





















That a material ended up being opened did not, naturally enough, mean that it was carefully read. Opening the material was only a step in the process. The next step was to glance at the material, to decide if it really was worth reading. If the article passed this impression-forming glance and the students started to read it, the judging process continued: *“I only bothered to read the article to the end if it gave me some new information or a new viewpoint, i.e. I felt it to be useful.”*

### 5.3 Evaluating materials

In 2012, evaluating a material involved an in-depth evaluation (title, star-rating, and text justification) while in 2011, compulsory evaluating consisted of simply giving a star rating while commenting was voluntary. Students viewed the possibility to evaluate materials positively, on average at 5.55 (SD=1.15) in 2012 and at 5.11 (SD=1.59) in 2011. The difference is not statistically significant; increasing the cost of evaluating did not reduce the positivity of student views. As reasons for liking evaluations, students mentioned social aspects and possibility for expressing opinions in addition to evaluations helping in selecting materials for reading.

Many students especially liked the text part of the evaluation: *“...the possibility of textual evaluation was very good, and it was also good that you could comment evaluations! It even led to some discussing.”* However, one student mentioned a negative aspect concerning the text comment: *“You saw what others had praised so you read it with interest, too, but I myself didn’t like to evaluate. Especially since I’m not a professional or somebody who knows a lot, so my comments may have appeared pretty bad for somebody who knew more.”* Two interviewed students touched the same theme, noting that it was easier to comment in LSRM than on the Internet because there are so much more knowledgeable people on the Internet. For them, a smaller community with the members more or less at the same level of knowledge made it easier to comment. Consequently, a small, closed community of peers can create a safer environment to encourage participation. This finding is in line with the idea that the sense of community is connected to a feeling of *membership* that includes boundaries that provide members with emotional safety [21].

Star-ratings were seen in a more problematic light: *“Comments enrich and give new viewpoints, stimulate discussion. Star-ratings I found somewhat unnecessary.”* One reason for disliking star-ratings was the difficulty in deciding the appropriate ratings: *“Occasionally it was hard to give stars because even if the material was really useful for me, it’s not necessarily that for everybody so you don’t want to rate it too highly, either.”* The problem was exacerbated by the fact that students were not told which specific aspect to rate. This was a clear design mistake; the interface should have made the rated aspect clear.

Although students in 2012 viewed on average slightly over 2.5 times more materials than in 2011, the number of evaluations on average was statistically the same in 2012 (M=4.14, SD=2.09) as in 2011 (M=3.89, SD=2.75). Also, on average materials had statistically as many ratings/evaluations in 2012 (M=1.45, SD=1.54) as in 2011 (M=1.8, SD=1.96). In effect, in 2012, students therefore read more materials but evaluated fewer in relation to the number read than in 2011. However, significantly,

on average students made more *honest* evaluations, i.e. viewed the material before evaluating it, in 2012 ( $M=4.14$ ,  $SD=2.09$ ) than in 2011 ( $2.86$ ,  $SD=2.65$ ). The difference is statistically significant,  $t(71)=2.28$ ,  $p=.026$ .

In fact, in 2012, not a single dishonest evaluation was made: The use log data shows unequivocally that on each occasion, the student had opened the link before evaluating it. Increasing social presence and the cost and complexity of evaluations removed dishonest attempts to get points without earning them entirely. Also, since the number of evaluations per student did not decrease, increasing the cost of evaluating did not reduce the number of evaluations. In effect, more students were motivated to do the required work when the 2012 design was used.

Not only did increasing cost and complexity of evaluations in comparison to ratings result in complete honesty but it also resulted in perceptions of honesty. Not a single student mentioned suspecting dishonesty in the 2012 feedback while such suspicion was entertained in the 2011 feedback (when dishonest rating in fact took place). While social presence likely played a significant role in this, too, given that changing the rating scale from binary (2010) to 5-star (2011) reduced dishonest rating almost by half (from 43% in 2010 to 26% in 2011), we conjecture that needing to write a textual justification for the rating made cheating simply too difficult: “*You had to read the materials to be able to evaluate it.*” Since student perceived others as doing the required work, they also ended up reciprocating by doing their own share.

The care with which materials were read before they were evaluated increased clearly. When we look at the time periods that passed between opening the link and adding the evaluation (when evaluating took place within the same session, as it typically did) for honest ratings/evaluations, we notice that the reading times almost doubled in 2012. If we examine the reading times that were shorter than 15 minutes (to filter out sessions that may have included other activities), there is still a clear statistical difference between 2012 ( $M=409$  seconds,  $SD=242.33$ ) and 2011 ( $M=231$  seconds,  $SD=203.94$ ),  $t(164)=5.133$ ,  $p < .001$ . The materials were clearly read longer in 2012 before they were evaluated. In fact, in 2012, only in two cases (1.8%) did a student evaluate a material after reading it for less than a full minute while in 2011, there were 21 (23.3%) such cases. Still, having to write a textual evaluation in addition to clicking a rating must also have been a partial reason for the increased time between open a link and evaluating it.

In 2012, only four students (21%) said that they had rated all the materials they had read while in 2011, eight students (42%) said the same. In both years, the main reason for not rating a viewed material was the same; students felt that they had not read the material carefully enough to rate it.

While in 2011 some students did refer to social factors as a rationale for not evaluating, it was in 2012 that social aspects were mentioned repeatedly in this context. Sociality inherent to the system clearly affected evaluating behavior: “*There were social aspects to evaluating, so I did not want to write an evaluation that just said ‘nice one’ or ‘interesting article’ but something more. For this reason I wanted to evaluate only articles on which I had a clear opinion and something a bit deeper to say—something that might inspire others to comments and something that others could comment.*” Besides a certain social pressure, there also was a sense of moral

duty towards others: “...I would have felt wrong about evaluating a material that I had not read entirely.” In effect, as with material additions, certain altruistic motivations were evident in many student comments concerning evaluations.

Also, a few students mentioned that they did not evaluate some materials they had read “because I had nothing new/significant to add to the comments by others.” While no student in 2011 mentioned thinking twice about rating a rated item, coupling ratings and comments made students feel that they had to have something significant to say, something that had not already been said: “Somebody else might have already said what was essential in his or her comment.”

Another reason for students not to evaluate a material they had read in 2012 was that “the materials had not aroused any big emotion.” Mediocre, bland articles simply did not garner evaluations: “I evaluated materials based on whether they stirred up thoughts or not. I selected for evaluating only materials on which I had some kind of an opinion. In the evaluating phase I simply skipped lackluster articles altogether.” If students did not have something to say about the material, they did not evaluate it.

Consequently, students read articles based on personal interests and need. If they read the whole article and felt that it was “useful” and “interesting” and they had something to say about it (that somebody else had not already said), they probably evaluated the material. As a result, students largely ended up evaluating good articles: “I didn’t really bother to read materials that I found worthless with the first glance, so I ended up choosing for evaluation only good materials.”

Usefulness was the most important criterion for students when they evaluated materials. However, how clearly written and presented and how illustrative the article was also affected the evaluation. Moreover, students appreciated learning something new from the material. “Usefulness and practicality, can I use the material in future in studies and at work. Also if I learned something new and if the materials was relevant to the course and its content.”

The above factors largely explain why over 50% of the star-ratings were four stars and 70% 4–5 stars in 2012. There is no statistical difference between the average star-ratings in 2012 ( $M=3.82$ ,  $SD=.83$ ) and 2011 ( $M=3.70$ ,  $SD=.86$ ),  $t(294)=1.17$ ,  $p=.244$  (Table 2).

**Table 2.** Distributions of star ratings in 2012 and 2011

	<b>1 star</b>	<b>2 stars</b>	<b>3 stars</b>	<b>4 stars</b>	<b>5 stars</b>
<b>2011</b>	0 (0%)	13 (9%)	42 (29%)	64 (44%)	25 (17%)
<b>2012</b>	4 (3%)	1 (1%)	41 (27%)	79 (52%)	27 (18%)

In 2012, three students gave 1-star evaluations (two once and one twice). Two materials ended up having one star as the average of its ratings (both had two 1-star evaluations). Based on the evaluation texts, it appears that the students giving the 1-star evaluations felt that the materials should not have been added to the system. Interestingly, it was only when nicknames were used that 1-star ratings were made; in 2011, no 1-star ratings were made. We conjecture that a heightened sense of social presence/sociality resulted in people showing disapproval for substandard materials.

Most appeared to have communal, even altruistic motivations, and they probably expected the same in return from the other members of the community. Thus, while students mostly ended up selecting good materials to evaluate, when the experience was strongly negative, they were ready to *vent negative feeling* [17], even if the comments connected to 1-star evaluations were still quite polite and matter-of-fact in tone.

Several student comments underline that students were aware of the benefit they got from reading materials and that they understood that evaluating materials did make them to read them more carefully: *“evaluating materials involved reading a lot of materials when 5 evaluations were required. 5 felt a lot but afterward I felt it was useful that it made me read so many articles.”* Students were also clearly aware of the information literacy benefits that evaluating materials and reading evaluations by others brought: *“Evaluating materials increased the teaching value of the articles. By reading the evaluations by others we got good feedback on how to apply scientific texts in studying. Also, finding and reading articles in our field is very important, especially for working life.”* This contrast with the 2011 comments where students focused on the ability of rating to guide them to better materials and to warn them against bad ones but did not discuss much the benefits of reading and evaluating.

Another social aspect of evaluations was curiosity of how others evaluated the materials one had added: *“Of course I checked out what kinds of evaluations the materials I had added had gotten.”* One interviewed student in fact mentioned she had just before coming to the interview (interviews took place after the course) checked if there had been any new evaluations on her materials. Another interviewed student said that the system gave him a feeling that the materials he had added were useful to others and that one of them had had *“a fair amount”* of evaluations. What others said clearly mattered to and interested students.

#### **5.4 Perceived social presence and its impact**

While contributions had previously been anonymous, in 2012 students were asked to choose a nickname for LSRM when registering to the course. The purpose was to increase perceived social presence and to allow reputation formation by giving students individual presence in the system. In effect, there was a significant effect for social presence,  $t(36)=2.06$ ,  $p=.047$ , with students reporting on average higher social presence in 2012 ( $M=3.53$ ,  $SD=1.22$ ) than in 2011 ( $M=2.75$ ,  $SD=1.15$ ).

Given that we were using a 7-point scale, average perceived social presence of 3.53 might not seem very high. However, LSRM is in a sense competing against such popular social systems as Facebook and instant messaging when it comes to how social students perceive it. In that sense, the seemingly low average of 3.53 may in fact actually indicate a relatively high perceived social presence for a system that does not support real-time presence. In fact, some comments from students who professed not to have felt others as present emphasized the slow rhythm of interaction: *“I didn’t find much [social presence] because there wasn’t that much activity and I didn’t get announcements of e.g. that my materials had been evaluated etc. The thought of there being other students affected so that I wondered what others thought about my evaluations....”* One interviewed student encapsulated the general perception by saying that

the feeling of sociality was clear but not very strong and that the level was appropriate to the system, as it was about as much as can be archived “unnaturally.”

Student comments indicate that the perceived social presence had a clear and positive impact on behavior: “*Sociality in the service affected my actions significantly: It affected so that I wanted to select as suitable articles as possible to add to the page and that I wanted to say something more deep than just ‘quite nice’ in the evaluation.*” Repeatedly, students mention having tried to find materials that would be useful to others and making evaluations that would help others. While there were similar trends also in 2011, altruistic aspects are more emphasized in the 2012 comments.

Some comments clearly connected nicknames to reputation: “*When adding materials I thought that I can’t add just any odd stuff ... because all the other students see them. The fact that my name was connected to the materials and evaluations I added also made me think twice what to say.*” Students appear to have felt that through the nickname they had an individual presence and reputation in the system, and that affected their behavior positively. Having an individual presence in the systems also made students consider their self-image in relation to the community, as one interviewed student explained: “*Also, building my self-image influenced it; I didn’t feel it satisfactory for myself to put there something that I wouldn’t want to read myself.*” The student comments indicate that most wanted to be responsible members of the community, and in this sense, the achieved social presence was high enough.

While some students professed not having felt the presence of others, their comments show that their actions were nevertheless influenced by awareness of others, e.g. “*I didn’t really feel others to be present that much. Still, the thought that others see what I add there affected what materials I added and what kind of evaluations I made. I.e., I did my job with care.*” The impact of this awareness on student behavior appears to have been larger than the numeric evaluation of social presence indicates.

Social presence had both activating and experience-enhancing impact: “*...the presence of others there activated me, too, and it was great to see that others actually read and evaluated materials.*” In particular, student comments underline that perceiving others as present improved the user experience, e.g. “*The presence of other students affected positively because [that way] you knew that somebody else is also reading these comments and not just the teacher alone.*”

The positive effects of increased social presence appear at least partially attributable to increased social pressure that drove students to do more than just bare minimum for earning points: “*[I read the materials I added] very carefully indeed exactly because of the sociality connected to evaluating materials. There was some ‘social pressure’ involved in evaluating articles because other students could read your evaluations, respond to them, disagree and comment, respond to the evaluation...*”

The effects of social presence are likely intertwined with the effects that the emerging sense of community had. Besides membership, using LSRM also had many other elements that McMillan and Chives [21] suggest as contributing to the sense of community, including *personal investment* (added evaluations and materials), *bidirectional influence* (students affecting the community and vice versa) and *integration and fulfillment of needs*.

## 5.5 Social presence of instructors and its impact

Students felt that the lecturer's presence had a positive impact in two ways. First, he maintained a feeling of activity in the system, e.g. "*A lot of added materials and evaluations on materials [by the lecturer]. I feel that it was good that the lecturer kept the page active when it occasionally got silent.*" Second, the lecturer's presence brought positive social pressure: "*The presence of the lecturer did encourage investing in the materials. I didn't have the nerve to add just any old dude's blog there and added instead content by recognized sites or known experts.*"

Some students mentioned having been nervous about evaluating materials added by the lecturer. In fact, student-added materials ( $M=1.88$ ,  $SD=1.54$ ) did get on average more evaluations than instructor-added materials ( $M=.61$ ,  $SD=1.18$ ),  $t(103)=4.34$ ,  $p < .001$ . Interestingly, there was no statistically significant difference between the average star ratings given to instructor-added materials ( $M=3.95$ ,  $SD=.72$ ) and student-added materials ( $M=3.79$ ,  $SD=.84$ ). The scale, in a sense, was the same, meaning that the 16 students who evaluated instructor-added materials (in contrast, 30 students evaluated student-added materials) did not give the materials special treatment.

Comments and evaluations from the lecturer were warmly welcomed: "*I especially liked how the lecturer commented on some evaluations and gave his own examples (e.g. on a material that I had added).*" The evaluations and comments that the instructors added were mainly positive or, in one case, only mildly challenging.

## 6 Discussion

**Using nicknames increases sociality and trust on materials and evaluations.** Using nicknames clearly increased social presence, as evidenced by the statistically significant increase in student evaluations and student questionnaire replies. Even the students who did not feel others as present described how the idea of other users affected their behavior positively. Many students reported altruistic motivations, and their comments show that they felt a certain sense of duty towards others. Also, students had a sense of individual presence in the system, which created social pressure that also affected their actions positively. Overall, students perceived others as taking online activity seriously, and this motivated them to approach it with due diligence.

**Increasing evaluation cost engenders trust on evaluations.** Students clearly trusted better that student evaluations were properly made than in 2011. How much of this is attributable to nicknames and how much to increased evaluating cost is an open question, but considering that in 2011, increasing the rating cost resulted in a significant improvement in honesty, we conjecture that coupling commenting and rating affected it significantly and also encouraged reading materials more carefully.

**Requiring more thorough evaluations increases pressure to add good materials.** There was less questioning of the motives of the students adding materials in 2012, indicating that at least some aspects of the perceived quality had improved. Also, many student comments show that students did approach finding materials to add very seriously, in part because they knew that they would be evaluated. We conclude that the more complex evaluations made students more careful about the links

they added but that the effect is again intertwined with the effects of the nickname use.

Overall, our working hypothesis concerning nicknames and evaluation cost worked out well. Using nicknames and a more complex evaluating approach removed dishonesty entirely from evaluations. Importantly, this was accomplished without the number of contributions falling; in fact, the number of honest contributions increased. This is a significant improvement to the system and gives other e-learning practitioners practical tools and approaches to root out dishonesty

### **6.1 A word of caution: Use and interface design are intertwined**

When applying our results in e-learning, and particularly in other contexts, it should be noted that our results are subject to specific contextual factors. Our system is designed for formal e-learning where compulsoriness can be used to encourage contributions, students form a small, closed community of peers, and the use period is short. In informal e-learning, for instance, using compulsoriness may not be possible or even advisable. Also, designing an e-learning space must go hand in hand with designing its use (compulsoriness, regulations etc.), as the two are intertwined. For example, our system used compulsoriness and high evaluation cost to improve evaluation/rating honesty. However, if there had been no compulsoriness, this design would likely have failed. With voluntary evaluating, it would have been advisable to lower the evaluation cost to encourage contributing; after all, there would have been little motivation for dishonesty. In 2011, while there were dishonest ratings (compulsory), there was no dishonest commenting (voluntary).

### **6.2 Enhanced user experience**

The 2012 use and interface designs led to students approaching their work more honestly and with more gusto. The space was significantly more social due to the tools bringing sociality and providing social texture and the nicknames providing sociality through individual presence. Student trust on materials and especially evaluations increased; students saw others as doing their work properly, which led to altruistic motivations and a sense of duty towards others, leading to deeper involvement. In effect, this perception was justified by actual changes in honesty and due diligence, as most students did work hard to add links that were meaningful and tried to make evaluations that would be useful. Also, in 2012, many students were more aware of the benefits—in particular, improved learning and information literacy skills—they accrued from using the system with due diligence. Seeing benefits in turn encouraged using the system, leading to a virtuous circle: Positive behavior led to positive experience and perception that in turn encouraged positive behavior.

### **6.3 How to develop LSRM further**

Further increases in evaluation cost do not appear necessary, as 100% honesty was already reached with the current approach. Nevertheless, the aspect to be rated needs



to be made clear in the interface so that all students are rating the same aspect. The most important criterion for students, *usefulness* of the materials, is the obvious candidate. While increasing the cost of adding materials by requiring a description might seem a logical step to induce further trust on materials, it may prove problematic; in an interview, one student said that this would be a “*miserable feature*” that would only lead to marketing one’s materials instead of objectively describing them.

In fact, the most promising approach to encourage positive behavior and improving the user experience further appears to be enhancing sociality and individual presence in the system. First, the system needs to be more connected to students’ everyday lives. LSRM should make it possible for students to subscribe to email notices so that they can maintain awareness of any development in the system that concerns them. In addition, the system should incorporate a private group in Twitter and Facebook, to mention two obvious candidates. This way, the information about new materials and evaluations would reach students without them having to log in the system. Making groups private is important, as the community being small, closed, and consisting of peers (same level of knowledge) were important factors for students.

The second approach to increasing sociality is to enhance the sense of individual presence by allowing viewing material additions, evaluations and comments by individual students. This would also increase social pressure, as one could not hope to be hidden in the mass of materials. Interviews gave indications that students would not find this intrusive. In all likelihood, this would further increase the care with which students add and evaluate materials.

## 7 Conclusion

We managed to enhance students’ user experience and remove dishonesty from additional reading material evaluations by increasing the evaluation cost and by replacing anonymity with nicknames, thus giving students an individual presence. This study also contributes to the field by providing much-needed experiences of using recommender features in e-learning in a genuine use context.

However, when applying our results in other systems and contexts, one should bear in mind that the results were obtained within formal e-learning context and that, consequently, contextual factors may limit their applicability elsewhere.

While we hope to develop LSRM further to the directions outlined here, we also encourage other practitioners to report their experiences of RS in various educational contexts. It is important that practice and theory go hand in hand in employing RS in e-learning instead of theories being developed independent of the ground realities.

## References

1. Tang, T., McCalla, G.: Beyond Learners' Interest: Personalized Paper Recommendation Based on Their Pedagogical Features for an E-learning System. In: LNCS. Vol. 3157, pp. 301--310. Springer, Heidelberg (2004).

2. Hage, H., Aïmeur, E.: Harnessing Learner's Collective Intelligence: A Web 2.0 Approach to E-learning. In: B.P. Woolf, E. Aïmeur, R. Nkambou, Lajoie (Eds.). LNCS, vol. 5091, pp. 438--447. Springer, Heidelberg (2008)
3. Kiliç-Çakmak, E.: Learning strategies and motivational factors predicting information literacy self-efficacy of e-learners. *AJET* 26(2), 192--208 (2010)
4. Schafer, J. B., Konstan, J., Riedl, J.: Recommender Systems in E-commerce. In: Proc. the 1<sup>st</sup> ACM Conf. on Electronic commerce, pp. 158--166. ACM, New York (1999)
5. Rashid, A.M., Ling, K., Tassone, R., Resnick, P., Kraut, R., Riedl, J.: Motivating Participation by Displaying the Value of Contribution. In: Proc. the SIGCHI conf. on Human Factors in computing systems, pp. 955--958. ACM, New York (2006)
6. Leino, J.: Case Study: Material Additions, Ratings, and Comments in a Course Setting. In: O.C. Santos, J.G. Boticario (Eds.), *Educational Recommender Systems and Technologies: Practices and Challenges*, IGI Global, pp. 258--280 (2011)
7. Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H. Koper, R.; Recommender Systems in Technology Enhanced Learning. In: P. Kantor, F. Ricci, L. Rokach, B. Shapira (Eds.), *Recommender Systems Handbook*, pp. 387--415. Springer, Heidelberg (2011)
8. Santos, O.C., Boticario, J.G.: Modeling Recommendations for the Educational Domain. *Procedia Computer Science* 1(2), pp. 2793--2800 (2010)
9. Ghauth, K.I., Abdullah, N.A.: Learning Materials Recommendation Using Good Learners Ratings and Content-based Filtering. *Journal of Educational Technology Research and Development* 58(6), 711--727 (2010)
10. Drachsler, H., Hummel, H.G.K., Koper, R.: Identifying the Goal, User model and Conditions of Recommender Systems for Formal and Informal Learning. *Journal of Digital Information* 10(2), 4--24 (2009)
11. Pu, P., Chen, L.: Trust-inspiring Explanation Interfaces for Recommender Systems. *Knowledge-Based Systems* 20, 542--556 (2007)
12. Konstan, J., Riedl, J.: Recommender Systems: From Algorithms to User Experience. *User Model User-Adap Inter* 22, 101--123 (2012)
13. Talwar, A., Jurca, R., Faltings, B.: Understanding User Behavior in Online Feedback Reporting. In: Proc. the 8<sup>th</sup> ACM conf. on Electronic commerce, pp. 134--142. ACM, New York (2007)
14. Lam, S.K., Riedl, J.: Shilling Recommender Systems for Fun and Profit. In: Proc. the 13<sup>th</sup> int. Conf. on World Wide Web, pp. 393--402. ACM, New York, (2004)
15. Sparling, E.I., Sen, S.: Rating: How Difficult is It? In: Proc. the 5<sup>th</sup> ACM conf. on Recommender systems, pp. 149--156. ACM, New York (2011)
16. Hu, N., Pavlou, P., Zhang, J.: Can Online Reviews Reveal a Product's True Quality? In: Proc. the 7<sup>th</sup> ACM conf. on Electronic Commerce, pp. 324--330 ACM, New York (2006)
17. Hennig-Thurau, T., Gwinner, K.P., Walsh, G., Gremler, D.D.: Electronic Word-of-Mouth via Consumer-opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet? *Journal of Interactive Marketing* 18(1), 38--52 (2004)
18. Leino, J., Rähkä, K.-J.: User Experiences and Impressions of Recommenders in Complex Information Environments. *IEEE Data Engineering Bulletin* 31(2), 32--39 (2008)
19. Nov, O., Ye, C.: Why Do People Tag? *Communications of ACM* 53(7), 128--131 (2010)
20. McNee, S.M., Riedl, J., Konstan, J.: Being Accurate Is Not Enough: How Accuracy Metrics Have Hurt Recommender Systems. In: CHI '06 Extended Abstracts on Human Factors in Computing Systems, pp. 1097--1101. ACM, New York (2006)
21. McMillan, D.W., Chavis, D.M.: Sense of Community: A Definition and Theory. *Journal of Community Psychology* 14(1), 6--23 (1986)