usage actions" that followed the recommendation outputs. Although the objective amount of data in QSIA's logs are fairly large (31,000 records, 10,000 items, 3,000 users), the relevant figures for analysis of recommendations are modest – 895 recommendations seeking records, accepted from 108 users, 3,000 rankings by 300 users, and 1,043 "usage actions" by 51 users. Our findings suggest that the perceived quality of the recommendations (measured in terms of "usage actions") is 14% to 24% higher ($\alpha \leq 0.05$) for user-controlled 'friends group' than for machine-computed 'neighbors group'. We almost felt that the ancient tribal friends "revived" in modern Information Systems.

1 INTRODUCTION

Our research concerns with computerized social collaborative systems known as Recommender Systems. The main task of a recommender system is to recommend, in a personalized manner, relevant items to users from large number of alternatives, for example: web resources, movies, books and ski resorts.

Little notice has been paid to the social aspects of recommender systems and to the unsuitability they impose to the natural process of seeking and providing recommendations. We chose to concentrate on the social aspects of user involvement in the recommendation process, specifically, in the formation of the advising groups.

We reported the data previously (Rafaeli, Dan-Gur, and Barak, 2005) and now we present the accompanying HCI process and implications.

We introduce the term "friends group" to describe a sub-group of the neighbors group that is not solely rank-dependent, as opposed to "neighbors" that are assigned by rating similarity. The 'friends group' is unique because of the user's involvement in its formation and the user's ability to choose the characteristics of its members. The latter aspect is in accordance with the "Social Comparison Theory" (Festinger, 1954) and the derived behavioral studies suggesting that 'neighbors' (like-minded group) are relevant for 'low-risk' domains whereas 'friends' (similar on personal characteristics) are more relevant for 'high-risk' domains.

2 RESEARCH QUESTIONS

Our first research question was concerned with users' preferences concerning control over the recommendation process as opposed to acceptance of recommendations from a "computerized oracle".

The second research question examined whether the attitude of the recommendation seeker obeys social rules, specifically, the "Social Comparison Theory". We also assumed that given the option, users will choose similar-to-themselves 'friends' for their advising group. The three corresponding hypotheses were:

H_1:Recommendation seekers will prefer to use

Dan-Gur Y..
THE STONE AGE IS BACK - HCI Effects on Recommender Systems .
DOI: 10.5220/0003303902630270
In *Proceedings of the 7th International Conference on Web Information Systems and Technologies* (WEBIST-2011), pages 263-270
ISBN: 978-989-8425-51-5

controlled 'friends groups' over automatically, machine-generated 'neighbors groups'.

H_2:Recommendations by user-controlled 'friends groups' will be better accepted and complied with by recommendation seekers than those produced by 'neighbors groups'.

H_3:Recommendation seekers will choose personally-similar 'friends' for their advising group.

3 RESEARCH METHODS

3.1 Research Tool – QSIA

QSIA™ (pronounced "QU-SHI-YA" and means QUESTION in ancient Hebrew language) is a collaborative system for collection, management, sharing and assignment of knowledge items for learning that was developed in the Center for the Study of the Information Society with the support of the Caesarea Edmond Benjamin de Rothschild Foundation Institute (CRI) for Interdisciplinary Application of Computer Science at the University of Haifa.

The QSIA system is built on four conceptual pillars: knowledge generation, knowledge sharing, knowledge assessment, and knowledge management (Rafaeli, Barak, Dan-Gur and Toch, 2003). We are mainly interested in the knowledge sharing aspect in which the QSIA sub-task is 'matching mates'- the system's capability of making matches among recommenders and those seeking recommendations via three phases:

- Uploading knowledge items – composing a question and allowing others to use it.
- Ranking knowledge items – answering a question and then grading it on an ordinal scale of 1-5, so others could benefit from ones' professional opinion, and letting the system revalidate the user's profile of preferences.
- Receiving recommendations – producing recommendations for the user by N-top nearest 'neighbors' or 'friends'.

QSIA's interface is multilingual to support users from a wide range of origins.

The system is a Web-based application with the 'business logic' operating from a central cluster of servers, enabling easy logging of user actions. The system's design allows administrators to download all data and user logs for research. Privacy is kept by maintaining arbitrary users'-id's in the data records and not recognizable personal details.

Since its release, QSIA has provided insights into knowledge sharing (Rafaeli et al., 2003), online question-posing (Barak and Rafaeli, 2004), communities of teachers and learners (Rafaeli, Barak, Dan-Gur and Toch, 2004) and the understanding of the potential of social recommender systems in support of E-Learning (Rafaeli, Dan-Gur and Barak, 2005):

- An arena of student-to-student and teacher-to-teacher information sharing was examined as well as the process of joint ranking and evaluations of knowledge items (Rafaeli et al., 2003).
- The creation of communities of teachers and learners that promote high-order thinking skills was discussed (Rafaeli et al., 2004), recognizing that web-based systems provide a prominent universe for learning (Rafaeli and Tractinsky, 1991).
- Online Question-Posing Assignment (QPA) was assesses by having students perform self and peer-assignments and take online examinations, all administered by QSIA (Rafaeli et al., 2004).

3.2 Conceptual Model of User-QSIA Interaction

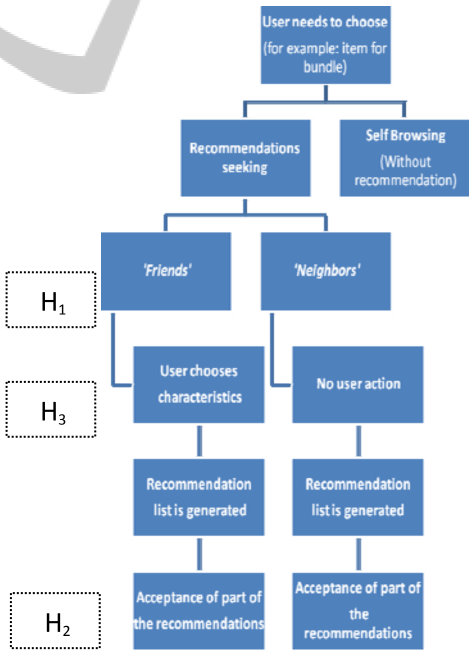

Figure 1: System's recommendation conceptual model.

We propose a five-stage conceptual model of user interaction with the recommendation module of QSIA, and define the variables, measures and involved computations accordingly. The model presented in figure 5 is relevant in each and every case that a user (teacher or student) has to make a selection (filtering) from the system's database (for example: a teacher is selecting items for a bundle or a student is practicing prior to an exam).

3.3 Procedure, Participants and Recorded Data

3.3.1 Procedure and Participants

This study includes data that was recorded in QSIA for two years: from July 2002 to July 2004. Since it was launched, QSIA was implemented in the following institutions and courses:

- Nesher High school, Nesher, Israel;
- Electronic Commerce course, Graduate School of Business, the University of Haifa, Israel;
- Electronic Commerce course, Industrial Engineering and Management, Technion, Haifa, Israel;
- Organizational Behavior course, Technion, Haifa, Israel;
- MIS course, the school for practical engineering, Ruppin College, Israel;
- Turkish Language course, the Faculty of Humanities, University of Haifa, Israel;
- General and systematic pathology course, the Faculty of Medicine, Tel-Aviv University; Israel;
- Electronic Commerce course, the Cyprus International Institute of Management, Nicosia, Cyprus;
- Electronic Commerce course, the University of Michigan, USA.

3.3.2 Recorded Data

During these two years, QSIA's database and logs presented us with the following figures:

- Number of users (teachers and students) – approximately 3000, most of them students.
- Number of items (either composed in QSIA or digitally imported) - approximately 10,000.
- Around 31,000 item-requests were served – mostly by self-browsing and a minor portion by recommendations seeking (friends or neighbors).
- Number of item rankings – approximately 3000, evaluated by around 300 users.
- Number of study groups/classes – 183.
- Number of knowledge domains – approximately 30.

When we filter out the data from recommendations seeking (either friends or neighbors), the figures downgrade to 895 recommendation requests (818 by students and 77 by teachers) generated by 108 active users.

3.3.3 Variables, Analysis and Measures

We classified major parts of this research as longitudinal (ageing effects and users' tendencies) and nonexperimental (an unobtrusive field study). We also noted that nonparametric analysis has to be applied to scores that violate the independence demand for parametric tests.

The main methods and tests that we used were the Wilcoxon Signed-Rank test, the Logistic Regression, the Generalized Estimating Equations (GEE) for analysis of longitudinal binary data using logistic regression and the Runs test for establishing randomness of a binary process.

Our field study was unobtrusive (Webb, Campbell, Schwartz and Sechrest, 1966; Kalman and Rafaeli, 2005), and we did not manipulate any variables. Data on users' behavior was collected retrospectively.
Our main dependent variables were:

Table 1: Main dependent variables.

Variable	Values	Number
SoR_i^j	The source of recommendation (friends or neighbors) for the j^{th} instance of recommendations seeking, by the i^{th} user: F_g or N_g.	(1)
R_i^j	The total number of items that the i^{th} user has rejected in the j^{th} instance, out of the produced recommendation list.	(2)
A_i^j	The total number of items that the i^{th} user has accepted in the j^{th} instance, out of the produced recommendation list.	(3)
DoU_j	Depth of Use - represents the maximum number of times that the jth user had asked for recommendations	(4)

4 RESULTS

We filtered out the records only to ones that were originated by recommendations and analyze the 895 records of recommendations seeking that were produced by the 108 users. The proportion of the recommendations seeking roles (teachers/students) is described in the following table:

Table 2: Students and teachers participation in recommendations logs.

	Users (N=108)	Records (N=895)
Students (or originated by students)	102	818
Teachers (or originated by teachers)	6	77
Total	108	895

When we examine the 895 records (108 users) which constitute the field experiment's log, we identify several aspects that require special attention.

The "Depth of Use" (DoUj), a variable that represents the maximum number of times that the j^{th} user had asked for recommendations, varies widely as the next figure presents. It should be noted that there are some users that asked for large instances of recommendations while many others presented us with a "cold start" behavior as presented in the following figure:

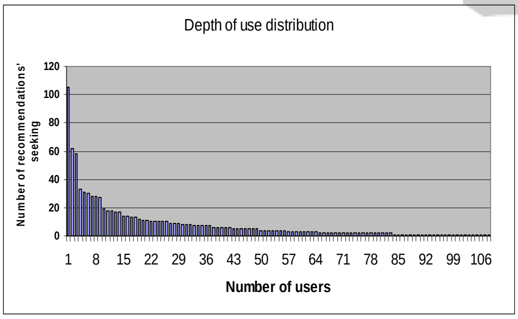

Figure 2: Depth of use (DoU) distribution.

4.1 H$_1$: Preference to Use 'Friends Groups' Over Machine-generated 'Neighbors Groups'

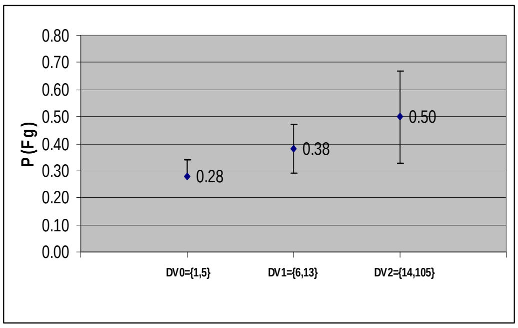

Figure 3: Results of GEE-105 model - Estimated values of 'friends group' choice at different instance ranges.

We ran six models all with different ranges of dummy variables. We report the results of a representative one, model GEE-105 that includes all 105 instances. The additional four models that also include all 105 instances presented similar results.

4.2 H$_2$: Acceptance of 'Friends Groups' Recommendations

The results of the "usage actions" (acceptances and rejections) for the same users who asked for recommendations from both sources (friends or neighbors) are presented in the following table:

Table 3: Acceptance ratios according to SoR.

	SoR=F$_g$	SoR=N$_g$
Number of records	264	377
Number of users	19	
Std. Dev.	0.29	0.3
Mean acceptance ratio	70%	56%
Mean difference	**14%**	
α (Wilcoxon, one tailed)	0.050	

The results show that acceptance ratio is 14% higher when users receive the recommendations from 'friends groups' rather than from 'neighbors groups' ($\alpha = 0.05$). These results represent 641 usage records by 19 users who sought recommendations from both sources. For exclusive users (who "experienced" only one source of recommendation), the mean acceptance ratio for those who chose only SoR=F$_g$ is higher by 24% from those who chose only SoR=N$_g$ (α=0.037).

4.3 H$_3$: Characteristics of the Chosen 'Friends'

We analyze a dataset of 335 records of 'friends group' recommendations seeking (SoR=F$_g$) from 32 users and examine their choices concerning each characteristic. The characteristics are considered statistically independent, (except for the impossibility of specifying a grade level when the chosen role was "teacher", because teachers do not have associated grades in QSIA).

Potentially we would have a maximum of 1,005 (335x3) non-zero field values but in reality we had only 270 such values. The maximum number of non-zero values decreases with any "no choice" of a user in any field and with any role = "teacher" choice because of the default grade value in such case.

We present in the following figure, the number of non-zero values in each distinct characteristic:

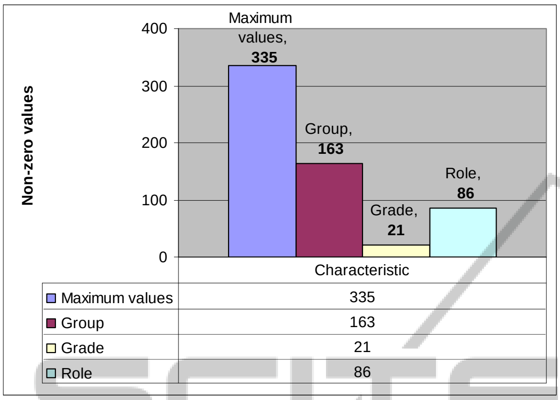

Figure 4: Number of non-zero values in the characteristics fields.

Several observations were evident even though the dataset was sparse:

- The sparseness of the data is high: approximately half of the records, although originating from the selection of 'friends group', do not include any specifications for the three possible characteristics
- Users were most likely to make group choices. We found a significant preference of users to include members of their own group in their 'friends group', than members of other groups. This result is also important because we have the largest amount of data concerning group choice – almost half the users assigned a value to this characteristic.
- Regarding role choice, we analyzed data only from students because teachers supplied only 5 records with this characteristic, without any choice in "student". The results present a preferred choice of students in teachers' advice rather than students' advice ($\Delta = 43\%, \alpha \le 0.0001$).

5 CONCLUSIONS

5.1 What is the Preferred Source of Recommendations for a User (H_1)?

Our findings suggest that users do develop a tendency to choose 'friends group' recommendations, and this tendency increases (in probability) as more recommendations are sought. Also, "experienced" users choose 'friends groups' significantly more than "new" users.

5.2 Are Recommendations from 'Friends Group' better Accepted (H_2)?

We found a 14% positive significant difference in the mean ratio of acceptance when we tested all users who had received and acted upon recommendations from both sources ('friends group' and 'neighbors group').

There was a higher positive significant difference in the mean acceptance ratios (24%, $\alpha = 0.037$) for users who received recommendations from only one source (either 'friends group' or 'neighbors group'). Also, when the same items were offered to users from both sources (N=36), the acceptance level was 6.5% higher when the recommendations were offered by 'friends groups' (P-value= 0.28).

For the most frequently recommended items that were recommended by both 'friends group' and 'neighbors group', the acceptance ratio was 15.2% higher (N=4, $\alpha = 0.034$) for the same items when they were recommended by 'friends groups'.

5.3 What Characteristics do Users choose for the 'Friends Group'? (H_3)?

There were many missing values in this part of our dataset: in almost half the records users made a group choice, in another quarter of the cases they made a role choice, and in only approximately 6% of the cases did users make a grade choice. We analyzed the characteristics independently and found that users significantly prefer their own group over other groups (76.6%, $\alpha<0.0001$).

5.4 What is New in Our Findings?

We addressed the HCI and social aspects of recommender systems by studying the uncharted domain of the advising group and the user's control over it. This attitude deviates from existing approaches that study algorithms (Breese, Heckerman and Kadie, 1998; Herlocker, Konstan, Borchers and Riedl, 1999; Fisher, Hildrum, Hong, Newman, Thomas and Vuduc, 2000; Goldberg, Roeder, Gupta and Perkins, 2000; Karypis, 2000), indices (Soboroff et al., 1999; Herlocker, 2000; Herlocker et al., 2004), items and technologies (Sarwar, Karypis, Konstan and Riedl, 2001).

Our findings suggest that there is a relationship between the perceived quality of the recommendations (measured in terms of "usage actions") and the formation of the advising group, and the control a user has over this process. We also addressed the issue of the inconsistency between preferences and behavior (Bacon, 1995; Minard, 1952; Wicker, 1969; Cosley, Lam, Albert, Konstan and Riedl, 2003) by introducing QSIA, a recommender system that enables immediate usage of the recommended items. This approach differs from studies that measure the accuracy of systems by measuring the accuracy of predicting users' ratings of items (Konstan et al., 1999; Herlocker, 2000; Sarwar et al., 2001).

We enabled users to rate the recommendations lists and thus, in future research, it will be possible to compare actual behavior (acceptances and rejections) and the users' explicit ratings of the recommended items. This comparison will be especially important for establishing relationships between attitude and behavior in recommender systems (Bacon, 1995; Cosley et al., 2003), and the characteristics of human taste (Freedman, 1998; Pescovitz, 2000).

5.5 Contribution of this Research

The findings may be of interest for further interdisciplinary research on collaborative filtering, bridging the gap between computerized oracles and social behavior.
We see potential contributions in the following aspects:

- Relating computerized collaboration systems and social theories.
- Motivation to conduct a field study of recommender systems, specifically in the 'high-risk' item domain (knowledge items), which users perceive as having a high value of predictive utility (Konstan, Miller, Malt, Herlocker, Gordon and Riedl, 1997).
- High validation of accepted recommendations, as we measure both implicit machine-collected data and explicit users declared attitudes.
- The economic implications of higher acceptance level of recommendations are substantial.
- A motivation to further examine one of the core pillars of 'social recommendation' – the advising group.
- Developers of recommender systems are advised to analyze deeply the design of interfaces and their influence on users.

6 WEAKNESSES AND LIMITATIONS

The current research on recommender systems has many limitations because of its uniqueness. The most important one to our view is that we do not have a relevant similar (or close to similar) comparable field study. Accordingly, we feel obliged, even more than in a "standard" study, to detail the main weaknesses and limitations as we recognize them.

6.1 Research Method

- We conducted a field study that inherently does not enable direct control of the independent variables. For a more detailed review of the characteristics of nonexperimental studies see Kerlinger (1986, p. 348-350).
- The statistical method we employed for longitudinal analysis of binary correlated data for finding ageing effects is the GEE extension of logistic regression. It is considered an area of statistics in which new developments occur on a regular basis (Hosmer and Lemeshow, 2000). Also, the Runs test (Bradley, 1968) that we tried to use for users' categorization requires sufficient data to test the degree of randomness, but due to low DoU's of users, we did not have enough data to employ the test for the majority of the users.
- The participating populations, except in one case, were homogeneous: students and teachers of academic institutions.
- The characteristics of the advising group that were possible for the recommendation seeker to control were very limited: groups, grade level and role.

6.2 Research Tool

- The QSIA system is unique in some aspects: to the best of our knowledge by enabling user's involvement in the determining the set of the 'neighbors group' for an automated collaborative filtering recommendation; QSIA is one of the few systems that enable immediate usage of the "liked" recommended items in the same system as the next step that follows suggestion of recommendations; and QSIA applies recommender technology to a novel domain – knowledge items for distance learning and online tests - that are not "natural" for recommender systems that are mostly applied to entertainment, commerce and news. Accordingly, we did not have other similar systems as a benchmark for these unique characteristics.

- We did not support the implementation and administration of QSIA to such an extent that builds significant trust and users' high expected utility, as could be done with larger resources (Gefen, 2004).

ACKNOWLEDGEMENTS

This completed research report was guided by Prof. Sheizaf Rafaeli to whom I own much of my humble research qualifications.

The QSIA system was designed by Dr. Eran Toch and Mr. Danny Shaham from the Research Center for the Study of the Information Society (INFOSOC at: http://infosoc.haifa.ac.il), under the guidance of Dr. Miri Barak with the support from the Caesarea Edmond Benjamin de Rothschild Foundation Institute for Interdisciplinary Application of Computer Science at the University of Haifa.

Finally, it must be noted (again) that throughout the research, we do not claim to prove causality; rather, we are aiming at relation establishment**.**

REFERENCES

Bacon, L. D. (1995). Linking attitudes and behavior - summary of literature. Paper presented at the *American Marketing Association/Edison Electric Institute Conference*, Chicago, Il.

Barak, M. & Rafaeli, S. (2004). Online question-posing and peer-assessment as means for web-based knowledge sharing in learning, *International Journal of Human-Computer Studies*, 61(1), 84-103.

Bradley, J. V. (1968). Distribution-Free Statistical Tests. New Jersey: Prentice-Hall.

Breese, J. S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive Algorithms for collaborative filtering. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence. Madison*, 43-52.

Cosley, D., Lam, S. K., Albert, I., Konstan, A. J. & Riedl, J. (2003). Is seeing believing?: how recommender system interfaces affect users' opinions. *Proceedings of the SIGCHI conference on Human factors in computing systems*, Ft. Lauderdale, Florida, 5(1), 585-592. New York: ACM Press.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 114-140.

Fisher, D., Hildrum, K., Hong, J., Newman, M., Thomas, M., & Vuduc, R. (2000). SWAMI: A framework for collaborative filtering algorithm development and evaluation, *Research and Development in Information Retrieval*, 366-368.

Freedman, S. G. (1998). Asking software to recommend a good book. *The New York Times*, 1998, June 20.

Gefen, D. (2004). What Makes ERP Implementation Relationships Worthwhile: Linking Trust Mechanisms and ERP Usefulness, *Journal of Management Information Systems*, 21(1), 275-301.

Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2000). Eigentaste: A Constant Time Collaborative Filtering Algorithm (*Technical Report* M00/41).

Herlocker, J. (2000). Understanding and improving automated collaborative filtering systems. *Unpublished Ph.D. dissertation*, UMI Order Number: AAI9983577, University of Minnesota.

Herlocker, J., Konstan, J., Borchers, A., & Riedl, J. (1999). An Algorithmic Framework for Performing *Collaborative Filtering, Research and Development in Information Retrieval* (pp. 230-237).

Herlocker, J., Konstan, A. J., Terveen, G. L. & Riedl, J. (2004). *Transactions on Information Systems. Communications of the ACM*, 22(1), 5-53. New York: ACM Press.

Hosmer, D. W. & Lemeshow, S. (2000). Applied Logistic Regression. *New York: Wiley*.

Kalman, Y. M. & Rafaeli, S. (2005). Email Chronemics: Unobtrusive Profiling of Response Times, Proceedings of the 38th International Conference on System Sciences, HICSS 38, 2005. Big Island, Hawaii. Ralph H. Sprague, (Ed.), 108. Available online: http://sheizaf.rafaeli.net/publications/KalmanRafaeliChronemics2005Hicss38.pdf

Karypis, G. (2000). Evaluation of Item-Based Top-N recommendation algorithms (CS-TR-00-46). *Minneapolis: University of Minnesota*, Department of Computer Science and Army HPC Research Center.

Kerlinger, F. N. (1986). Foundations of behavioral research. *Orlando: Holt, Rinehart and Winston, Inc*.

Konstan, J., Miller, B. N., Malt, D., Herlocker, J., Gordon, L. R., & Riedl, J. (1997). GroupLens: applying collaborative filtering to Usenet news. *Communications of the ACM*, 40(3), 77-87.

Konstan, J., & Riedl, J. (1999). Research Resources for Recommender Systems. Paper presented at the ACM SIGIR: *Workshop on Recommender Systems-Algorithms and Evaluation*, University of California, Berkeley.

Minard, R. D. (1952). Race relations in the Pocahontas Coal Field. Journal of Social Issues, 8, 29-44.

Moon, Y. (1998). The Effects of Distance in Local versus Remote Human-Computer Interaction. In *proceedings of the CHI 98'*, Los Angeles, CA. 103-108.

Moon, Y., & Nass, C. (1998). Are computers scapegoats? Attributions of responsibility in human-computer interaction. *International Journal of Human-Computer Studies*, 49, 79-94.

Pescovitz, D. (2000). Accounting for taste. Scientific American, June 2000.

Rafaeli, S., Barak, M., Dan-Gur, Y. & Toch, E. (2003). Knowledge sharing and online assessment, E-Society *Proceedings of the 2003 IADIS conference IADIS e-Society 2003*, 257-266.

Rafaeli, S., Barak, M., Dan-Gur, Y. and Toch, E. (2004). QSIA - A web-based environment for learning, assessing and knowledge sharing in communities, *Computers and Education, 43*(3), 273-289.

Rafaeli, S., Dan-Gur, Y. & Barak, M. (2005). Finding friends among recommenders: Social and "Black-Box" recommender systems", International Journal of Distance Education Technologies (IJDET), Special Issue on Knowledge Management Technologies for E-learning: Exploiting Knowledge Flows and Knowledge Networks for Learning, 3(2), 30-47.

Rafaeli, S. & Tractinsky, N. (1991). Time in computerized tests: A multi-trait multi-method investigation of general knowledge and mathematical reasoning in online examinations. *Computers in Human Behavior*, 7(2), 123-142.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). Item-Based collaborative filtering recommendation algorithms. In Proceedings of the 10th International *World Wide Web Conference (WWW10)*, Hong Kong. Available: http://citeseer.ist.psu.edu/sarwar01itembased.html.

Webb, E. Campbell, D. Schwartz, R. & Sechrest, L. (1966). Unobtrusive measures: Nonreactive research in the social sciences. *Chicago: Rand McNally*.

Wicker, A. W. (1969). Attitudes versus actions: The relation of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues,* 25, 41-78.