Mixed-Approach Social Comparison for Improving Online Discussion Efficacy: Insights from Field Experiments

Elahe Javadi
ejavadi@ilstu.edu
School of Information Technology
Illinois State University
Normal, IL 61790-5150

Judith Gebauer
gebauerj@uncw.edu
Department of Information Systems and Operations Management
University of North Carolina Wilmington
Wilmington, NC 28403-5611

Nancy L. Novotny
nlnovot@ilstu.edu
Mennonite College of Nursing
Illinois State University
Normal, IL 61790-5810

Abstract

This manuscript reports results of field experiments that investigated the impact of a mixed-approach social comparison on quality and quantity of student interactions in course-related online discussions. The approach was mixed because firstly students had access to scores of both lower and higher performing peers (upward and downward social comparisons); and secondly students had access not only to their peers' scores but also to the goal-specific informational evaluations associated thereto, i.e., notes on why a score was earned. Student interactions manifested in commenting behavior and were compared across two consecutive online discussions in which students shared their analysis of a topic specified by the course instructor. Quantity and quality of student comments on each other's posts were measured as dependent variables in the experiment; the field experiment included 12 sections over a span of four semesters (total of 24 online discussions). Half of the online discussions involved the use of a mixed-approach social comparison. The mixed approach was effective in advancing quality and in decentralizing commenting networks.

Keywords: Online discussion, social comparison, informational evaluation

1. INTRODUCTION

In traditional course settings (face-to-face classes), online discussion can extend opportunities for critical thinking, self-expression, and peer-learning beyond classroom hours (Waters & Gasson, 2012). To enhance critical thinking and peer-learning, online discussions must encourage interactivity. Interactivity is usually achieved by requiring students to read, analyze, and respond to their peers' ideas. Therefore, interactivity may be measured by quantity, distribution, and quality of the comments exchanged. Quality of comments can be measured by their level of integrative complexity, i.e., the extent to which they have examined different dimensions of the topic.

Evidence from prior research and practice suggests that electronic tools (e.g., group brainstorming tools) can create an illusion of productivity without yielding much gain in terms of quantity or quality of the ideas generated by individuals (Pinsonneault et al., 1999). Ceteris paribus, increased quantity and quality of posts and comments, as well as decentralized interaction networks (who responded to whom) are desirable in educational settings. To achieve the three desirable dimensions of interactive discussion, this research study applied a mixed-approach social comparison as an enabler of interactivity. The mixed approach included elements of both upward and downward social comparison as well as goal-specific informational evaluation. The impact of the mixed approach, created based upon Social Comparison (Festinger, 1954) and Cognitive Evaluation (Deci & Ryan, 1980) theories, was examined through field experimentation in twelve information technology course sections.

2. THEORETICAL BACKGROUND AND RESEARCH MODEL

Effective online discussions are interactive and involve both original ideas and responses thereto. To achieve interactivity in online discussions, underpinning group processes must be strengthened. Prior research on group processes has identified factors that contribute to or hinder productivity in group settings. Examples of enabling factors are cognitive stimulation and observational learning; and examples or obstacles are evaluation apprehension and social loafing (Pinsonnault et al., 1999). Evaluation apprehension occurs when fear of being evaluated hinders contributions or creativity. Social loafing occurs when individuals in in a group underperform and their performance matches that of lowest-performing peer in the group. The current study focuses on these two group productivity obstacles by applying Cognitive Evaluation and Social Comparison theories as theoretical lenses (Figure 1) (Deci & Ryan, 1980; Festinger, 1954).

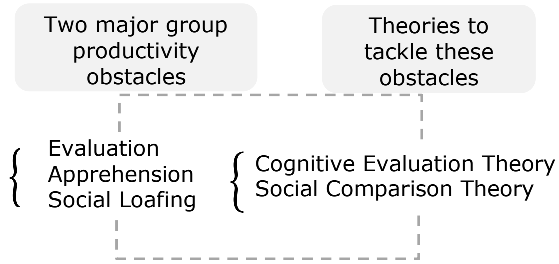

Figure 1: Group productivity obstacles

Prior research posits that the existence of a discrepancy in a group with respect to opinions or abilities will lead to action by the members of that group to reduce the discrepancy (Festinger, 1954). Social comparison can take many forms and can be implemented through mechanisms, such as charts or leaderboards. Upward or downward social comparison happens when individuals are exposed to the process outcomes of higher and lower performing competitors, respectively. Research indicates that social comparison and its saliency influence outcomes in brainstorming and electronic brainstorming systems (Dugosh & Paulus, 2005). Shepherd and colleagues (1996), for instance, examined the impact of social comparison and the saliency of comparison tools on the brainstorming performance in an electronic setting. In their lab experiments, the authors observed a 63% increase in the number of unique ideas generated in the treatment groups which used a highly salient social comparison tool. The 63% gain was compared to only a 22% gain in the low salience social comparison treatment group. Dugosh & Paulus (2004) observed higher productivity, as measured by the number of ideas generated, in social comparison treatment; in their experiments, social comparison was manipulated through instructional sets. In another related study, Michinov & Primois (2005) found that social comparison via the use of a shared table showing the contributions of each member positively influenced productivity and creativity; their experimental design allowed communication among brainstormers through a newsgroup feature. The authors noted that even when the brainstormers could publicize their contributions in the newsgroup, the publicizing did not have the same impact as having a highly salient shared contribution-tracking table, i.e., social comparison mechanism.

Informational Evaluations & Goal-Specificity
Individuals are more likely to generate creative ideas when they are intrinsically motivated (Deci & Ryan, 1980). Intrinsic motivation proves to be higher in experimental groups when individuals expect informational evaluation (Shalley & Perry-

 ISSN: 2473-3857
v4 n4642

Smith, 2001). In scholarly work on teaching and learning, informational evaluation is labeled formative assessment. Research studies on formative assessment suggest that goal specificity is a crucial component of formative evaluation methods (Ambrose et al*.,* 2010). Goal specificity facilitates effectiveness of deliberate practice which leads to expert-level performance (Ericsson & Charness, 1994). Goal specificity for discussions can be achieved by clearly identifying learning goals on which discussion participants are expected to excel and providing feedback that directly assesses the extent to which students have achieved said goals. Therefore, goal specificity provides a focus for participant's efforts. Goal specificity can be included in assignment instructions and feedback, for example by providing concrete examples of successful performances. This study operationalizes mixed-approach social comparison based on three elements, namely (1) goal specific instructions, (2) goal-specific feedback on individual as well as peer performances, and (3) concrete examples of successful and unsuccessful performances by sharing scores and feedback on the contributions of all peers.

Online Discussions' Efficacy

This study uses levels of participation, integrative quality of discussion posts, and the dynamic of interactions among participants as measures of online discussion efficacy. While each student was expected to submit one initial post and four subsequent comments, variations were observed in the levels of students' activities (whether or not they posted an original idea or four comments) and their choices of where to post their comments (in response to whose posts).

In the brainstorming and online discussions literature, most experimental studies have focused on individual idea-sharing behavior in electronic settings (e.g., Wasko & Faraj, 2005). Comparatively little research has been done to examine the extent to which individuals build on the ideas shared by others. This study measures integrative quality of the posts, i.e., the extent to which discussion participants take into account and analyze different dimensions of the topic discussed. An idea is defined as a basic element of thought that consists of at least one testable proposition (Simon, 1947). We conceptualize and measure integrative quality of the posts based on the well-studied concept of integrative complexity in social psychology (Baker-Brown et al., 1992; Suedfeld et al., 1992). More details on the measurements are shared in the section on field experiments.

Mixed-Approach Social Comparison

The mixed-approach social comparison in this study was operationalized by allowing and even encouraging discussion participants to view both controlling and informational evaluation that their peers received on the discussion posts. Controlling (summative) evaluations focus on the outcome whereas informational (formative) evaluations provide information on how to improve said outcome. Viewing other students' scores and comments associated with those scores, implies exposure to both lower performing and higher performing peers, thus yielding a mixed upward/downward social comparison. According to Cognitive Evaluation Theory, individuals are more likely to generate creative ideas when they are intrinsically motivated (Deci & Ryan, 1980); and this study proposed that intrinsic motivation can be higher in experimental groups in which individuals view and process informational evaluation associated with their scores and those of others (Shalley & Perry-Smith, 2001). As summarized in Figure 2, we propose:

Proposition: The presence of mixed-approach social comparison is associated with higher quality of integrative ideas.

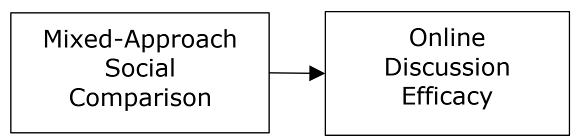

Figure 2: Research model

4. FIELD EXPERIMENTS

The field experiments involved twelve course sections, with three sections each taught during four semesters. Each course section included two discussions, i.e., twenty-four discussions total. Half of the course sections were used as control groups (C) and the other half as treatment groups (T). Table 1 indicates the sample sizes for each section.

Table 1: Control and treatment group sizes

Condition	Semester	Section sizes	Total Sample
Control	Fall 2014	30, 22, 21	139
	Spring 2015	24, 20, 21	
Treatment	Fall 2015	30, 25, 18	136
	Spring 2016	30, 22, 11	

In the control sections, after the first discussion, students were given their individual scores, and were reminded of the general scoring rubric. In the treatment sections, students were given goal-specific instructions. Goal specific instructions were posted on the course's learning management system and were reiterated in the class by the instructor. An excerpt from the instruction is included below:

"… *your goal is to generate synthetic ideas. It is vitally important for the purpose of this assignment that you generate ideas that synthesize your ideas and those that you read. I expect you to prepare analyses that combine your ideas with ideas presented in the articles that I listed or other articles that you read during your independent research. Your posts will be will be carefully reviewed for their SAD (systems analysis and design) content and synthetic quality*…"

Also, after the first discussions, students were given an annotated transcript of the whole discussion which contained each student's discussion score along with the instructor's goal-specific feedback associated thereto. To alleviate privacy concerns, students' names were removed from the transcripts; and at the time discussion transcripts with feedback were released to students, online discussion forums were closed for viewing. Both instructions and informational evaluation for the treatment groups were goal-specific, in that students were clearly instructed to focus on integrating ideas and were given feedback on the annotated transcript on how they performed with respect to that goal. Following guidelines created by Shalley and Perry-Smith (2001) in their research study on creativity, the instructions were formulated as below:

"…you will be told how your discussion post compared to other students' posts. A transcript of all students' posts & comments annotated with scores and comments for each score was shared with students after each discussion."

To measure the quality of posts, we modified the integrative complexity measure developed by Baker-Brown and colleagues (1992). The integrative complexity measure is a 0-5 scale which rates comments that show "no conceptual differentiation or integration" as 1; and comments in which "the nature of the relationship or connectedness between alternatives are clearly delineated and are described in reasonable detail" as 5. In this study's measurement scale, integrative complexity measurement scores 1-5 were used to represent different levels of integration from non-existent to emergent to fully developed. Examples of comments given to students are included in Table 2. One instructor, who is also the principle investigator of this study, taught all of the sections involved in this study and two trained students coded the discussion transcripts. The inter-coder reliability was high at an average level of .87.

Table 2: Scores and sample feedback

Discussion topic: Project Manager and Business System Analyst roles, collaborate or combine?	
Score	**Sample Feedback**
0	*'I agree' or 'I like' do not contribute the discussion.*
1	*The post includes only acknowledgements; and repeats ideas in the paper.*
2	*The post includes mostly acknowledgements; new ideas or perspectives are emerging but not well developed.*
3	*A valid point on contingencies, but post focuses on summarizing/repeating ideas in the paper rather than presenting a rationale for the given point.*
4	*There is a good point on small vs. large organizations but needed more elaboration, remove the last statement which is unclear and avoid repetitions.*
5	*New ideas, well connected and sufficient reasoning.*

5. DATA ANALYSIS

Discussion networks were created based on binary discussion matrices in which cell (i, j) was 1 if student i commented on student *j*'s posts, and 0 otherwise. Non-binary discussion matrices stored in cell (i, j) the score that student *i* received for the comment posted on student *j*'s post. In the following analyses, both binary and score matrices are used.

The first comparison was conducted on the density of interactions among students in the online discussion forum. Density measures the number of connections among nodes in a given network. For a binary directed network density is calculated by number of ties divided by $n \times (n-1), i.e.$ all possible (directed) ties. For a score matrix, density is the average value of all cells (Borgatti et al., 2002). Denser discussion networks include a higher number of comments between students, and less dense discussion networks include a smaller number of comments. While the discussions expected students to post one original idea and four comments, not all students completed the requirements of the discussion, therefore variations exist in the density levels of twenty-four discussion networks. Two relatively consistent patterns were observed in the control and treatment sections (Figure 3). All of the control groups, in which students only received their own scores, showed a decrease in density from the first to the second discussion, implying that there might be an evaluation apprehension mechanism in play when students receive only their scores. Evaluation apprehension occurs when students' perceptions on how their contributions is to be scored adversely impacts their motivation to contribute or create high quality contributions. In contrast, the density of all sections in the treatment groups increased from the first to the second discussion. The rates of change in density levels, measured as $\frac{Density_{D2}-Density_{D1}}{Density_{D1}}$, are listed in Table 3 for each of the six groups.

Insert Figure 3 Here

Table 3. Changes of Discussion Network Density

Condi -tion	**Semeste r**	**Density change rate**		
Contro l	Fall 2014	-25.7%	-27.0%	-18.0%
	Spring 2015	-34.3%	-33.2%	-40.5%
Treat-ment	Fall 2015	23.2%	17.3%	18.6%
	Spring 2016	17.1%	25.1%	18.0%

Next, we examined changes of in-degree centralization of each course section's discussion network normalized over the changes in density (Table 4 in Appendix). At the node-level, the in-degree measure shows the number of comments that each student received. At the class-level, the in-degree measure shows the extent to which the total number of comments exchanged in the discussion are distributed among different posts by different students. For a given binary network, the network–level in-degree centralization measure is the sum of $\sum max_{in-degree} - actual_{in-degree}$ divided by the maximum value possible (Borgatti et al., 2002). A more centralized discussion indicates that a few students receive the bulk of the comments and a less centralized discussion implies that the comments are more evenly distributed among different posts in the discussion. Class-level in-degree centralization measures were normalized by density in order to eliminate the impact of variations in activity levels of each specific cohort. The numbers listed in Appendix Table 4 show the change in centrality assuming equal levels of activity across sections.

Insert Appendix Table 4 Here

Comparison of means with t-test was performed to the normalized in-degree centralization and resulted in a p-value of <0.001. Results shown in tables 3 and 4 indicate a more desired online interaction dynamic observed in the treatment groups: students are more active (higher density) and discussion comments are more broadly distributed (instead of having a few students receiving more attention). It is important to note that while five contributions were expected, students ultimately chose how many contributions they made. Students also chose whose posts they commented on. Thus, variations are observed in both density and in-degree centralization.

After examining density and centralization, we investigated reciprocity. A desired tendency in discussion networks is a low level of reciprocity, which implies that students do not necessarily comment on their peers who have commented on their post, but instead focus on the content of a given post and choose which one to comment on. Reciprocity may be impacted by factors external to the discussion dynamics, such as students' familiarity with each other, as well as internal factors, such as the timing of posts. While in this specific research project we did not measure familiarity at the class- or dyad-level, the second confounding factor is not present due to the set-up of the discussions that separated the posting of original ideas and responding comments. The rate of change in reciprocity from Discussion 1 to Discussion 2 was calculated as $\frac{Reciprocity_{D2}-Reciprocity_{D1}}{Reciprocity_{D1}}$ for each of the six control and six treatment discussion networks. The rates of change in reciprocity were normalized by density to account for variations in level of participation in each cohort. Then the six normalized values were compared with a t-test (Table 5).

Table 5. Comparison of differences in group-level reciprocity normalized by density from D1-D2

Condition	**Mean**	**Variance**
Control	.902	0.566
Treatment	-0.371	0.11
***t*-Stat : 3.795 (*df*=10) *p*-value: 0.003**		

In the next step of the analysis, trends of quality improvement were examined in all twelve course sections. For this purpose, quality matrices were

 ISSN: 2473-3857
v4 n4642

used. In the non-binary quality matrices, cell (i,j) would store the score (0-5) that student i received for the comment posted on student j 's idea, if such comment exists, and cell (i,j) would store zero if such comment does not exist. Comments that do not convey any useful information will also be given 0 (Table 2). To compare quality improvements from Discussion 1 to Discussion 2 in control and treatment groups, the average score for each student was calculated (across all posts by said student); then the average scores were normalized in each section; then the normalized average quality of posts was compared for the two discussions in each section to calculate a measure called Integration Improvement Factor (IIF):

$$\text{Normalized scores } NS \text{ in section } s = \frac{score - Min_{scores\ in\ s}}{Max_{scores\ in\ s} - Min_{scores\ in\ s}}$$

$$\text{Integration Improvement Factor } (IIF) = \frac{NS_{D2} - NS_{D1}}{NS_{D1}}$$

Each course section had one IIF vector (one vector element for each student), and a total of twelve integration improvement factors for all sections. The sections in Fall 2014 and Spring 2015 did not apply social comparison (C: control) groups, whereas the sections in Fall 2015 and Spring 2016 were mixed-approach social comparison (T: treatment) groups in the experiment. IIF vectors for the six sections in the control group were concatenated to create IFF_C. Similarly, IIF vectors for the six sections in the treatment group were concatenated to create IFF_T. A t-test was performed to compare the mean value of each. The summary is included below (Table 6).

Table 6. Quality comparisons

Treatment	**N**	**Mean**	**Variance**
Control	139	.14	1.83
Treatment	136	.36	1.46
t-Stat : -1.4 (df=271)		**p-value: 0.08**	

Node-level analyses were performed to assess the extent to which each student's improvement in the discussion posts quality was correlated with their structural measures in their discussions' interaction network (e.g., in-degree, reciprocity) and if the level of correlation was different for control and treatment groups. The IIFs calculated previously were correlated with normalized student-level (node-level) in-degree centralizations for discussions in treatment groups. All but one of the treatment groups showed a negative correlation implying that the students who received fewer comments were more likely to improve the average quality of the posts and comments they shared in the subsequent discussion. The correlations were negative for only one section of the control group; the correlations are depicted in Figure 4. This implies that a 'winners keep winning' mechanism was prevalent in control groups; students who received more comments (whose posts received more attention), improved the quality of their posts. An opposite phenomenon is prevalent in treatment groups, perhaps because informative nature of the comments that helped posters of less popular ideas to work harder on improving quality of their future posts or because informational evaluation has created stimulated upward social comparison in class.

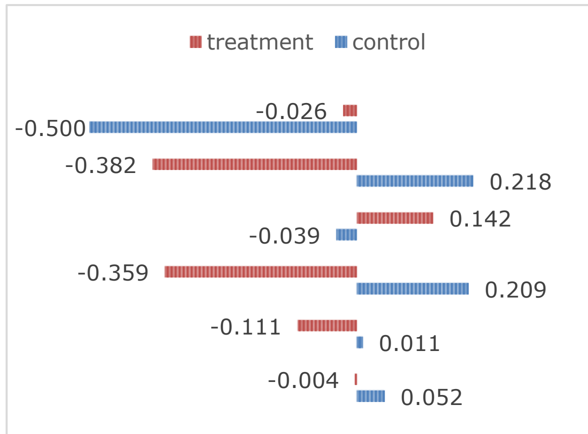

Figure 4: Correlation between quality improvements and normalized in-degree centralization in Discussion 1

The network- and node- level analyses were followed by dyadic analysis. Dyadic analyses would reveal whether or not the interactions at dyad-level persist from discussion 1 to discussion 2. For instance, whether the same pair continue commenting on (or ignoring) each other's posts. We examined Jaccard's coefficient for similarity between the two discussions' binary networks in each of the 12 sections. We also examined QAP (Quadratic Assignment Procedure) correlations between the two discussions' non-binary networks. QAP helps assess the extent to which patterns observed in a given network are unique observations as opposed to being commonly observed patterns in similar networks. The Jaccard's coefficients and QAP correlation numbers for the six treatment groups were not significantly different from those of the control groups. Therefore, while network-level changes in the discussions were observed, those changes are not discernible at dyadic level when control and treatment groups are compared. In general, a low QAP correlation and Jaccard's coefficient are desirable, they show students treat each discussion independently when it comes to whom they choose to comment on. QAP correlations for control and treatment groups ranged from [.03, .153] to [-.009, 186] respectively; and Jaccard's coefficient ranged from [.087, .155] to [.086, .241].

6. SUMMARY AND CONCLUSION

This study aspires to contribute to the literature on productivity and effectiveness of online discussions by advancing integrative quality of posts through use of a mixed-approach social comparison. The proposed mixed-approach social comparison had built in it elements of upward and downward comparisons with goal-specific informational evaluations. The mixed-approach social comparison was used in six of the twelve course sections in the reported field experiments. Treatment groups had higher rates of increase in activity levels (density) from the first to the second discussion (Figure 3 in Appendix), indicating the social comparison method accompanied by informative feedback is an enabling factor for students' participation in dialogue with their peers on course-related topics. While the control groups entailed a 'winners keep winning' mechanism, the treatment groups were successful in encouraging students with less popular posts to make improvements in quality in their second discussion's posts. While causal links have not been examined or established, we believe that the informative nature of the comments has helped posters of less popular ideas to work harder to improve the quality of their future posts and the sharing of classroom posts (scores & feedback) has stimulated upward social comparison in class. Popularity (number of comments received) was a more equally distributed commodity in the treatment groups (using in-centrality measures). Also, at class-level, treatment groups showed higher levels of quality improvement and lower levels of centralization in commenting networks when two consecutive discussions were compared therein. All these factors contribute to a healthier, more engaging, and opener discussion dynamic, thus the findings are consistent with this paper's proposition. At the end, limitations of field experiments apply to this study as well; we are not certain which students did or did not read the transcript (to actively engage in social comparisons) and how other online and in-class dynamics impacts students' commenting behavior in course discussions. The findings of this study, however, are consistent with literature on social comparison and informational evaluation. The mixed-approach social comparison employed in the treatment groups of this study can inform design of online discussions and electronic brainstorming features, and creativity support tools.

7. REFERENCES

Ambrose S.A., Bridges, M.W., Lovett, M.C., Dipietro, M., & Norman, M.K. (2010*). How Learning Works*. New York, NY: Jossy-Bass.

Baker-Brown, G., Ballard, E.J., Bluck, S., de Vries, B., Suedfeld, P., & Tetlock, P.E. (1992). The conceptual integrative complexity scoring manual, In C.P. Smith, J.W. Atkinson, D.C. McClelland, & J. Veroff (Eds.), *Motivation and Personality: Handbook of Thematic Content Analysis* (pp. 393-400), Cambridge, NY: Cambridge University Press.

Borgatti, S.P., Everett, M.G., & Freeman, L.C. (2002). Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Deci, E. L., & Ryan, R. M. (1980). The empirical exploration of intrinsic motivational processes. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 13, pp. 39–80). New York, NY: Academic Press.

Dugosh K. L., & Paulus, P.B. (2005). Cognitive and social comparison processes in brainstorming. *Journal of Experimental Social Psychology*, *41*(3), 313-320.

Ericsson K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American psychologist, 49*(8), 725: 747.

Festinger, L. (1954). A Theory of Social Comparison Processes, *Human Relations*, 7 (2), 117-140.

Michinov, N., & Primois, C (2004). Improving productivity and creativity in online groups through social comparison process: New evidence for asynchronous electronic brainstorming. *Computers in Human Behavior*, *21*(1): 11-28.

Pinsonneault, A., Barki, H., Gallupe, R. B., & Hoppen, N. (1999). Electronic brainstorming: The illusion of productivity. *Information Systems Research*, *10*(2): 110-133.

Shalley, C. E., & Perry-Smith, J. E. (2001). Effects of social-psychological factors on creative performance: The role of informational and controlling expected evaluation and modeling experience. *Organizational Behavior and Human Decision Processes*. *84*(1), 1-22.

Shepherd, M.M., R.O. Briggs, B.A., Reinig, J. Yen, & J.F. Nunamaker (1996). Social comparison to improve electronic brainstorming: Beyond anonymity. *Journal of Management Information Systems*, *12*(3), 155-170.

Simon, H. A. (1947). *Administrative behavior: A study of decision-making processes in administrative organization*. New York, NY: Macmillan Co.

Wasko, M.M. & S. Faraj. (2005). Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, *29*(1), 35-57.

Waters, J., & Gasson, S. (2012). Using asynchronous discussion boards to teach IS: reflections from practice, *Proceedings of the 32nd International Conference on Information Systems (ICIS)*, Orlando, Florida, December 2012.

ISSN: 2473-3857
v4 n4642

Appendices and Annexures

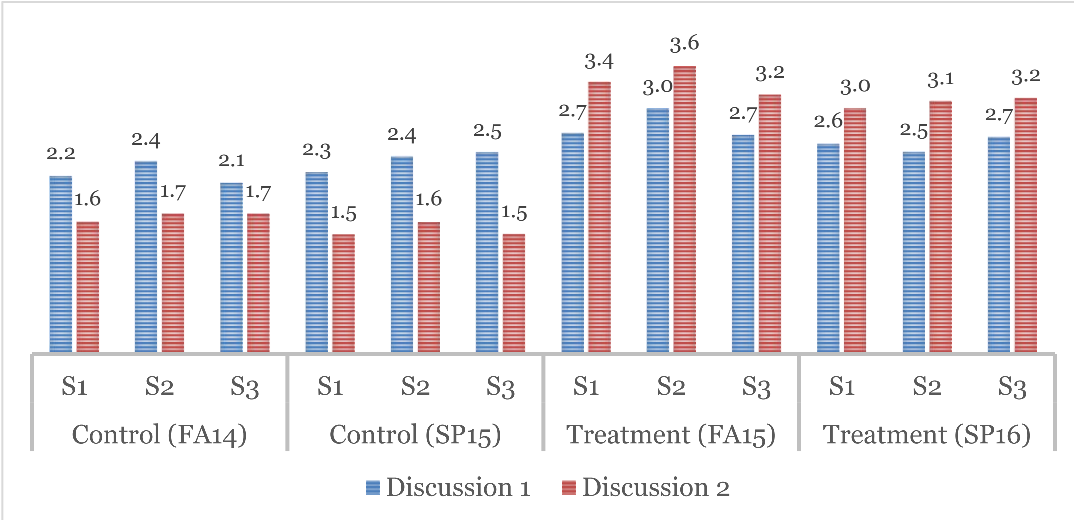

Figure 3: Density in control (left) and treatment (right) groups *

*: The numbers in Appendix Figure 3 were used to calculate the change rates reported in Table 3; because of rounding, the results may be slightly different from those calculated manually.

Table 4. Normalized in-degree measures

Condition	Semester	In-degree centrality change rate normalized by density change			*t*-test comparison
Control	Fall 2014	0.10	0.11	0.09	Mean: 0.11 Variance: .00037
	Spring 2015	0.14	0.13	0.11	
Treatment	Fall 2015	0.02	0.04	0.05	Mean: .047 Variance: .00027
	Spring 2016	0.04	0.07	0.06	
t-Stattatistic: 6.43 (*df*=10) *p*-value: <0.001					