ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Goethals, George R.

Working Paper

Social comparison and peer effects at an elite college

WPEHE Discussion Paper, No. 55

Provided in Cooperation with:
Williams Project on the Economics of Higher Education, Williams College

Suggested Citation: Goethals, George R. (2000) : Social comparison and peer effects at an elite college, WPEHE Discussion Paper, No. 55, Williams College, Williams Project on the Economics of Higher Education (WPEHE), Williamstown, MA

This Version is available at:
https://hdl.handle.net/10419/23502

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Discussion Paper No. 55

Williams Project on the Economics of Higher Education
Denison Gatehouse
Williams College
Williamstown, MA 01267

SOCIAL COMPARISON AND PEER EFFECTS AT AN ELITE COLLEGE

George R. Goethals

May, 2000
DP-55

An earlier version of this paper was presented at "Lessons from CUNY: A Forum on Clashing Visions of Higher Education" City University of New York, New York, New York, March 18, 2000

This paper is intended for private circulation and should not be quoted or referred to without the permission of the author. The author would to thank Gordon Winston and David Zimmerman for their wisdom and advice in planning and conducting the research reported here, and Sally-Thomas Daigneault, Edward Murphy, Melissa Murphy, Gillian Weitz, and Sarah White for their help in collecting and analyzing the data. Funding for this paper was provided by the Andrew Mellon Foundation through the Williams Project on the Economics of Higher Education.

George R. Goethals, Department of Psychology, Williams College,
Williamstown, MA 01267. 413 597-2443 phone; 413 597-4435 fax.
agoethals@williams.edu

ABSTRACT

A study was conducted to see whether peer effects could be observed among undergraduates at Williams College, an elite four-year liberal arts school. Specifically, the study explored whether students in the bottom third of their class, with average SAT's of about 1300, would perform better in writing about newspaper articles they read and discussed in groups of three if the two others in the group were academically superior -- from the top third of their class, with SAT's averaging about 1500 -- rather than similar -- also from the bottom third of the class. The results showed that women subjects performed better if their discussion partners were from the top third of the class, but men did better if their discussion partners were from the bottom third. Alternative analyses comparing subjects who had better or worse discussion partners as determined by the quality of their peers videotaped discussion statements, showed that across gender subjects did better written work when their discussion partners were better. The results were interpreted in terms of the principles of social comparison theory (Festinger, 1954).

Social Comparison and Peer Effects at an Elite College

Recent research by Winston and his colleagues (Goethals, Winston, & Zimmerman, 1999;Winston, 1997) suggests that colleges and universities refrain from expanding their student body, even though they have many more talented applicants than they can accept, because they want to maximize student quality. There are a number of reasons that schools may want to do this, importantly including the belief that students get a better education if their fellow students, their peers, have higher degrees of academic talent. In that sense, one of the things that students buy when they attend college is the other students who form the peer environment. Students are both customers and a key component of the product they are buying. That is, one key aspect of the technology of producing higher education is a "customer-input technology."

Are colleges correct in believing that peer quality makes a difference in student education? Can such peer effects be demonstrated? What intricacies and qualifications complicate a simple story about the value of having more rather than less talented fellow students? Clearly there is a good deal of evidence that young people are influenced by peers and some evidence that peer effects in education operate at elementary and secondary school levels (e.g., Coleman et al., 1966). A few econometric studies support the idea of peer effects among college students (Hoxby, 1999). The purpose of this study is to try to show experimentally that peer effects are operative in a college environment, and to place the results of such a study in the context of relevant social psychological theory.

Social comparison and peer effects. One influential social psychological framework that provides a useful framework for thinking about possible peer effects is Leon Festinger's theory of social comparison processes (Festinger, 1954; Suls & Wheeler, in press). This theory sets forth several principles with implications for both the prospects and problems associated with student interacting with highly capable peers.

Social comparison theory is a theory of self-evaluation and begins with the proposition that people have a drive to evaluate their opinions and abilities. Decades of research have shown that people compare on many other personal characteristics, such as income, attractiveness, and health, but the theory's original emphasis on opinions and abilities is extremely relevant to a consideration of peer influences among college students (Suls & Miller, 1977; Suls & Wills, 1991; Wood, 1996). Festinger argued that people evaluate their opinions and abilities through comparison with other people and that they can make much more stable evaluations by comparing with other people who are similar. People check their opinions against those of peers who have generally similar opinions and world views. Similarly, they compare their performances against those of people whose ability levels and training and experience are similar. In the absence of similar others for comparison, people are not able to satisfy adequately their need to evaluate themselves.

An important consequence of the need for similar others to satisfy evaluation needs is strong pressure within groups toward uniformity of opinions and abilities. Those pressures are stronger when the opinion or ability in question is important and relevant to the group's immediate situation. When opinions are at issue the pressures toward uniformity are unalloyed, and there is discussion until talk has produced uniformity, or

until those with deviant opinions are rejected from the group, usually with some degree of hostility. When abilities are being evaluated, pressures toward uniformity combine with pressures toward excelling and being better than others. Individuals compete until a ranking evolves, marked by differences within a narrow range. Those with highly different ability levels become defined as non-comparable – comparison with them ceases – although they are not rejected in a hostile way, as is the case for opinions. They simply cease being a part of the individual's reference group, and they are largely ignored. In short, pressures toward uniformity produce talk and competition, and ultimately, marked homogeneity, if not uniformity.

What are the implications of the dynamics of talk and competition produced by pressures toward uniformity? They clearly have the potential to produce peer effects of the positive kind imagined by the schools that attempt to maximize student quality. But they generate some perils as well.

The cognitive consequences of talk. On the way to opinion uniformity, a great deal can happen that is of direct relevance to the concern with peer effects in college. While it is not always the case that people achieve consensus by talk--sometimes conformity pressures produce opinion and behavior change without any need for persuasion or rationale--there often is a great deal of discussion in groups. These discussions can affect the way people think in several ways.

First, information is transmitted. This information can affect people's beliefs by affecting the knowledge that underlies those beliefs. In some cases the new knowledge may simply add to an individual's general way of thinking. In that case, the new information is simply assimilated into the person's general knowledge structures, or

schemas. Their schemas may become relatively more detailed, and slightly more complex, but basic viewpoints do not change. They simply become more elaborate. Theories are confirmed, not challenged. In other cases, the new information cannot be assimilated to existing schemas. It doesn't fit and cannot be understood within existing categories, theories, or beliefs. Then the knowledge structures must change to fit the data. They actually accommodate to the information, and become entirely reshaped (Piaget, 1937). When new or highly revised schemas are produced by new information, the result is more than just an accumulation and cataloging of new information. The result is new theories and new conceptualizations which facilitate the absorbing of new information and further cognitive development.

The importance of talk among peers in producing new conceptualizations is argued powerfully in the work of developmental psychologist L.S. Vygotsky (1935). Vygotsky notes that "*human learning presupposes a specific social nature and a process by which children grow into the intellectual life of those around them*" (1935, p. 88, italics in original). Furthermore, Vygotsky argues that learning specific points, ideas, facts, techniques, approaches etc. fosters increased cognitive development. He notes that "in making one step in learning, a child makes two steps in development" (1935, p. 84). Learning fosters development and "sets in motion a variety of developmental processes that would be impossible apart from learning" (1935, p. 90).

One clear implication for peer effects education is that the potentially highly educational impact of talk will be maximized to the extent that the talkers whom students hear are intelligent and well-informed, and use that intelligence in their discussion. One compelling line of research supporting this notion concerns the intellectual impact of an

extremely important peer environment, that constituted by a child's siblings, and his or her parents (Zajonc, 1976; Zajonc & Mullally, 1997). This research, supporting what Robert Zajonc (1976) termed the confluence model, suggests that SAT and IQ scores in the adolescent and adult years are influenced by the quality of the intellectual environment of the family during the child's formative years. The quality of this environment is in turn affected by the number and ages of the children and adults in the household, and thus the average developmental or intellectual level of the individuals in the home.

Thus far we have considered just the potential benefits of listening to talk. There are also benefits to speaking rather than listening. Two highly divergent lines of research by Zajonc support this position. First, in an important paper on "cognitive tuning", Zajonc (1960) noted that people process information differently depending on whether they are in "transmission tuning" or "reception tuning." In reception tuning they simply expect to receive more information. They remember the complex details of what they see, hear, or read. When they are in transmission tuning they expect to tell other people about what they learn. In this case they develop a more coherent account of the information, one that is perhaps simpler but more internally consistent. While it may omit all the relevant information, it tells a better story. Explaining something to another person induces a more active, organized cognitive integration that itself produces learning.

Second, in his research on the effects of childhood family configuration on adult intelligence, Zajonc found that the last born child, whether that one is the only child or the youngest of a set of siblings, shows lower SAT or IQ scores than would be predicted by the simplest version of his confluence model. For example, only children score lower

than the elder child in a pair, and the third child who is last-born has lower scores than the third of four children. Zajonc's explanation for these somewhat anomalous findings is that the last child is deprived of the benefit of teaching younger siblings. Children seem to benefit from two things: one, being raised in an intellectually alive and sophisticated social environment of older siblings and parents, and, two, having a younger sibling to teach. For the most part children are in reception tuning in relation to parents and their older siblings, and transmission tuning in relation to young siblings. Both tunings foster intellectual growth.

In sum, talk produces intellectual growth in a variety of ways. At the same time, it is important to remember the context in which we are discussing talk, that is, conversations directed toward achieving consensus and uniformity of opinions in groups. Uniformity is sought, according to social comparison theory, to enable individuals to develop stable evaluations of their opinions. That is, talk can produce distinct cognitive development. It is also likely to produce uniformity of opinion through combinations of influence, conformity, and rejection of those who hold deviant opinions. In the case of rejection, opinion uniformity is achieved by defining group boundaries in a way that only those who agree are considered to be the group. We need to be vigilant about the consequences for colleges of the strong tendencies to evolve many small, highly homogeneous groups of like-minded individuals.

The performance consequences of competition. Social comparison theory addresses the evaluation of abilities as well as opinions. In fact, when originally published the theory was quite startling in focusing on these two human attributes, since the processes flowing from their evaluation produce some very different consequences.

However, Festinger's attempt to highlight the similarities among the evaluation processes for opinions and abilities can be understood in terms of his interest in level of aspiration for performances, his first area of research, and his interest in social communication and conformity, the area he was working on just before developing social comparison theory.

Despite the similarities, the theory highlights two important differences between opinions and abilities. First, people consistently try to raise their performance level. Second, there are nonsocial constraints on changing abilities which do not apply to opinions. People can't change their ability to serve aces in tennis like they can their opinion of Chris Evert. That is, people want to improve but that may be very difficult. Social comparison research has shown that the drive to improve and pressures toward uniformity combine to produce competition, at first, and then tendencies to define groups so that they are composed of people with similar ability levels. They can also produce efforts to prevent peers from performing significantly better than most others in the group. For example, people form coalitions to prevent their peers from excelling when important abilities are implicated by relative performance (Hoffman, Festinger, & Lawrence, 1954).

We noted above that Festinger's interest in ability comparison reflected a very long-standing interest in the way people set their level of aspiration for performance. In developing social comparison theory he discussed the effects of level of aspiration on the cessation of comparison on abilities. When people cease comparing with superior others, and define their own group as consisting only of those with more modest ability levels, their level of aspiration often drops. When they cease comparing with inferior others, their level of aspirations correspondingly rises.

As with the social comparison of opinions, the social comparison of abilities produces effects with both beneficial and worrisome consequences for learning and education. Competition may spur productive academic involvement. That depends on academic performance and ability being an important value in any particular reference group. However, competition may produce distinctly uncooperative behaviors designed to undermine superior performances by others. Also, when people cease comparing and competing, and define their reference groups as a more homogeneous set of individuals with similar ability levels, there can be increases or decreases in their levels of aspiration. These changes may help or hurt academic performance.

Implications for observed peer effects. The considerations above suggest that we might well find peer effects among college students. The present study attempts to discover them in a setting where they might, however, be quite limited. It looks for them among undergraduates at Williams College who spend twenty-minutes in the context of a psychology experiment discussing articles from The New York Times with two classmates, then twenty minutes writing about what they've learned from reading and discussing the articles.

In all cases, the "subject" in the study is in the bottom third of the class according to data from the college's admission office. In the High Peer condition, the two peers are in the top third of the class while in the Low Peer condition the two peers are both, like the subject, in the bottom third of the class. The average SAT's for students in the top third of the class are approximately 740 Verbal and 755 Math. The approximate averages for students in the bottom third of the class are 640 Verbal and 640 Math. That is, the top third students have an average combined SAT of about 1500 while the bottom third have

a combined average of nearly 1300. What can we expect? Will our lower third subjects show the benefits of interacting with superior peers after only 20 minutes of group discussion? Is there enough difference between the High and Low Peer environments to make a difference in individual learning after so short a discussion?

Our hypothesis was that the answer would be "Yes." We conducted the study to show that peer effects can be observed even in this most minimal context. If students learn from each other, that should be observable, and if they learn more from more able peers than less able peers, we ought to be able to show it. Specifically, we predicted that subjects in the High Peer groups would perform better in the discussion and in their written reports of what they learned from reading and discussing. However, there were at least two kinds of reasons not to be optimistic about this intended demonstration, one methodological and the other theoretical. The methodological concerns were the ones noted above -- the time is too short, and the high and low peers are not that different, given the full range of SAT scores. The theoretical concern was that the high and low peers may, in fact, be too different. If social comparison is to be fully engaged, and students are to compare opinions and compete to perform well in discussion, they must regard each other as similar. It is possible that our subjects, from the bottom third of the class, will find the high peers too intelligent, and will cease comparing with them and taking any interest in their views or analyses. In fact, they might be intimidated by them in a way that would make them lower their aspiration for performing well in the discussion and in their written reports of what they learned from the discussion.

The study described below was conducted to find out. We predicted that subjects in the High Peer groups would show that they had learned more from the discussion, but we realized that several difficulties make this result somewhat unlikely.

Method

Participants

One hundred and two Williams College first-year students and sophomores volunteered to participate in this study. They were paid $15.00 or received one-hour of extra-credit in an Introductory Psychology course. The study was called “College Students and Public Affairs". All the participants were in the top third or bottom third of their class in academic potential according to College ratings made at the time the students applied for admission to Williams.

Procedure

Participants were scheduled in groups of three, such that all three participants were in the same class (freshman or sophomore) and such that all three were either in the bottom third of their class (Low Peer condition) or one participant was in the bottom third and the other two were in the top third (High Peer condition). Participants were greeted by an experimenter who explained briefly that the study entailed reading three articles from the New York Times, discussing those articles as a group, and answering questions about what they had read and discussed. The participants sat at a round table with a microphone in the center. The experimenter explained that they would be observed through a one-way mirror and that their discussion would be video-taped using the microphone and a ceiling-mounted camera.

Participants were given twenty-minutes to read three articles, twenty-minutes for discussion, and twenty-minutes to answer a questionnaire asking about what they had learned from reading and discussing the articles. The students were instructed to discuss initially just the first two articles, and the discussion was stopped before the third was discussed. This procedure was adopted to allow comparison of responses to discussed vs. not discussed articles. After giving instructions the experimenter left the room, and subsequently returned twice, first to ask the participants to begin the discussion and then to ask them to stop the discussion and complete the questionnaires. When the experimenter was out of the room she was partially visible in an adjoining room through the one-way mirror.

Academic Ratings

As mentioned above, students were recruited for the study and assigned to groups on the basis of academic ratings assigned by the Office of Admission when students apply. The academic rating is based on students' secondary school grades, the quality of their secondary school academic program, their SAT's, and information in recommendations that seems to reveal academic potential. The academic ratings have been used for many years and are, at Williams, the best available predictors of student grades. While the academic rating predicts student grades better than any of its components, the best single predictor among the components is Verbal SAT.

Materials

Participants read three articles published in the New York Times in August of 1998. The first ("Brave New Worlds") was a review of a book by Bryan Appleyard called Brave New Worlds (Kass, 1998). It discussed benefits and dangers in the new

world of genetic engineering, making reference to ways the science fiction of Aldous Huxley and George Orwell has become increasingly realistic. A second ("More Hits") entitled "More Hits, More Runs, and More Concerns" discussed baseball bat technology, and the resulting increase in both home runs and injuries in the national pastime (Guzman & Johnson, 1998). The third ("Popeye Spikes His Spinach") called "The News is Out: Popeye Spikes His Spinach" was also about baseball and questioned slugger Mark McGwire's use of androstenedione and other performance enhancing drugs (Araton, 1998).

The questionnaire asked students to rate on seven-point scales how much they learned from reading and discussing each article, how interested they would be in reading or discussing such articles in the future, and how much they learned from each of the other two students. It also asked them to write on one page the ideas or information they learned from reading or discussing each article. The page listed the numbers one through ten, to provide space to write ten statements, but said that the reverse side could be used as well.

Coding of written responses

Undergraduate raters coded each participant’s statements of ideas and information. A quantity rating gave credit for each idea or piece of information the participant stated. A quality score rating from one to three was given to each statement on the basis of its specificity, detail and elaboration. A total quality points score and an average quality score per statement were calculated for each article, for the first two articles (the ones that were discussed), and for all three articles. Inter-rater agreement on

quantity scores was virtually 100%. For quality scores it was 92% and all disagreements were reconciled through discussion.

Coding of discussion videotapes

Undergraduate raters coded each statement in the videotape of each discussion. First each rater proposed a written "order of talk" that listed who spoke when on the tape. There was near 100% agreement on who was speaking and any disagreements were resolved through discussion. Then each statement was given a length rating from one to four, depending on whether the statement was less than 5 seconds, from 6 to 10 seconds, 11 to 15 seconds, or greater than 15 seconds. Inter-rater agreement for quantity ratings was 95%. Disagreements were resolved through discussion, or, in rare cases, averaging.

Each statement was also given a quality rating of negative one to three, based on how effectively the statement advanced the discussion, and contributed to the intellectual quality of the discussion. Negative one scores were given to statements that halted or derailed discussion. Zero was given to statements that were neutral or bland, one was given to remarks that advanced the discussion through simple statements or questions, two was given to remarks that were more thought provoking, and three was given to those rare statements that advanced the discussion productively and that were exemplary in thought and expression. The two raters agreed on 88% of the quality ratings. Three quarters of the remaining ratings were resolved through discussion and the others were given averaged ratings.

For each participant and each group, total and average quantity and quality scores were calculated. Also, each subject's "peer environment" was calculated by averaging the total quality scores of his or her two fellow participants.

Results

There were a total of 34 groups. Twenty-two were High Peer groups, with one participant in the bottom third of the class and two in the top third. Twelve groups were Low Peer groups, with all three participants in the bottom third of the class. In the High Peer groups, only the participant in the bottom third is used as a “subject” in the analyses reported below. In the Low Peer groups, all three participants are used as “subjects” since all three interacted with two peers in the bottom third of the class. All results reported below are statistically significant at the .05 level unless otherwise noted.

Subjects vs. High Peers

Initial analyses were done to compare the 44 high academic rating "peer" participants with the 58 low academic rating "subjects" from both High Peer and Low Peer groups. These analyses provide some indication of whether the students with high academic ratings behaved differently in the study than students with low academic ratings. Analyses of variance showed a large number of significant effects, all consistent with the “peers” having more academic promise than the “subjects.” The peers reported more past reading of articles like those in the study, more interest in such articles, and more interest in reading such articles in the future. They also reported learning more from “Brave New World.” Furthermore, the peers videotaped discussion statements had higher individual total quantity and quality ratings. This result is important. It provides evidence that the students with high academic ratings did perform better in the discussions, a necessary condition for the emergence of peer effects in this study.

Somewhat surprisingly, the peers did not show any superiority to the subjects in their written reports of what they learned from reading and discussing the articles.

The effects of peers with high vs. low academic ratings

There were relatively few overall differences between subjects interacting with highly rated peers and those interacting with low rated peers. One significant difference was that subjects in high peer groups reported learning more in the discussion of “Brave New World.” There was also a trend indicating that subjects in high peer groups reported learning more from their peers ($p<.12$). Finally, the video-taped “peer environment,” the average total quality of each subjects’ two peers’ discussion statements, was superior for High Peer subjects. This effect was highly significant ($p<.02$), and shows again that the "manipulation" of peer environment through group composition based on academic ratings worked as intended.

While the results above show very little of the peer effects we expected, the analyses below, which consider gender and actual group performance, give a clearer and brighter picture.

Gender and peer effects

A number of results showed that there were important significant differences between men and women subjects in their self-reports and their written and verbal behavior. First, men report having read more articles like those used in the study, and greater interest in reading more in the future. Second, the women report having learned more from the two articles dealing with baseball, “More Hits” and “Popeye Spikes His Spinach.” Third, women’s written statements have significantly higher total quantity and total quality ratings across all three articles, especially for “More Hits,” the baseball

article that was both read and discussed. Women wrote on average 15.0 statements with an average quality rating of 1.52. Men wrote an average of 12.5 statements with an average quality rating of 1.49. Thus the women wrote more, and didn't sacrifice quality in doing so. Finally, the analysis of the discussion videotapes revealed that men had significantly more total quality points than women, and, quite dramatically, a higher average quality rating per statement than women (average quality ratings: men = .688, women = .457; $p<.002$). Women actually made slightly, but not significantly, more statements than men, 24.9 vs. 22.1, but the higher average quality scores for men translate into significantly more total quality points for men, 15.2 vs. 11.4.

In sum, men's self-reports about their past and future reading were more positive than women's and they spoke more effectively. Women reported learning more from reading and discussing the two baseball articles, and wrote more with slightly higher average quality than men.

Most important for the present study were several findings indicating that men and women reacted quite differently to high vs. low peer environments. First, each subject rated how much he or she had learned from each of their two peers. Thus there were ratings of how much subjects felt they learned from Participant A, B, or C. If the subject was participant A, he or she rated B and C, and so forth. On the rating of how much the subjects learned from participant A, there was a highly significant interaction ($p<.02$) indicating that women with high peers gave substantially higher ratings than women with low peers, while men with high peers gave slightly lower ratings than men with low peers. On a 1 – 7 point scale, the means are as follows: women with high peers, 5.2; men with low peers, 4.5; men with high peers, 4.3; women with low peers,

3.5. Similar effects were not found for reports of the amount learned from participants B and C. However, this one finding is consistent with several indications of a similar pattern in the quality of men's and women's written performance with high and low peers. The total quality scores of what subjects wrote on "Brave New Worlds" and "More Hits," the two articles which were both read and discussed, produced a nearly significant interaction showing that women subjects did better with high peers while men did worse ($p = .06$). Their total quality scores (number of statements multiplied by average quality scores) were as follows: women with high peers, 19.46 (11.8 X 1.65); women with low peers, 16.29 (10.6 X 1.54); men with low peers, 15.18 (9.7 X 1.57) men with high peers, 11.28 (7.9 X 11.28). The total quality scores for women with high peers were significantly higher than the total quality scores for men with high peers ($p < .01$).

In sum, women in High Peer groups report learning more than women in Low Peer groups, while the men do the reverse. Also, women seem to perform better in their written accounts of what they learned from the two articles that they discussed when their peers had high academic ratings, while men performed better when their peers had lower academic ratings.

Video Peer Environment Effects

An initial concern in this study was whether the peers with high academic ratings would actually provide a superior discussion environment than peers with low academic ratings. The findings reported above regarding the quality of the videotaped discussion statements made by highly rated students vs. low rated students indicate that, on average, the highly rated students did a better job in discussion than the low rated students, thereby providing a superior discussion environment. To further explore the effect of a

demonstrably superior peer environment, subjects were divided into two same sized groups based on the total quality ratings of the discussion statements of their two peers, regardless of the peers' academic ratings. This division of the subjects based on their peers' video scores created one group of subjects that had better peer environments and another that had worse.

The results were similar, though clearer, to those found from the analysis of peer effects based on academic ratings. Non-significant statistical trends showed that subjects with superior video peer environments reported learning more from Participant A and B, but not C. Also, a nearly significant effect showed that subjects with superior video peers reported that they would be more inclined to read more on the topic of "Brave New Worlds" ($p = .058$). Thus subjects with superior video peers reported learning more from their peers and more interest in reading about one of the topics discussed in the articles. Furthermore, and most important, the average quality and total quality scores of subjects written statements about "Brave New Worlds" were significantly higher in the superior video peer group, as were the combined total quality scores for the two articles that were discussed, "Brave New Worlds" and "More Hits." For the two articles, the combined average total quality score (number of statements times average quality) for subjects with high peers was 17.12 (10.9 X 1.56); for subjects with low peers the score was 13.38 (9.2 X 1.46). Subjects in groups with peers who were the most articulate wrote superior reports about the articles that the group discussed.

Finally, the length of subjects' statements tended to be shorter in the superior video peer environments ($p < .09$) and the proportion of group statements made by subjects in the superior video peer groups was significantly less than the proportion made

by subjects in the inferior video peer groups (30% vs. 35%,). In short, subjects in the groups with high quality peers, as measured by the quality of their verbal statements, talked less and, presumably, listened more.

Discussion

These findings taken together support two major conclusions. First, peer effects as measured by self-reports of what students learned and their actual written statements of what they learned can be observed among college students even after as little as twenty minutes of discussion. Second, men and women react quite differently to superior peers. Women seem to soak up ideas and information from discussion with superior peers. Men seem to do better with similar peers. Clearly, these results must be regarded as suggestive rather than conclusive. More research needs to be done, especially on gender differences. But it is useful to know that peer effects can be generated in the specific context of this study.

What are the implications for colleges thinking about trying to harness the positive benefits of peer effects? One is that peer effects happen, even if the two peers environments compared are quite similar, as they are in this study. If one is a college student with SAT's around 1300, it makes a difference whether one's peers also have SAT's around 1300 or have scores more like 1500. Second, peer effects can happen, but that doesn't necessarily mean they will. There are many other variables that can affect whether and how one responds to peers of superior intellectual ability. Our women

subjects were clearly more willing and/or able than the men students to reap the benefits of interacting with superior peers.

The difference we found between women and men alerts us to an important consideration from social comparison theory that is highly relevant to this study. How much individuals compare and engage each other in discussion, either collaboratively to compare opinions or competitively to perform better in analyzing issues, depends on whether they feel similar and comparable to the other members of the group. While we were somewhat doubtful in designing the study that there would be enough difference between peers with high academic ratings and low academic ratings to allow us to demonstrate peer effects, we were also aware that the difference might be too much. Students with verbal SAT's of 640 might feel quite different from students with Verbal SAT's of 740. There may be an "intimidation effect" where the students feel outclassed by students with SAT's 100 points higher, and then cease comparing and conscientiously engaging in discussion with them.

Also, while students might not feel intimidated by students with higher Verbal SAT's, they may feel dissimilar on this dimension, and may also perceive other differences correlated with the difference in verbal ability that reinforces the tendency to cease comparing. For example, in our study there seemed in a few groups to be something of a "jock vs. geek" dynamic. At Williams, as at any college campus, there are lots of subgroups and lots of intergroup stereotypes which create barriers to constructive interaction and engagement across subgroup lines. It might have been easy for subjects who were athletes to view superior peers as "geeks." One of the key hypotheses of social comparison theory is relevant here. Festinger (1954) wrote "if

persons who are very divergent from one's own opinion or ability are perceived as different from oneself on *attributes consistent with the divergence*, the tendency to narrow the range of comparability becomes stronger (p. 133, emphasis in original). Some of our students, especially perhaps, athletes, may have perceived that the verbal ability of their high rated peers was "very divergent" from their own and that those divergences were related to their more able peers being "geeks." This perception would reduce comparison, and the opportunity to benefit from interaction with superior peers. More males than females at Williams are athletes, and male athletes may be less willing to compare with brighter students than male non-athletes, female athletes, and female non-athletes. This "jock/geek" dynamic may be part of the story of the sex differences found in our study.

Another difference between men and women may be related to a somewhat puzzling pair of findings in this study. On the questionnaire, the female subjects wrote more statements and had slightly higher average quality scores for each statement than the male subjects. On the videotapes, women subjects had significantly lower average quality scores per statement. For what reason might women do a better job in writing but a poorer job in speaking? One explanation may be the role that women played in the group discussions. Women are more likely to have interdependent as opposed to independent self-construals (Cross & Madison, 1997). Their behavior often emphasizes connectedness with others. For this reason they may be much more likely to verbally reinforce their discussion partners. Consistent with this hypothesis, we found that women were much more likely to say, "yah" or "aha" than men, encouraging their partners to express their ideas. Women on average made this kind of utterance 28.7 times in each

discussion, while the men made it 17.6 times (<.006). These statements were given quality scores of zero. In terms of the content of the discussion they contributed very little. However, they may have played an important role in creating a positive and welcoming discussion environment. Thus women's approach to discussion may be more focussed on maintaining a positive group atmosphere than in generating interesting ideas or insights.

This discussion simply underlines the fact that comparison and engagement among peers has many dimensions. We have demonstrated peer effects in this study. We know they can occur in situations like the one we created. But our findings suggest that we need to be extremely thoughtful about the range of factors which in any given situation can raise or lower the probability that beneficial peer effects will emerge.

References

Araton, H. (1998) The news is out: Popeye spikes his spinach. The New York Times, August 23, 1998. Section 8, p. 1,3.

Coleman, J.S. et al. (1966) Equality of Educational Opportunity, U.S. Department of Health, Education, and Welfare, Office of Education.

Cross, S.E., & Madson, L. (1997) Models of the self: self-construals and gender. Psychological Bulletin, 122, 5-37.

Festinger, L. (1954) A theory of social comparison processes. Human Relations, 7, 117-140.

Goethals, G.R., Winston, G.C., & Zimmerman, D.J. (1999) Students educating students: the emerging role of peer effects. In Forum futures: Exploring the future of higher education, M. Devlin & J. Meyerson, (eds.), pp. 25-45. Forum Publishing

Guzman, E., & Johnson, K. More hits, more runs, and more concerns: high tech bats overpower the game. The New York Times, August, 1998.

Hoffman, P.L. Festinger, L., & Lawrence, D.L. (1954) Tendencies toward comparability in competitive bargaining. Human Relations, 7, 141-159.

Hoxby, C., & Long, T.L. (1999) Explaining rising income and wage inequality among the college educated. Unpublished mimeo, Harvard University, April 1999..

Kass, L.R. (1998) Brave New Worlds A review of Brave new worlds: Staying human in the genetic future by Bryan Appleyard. The New York Times, Sunday Book Review, August 23, 1998.

Piaget, J. (1937) The construction of reality in the child. New York: Basic Books, 1954.

Suls, J.M., & Miller, R.L. (1977) Social comparison processes: theoretical and empirical perspectives. Washington, DC: Hemisphere.

Suls, J.M., & Wills, T.A. Social comparison: contemporary theory and research. Hillsdale, NJ: Erlbaum.

Vygotsky, L.S. (1935) Interaction between learning and development. In M. Cole, V. John-Steiner, S. Scribner, & E. Souberman. Mind in society: the development of higher psychological processes: Cambridge, MA: Harvard, 1978.

Winston, G.C. (1997) Why can't a college be more like a firm? Williams Project on the Ecnomics of Higher Education, Discussion Paper #42.

Wood, J.V. (1996) What is social comparison and how should we study it? Personality and social psychology bulletin, 22, 520-537.

Zajonc, R.B. (1960). The process of cognitive tuning in communication. Journal of Abnormal and Social Psychology, 61, 159-167.

Zajonc, R.B. (1976). Family configuration and intelligence. Science, 192, 227-236.

Zajonc, R.B., & Mullally, P.R. (1997) Birth order: reconciling conflicting effects. American Psychologist, 52, 685-699.