Dare to Compare: Motivating Expertise Building in the Enterprise through Intelligent User Modeling Interfaces

Mercan Topkara*, Justin D. Weisz*, Shimei Pan#, Jie Lu*, Jennifer Lai*
*IBM Research, Yorktown Heights, NY, 10532, USA
#University of Maryland, Baltimore County, Baltimore, MD 21250, USA

ABSTRACT
Expertise and skill assessments are a common aspect of working in an enterprise, but manual assessments are onerous and quickly outdated. Automated assessments can alleviate these problems, albeit at the risk of being inaccurate. In this short paper, we focus on the problem of how to design an engaging learning system in the presence of potentially inaccurate automated expertise assessments; especially when the users are in their early stage of using the system. We explore two dimensions associated with reporting automated expertise assessments to users: i) the inclusion of a social comparison, and ii) the precision of how expertise scores are presented. In a controlled experiment (N=60), we examined the impact of these dimensions on the perceived accuracy of the assessments, the perceived utility of the system, and peoples' willingness to share expertise scores within the enterprise.

Author Keywords
Enterprise, expertise, social comparison, assessment, user study, user interface design.

ACM Classification Keywords
H5.m. Information interfaces and presentation

General Terms
Human Factors; Design; Measurement.

INTRODUCTION
Continuous knowledge acquisition is critically important for knowledge workers who must regularly expand the scope of what they know in order to remain competitive. Given the multitude of demands on their time, informal learning is an important method by which workers can efficiently increase the depth and breadth of their expertise [4]. However, almost by definition, there is little formalized support for informal learning in enterprise settings and few systems that structure feedback on progress. To address this deficiency, we are building a system that creates a personalized learning environment for expertise building.

A core feature of this system is to automatically assess an individual's expertise level based on the content s/he has posted to an enterprise online community. This assessment helps the system identify learning needs for a user so it can create personalized recommendations and guidance.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
DUBMOD'14, November 3, 2014, Shanghai, China.
Copyright © 2014 ACM 978-1-4503-1303-2/14/11…$15.00.
http://dx.doi.org/10.1145/2665994.2665996

Measuring general expertise is a complex task that requires taking into account multiple aspects of a worker's skills, activities, and experiences. However, methods for assessing expertise in a given topic based on online content created by that person have been used successfully in expert identification systems [2]. Our system takes a similar data-driven approach of assessing expertise by examining data available from digital content repositories [5,7], and analyzing each user's activity, word usage, and content contributions. Unlike expert identification systems, expertise assessments in our system are used for self-evaluation, for providing ongoing feedback, and for personalizing learning recommendations.

One concern with data-driven expertise assessments is whether they are accurate. Since accuracy is a function of data, we do not wish to discourage usage due to inaccurate assessments. Thus, we address the question of how users react when presented with a score that may or may not agree with their own beliefs about their expertise. We do this in a controlled laboratory experiment that compares a precise representation of an expertise score with one that shows the score as a range of values.

In order to further motivate knowledge workers to use our system, we utilize from Social Comparison theory. Identified by Festinger in 1954, it suggests that individuals feel the need to self-evaluate by comparing themselves directly with others in a similar situation [3]. Comparing

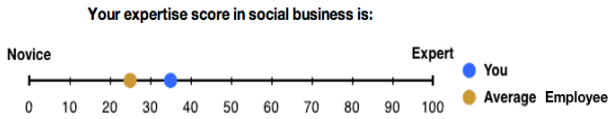


Figure 1. Positive social comparison with point presentation.

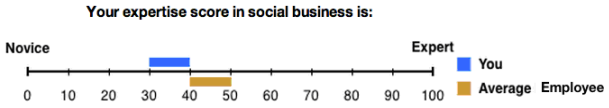


Figure 2. Negative social comparison with range presentation.

ourselves to those better off has been shown to have motivating effects, and comparing ourselves to those worse off makes us more likely to feel better about ourselves [1].

This paper makes several important contributions: we demonstrate that social comparisons improve perceptions of utility, we discover a counter-intuitive willingness to share potentially inaccurate expertise assessments within the enterprise, and we find that representing uncertainty may not have an effect on perceptions of score accuracy.

HYPOTHESES

Numerous studies have examined the effects of social comparison on users' motivation to use interactive systems. Mumm & Mutlu [8] found that feedback in the form of praise or social comparison increased intrinsic motivation; users that did not receive either type of feedback showed the lowest levels of motivation. In our study, we examine whether social comparisons can engage learners in completing expertise-building activities in an enterprise setting. We hypothesize that even in the presence of potentially inaccurate assessments, users that see social comparisons will have a higher level of perceived utility and engagement with the system (H1). In addition, due to the evaluative nature of the assessments, we hypothesize that users in a corporate environment will not be willing to share their expertise scores publicly when they are below average (H2).

EXPERTISE-BUILDING SYSTEM STUDY

We performed a controlled laboratory experiment with 60 knowledge workers to address questions about how social comparisons and expertise score representations affect the utility of and engagement with our expertise system.

Experimental Design

The independent variables in our study were representational precision and social comparison. To explore precision, we either rendered the expertise score as a point on a line scale or as a bar spanning a range on a line scale (Figs. 1 & 2). The scale was anchored at the endpoints by the labels "Novice" and "Expert." To explore social comparison, we showed either no comparison, a positive comparison (participant is 10% above average), or a negative comparison (participant is 10% below average). Our experiment used a 3 x 2 factorial design with precision and social comparison as between-subjects factors.

Participants

The 60 participants were randomly assigned to a condition and were distributed evenly across the conditions. Participants were employees at a corporate research laboratory and consisted of researchers, software engineers, summer interns, and managers. Participants' ages ranged from the mid-20s to over 65 years of age. Sixteen (27%) participants were female.

Procedure

Our study took approximately 45 minutes. Participants were informed that the system created a personalized learning environment that would enable them to enhance their expertise in the area of Social Business. We chose this area as it is a new, rapidly growing field in which our participants might not yet have built much expertise or published any work or joined a discussion on the social collaboration tool we monitored. This aspect is important as it enabled us to capture a broad distribution of prior expertise.

Our study consisted of four phases. First, participants logged into the system using their intranet ID and reviewed their expertise profile, which contained information about their activities in an internal collaboration tool, as well as papers and patents they published.

Next, participants were presented with an expertise assessment in the area of Social Business. This assessment was held constant across participants to control for the confounding effects the actual score (and not its presentation) might have on motivation, and to enable us to make controlled social comparisons. Thus, all participants in the point condition saw a point centered at 35% (Figure 1) and all participants in the range condition saw a bar covering 30-40% (Figure 2).

Using a fixed, low score provided several benefits over using the actual expertise assessments produced by our system. Because the score was on the lower end of the scale, we felt it provided participants with the same sense of opportunity for improvement. Using a score that may not necessarily match the expectations of participants added a degree of realism, as the real expertise evaluation system could also present inaccurate evaluations due to incomplete information or inaccurate expertise computation.

After seeing the expertise score, participants were shown a reading list consisting of 15 learning items (research papers, news articles, web sites) to read. Participants were told that this list was personalized for them; however these items were identical across participants to observe the effects of presentation of the expertise score. Participants were asked to identify at least three items they felt would improve their expertise in the area of Social Business. Participants were told that they could take as much time as they wanted. After this last task, participants filled out a survey.

Measures

To assess the perceived utility of the expertise system, we asked participants to rate the degree to which they felt the system would be useful and the likelihood that they would use it or recommend it to others. These three items were measured on a 5-point scale (1=Not at all, 2=Slightly, 3=Moderately, 4=Very, 5=Extremely/Completely). This scale was highly reliable (Cronbach's $\alpha = .90$), and all items loaded on one factor in a confirmatory factor analysis. In addition, we asked three questions focused on the utility of the recommended reading list. These questions asked the degree to which participants felt the items increased their expertise, were interesting, and were relevant to Social

Measure	Point (N=30)	Range (N=30)	df	F	p	η^2
Trust	3.30 (1.34)	2.77 (1.30)	1,46	3.18	**.08**	.004
Accuracy	3.60 (1.59)	2.67 (1.18)	1,45	2.73	.10	.006
	No SC (N=20)	**SC (N=40)**	**df**	**F**	**p**	**η^2**
Utility						
Full System	2.72 (.80)	3.17 (.94)	1,56	3.40	**.07**	.005
Reading List	3.22 (.77)	3.51 (.64)	1,56	4.55	**.04**	.003

Table 1. Summary of results. Cell values are shown as M (SD). SC=Social Comparison.

Business. These items were measured on the same 5-point scale. This scale was reliable (Cronbach's $\alpha = .74$). In a confirmatory factor analysis, all items loaded on one factor.

Finally, in order to gauge the effectiveness of reporting a fixed expertise score to participants, we assessed the degrees to which participants trusted the score and felt it was accurate on a 5-point scale. Although these items were significantly correlated ($r = .47$, $p < .01$), this scale was of questionable reliability (Cronbach's $\alpha = .63$); some participants felt that although their profile was incomplete, they trusted the underlying assessment algorithm. Thus, we analyze these measures separately.

RESULTS

We analyzed the data using correlations and analyses of variance (ANOVA). The ANOVA models used precision (point, range) and social comparison (present, absent) as explanatory variables. There were no significant differences between the responses from users that saw positive or negative social comparisons, so we analyze these groups together. We report effect sizes as $\eta^2=SS_{between}/SS_{total}$ [6]. A summary of our significant results is shown in Table 1.

Trust in and Perceived Accuracy of the Assessment

Participants self-assessed their expertise in Social Business on a scale of 'No knowledge,' 'Passing knowledge,' 'Knowledgeable,' and 'Expert.' There were no significant differences in these assessments across conditions. Most of the participants rated themselves as having passing knowledge (31/60) or being knowledgeable (25/60).

Participants' ratings of trust and accuracy of the expertise score fell below the midpoint of the scale (trust M (SD) = 2.63 (.89); accuracy M (SD) = 2.58 (.70)). Participants with the point presentation tended to trust the score more than participants with the range presentation, $p = .08$. In addition, participants with the point presentation tended to feel the score was more accurate than participants with the range presentation, $p = .10$. Social comparison did not have an effect on ratings of trust or accuracy.

We observed a significant negative correlation between participants' expertise self-assessments and their ratings of the accuracy of the system's expertise assessment ($r = -.419$, $p < .01$). Thus, as expected, participants who rated themselves as expert felt the assessment was inaccurate.

Utility

Expertise System

Overall, participants felt that the system had moderate utility (M (SD) = 3.02 (.91)). Participants with social comparison tended to feel that the system provided greater utility than participants without social comparison, $p = .07$, $\eta^2 = .005$ (partially supporting H1). There was no main effect of precision on utility.

Reading List

Participants felt that the reading list feature had moderate utility (M (SD)=3.41 (.69)). Participants with social comparison felt that the reading list provided greater utility than participants without social comparison, $p = .04$. There was no main effect of precision on reading list utility.

Participants spent an average of 11.6 minutes (SD = 8.2 minutes) reviewing the reading list and they identified an average of 4.2 items (SD = 1.8 items) as being helpful.

Precision did not have an effect on the amount of time participants spent reviewing the reading list, nor on the number of items identified as helpful. Similarly, social comparison did not have an effect on these measures either. However, participants with social comparison tended to find more items helpful than participants without it, $p = .10$, eventhough they all saw the same reading list.

Willingness to Share Expertise Scores

We asked participants to report on their willingness to share their score with other people in their organization under several conditions (Table 2).

In general, there were no significant differences in participants' responses across the different precision and social comparison conditions. Rather, willingness to share scores was strongly associated with whether a colleague asked them to share their score, or with whether a culture of sharing existed. In support of H2, participants were more willing to share their score if it was above average (23/54) than below average (12/55), $t(53) = -3.68$, $p < .001$.

Thirty-eight participants provided qualitative feedback on their willingness (or unwillingness) to share their expertise scores. For some participants, sharing a potentially inaccurate expertise score would not be acceptable in a workplace. *"The system has to be accurate. That is critical. Otherwise it can raise lots of issues both legal and monetary. It is very hard to be accurate without personal information or private communications. If you know you are inaccurate, don't use scores."(P32)*

Some participants demanded to know more about the algorithm used for calculating the expertise score and to have a clear explanation of how to improve their scores. *"[my] answer is conditional upon some verification of system accuracy and guidance on how to improve the score. In other words, customization and the configurability of the system is utmost important. I would check 'not visible to anyone' to all questions if I do not trust the system. On the other hand, if I trust I make my info visible to all." (P11)*

Several other participants expressed concern from making scores public because they felt it would either embarrass them or make them the target for questions. *"There are two problems with its general release. (1) if I am below par I feel embarrassed, and (2) if it is above par I may not want to be hounded by advice seekers." (P43).* To protect against being hounded by others, several participants expressed a desire to be able to specifically designate who could see their expertise score.

Contrasting these opinions, 13 participants (23%) would make their scores public under any condition. *"I don't have a problem with people knowing my expertise score." (P5)*

There were three main points raised by the participants in their response to this question: (i) the implications of inaccurate assessments on their performance reviews; (ii) the need to know details of back-end expertise assessment analytics in order to have an assurance of their accuracy, as well as a clear definition of how to improve scores; and (iii) the concern of opening themselves to a flood of questions.

DESIGN IMPLICATIONS AND CONCLUSIONS

Knowledge workers have a strong need to continually increase their knowledge and expertise in order to remain capable in their professions and competitive with their peers. This fact is emphasized by the large number of participants who indicated that they read professional material online either 'more than once a day' or 'almost every day' (50/59). But perhaps this sentiment was best summarized by P31, who said: *"no job is permanent. As the business moves we all need to improve skills."*

In creating our expertise-building system, we have uncovered a multitude of challenges that could cause it to not be adopted and used, such as missing data, inaccurate expertise assessments, or irrelevant content recommended to the learners. To mitigate some of these effects, we have examined how the presentation of expertise assessments affects perceptions of utility, accuracy, and trust.

We have found some evidence that the presence of social comparisons, either upward or downward, increased ratings of

utility and engagement with the learning system (H1 generally supported). In addition, more users were willing to share their expertise scores publicly when they were above average (H2 supported).

Overall, we were surprised to find that many participants would opt for sharing their scores in the enterprise under some condition (e.g. when it was above average or if others generally shared their scores) given the large literature on the negative effects of competitive stress on organizational behavior [9]. The points participants raised suggest that

	Not visible	**Visible to people of my choosing**	**Publicly visible**
No matter what the score is	23 (40%)	21 (37%)	13 (23%)
Score < Average	24 (44%)	19 (34%)	12 (22%)
Score = Average	17 (31%)	23 (42%)	15 (27%)
Score > Average	12 (22%)	19 (35%)	23 (43%)
If a colleague asked	8 (15%)	32 (58%)	15 (27%)
If others share	9 (16%)	18 (33%)	28 (51%)

Table 2. Expertise score sharing preferences. Cell values show the people that would present their score in the given format.

users would prefer to be able to control when their assessments can be shared. Our results motivate the inclusion of sharing and comparison features in enterprise expertise-building systems.

REFERENCES

1. Buunk, B. P., Collins, R. L., Taylor, S. E., VanYperen N. W., & Dakof, G. A. (1990). The affective consequences of social comparison: either direction has its ups and downs. *JPSP*, 59 *(6)*, 1238-1249.
2. Ehrlich K., Lin C., and Griffiths-Fisher V. 2007. Searching for experts in the enterprise: combining text and social network analysis. (GROUP '07).
3. Festinger, L. (1954) A theory of social comparison processes. *Human Relations*, 7.
4. Gray, B. (2004). Informal learning in an online community of practice. *Journal of Distance Education*, 19 *(1)*, 20–35.
5. Lai J., Lu J., Pan S., Soroker D., Topkara M., Weisz J., Boston J., and Crawford J. Expediting expertise: supporting informal social learning in the enterprise. IUI '14.
6. Levine, T. R. and Hullet, C. R. (2002). Eta squared, partial eta squared, and misreporting of effect size in communication research. *HCR*, 28 *(4)*, 612–625.
7. Matthews, T., Whittaker, S., Badenes, H., Smith, B. A., Muller, M., Ehrlich, K., Zhou, M. and Lau, T. (2013). Community Insights: Helping Community Leaders Enhance the Value of Enterprise Online Communities, CHI 2013.
8. Mumm, J. and Mutlu, B. (2011) Designing motivational agents: the role of praise, social comparison, and embodiment in computer feedback. *Computers in Human Behavior*, 27 *(5)*, 1643-1650.
9. Tsai, W. (2002). Social structure of 'coopetition' within a multiunit organization: Coordination, competition, and intraorganizational knowledge sharing. Organization Science, 13(2), 179-190.