The Impact of Group Diversity on Performance and Knowledge Spillover -- Evidence from College Classroom

Zeynep Hansen*
Washington University in St. Louis
and National Bureau of Economic Research
zhansen@wustl.edu

Hideo Owan
Graduate School of International Management
Aoyama Gakuin University
owan@gsim.aoyama.ac.jp

Jie Pan
Washington University in St. Louis
jpan@artsci.wustl.edu

September 12, 2006

Abstract
We combine class performance data from an undergraduate class with students' personal records to explore diversity and knowledge spillover effects. A major advantage of our dataset is the exogenous assignment of groups, which rules out the self-selection problem. Our results indicate that male-dominant groups performed worse both in group work and in individually taken exams than female-dominant and equally-mixed gender groups. Individual members from a group with more diversity in age and gender scored higher in exams. Another novel aspect of this natural experiment is that each group chooses their own group contract form – members of "autonomous" groups receive equal grade for their group work while those in "democratic" groups can adopt differentiated point allocation, thus, providing a proper mechanism to punish free riders. Our estimation results show a significant correlation between the choice of a democratic contract and the group and individual performance.

Keywords: Teams, diversity, productivity, learning, collaborative skills, education
JEL Classifications: D2: Production and Organizations, I2: Education

* Corresponding author. Mailing address for corresponding author: John M. Olin School of Business, Washington University in St. Louis, Campus Box 1133, One Brookings Drive, St. Louis, MO 63130. Telephone: (314) 935 6323; fax: (314) 935 6359.

I. Introduction

There is a growing public interest in the impact of workforce diversity on firm performance. When the US Supreme court reviewed the legality of affirmative action programs at the University of Michigan in the spring of 2003, 65 leading American businesses submitted a legal brief as *Amici Curiae* supporting the University's policy, which argued:

> "In the practical experience of the *Amici* businesses, the need for diversity in higher education is indeed compelling. Because our population is diverse, and because of the increasingly global reach of American business, the skills and training needed to succeed in business today demand exposure to widely diverse people, cultures, ideas and viewpoints. Employees at every level of an organization must be able to work effectively with people who are different from themselves." [1]

Educators also have debated for decades about the educational value of racial and ethnic diversity since the US Supreme court of ruling in the *Brown v. Board of Education of Topeka, Kansas* in 1954 (See Appel, Cartwright, Smith and Wolf (1996); Chang (1999); Chang, Witt-Sandis and Hakuta (1999) for a survey and examples of recent research). Despite significant welfare and public policy implications of the debates about how diversity affects productivity in workplace and learning in schools, there has not been extensive theoretical or empirical research on the potential mechanisms of this possible linkage. Furthermore, too much emphasis has been put on ethnic and racial diversity and other group characteristics such as age, gender and skill diversity, which can also be beneficial in workplace, have been left under-explored.

Researchers in the field of labor and organizational economics have assessed both the pros and cons of teamwork on productivity and have also explored the impact of team

[1] See "Brief for *Amici Curiae* 65 Leading American Businesses in Support of Respondents" in the case of Gratz v. Bollinger, University of Michigan, et al. (No. 02-516) and Grutter v. Bollinger, University of Michigan, et al. (No. 02-241) submitted to Supreme Court of the United States by attorneys of *Amici Curiae* on February 18, 2003.

composition on performance. Teamwork utilizes members' comparative advantages and skill specializations and facilitates knowledge transfer, thus improving productivity.[2] On the other hand, free-riding[3] and coordination problems[4] arise and tend to hinder efficient operation of a team. Diversity may improve productivity through knowledge sharing and task coordination if diversity exists in skill dimensions or knowledge sets. Non-skill-related demographic diversity, however, may harm productivity by raising communication costs and making peer pressure less effective. Evaluating the effect of diversity on overall performance thus requires careful empirical examination of how group characteristics can change the tradeoff between the benefits and costs of teamwork by affecting the way in which group members interact with each other. Such relationships are likely to be dependent on specific contexts, namely, the nature of technology, the institutional environment and other practices adopted in the workplace. Therefore, the first step to analyze the relationships is to measure the effectiveness of group work and understand the specific context in which groups are operating. Unfortunately, empirical research on teamwork in different contexts is scarce.

Several factors have contributed to the absence of empirical research on the relationship between workforce diversity and productivity. First, group-level or work-unit-level data are needed to compare productivity with group characteristics. Such data are usually hard to obtain due to the lack of systematic record of group activities either in educational or work environments. As a result, the effect of workforce diversity on performance has been empirically examined mostly using aggregated statistics at the industry or firm level (e.g. Heckman and

[2] Some of these gains or tradeoffs among the benefits of teamwork are discussed in Lazear (1999) and Garicano (2000).

[3] For a thorough discussion of free-riding, see Alchian and Demsetz (1972), Holmstrom (1982), and Kandel and Lazear (1992). Free-riding problems arise especially when team members' actions are unobservable or team members do not have means to punish free-riders (e.g. social pressure).

[4] For a theoretical model of how coordination costs affect the degree of specialization, see Becker and Murphy (1992).

Payner (1989), Leonard (1984), Conrad (1995), Holzer and Neumark (2000)). One exception is Hamilton, Nickerson and Owan (2003, 2004) where detailed data on worker productivity and demographic information from a garment factory are used to analyze the effect of the change in the factory work organization and pay scheme from individual piece rate to team piece rate. Second, even with available data set, there are further challenges in correctly identifying team effects or peers effects, mostly due to the self-selection problem. Unless randomly assigned, subjects of teams usually self-select into teams or groups. For example, workers might self-select into teams with members from same ethnic background or with members of similar productivity level; students (or their families) self-select into districts, schools, or classes based on their capabilities and expectations.

In this paper, we contribute to the literature by providing empirical evidence that age and gender diversity matter in group performance and group learning in certain team environments. We do not find any significant effect of ethnic or racial diversity on performance. We use class performance data from an undergraduate introductory management course and the students' personal records from the registrar's office at a private US university to analyze diversity and knowledge spillover effects. Individual-level performance is measured by the scores on three exams while group-level performance is assessed by the grades on group homework and other group projects. A major advantage of our dataset is the exogenous assignment of groups, which rules out the troublesome yet common self-selection issue in team literature.

According to our results, male-dominant groups performed worse both in group work and in individually taken exams than female-dominant and equally-mixed gender groups, after controlling for other group characteristics. Individual members from a group with more diversity in age and gender scored higher in exams. We also find that younger students and those students

with high SAT scores benefit more from gender diversity while older students and students with low SAT scores benefit more from age diversity.

We also contribute to the contract literature by demonstrating that the governance choice by group members can have a significant impact on group performance. In our classroom setting, the role of governance form is to create the expectation that free-riding behavior may be punished by peers. Each group chooses their own group contract form – members of "autonomous" groups receive equal grade for their group work while those in "democratic" groups can adopt differentiated point allocation approved by the majority of the members thus providing a proper mechanism to punish free riders and reward hard workers. Our OLS estimation results show a significant correlation between the choice of a democratic contract and the group and individual performance. Because the same unobservable factors that influence group performance may also influence a group's contract choice, we adopted treatment effects maximum likelihood estimation method to deal with the endogeneity problem. We asked teaching assistants (TAs) whether they discussed the positive aspects of the democratic group contract and talked about their positive experience with the contract during the small-class sessions led by TAs when groups were asked to select their group contracts. Then, we used the TAs' responses to the question as an instrument in the first stage and found a significant positive impact of the democratic group contract on the group performance.

The rest of the paper is organized as follows. Section 2 discusses the related literature; Section 3 introduces the content and structure of group work in the introductory management class; Section 4 explains data and variables constructed; Section 5 discusses empirical analysis of the data; Section 6 concludes.

II. Literature

Sociology and psychology have a long history of studying the effects of diversity. In various studies, the word *diversity* has been used to describe many types of differences among people. Organizational behavior researchers mostly focused on characteristics of group composition that are salient such as age, race, and gender (See Kanter 1977, Konrad and Gutek 1987, Bantel and Jackson 1989, Tsui, Egan, and O'Reilly 1992, Stangor, Lynch, Duan and Glass 1992, Cummings, Zhou, and Oldham 1993, Rothbart and John 1993, Pelled 1996, Williams and O'Reilly 1998 for theoretical frameworks and survey). Visibility is important when people use these characteristics to categorize others and form expectations for their behavior based on the categorization. Decision-making researchers, however, typically define diversity in terms of variation in expertise or information (See Gruenfeld, Mannix, Williams, and Neale 1996, Stasser, Stewart, and Wittenbaum 1995, Wittembaum and Stasser 1996, Robinson and Dechant 1997 for example for some leading works and survey). There are two major difficulties in the study of diversity. First, it is prohibitively difficult in most contexts to separate the effect of salient demographic differences from those differences in other, but likely correlated personal attributes such as ability, knowledge, and personality. As a result, it is often not feasible to identify the mechanism that mediates the effect of diversity. Second, in order to identify the mediating mechanism, most studies focus on a particular group process such as commitment, communication, and conflict. As a result, these studies in organizational behavior and psychology generally fail to derive clear implications for the overall impact on the group performance. Mannix and Neale (2005) describe the current state of diversity research in their survey as follows:

> As we disentangle what researchers have learned from the last 50 years, we can conclude that surface-level social category differences, such as those of race/ethnicity, gender, or

> age, tend to be more likely to have negative effects on the ability of groups to function effectively. By contrast, underlying differences, such as differences in functional background, education, or personality, are more often positively related to performance—for example, by facilitating creativity or group problem solving—but only when the group process is carefully controlled. The majority of these effects have typically been explained in terms of potential mediators such as social integration, communication, and conflict. However, the actual evidence for the input-process-output linkage is not as strong as one might like.

Some evidence for the "economic" value of diversity in both demographic and non-demographic dimensions can be found in Hamilton, Nickerson and Owan (2003, 2004). They find a significant impact of skill diversity on team productivity: teams with more heterogeneous composition in member's individual skill level were more productive holding other characteristics constant.[5] They also show that both age and ethnic diversity had slightly negative effects in team productivity, but these effects became insignificant once team fixed effects were included. These results could indicate higher communication costs in demographically diverse teams. The negative effects of diversity on communication among employees have been noted by a couple of studies. Zenger and Lawrence (1989) examined the effects of age diversity on technical communication and found a negative correlation between age heterogeneity and frequency of technical communication. Hoffman (1985) also found that increasing black representation among supervisors in the federal government offices was negatively associated with interpersonal communication. Ethnic diversity could also affect the trust or effectiveness of peer pressure. In a field study conducted by Bandiera, Barankay and Rasul (2005) on the

[5] They also find that the most able team member had more influence over team productivity than the least able team member. Two explanations were offered: (1) knowledge possessed by the most able workers was most broadly and most effectively shared among team members and (2) the high-productivity member had stronger bargaining power in setting the work pace because participation in teamwork was voluntary.

productivity of fruit farm workers in a teamworking environment, workers cooperated more in a smaller groups with co-workers who had the same nationality.

One important finding in our paper is the significant role of gender composition on both group and individual performances. The finding that gender diversity has a favorable impact on group performance and learning is striking. Other work has found that gender heterogeneity led to a higher level of conflicts (Alagna, Reddy and Collins (1982), Pelled (1997)), a lower performance in some cognitive team activities (Clement and Schiereck (1973)), and less creative works (Kent and McGrath (1969), Hoffman, Harburg and Maier (1962)). The only exception is Hoffman and Maier (1961) where sex diversity improved the quality of the groups' solutions on five cognitive tasks in their study of 41 four-person groups.

Also related to our study is the literature on peer effects. In the field of education research, whether students are affected by the achievement of their classmates has long been of great interest. If peer effects exist, group composition in terms of skill and knowledge levels should matter for the performance of individual students. Arcidiacono and Nicholson (2002) consider classes of a medical school as the units of peer groups and used students' MCAT scores as the outcome variable to test peer effects. The formation of such peer groups was highly endogenous since students' choices in school applications depend mainly on their ability and preferences. Arcidiacono and Nicholson (2002) find that the average ability (measured by average MCAT scores of the school) of a medical student's peer group did not affect the student's board score when school fixed effects were used. In her study of peer effects in elementary school classrooms, Hoxby (2000) deals with self-selection problem by using exogenous variations in peer group composition – changes in gender and racial compositions of a grade in a school over adjacent years. She finds positive and significant peer effects on a

student's performance on standard tests, and such effects appeared to be stronger inside a racial group than among racial groups.

The literature on workforce and school diversity (in race and ethnicity) often implicitly assumes that: (1) diversity in demographics naturally generates various sets of knowledge and perspectives; and (2) diverse sets of knowledge and perspectives can be effectively used to raise productivity. Arcidiacono and Vigdor (2003), for example, test the hypothesis that a diversified student body in college contributes to the accruement of "diversity capital" [6] and improves the students' post graduate labor market outcomes, which are measured by earnings, postgraduate education level and individual's subjective evaluation of life and employment satisfaction.[7] They find no positive effect of collegiate diversity on above outcomes for an individual[8].

Empirical studies in higher educational environment show that peer effect may be expressed differently in female peer groups and in male ones (see for example, Arcidiacono and Nicholson (2002)). Gneezy, Niederle and Rustichini (2003) conducted controlled experiments to test both genders' productivity changes when the working environment shifted from a non-competitive one (piece rate payment scheme) to a competitive one (tournaments). Significant gender gap, with men performing better than women, was observed in a competitive environment even though such gender difference did not occur under non-competitive

[6] Arcidiacono and Vigdor (2003) introduced the concept of "diversity capital" which measures an individual's ability to generate surplus in interacting with other agents from different ethnic, racial, or socioeconomic backgrounds.

[7] In order to control for unobservable factors and possible sorting phenomenon, Arcidiacono and Vigdor (2003) assume that racial diversity could vary significantly within an institution and controlled for both college and major fixed effects.

[8] Arcidiacono and Vigdor (2003) provide several interpretations for their empirical result: First, "diversity capital" may not exist or at least does not play an important role in college education. Second, since students are more likely to interact with peers with similar college preparation levels, diversity capital may be accumulated only in colleges where knowledge levels across racial groups overlap to some extend.

environment.[9] These findings may imply that women respond differently to teamwork environments and governance.

Finally, our study of how the group governance form affects the group performance could be evaluated in comparison with other works on team incentive pay. In their study of a garment factory, Hamilton, Nickerson and Owan (2003) find that workers' productivity improved significantly after a team production system with team pay replaced the traditional individual piece rate scheme. Different results were obtained in the study conducted by Bandiera, Barankay and Rasul (2005) - productivity of workers from a UK fruit farm was observed both in teams with a relative pay scheme[10] and under individual piece rate scheme. Individual productivity was significantly lower under teamwork with relative pay scheme than under individual piece rates. Workers withheld production cooperatively to avoid a decrease in piece rate, internalizing the negative externality their efforts impose on others. Finally, Knez and Simester (2001) find that the incentive bonus program based on a firm-wide performance goal introduced at Continental Airlines raised employee performance despite the threat of free-riding- possibly due to mutual monitoring facilitated by autonomous work groups.

III. Group Work in Management 100

We study the effect of group composition and knowledge spillovers on individual and group performance using class performance data from an undergraduate introductory management course (MGT100) that two of the co-authors taught during 2001 – 2004 in Olin

[9] In a similar study, Gneezy and Rustichini (2002) administered experiments to test gender differences in young children's running performance in a competitive environment and found that competition increased boys' performance more than girls' performance; moreover, a girl's running improved only after being paired to compete with a boy (instead of a girl).

[10] Workers were paid according to their relative fruit picking rate as compared with the average picking rate of their team. Thus, even a low productivity worker was paid well if his team's overall productivity was low.

School of Business, Washington University in St. Louis. MGT 100 is a course designed to offer a thematic introduction to the world of business and to provide an overview of many key economic concepts covered by various business courses including equilibrium, efficiency, competition, monopoly, asymmetric information, externalities, transaction costs, asset specificity, human capital, strategic positioning, product differentiation, commitment, etc. The course is required for all business-major and business-minor students and its instruction format has been largely constant over time. All entering business-major students as well as other interested freshmen from other schools take the course together in a large auditorium in the fall of their first year. MGT100 typically has an enrollment of 150 to 200 students in the fall semester of each year. Students with a minor in business and students who are transferring to the business school generally take the course in spring semesters. The course, which was initially an elective for business-minor students, became a required course as a part of the curriculum reform beginning with the class of 2004. This change caused the class size in spring semesters to grow in our observation period from around 40 to 100.

In addition to attendance in lectures at regularly scheduled class hours, MGT100 requires participation in discussion sessions every Friday where teaching assistants (TAs) lead discussions and run classroom games in small subsections of 12 to 26 students. These sessions are not designed to be lecture reviews. TAs lead discussions of business cases or run simulation games that are related to lectures to enhance student understanding of class concepts in a small subsection setting. In the second week of scheduled classes, TAs assign students in their subsections into four or five member teams. TAs are instructed to make this team assignment procedure random with the possibility of switching team members to maintain gender diversity

within groups.[11] Despite this instruction, gender makeup varies substantially in groups due to differences in gender composition within and across subsections and due to the attrition of group members after the first week.[12]

MGT100 requires a substantial amount of group work including three assignments and a group project. The group work requirement in the course accounts for 28 percent of total course grade. For their group project, each group selects one of the business books listed by the instructor (or a substitute approved by the instructor) on which to write a paper and to make an oral presentation.[13] Students are instructed to use course concepts in their paper to analyze the market environment of the firm, to discuss features of the firm as a response to the market environment, and to examine the strategic options that are available to the firm. This group project is completed by the end of the semester. Since instructors show students the evaluation forms used in grading of their group projects at an early stage of their work, students know how their work will be evaluated before they actually start writing their papers and preparing for oral presentation. The evaluation forms were not changed during our observation period; thus, the same evaluation criteria were applied to all five classes in our sample. In addition, each group also works together on three assignments during the semester. Despite the relatively heavy workload, each assignment is allocated only one-percentage weight of the total course grade and

[11] Although this exogenous assignment of students into teams rules out the problems associated with self-selection, there may be two possible sources of bias. First, TAs have some discretion about changing the composition of teams to raise gender diversity. However, we believe it is unlikely that TAs use this discretion to make any other consideration or give favorable treatments to some students (e.g. let friends form teams etc.). TAs may know only a few students in the first week of the semester especially since all students are freshmen in the fall semester and those in the spring semester are all from different schools and programs. Second, a few students drop the course after the team selection. This self-selection could cause serious bias if this decision is related to the composition or the performance of teams. This is not likely to be the case since most attrition takes place very early in the semester before students start working on their group projects. Although we do not have a precise figure, the average attrition rate is less than 10 percent.

[12] When students drop the course after the first week, we may have teams with only three or even two members. There were three 2-member groups out of 102.

[13] In fall 2004, they were *Liar's Poker* (Salomon Brothers); *Hard Landing* (airline industry); *Swoosh* (Nike); *The Merck Druggernaut; McDonalds Behind the Arches*; and *getting the bugs out* (Volkswagen).

it is graded only as "satisfactory" (1 point) or "unsatisfactory" (0 point).[14] Groups are given one chance to re-do their homework if they receive an unsatisfactory grade in the first grading. A typical time frame of the group work is presented in Table 1.

Clear logical and analytical thinking, a deep understanding of class concepts and good exam preparation and taking skills are necessary and largely sufficient for successful performance in exams in MGT100, but success in group work requires somewhat different and additional capabilities. The group paper requires good writing skills in addition to understanding and applying class concepts to a specific firm and industry. The group presentation necessitates a lot of practice and speaking at ease in front of an audience. Most importantly, successful completion of the group project requires responsibility to meet regularly, effective organization, coordination and communication among group members and good time management skills. Regularly assigned group homework entails long hours of hard work, and is very useful for understanding of class concepts and for exam preparation.

A novel aspect of the group work in MGT100 is that each group selects one of the following three team organizational structures to run and manage its team throughout the semester: autonomous team, democratic team, or manager-led team. The choice of the organizational form defines the set of voting rules that team members must follow to coordinate their actions. A "consensus-based" autonomous team requires that all decisions made within the team receive unanimous support and all team members are awarded the same grade for their group projects. A "majority rule-based" democratic team requires that all decisions require

[14] The reasons for the low weight on homework assignments are explained to the students as follows: first, problems similar to those in the homework will appear in the exams and, therefore, the students who work hard on the assignments will be rewarded by good exam grades while those who shirk will be punished by bad exam outcome. Second, in case some members free-ride on the others' work, low weight will help ease the stress caused by resentment. But one substantial reason that the instructors do not tell students is that low weight discourages students from illegitimately obtaining the answers to homework assignments that are used every year with very little modifications.

majority (i.e., more than 50% of votes). The members of democratic team vote at the end of the semester on the allocation of grades to individual team members. A manager-led team assigns coordination responsibility to one member who is elected by the team at the beginning of the course. The manager evaluates other team-member contributions and assigns grades to team-members at the end of the course. Similarly, team-members evaluate their manager's performance at the end of the course and assign a grade to the manager. The instructor assigns an overall grade for the group project, but individual grades of group members then can be modified by at most one letter grade, depending on the team structure adopted and team's self-grading.[15] The differentiated point allocation rule allows the democratic and manager-led team group members to punish free-riders. After TAs explain the rules of each team structure and discuss their implications, groups are asked to discuss and choose their team governance form in the second week of the semester. When all members agree and sign the team contract, they cannot change their contract type under any circumstance. If group members cannot reach the agreement during the discussion section, the democratic team contract is automatically assigned.[16]

IV. Data Description

In this paper, we utilize a novel data set constructed from academic and demographic records of 431 undergraduate students who took MGT100 over five semesters during 2001 – 2004 (Spring 2001, Spring 2002, Spring 2003, Spring 2004, and Fall 2004). Our data include grade sheets that provide detailed information on students' grades on two midterm exams and a

[15] For example, a team could receive a B (or 85 in the 100 point score) from the instructor. If the team is autonomous, all members would receive a B (85) for the team project. If the four-member team is democratic or manager-led, one member could conceivably receive an A (95), two members a B (85), and one member a C (75).
[16] Classes in spring 2003 and 2004 were allowed to have one week extension when they were unable to reach agreement.

final exam as well as group assignments and the group project that includes a paper and oral presentation. In addition, each student is matched to his or her corresponding group and each group's contract choice (democratic, manager-led, or autonomous) is also recorded. We have 102 student groups, out of which only 3 had chosen the manager-led group structure. Since both manager-led and democratic groups adopt proper mechanisms to prevent free-riding and reward hard working members, groups that chose the manager-led contract type were categorized as democratic type group in our empirical analysis. Unfortunately, the group contract choice information is missing for the Spring 2002 class; thus the observations from this class (64 individuals and 15 groups) were excluded from analysis whenever the contract choice variable was involved.

We augmented the class performance data with the records we obtained on students' academic and demographic backgrounds from the university registrar office. Information on a student's academic background includes: college status (freshman, sophomore, etc.) at the time of enrollment in MGT100, semester and cumulative GPA[17] prior to taking MGT100, SAT score[18], courses taken prior to MGT100, and the primary division[19] of the student. Demographic information consists of age, gender, ethnic origin, and housing location (living on or off campus).

Based on individual members' information, we constructed several group level variables to measure demographic and knowledge/skill composition of groups. These group characteristic variables are used to analyze the effect of group composition on group performance and group's

[17] In one of the classes (the Fall 2004 class), students enrolled in MGT100 were predominantly freshmen; thus no GPA information was available for these students and it was not used in our analyses.

[18] For those students who took ACT instead of SAT, the ACT scores were converted into SAT according to the Comparison Table provided by College Educational Board. The comparison table was constructed based on 103,525 test-takers who took both SAT and ACT between October 1994 and December 1996. Equivalent SAT and ACT scores are those with the same percentile ranks for a common group of test takers.

[19] The primary divisions of the students in our sample pool are Schools of: Art and Sciences, Fine Art, Architecture, Business, and Engineering.

contract choice. Most group level variables were constructed by taking the average and standard deviation of individual level variables for the group members. Thus, we created variables measuring average age level in groups, average class status, group male and female dominance measures, average SAT levels, proportion of domestic Caucasian students in groups as well as standard deviations in age and in SAT scores in each group. Group age and status variables are used to indicate a group's knowledge level and knowledge diversity. SAT scores at group levels are intended to capture group's skill composition. Table 2 presents descriptive statistics for all the variables used in the empirical analysis.

V. Empirical Analysis

In this section, we analyze three issues regarding the benefits and costs of teamwork. First, we investigate whether and how a group's composition affects its performance in group work. We focus on four measures of group composition: heterogeneity in gender, age, race and capabilities. We study whether heterogeneous groups perform better than comparable but more homogeneous groups. Heterogeneous groups may face setbacks caused by communication and coordination problems but they may also enjoy the benefits of an enhanced knowledge and skill set. Second, we address the extent of knowledge spillover from groups to their members. Which group characteristics contribute to learning and personal growth of individual members? Which individuals are likely to be benefited most from group learning? We answer these questions by analyzing the impact of group characteristics on exam performance of students. Finally, we investigate how a group's governance structure may have affected its operation and performance. As discussed earlier, every group chose a form of governance shortly after group formation by signing one of the three types of contract. We study what group characteristics were influential

in a group's contract decision, and how important the contract choice was in the performance of students in their group projects. We now address these issues in turn.

A. How does group composition affect group performance?

We first analyze the impact of group composition characteristics on group performance. We model the group performance as:

$$(1)\ GroupScore_{jk} = X_j\beta + Dvsty_j\alpha + Class_k + \varepsilon_{jk}$$

Subscript j and k are the indexes for the j^{th} study group in the k^{th} semester. $GroupScore_{jk}$ is the group performance measure constructed by taking the weighted average of a group's homework and project scores. X_j is a vector of control variables for group j including average age of members and average SAT scores of members. $Class_k$ is a vector of dummy variables to account for class fixed effects; including such fixed effects in the regression can control for factors including variations in two instructors' grading standards. In addition, we constructed an indicator variable for groups with 5 members to control for the potential differences in group work within "large" groups. While estimating all our models, we used the demeaned forms of age, SAT scores, and their averages to avoid reporting distorted coefficient estimates for other variables that are used in interaction terms with age and SAT scores.

Vector $Dvsty_j$ consists of variables that capture two different types of heterogeneity in group j's member characteristics – in demography and in knowledge, ability, or skill. We use two variables to measure gender composition of a group, which are defined as *Malegroup* = max {0, proportion of male students- 50%} and *Femalegroup* = max {0, 50% - proportion of male students}. These variables are constructed to indicate the degree of male dominance and female

dominance within a group[20]. We use the standard deviation of group members' ages, to measure age diversity within a group. This variable also reflects the differences in members' knowledge levels within a group[21]. The standard deviations of SAT scores are computed for every group as measures of skill diversity, and the proportion of domestic Caucasian students within a group is used to indicate racial or ethnic heterogeneity[22].

The first two columns in Table 3 present the estimates of the impact of group composition characteristics on group performance. Group average SAT scores had positive effect on group performance under both specifications but it is not significant in any of the specifications. Male dominant groups tend to perform worse in their group projects than mixed-gender groups but female dominance did not affect group performance. These results indicate that gender diversity in groups possibly contributed to the successful performance in group projects.[23]

Three interpretations are possible for this gender effect. First, it may simply mean that male students are lazier than female students on average. Groups perform well if they have a couple of hard workers and mixed-gender or female-dominant groups may tend to satisfy this condition. This interpretation is probably not completely true because male students do not perform significantly differently in exams as we show later. Second, it may be the case that women are better in cooperating, organizing and coordinating activities and male-dominant

[20] Initially, we constructed a variable to measure group gender diversity by taking the absolute value of the difference between proportion of male members in a group and 50%. Since this variable showed significant impact over group performance, we further constructed these two variables to separate the effects of male and female dominance on group performance.

[21] We also used students' college status to reflect on knowledge diversity but age variable seems to be a better proxy of general experience and maturity and has more variation. Age and college status are highly correlated in our data set and we do not have any older "non-traditional" students.

[22] We used other measures of ethnic diversity as well, such as the number of different ethnic backgrounds within a group but did not see any substantial changes.

[23] We also estimated models with dummy variables controlling for all male and all female groups as well as groups with only one female and one male member instead of using our gender dominance variables. Results are similar to the reported ones, indicating that all male groups and groups with only one female member perform worse in their group projects.

groups are more likely to suffer coordination failures or free-riding. [24] Third, male students may be more motivated to work hard in the presence of female students. This may be due to competing for the attention of female students in the group or being more susceptible to peer pressure from female students when students socialize within groups.

Although we cannot directly differentiate between these three possible interpretations of gender effect found in our results, we conjecture that the gender effect explained by the second interpretation may diminish as students get older. More mature and advanced male students may have better coordination and organization skills than their yournger counterparts. Thus, we may expect the impact of gender diversity on group performance to be more prominent for younger male students who need the assistance of female students in organizing teamwork. In order to test this hypothesis, we add interaction terms between *malegroup, femalegroup* and average age of a group. The second column in Table 3 reports the OLS regression results obtained by adding these interaction terms. According to the results, the coefficient for the interaction term of male group and group average age is positive and statistically significant; showing that male dominance in younger groups was especially damaging to group performance, thus, indicating the importance of maturity in successful teamwork. The role of maturity is also apparent in the female dominant groups as indicated by the positive and statistically significant coefficient on the interaction term between female group and group average age. The latter result, however, implies that female dominance is particularly valuable in older teams. Additional results presented in the first two columns of Table 3 show that age and racial diversity measures do not have any significant effect on group performance. Further, we do not find any significant effect

[24] We expected that, in addition to the teamwork capabilities discussed, female students might have better writing, composition and verbal skills that may be especially helpful in writing the group paper. A cursory look at the differences among genders in SAT verbal and SAT math scores in our sample indicated that female and male students performed similarly in the verbal section of the SAT, however, male students' math scores were statistically significant and higher than the female students' math scores.

of group size on group performance. Thus, the claim that groups with 5 members may face greater danger of free-riding and coordination difficulties relative to smaller groups is not supported.

An important feature of our data set is that groups choose their governance structures by signing either democratic or autonomous contract types. It is thus natural to ask whether the contract types have significant impact over group performance. The democratic governance form implements a mechanism to punish free-riders and reward hard workers within a group. It defines an incentive system for group members to work harder; shirking can result in lower project grade based on other team members' evaluations in the end of the semester. Such punishment and reward mechanism does not exist in an autonomous group. The last two columns in Table 3 present results including the contract choice variable. According to these results, teams with democratic governance form performed better in group projects. However, we are concerned that the effect of contract choice on group performance may be overestimated due to the likely endogeneity of the contract choice variable – the same unobservable factors that affect group performance may also influence a group's contract choice. For example, an individual who expects to work less for the course may insist on the choice of autonomous contact type where free riders are less likely to be punished; at the same time, groups with such potential free-riders are likely to perform poorer. This endogeneity issue will be addressed in a later section by utilizing an endogenous treatment effect regression model.

B. Knowledge spillovers and group learning

In this section, we address the extent of knowledge spillovers from groups to individual group members. We investigate which group characteristics may have contributed to the learning

and personal growth of individual members. We model individual performance in exams using the following Ordinary Least Squares (OLS) specification with class fixed effects:

$$(2)\quad IndScore_{ijk} = Z_i\gamma + X_j\beta + Dvsty_j\alpha + Class_k + \eta_{ijk}$$

Subscript i is the index for the i^{th} student in the sample. Vector Z_i consists of individual level control variables including age, gender dummy, SAT score, and off campus housing dummy variable. [25] X_j is a vector of group level characteristics for group j which individual i belongs to. As in equation (1), the vector of $Dvsty_j$ includes group diversity measures in demography and knowledge. $Class_k$ is a vector of dummy variables accounting for class fixed effects. Individual performances of students are measured by $IndScore_{ijk}$, which is the weighted average of the three exam scores – two midterms and one final.[26]

The first column in Table 4 reports the baseline regression results. An individual's SAT score is a good indicator of how well he or she performed in the MGT100 exams. The effect is significant at 1% statistical level and robust across all alternative model specifications. A female student performed slightly better in exams than a comparable male student, although the difference is not statistically significant. Individual's age also has a positive and significant effect over exam performance. Since age is highly correlated with college year level, the positive effect of age over course performance reflects an individual's knowledge level– a higher level student has accumulated more relevant knowledge and experience, enabling him or her to outperform younger students in exams. An engineering dummy variable was included in the model, indicating students majoring in engineering fields. Results show that engineering students

[25] Gender dummy = 1 for a male student and 0 for a female student.
[26] We also estimated an ordered probit model to take into account the possible non-linearity between performance measures and group and individual characteristics. Efforts to raise grades from 60 to 70 percent are likely to be different that efforts to raise grades from 80 to 90 percent. However, results of ordered probit estimation are qualitatively very similar to the OLS estimation and, thus, are not reported.

performed better in exams compared to students from other majors.[27] These students generally performed better at answering math-intensive questions and were likely to have higher analytical skills that were beneficial in exam performance.

Among variables indicating group characteristics, gender composition of the group an individual belonged to had noticeable influence over individual performance. Specifically, a student in a male-dominant group tended to do worse in exams compared to another student in a mixed-gender group. This finding is consistent with the previous results obtained using group level data. Mixed-gender teamwork improves performance in both group and individual tasks. In accordance with our earlier interpretations, it may be the case that enhanced organization and coordination within groups in the presence of female peers increase the knowledge spillovers. Or, this gender effect may be due to cross-gender socialization that better motivates both male and female students and leads to more learning through group work. In addition, similar to the results obtained regarding the group work, we find that racial diversity of groups do not affect individual's performance in the exams.

We find evidence of knowledge spillovers from groups to individuals. In particular, the standard deviation of members' ages within a group has a positive and statistically significant impact on member's performance in exams. Age heterogeneity within a group reflects the differences in group members' college year levels (i.e. knowledge accumulation in college), as

[27] None of the other groups of students majoring in arts and sciences, fine art, architecture or business perform significantly differently from the rest of the students. In other specifications we included a count variable indicating the number of different schools students in each group were majoring in, but this variable was never statistically significant. Our initial intuition was that group heterogeneity within major schools may increase the coordination and communication problems within groups, or on the contrary may contribute to the knowledge diversity.

well as the maturity of individuals.[28] When a group is composed of students from various knowledge levels, knowledge transfer is more likely to occur. Students who are more advanced can also take leadership and mentorship roles in their groups and facilitate knowledge transfer and contribute to improved exam performances. In contrast, the standard deviation of SAT scores within groups has no impact on individual or group performance. SAT scores reflect very basic analytical and cognitive skills which are much more correlated with students' natural ability. Such skills cannot be easily transferred through teamwork.

While it may be trivial that SAT score is a good indicator of individual's college coursework performance, we suspect that the impact of an individual's knowledge and skill set prior to entering college (measured by SAT score) on his or her college performance diminishes over time and is replaced by newer knowledge set obtained through the learning process in college. To test this hypothesis we added an interaction term between individual's SAT score and age while taking MGT100. This interaction term has a negative and statistically significant effect on exam performance, as presented in the second column of Table 4, providing evidence that the impact of SAT score over coursework performance diminishes over time as an individual advances in college.

The third column in Table 4 reports regression results where the model includes the binary contract choice variable. Results indicate that individuals in democratic groups scored significantly higher in exams compared to students in autonomous groups, holding other observable factors constant. This is consistent with our earlier results on the impact of contract choice on group performance. Contract choice, however, is likely to be an endogenous variable and the estimated coefficient is likely biased upward when more able students are also more

[28] We also measured age heterogeneity (and knowledge diversity) within groups using the ratio of maximum and minimum age among group members. This measure did not perform as well in our models.

likely to choose a democratic governance structure where free-riders are punished and hard workers are awarded. We investigate the contract choice decision and its impact in a later section in more detail.

C. Who benefits most from working in groups?

As discussed earlier, we observe that students benefit from group work; specifically students who are in groups with age heterogeneity and with mixed-gender composition perform better in their exams. In this section, we investigate which students are likely to benefit most from working in heterogeneous groups.

As shown in the first column of Table 4, we observe evidence of knowledge spillovers from groups to individuals. Age heterogeneity in groups, reflecting differences in knowledge accumulation and in student maturity, has a significant positive impact in individual exam performance. To further explore the effect of group age heterogeneity on individual members' performances, we raise the following two hypotheses: (1) knowledge flows from older students to younger students; and (2) more knowledge is transferred to students with lower skills or aptitude. Thus, although everyone benefits from age and knowledge diversity, younger and less able students are likely to benefit more from this experience.

We expect the first hypothesis to be true for the following reason: older students may have already taken other courses with similar subject matter and they are more comfortable with the concepts that are taught in MGT100. Thus, they are able to assist younger students while working on homework and preparing for exams. In order to test this hypothesis, we included an interaction term between age of a student and the standard deviation in age in the group he or she participated. The results, presented in the fourth column of Table 4, shows that this interaction

term is positive and statistically significant, thus, suggesting that *older* students, not the younger ones, benefit more from age heterogeneity in groups and that knowledge flow within groups is not as simple as hypothesized. One explanation for this effect is that older students may take more of a mentorship and leadership roles within groups and assisting other students actually may create bigger benefits to these students in terms of better understanding of the material. It is also possible that older students are expected to take more of a leadership role and thus work harder to avoid looking incompetent.

The second hypothesis, more knowledge is transferred to students with lower skills or aptitude, is not so trivial because it is also possible that smarter students pick up more knowledge from group work. But casual conversation with a number of students revealed a belief that group discussion of homework assignments was a helpful way to prepare for exams. If a group makes sure that its members all understand the solutions to homework, the students with lowest analytical and problem-solving skills are likely to benefit most from such group work. To test this hypothesis, we added an interaction term between SAT score of a student and the standard deviation in age in the group he or she participated. The results presented in column 5 of Table 4 shows that the coefficient of this interaction term is negative and significant, thus supporting our hypothesis.

Next, we examine who benefits most from high knowledge spillover in mixed-gender groups. In Section A, we already obtained a support for our hypothesis that the negative effect of male dominance decreases as students get older possibly because older and more advanced male students may have better coordination and organization skills that female students might generally bring in to group work, or because more mature male students are less likely to be motivated just by socialization with female students in groups. In accordance with this earlier

result, we expect to see a positive coefficient for the interaction term between age of a student and male dominance in the group he or she belonged. The results presented in Column 6 of Table 4 confirm our prediction.

Finally, we raise the following hypothesis: If some male students are motivated by the presence of female students, then smarter male students may be most affected by this gender incentive since smarter male students are more likely to out-perform their male peers. To test this hypothesis, we add an interaction term between male group and individual SAT score. The coefficient of this interaction term, as shown in the last column of Table 4, is negative and statistically significant at 5% level supporting our hypothesis. Note, however, that this result is not necessarily inconsistent with our other explanation of the gender effect: female students are more likely to bring in organizational and coordination skills that facilitate group work than male students. If analytical and problem-solving skills are complementary with organization and coordination skills because the latter facilitates knowledge transfer, students with high SAT scores benefit more by having someone who can better organize and coordinate study group activities.

D. What group characteristics influence groups' governance choice and what is the impact of governance choice on group performance?

As discussed earlier, a novel feature of our dataset is the choice of governance structures by group members at the start of the semester. We first find out which group characteristics influenced the group governance choice, then, we estimate the impact of group contract choice on performance. In order to identify the group characteristics that affected the choice of governance form, the following probit model is estimated:

$$(3)\ Contract_j^* = X_j\beta + Dvsty_j\alpha + TA_j + \omega_j$$

Let $Contract_j$ be the dummy variable for group *j*'s contract choice; $Contract_j = 1$ if group *j* chose democratic or management governance structure, $Contract_j = 0$ if the group chose autonomous contract. The Probit model assumes that $Contract_j = 1$ if $Contract_j^* \geq 0$ and $Contract_j = 0$ if $Contract_j^* < 0$.

As discussed earlier, the same unobservable factors that influence group performance may also influence a group's contract choice. Thus, to address this endogeneity issue, we constructed an instrumental variable, *TA influence,* for the group contract choice, using TAs' responses to the following brief questionnaire we sent in October 2005 to all the TAs of the MGT100 classes in 2003 and 2004[29]:

> *During Friday sessions, did you discuss the positive aspects of the democratic or manager-led group contract (contract choice that allows unequal point allocation within group) or talk about your positive experience with these types of contracts (e.g. you actually punished your teammate who did not work well)?*

TAs were asked to rate their answers in the 5-point Likert scale from 0 (strongly disagree) to 5 (strongly agree). Such variable can serve as a potential instrument because TAs' discussions about the effectiveness of different group contracts and their experience in previous years are less likely to be correlated with factors that affect students' performance or their group characteristics. The TA influence variable was constructed by adding up the responses from the two TAs who took charge of the subsection that each group belonged to. Hence, the variable takes integer value in the range between 0 and 10 – the higher value indicting higher encouragement by TAs to choose the democratic contract.

[29] In order to make sure that the TAs' responses were precise reflections of their influence, we only considered the responses from the TAs for the three most recent classes (one class in 2003 and two classes in 2004).

Since TA influence was shown to be a significant indicator of groups' contract choices in our Probit model estimation, we use it as an instrument to the contract choice in the group performance models. We consider the contract choice as an endogenous binary treatment variable over the outcome variable of group performance, and hence, used robust maximum likelihood estimation method.[30]. This method considers the effect of an endogenously chosen binary treatment (contract choice) on another endogenous continuous variable (group performance), conditional on two different sets of independent variables.

Results of the Probit model, without the TA influence as an explanatory variable, are reported in the first two columns of panel A of Table 5. Average age has a negative and significant coefficient in these first specifications implying that groups with younger students are more likely to choose the democratic team contract. One explanation is that younger students who have had less experience in group work and typically are more concerned about their grades may be more risk-averse and choose democratic governance form as a safeguard to protect themselves from possible free-riding. Another plausible story is that older students knowing that they have many other competing tasks (e.g. job hunting, leadership roles in student organizations or fraternities, etc.) may expect to put forth less time and effort into the course than younger students and selfishly support the choice of the autonomous team contract.

The second column reports the Probit estimation results including the interaction term between group average SAT score and average age. The inclusion of this interaction term reflects our expectation that the age effect discussed above will be mitigated if group members are more capable. Smart young students may not require a safeguard against free-riding knowing that they can finish projects alone even if they are aware of the risk. Smart older students might prefer the democratic team contract knowing that they can still contribute to the group work with

[30] Heckman's two-step method was also adopted with similar results.

their greater knowledge and skills even if they have tighter time constraint. The coefficient for this interaction term is positive and significant at 10% level supporting our prediction.

Results in the second column of Panel A indicate that groups with 5 members were more likely to choose the autonomous governance form, holding all other factors constant. This result appears to be counter-intuitive at first glance: we may expect group members in a larger group to be keener on selecting the contract type that punishes free-riding behavior, i.e., a democratic type since free-riding behavior tends to be more prominent in larger groups. However, the consequences of free-riding behavior are likely to be dissimilar for different size groups. In a three-member group, cooperation of all members is critical and the effect of potential free riding by any member is detrimental. As a result, a small group may prefer to choose the explicit punishment mechanism that safeguards its members against free riding. In contrast, in a five-member team where students might feel that they have more members than necessary, they may conclude that trust, implicit agreement, or peer pressure are sufficient to achieve their goal and get the task completed because shirking by one or two students is not so damaging. Such groups will choose the autonomous type to promote the sense of unity or partnership.

The results of the first stage of the treatment effect MLE (*i.e.* contract choices based on group characteristics) are in columns 3 and 4 and show that TA influence was an important factor in the choice of the group governance structure. If the TAs discussed the positive aspects of a democratic (or manager-led) type of governance form before students chose group contract, groups were more likely to choose the democratic governance form. TA influence was especially prominent in one professor's class (professor dummy = 0) which was held in fall 2004 while the variable does not have much impact in the other professor's class (professor dummy = 1) which were offered in spring 2003 and spring 2004. We believe that TA influence is a much noisier

measure for earlier classes because most TAs in spring 2003 and some TAs in spring 2004 noted that they did not remember the details of their discussion on group governance structures although they answered the question to the best of their knowledge.[31]

Panel B of Table 5 reports the results of the second stage that includes the similar set of control variables as in the first stage plus the contract choice as a treatment variable. After controlling for the endogeneity of the treatment variable (contract choice) using TAs' responses, a democratic group performed significantly better than its autonomous counterpart. The coefficient of the treatment variable (contract choice) is statistically significant at 5% level, and on average a democratic group obtained about 5 points (out of 100 points) higher on group project work compared to an autonomous group, holding all other group characteristics constant. Group members had motivation to work harder in order to avoid punishment associated with shirking under democratic governance form that punishes free riders and rewards hard workers; while in autonomous groups, such motivation does not exist and individual productivity could be further discouraged by potential free riding activities.

The rest of the results presented in Panel B of Table 5 are mostly consistent with our earlier results on group performance. Male dominant groups tend to perform worse in their group projects than mixed-gender groups, although the result is not always significant. Based on our earlier hypothesis that the effect of male dominance decreases as students get older because older and more advanced male students may have better coordination and organization skills similar to female students, we included the interaction terms between *malegroup*, *femalegroup* and average age of a group. The coefficient for the interaction term between *malegroup* and group's average age was positive and statistically significant; indicating that negative effect of

[31] In the questionnaire, we requested the TAs to state honestly if they could not remember or were not sure about the answer to the question.

male dominance is moderated for groups with older students. Again, racial diversity measure has no impact on group performance. An interesting result is that group average SAT scores now have positive and statistically significant effects on group performance under both specifications, indicating that smarter students tend to perform better in group assignments. This measure did not have a significant effect on group performance in OLS regressions.

VI. Conclusion

Our research implies that forming groups with mixed gender and some age differences is desirable in certain educational environments. If this result holds in many classroom environments, the shift to coeducation in American higher education and secondary schools in the second half of the 19th century likely facilitated learning and was beneficial to students, especially to male students, under the assumption that students were often given chances to participate in teamwork and encouraged to study in groups.[32] One current successful example that is consistent with our story of learning through age diversity is Montessori schools. The Montessori method employs multi-age classes, where older children may help younger ones, and the younger children are able to observe, learn from, and be inspired by their older peers. Although we find significant gender and age diversity in group work, we detect no effect of racial diversity.

Our study also provides interesting evidence on how diversity improves the learning process. First, we find that older students benefit more from age diversity. This is counter-intuitive because we normally expect knowledge to flow from older and more experienced people to younger and less experienced ones. We offer two explanations: (1) older students

[32] Details on the history of education in the United States can be found in Goldin (1999).

might learn more through their mentorship and leadership roles; and (2) older students work harder to avoid looking incompetent in their leadership role. Second, we find that smarter students benefit more from gender diversity. Although it is not necessarily counter-intuitive, this again allows multiple interpretations: (1) the analytical and problem-solving skills smart male students bring in to group work may be complementary with organization skills many women possess; or (2) smarter male students, knowing that they can outperform other male students, may be more motivated to work hard to attract attention of female members. Although these explanations are speculative, these findings at least suggest that learning process and learning incentives are not so simple and require further research.

Our paper also provides additional empirical evidence for the importance of governance form chosen for teamwork. Although previous work mostly studied the effect of pay schemes, what is important is the expectation that free-riders will be punished. According to our results, adopting the democratic group contract, which enables group members to punish free-riders effectively by differentiated point allocation, improved the scores on group project and group homework assignments by approximately 5 points (out of 100 points) on average compared to an autonomous group, holding all other group characteristics constant. What is important, however, is the expectation of punishment rather than actual punishment. Our data show that 32 out of 60 democratic groups actually adopted equal point allocation.[33]

Overall, our study confirms the notion that teamwork output is the result of multiple mechanisms that may interact, and that diversity in skills, knowledge, and demographic characteristics affect individual and group performance through multiple channels. Therefore,

[33] In most instances of differentiated point allocation, the maximum punishment (10 points deduction) was imposed on one student and these 10 points were reallocated equally among the rest of the team members.

managers and teachers need to choose group member composition and governance mechanisms carefully since these factors affect both the benefits and costs of teamwork.

References

Alagna, S.; Reddy, D.; and Collins, D. "Perceptions of Functioning in Mixed-Sex and Male Medical Training Groups" *Journal of Medical Education*, Vol. 57 (1982): 801-803.

Alchian, Armen A.; and Demsetz, Harold. "Production, Information Costs, and Economics Organization" *American Economics Review*, 62 (December 1972): 777-95.

Appel, M., Cartwright, D., Smith, D. G., and Wolf L. E. "The impact of Diversity on Students: A Preliminary Review of the Research Literature" Washington, DC: Association of American College and Universities, 1996.

Arcidiacono, Peter; and Nicholson, Sean. "Peer Effects in Medical School" NBER Working Paper 9025, July 2002.

Arcidiacono, Peter; and Vigdor, Jacob L. "Does the River Spill Over? Estimating the Economic Returns to Attending a Racially Diverse College", March 2003.

Bandiera, Oriana; Barankay, Iwan; and Rasul, Imran. "Social Preferences and the Responses to Incentives: Evidence from Personnel Data" *Quarterly Journal of Economics*, Vol. 120 (3) (August 2005), 917-962.

Bantel, K.; and Jackson, S. "Top Management and Innovations in Banking: Does the composition of the team make a difference?" *Strategic Management Journal*, Vol. 10 (1989): 107-124.

Becker, Gary S.; and Murphy, Kevin M. "The Division of Labor, Coordination Costs, and Knowledge" *The Quarterly Journal of Economics*, Vol. 107, No. 4 (Nov., 1992), 1137-1160.

Chang, M. J. "Does Racial Diversity Matter? The Educational Impact of a Racially Diverse Undergraduate Population" *Journal of College Student Development*, (1999), 40, 377-395.

Chang, M.J.; Witt-Sandis, D.; and Hakuta, K. "The Dynamics of Race in Higher Education: An Examination of the Evidence" *Equity and Excellence in Education*, 32(2), (1999), 12-16.

Clement, D.; and Schiereck, J. "Sex Composition and Group Performance in a Signal Detection Task" *Memory and Cognition*, Vol. 1 (1973): 251-255.

Conrad, Cecilia. "The Economic Cost of Affirmative Action" *Economics Perspectives on Affirmative Action*, (1995), Margaret Simms, Editor (Washington, DC: Joint Center for Political and Economics Studies).

Cummings, A.; Zhou, J.; and Oldham, G. "Demographic Differences and Employee Work Outcomes: Effects of Multiple Comparison Groups." Paper Presented at the Annual Meeting of the Academy of Management (1993), Atlanta, GA.

Garicano, Luis "Hierarchies and the Organization of Knowledge in Produciton" *Journal of Political Economy*, Vol.108, No.5, (2000): 874-904.

Gneezy, Uri; and Aldo Rustichini, "Gender and Competition at a Young Age," mimeo, 2002.

Gneezy, Uri; Niederle, Muriel; and Aldo Rustichini, "Performance in Competitive Environments: Gender Differences" *The Quarterly Journal of Economics*, August 2003, 1049-1073.

Goldin, Claudia. "A Brief History of Education in the United States" NBER Working Paper Series on Historical Factors in Long Run Growth. Historical Paper 119 (1999).

Gruenfeld, D.; Mannix, E.; Williams, K.; and Neale, M. "Group Composition and Decision Making: How Member Familiarity and Information Distribution Affect Process and Performance" *Organizational Behavior and Human Decision Processes*, Vol. 67 (1996): 1-15.

Hamilton, Barton H.; Nickerson, Jack A.; and Owan, Hideo. "Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation." *Journal of Political Economy*, Vol. 111, no. 3., (2003), 465-497.

Hamilton, Barton H.; Nickerson, Jack A.; and Owan, Hideo. "Diversity and Productivity in Production Teams", manuscript, Washington University in St. Louis, November 2004.

Heckman, James J.; and Payner Brook S. "Determining the Impact of Federal Antidiscrimination Policy on the Economic Status of Blacks: A Study of South Carolina" *The American Economic Review*, Vol. 79, No. 1 (March, 1989), 138-177.

Hoffman, L.; Harburg, E.; and Maier, N. "Differences and Disagreement as Factors in Creative Group Problem Solving" *Journal of Abnormal and Social Psychology*, Vol. 64 (1962): 206-214.

Hoffman, L.; and Maier, N. "Quality and Acceptance of Problem Solutions by Members of Homogeneous and Heterogeneous Groups" *Journal of Abnormal and Social Psychology*, Vol. 62 (1961): 401-407.

Holmstrom, Bengt. "Moral Hazard in Teams" *Bell Journal of Economics* 13 (Autumn 1982): 324-40.

Holzer, Harry; and Neumark, David. "Assessing Affirmative Action" *Journal of Economic Literature*, Vol 38, No. 3 (Sep. 2000), pp. 483-568

Hoxby, Caroline. “Peer Effects in the Classroom: Learning from Gender and Race Variation” NBER Working Paper 7867, August 2000.

Kandel, Eugene; and Lazear, Edward P. “Peer Pressure and Partnerships” *The Journal of Political Economy* Vol. 100, No. 4 (Aug., 1992), pp. 801-817

Kanter, R. “Some Effects of Proportions on Group Life: Skewed Sex Ratios and Responses to Token Women” *American Journal of Sociology*, Vol. 82 (1977): 965-990.

Kent, R: and McGrath, J. “Task and Group Characteristics as Factors Affecting Group Performance” *Journal of Experimental Social Psychology*, Vol. 5 (1969): 429-440.

Knez, Marc, and Simester, Duncan “Firm-Wide Incentives and Mutual Monitoring at Continental Airlines” *Journal of Labor Economics* Vol.19, No.4 (2001):743-72.

Konrad, A.; and Gutek, B. “Theory and Research on Group Composition: Applications to the Status of Women and Minorities” In S. Oskamp and S. Spacapan (Eds.), *Interpersonal Processes* (1987): Newbury Park, CA: Sage.

Lazear, Edward P. “Globalization and the Market for Team-Mates” *Economic Journal* 109 (1999): C15-40.

Leonard, Jonathan S. “Antidiscrimination or the Reverse Discrimination: The Impact of Changing Demographics, Title VII, and Affirmative Action on Productivity” *The Journal of Human Resources*, Vol. 19, No. 2 (Spring 1984), 145-174.

Mannix, Elizabeth; and Neale, Margaret A. “What Differences Make a Difference?” *Psychological Science in the Public Interest*, Vol. 6, No. 2 (October 2005): 31-55.

Pelled, Lisa H. “Demographic Diversity, Conflict, and Work Group Outcomes: An Intervening Process theory” *Organization Science*, Vol. 7 (1996): 615-631.

Pelled, Lisa, H. “Relational Demography and Perceptions of Group Conflict and Performance: A Field Investigation” *International Journal of Conflict Management* Vol. 7, Iss. 3; 230-246.

Robinson, G.; and Dechant, K. “Building a Business Case for Diversity” *Academy of Management Executive*, Vol. 11, No. 3 (1997): 21-31.

Rothbart, M.; and John, O. “Intergroup Relations and Stereotype Changes: A Social-Cognitive Analysis and Some Longitudinal Findings” In P. Sniderman and P. Tetlock (Eds.), *Prejudice, Politics, and Race in America* (1993), Stanford, CA: Stanford University Press.

Stangor, C.; Lynch, L.; Duan, C.; and Glass, B. “Categorization of Individuals on the Basis of Multiple Social Features” *Journal of Personality and Social Psychology*, Vol. 62 (1992): 207-218.

Stasser, G.; Stewart, D.; and Wittenbaum, G. “Expert Roles and Information Exchange during Discussion: The Importance of Knowing Who Knows What” *Journal of Experimental Social Psychology*, Vol. 57 (1995): 244-265.

Tsui, A.; Egan, T.; and O’Reilly, C. “Being Different: Relational Demography and Organizational Attachment” *Administrative Science Quarterly*, Vol. 37 (1992): 549-579.

Wittenbaum, G. and Stasser, G. “Management of Information in Small Groups” In J. Nye and M. Brower (Eds.), *What’s Social about Social Cognition? Social Cognition Research in Small Groups.* Thousand Oaks, CA: Sage.

Williams, Katherine Y. and O’Reilly, Charles A. “Demography and Diversity in Organizations:

A Review of 40 Years of Research” *Research in Organizational Behavior*, Vol. 20 (1998): 77-

140.

Table 1
Group Work Schedule for Management 100 in Fall 2004

Friday, September 3	Group members are selected.
Friday, September 10	Choose and sign a group contract.
Friday, September 17	Work on the group assignment #1.
Friday, September 24	Group assignment #1 is due.
Friday, October 1	Select team books and develop project work plan.
Thursday, October 7	Midterm #1.
Friday, October 15	Work on the group assignment #2.
Friday, October 22	Fall Break
Friday, November 5	Start working on a group outline for your paper.
Thursday, November 11	Midterm #2.
Friday, November 12	Have a group outline ready to discuss with TA's.
Friday, November 19	Start writing the introduction of the paper. Work on the group assignment #3.
Friday, November 26	Thanksgiving Holiday
Friday, December 3	Discuss the introduction and the first draft of group paper.
Friday, December 10	Paper and Presentations due
Monday, December 20	Final exam

Table 2
Summary Statistics

Individual Level Variables (Number of Observations: 431)

	Mean	Standard Deviation	Minimum	Maximum
Individual Exam Score	81.60	8.63	43.15	99.77
Age	19.25	1.08	17.17	25.08
SAT Score	1377.73	93.48	1010	1600
Gender* (male=1)	0.56	0.50	0	1
Ethnicity* (white=1)	0.67	0.47	0	1
Engineering student*	0.08	0.28	0	1
Off Campus*	0.11	0.31	0	1
Contract*(Democratic=1)	0.59	0.49	0	1
College Status**	1.46	0.81	1	4

*: Dummy variables, taking the value of 1 if an observation satisfies the criterion

**: The variable "college status" takes values 1 to 4 corresponding to the status of freshman to senior.

Group Level Variables (Number of Observations: 102)

	Mean	Standard Deviation	Minimum	Maximum
Group Performance Score	88.89	5.11	70.26	100
Group Size	4.25	0.74	2	5
Average Age	19.29	0.74	18.25	21.48
Standard Deviation of Age	0.67	0.45	0.06	2.46
Average Gender Composition	0.56	0.19	0	1
Average SAT Score	1375.61	55.55	1175	1480
Standard Deviation of SAT Scores	71.09	30.88	17.85	169.17
Proportion of White Students	0.68	0.24	0.20	1
Off Campus Dummy*	0.35	0.48	0	1
Engineer Dummy*	0.30	0.46	0	1
Contract (Democratic = 1)	0.60	0.49	0	1
TA influence**	7.93	1.08	6	10

*: Off campus dummy takes the value of 1if a group has at least one member living off campus.

*: Engineer dummy takes the value of one if a group has at least one Engineering student.

**: TA influence measure takes the value from 0 to 10 for each group; higher value corresponds to higher influence.

Table 3
Relationship between Group Characteristics and Group Performances
OLS Regression Results
Dependent Variable: Group Performance Score ($GroupScore_i$)

Variable	(1) Baseline Model	(2) Interaction between Age and Gender Composition	(3) Including Contract Choice	(4)
Average SAT	0.021	0.027	0.018	0.026
	(0.016)	(0.017)	(0.016)	(0.016)
Standard Deviation of SAT	-0.006	-0.008	-0.013	-0.014
	(0.019)	(0.019)	(0.021)	(0.020)
Male Group	-6.959*	-8.283**	-5.615	-6.797*
	(4.116)	(3.997)	(4.242)	(4.066)
Female Group	2.939	-2.879	9.071	3.991
	(5.063)	(5.135)	(5.719)	(7.001)
Proportion of White Students	1.700	0.822	1.100	0.294
	(2.550)	(2.585)	(2.470)	(2.473)
Average Age	-0.520	-1.486	0.156	-0.676
	(1.572)	(1.724)	(1.673)	(1.827)
Standard Deviation of Age	-0.154	-1.457	0.785	-0.909
	(1.651)	(1.886)	(1.626)	(1.840)
Group Size Dummy	0.769	-0.866	0.968	1.230
	(1.065)	(1.061)	(1.202)	(1.207)
Engineer Dummy	0.872	0.919	0.431	0.341
	(1.176)	(1.129)	(1.198)	(1.173)
Malegroup*Average age		10.663**		13.045**
		(5.299)		(5.180)
Femalegroup*Average age		12.832**		10.171
		(6.389)		(7.056)
Contract Choice			3.231**	3.505**
			(1.464)	(1.512)
Prob>F	0.0852	0.0042	0.0111	0.0000
R-squared	0.1950	0.2339	0.2810	0.3303
Root MSE	4.9138	4.8491	4.8384	4.7349
Number of observations	102	102	87	87

Robust standard errors are in the parentheses. *, ** and *** indicate significance levels of 10%, 5%, and 1% respectively.
Each regression also includes dummies for each class.
The dummy variable contract choice takes the value of 1 for Democratic groups and 0 for Autonomous groups.
Both Average SAT and Average Age are demeaned.

Table 4 (part 1)
Knowledge Spillovers and Group Learning
OLS Regression Results
Dependent Variable: Individual Exam Score ($IndScore_i$)

Variable	(1) Baseline	(2) SAT*age	(3) Contract
SAT	0.029***	0.030***	0.029***
	(0.005)	(0.005)	(0.005)
Age	1.099**	0.757	0.981*
	(0.545)	(0.555)	(0.551)
Gender	-0.412	-0.403	0.314
	(0.876)	(0.872)	(0.918)
Ethnicity	0.369	0.475	-0.082
	(0.916)	(0.909)	(0.955)
Off Campus	1.788	1.451	2.001
	(1.507)	(1.490)	(1.574)
Engineering Student	2.545**	3.002***	1.934
	(1.120)	(1.092)	(1.193)
Average SAT	0.000	0.000	-0.008
	(0.013)	(0.013)	(0.014)
Standard Deviation of SAT	0.008	0.007	-0.001
	(0.014)	(0.013)	(0.014)
Average Age	-1.455	-1.148	-1.277
	(0.941)	(0.958)	(0.963)
Standard Deviation of Age	2.618**	2.154*	3.470***
	(1.205)	(1.160)	(1.255)
Male group	-9.824**	-10.833***	-9.167**
	(3.843)	(3.824)	(3.722)
Female Group	-6.724	-7.367	-0.236
	(4.650)	(4.556)	(6.676)
Proportion of White Students	-2.027	-2.044	-1.384
	(2.266)	(2.232)	(2.344)
Group Size Dummy	0.649	0.704	1.195
	(0.932)	(0.915)	(1.094)
SAT*Age		-0.007***	
		(0.002)	
Contract			2.559***
			(0.900)
Prob>F	0.0000	0.0000	0.0000
R-squared	0.2090	0.2204	0.2287
Root MSE	7.8405	7.7934	7.7692
Number of observations	431	431	367

*, ** and *** indicate significance levels of 10%, 5%, and 1% respectively.
Robust standard errors (in parentheses) are reported by accounting for potential dependence of individual observations within group.
Each regression also includes dummies for each class.
The dummy variable contract choice takes the value of 1 for Democratic groups and 0 for Autonomous groups.
Both SAT and age are demeaned.

Table 4 (part 2): Knowledge Spillovers and Group Learning

OLS Regression Results

Dependent Variable: Individual Exam Score ($IndScore_i$)

	(4)	(5)	(6)	(7)
	Who benefit most from group work?			
Variable	Age*dage	Sat*dage	Age*Malegroup	SAT*Malegroup
SAT	0.030***	0.036***	0.030***	0.037***
	(0.005)	(0.007)	(0.005)	(0.006)
Age	-0.488	0.970*	0.243	0.970*
	(1.036)	(0.566)	(0.700)	(0.548)
Gender	-0.409	-0.511	-0.339	-0.457
	(0.879)	(0.883)	(0.880)	(0.884)
Ethnicity	0.482	0.386	0.460	0.222
	(0.920)	(0.922)	(0.902)	(0.910)
Off Campus	1.636	1.779	1.961	1.758
	(1.502)	(1.502)	(1.496)	(1.499)
Engineering Student	2.678**	2.904***	2.469**	2.716**
	(1.133)	(1.093)	(1.069)	(1.078)
Average SAT	0.000	-0.000	0.002	0.000
	(0.013)	(0.013)	(0.013)	(0.013)
Standard Deviation of SAT	0.007	0.007	0.005	0.006
	(0.014)	(0.013)	(0.013)	(0.013)
Average Age	-0.836	-1.433	-0.753	-1.480
	(1.035)	(0.941)	(1.032)	(0.943)
Standard Deviation of Age	1.995	2.418**	1.836	2.555**
	(1.254)	(1.187)	(1.248)	(1.194)
Male group	-10.742***	-10.130***	-10.282***	-10.425***
	(3.900)	(3.780)	(3.669)	(3.807)
Female Group	-7.030	-7.076	-7.020	-7.160
	(4.480)	(4.586)	(4.497)	(4.560)
Proportion of White Students	-2.205	-1.866	-2.250	-1.807
	(2.247)	(2.263)	(2.201)	(2.217)
Group Size Dummy	0.757	0.672	0.703	0.551
	(0.917)	(0.927)	(0.921)	(0.925)
Age* Standard Deviation of Age	1.270*			
	(0.671)			
SAT*Standard Deviation of Age		-0.009*		
		(0.005)		
Age*Male Group			5.506*	
			(2.850)	
SAT*Male Group				-0.060**
				(0.028)
Prob>F	0.0000	0.0000	0.0000	0.0000
R-squared	0.2163	0.2121	0.2182	0.2165
Root MSE	7.814	7.8347	7.8046	7.8129
Number of observations	431	431	431	431

*, ** and *** indicate significance levels of 10%, 5%, and 1% respectively.

Robust standard errors (in the parentheses) are reported by accounting for potential dependence of individual observations within group.

Each regression also includes dummies for each class. Both SAT and age are demeaned.

The dummy variable contract choice takes the value of 1 for Democratic groups and 0 for Autonomous groups.

Table 5 (Panel A)
The impact of Group characteristics on Contract Choice
Treatment Effect Model Using Maximum Likelihood Estimation

Panel A: Treatment Effect Estimation Results
Dependent Variable: Contract Choice (1=Democratic, 0=Autonomous)

	(1)	**(2)**	**(3)**	**(4)**
	Probit Model	Probit Model	Treatment Effect MLE	
Variables	Baseline	Asat*aage	Include TA influence	
Average SAT	0.001	0.001	0.004	0.004
	(0.003)	(0.004)	(0.004)	(0.004)
Male Group	-0.527	0.231	0.204	0.259
	(1.305)	(1.406)	(1.533)	(1.538)
Female Group	1.470	1.510	2.076	2.550
	(1.613)	(1.622)	(3.378)	(2.772)
Average Age	-0.647*	-0.700**	-0.486	-0.459
	(0.346)	(0.353)	(0.505)	(0.518)
Standard Deviation of Age	-0.681	-0.560	-0.438	-0.478
	(0.426)	(0.425)	(0.457)	(0.453)
Group Size Dummy	-0.479	-0.637*	-0.971**	-0.987**
	(0.362)	(0.372)	(0.394)	(0.396)
Engineering Student Dummy	0.668*	0.575	0.994**	0.967**
	(0.342)	(0.361)	(0.461)	(0.453)
Professor Dummy	0.913	0.726	4.247**	4.374**
	(0.587)	(0.632)	(2.072)	(2.193)
Average SAT* Average Age		0.010*	0.009*	0.010*
		(0.005)	(0.006)	(0.005)
TA Influence††			0.463**	0.454**
			(0.191)	(0.198)
TA*Prof Dummy			-0.510**	-0.532**
			(0.237)	(0.256)
Prob>chi2	0.0104	0.0133	0.1889	0.2669
Pseudo R-squared	0.1743	0.2057		
Number of Observations†	87	87	81	81

†: For the treatment effect model the number of observations is 81 groups (instead of 87 groups for the Probit model) since only data from three most recent classes were included in the regression analysis.

††: The variable "TA influence" measures how much the TAs discussed the positive aspects of Democratic governance form.

Robust standard errors are in the parentheses. *, ** and *** indicate significance levels of 10%, 5%, and 1% respectively.

The dummy variable contract choice takes the value of 1 for Democratic groups and 0 for Autonomous groups.

The results are robust using two-step estimates instead of MLE method.

Both Average SAT and Average Age are demeaned.

Table 5 (Panel B)
The Impact of Contract Choice on Group Performance
Treatment Effect Model Using Maximum Likelihood Estimation

Panel B: Group Performance Estimation Results
Dependent Variable: Group Performance Score ($GroupScore_i$)

Variables	**(3)** Treatment Effect MLE	**(4)** Including Asat*aage in first step
Average SAT	0.022*	0.030**
	(0.013)	(0.013)
Standard Deviation of SAT	-0.014	-0.013
	(0.018)	(0.017)
Male Group	-4.516	-6.796*
	(4.032)	(3.914)
Female Group	10.077	2.676
	(6.250)	(6.454)
Proportion of White Students	0.250	-0.434
	(2.482)	(2.431)
Average Age	-0.176	-1.087
	(1.625)	(1.709)
Standard Deviation of Age	0.290	-1.109
	(1.626)	(1.632)
Group Size Dummy	0.953	1.162
	(1.021)	(1.022)
Engineering Student Dummy	1.048	0.953
	(1.150)	(1.128)
Professor Dummy	-0.120	0.398
	(2.076)	(2.184)
Contract Choice	4.987**	4.837**
	(2.201)	(2.226)
Male Group*Average Age		11.779**
		(4.660)
Female Group*Average Age		13.759**
		(5.659)
Wald Chi2	47.27	84.77
Prob>Chi2	0.0000	0.0000
Number of Observations†	81	81

†: For the treatment effect model the number of observations is 81 groups (instead of 87 groups for the Probit model) since only data from three most recent classes were included in the regression analysis.
††: The variable "TA influence" is the sum of two TAs' responses, each one being a 5-point Likert scale that measures how much the TA discussed the positive aspects of Democratic governance form.

Robust standard errors are in the parentheses. *, ** and *** indicate significance levels of 10%, 5%, and 1% respectively.

The dummy variable contract choice takes the value of 1 for Democratic groups and 0 for Autonomous groups.

Both Average SAT and Average Age are demeaned.