

**The Greatest Good for the Greatest Number:
An Experimental Study of the Effects of Racial and Ethnic Diversity on
Liberal Arts College Discussions**

2009

Robert J. Beck (robert.beck@lawrence.edu)

Lawrence University

Abstract

The study tested the Supreme Court majority opinion in the affirmative action case, *Grutter v. Bollinger* (2003) that racial and ethnic diversity contributes to benefits for all students in classroom discussions. Two randomly selected classes with 25% racial and ethnic diversity were compared with two randomly selected all-white classes (N = 61). The data were transcripts of 16 discussions (70 minutes in length each) in which students made interpretations about “Great Books” including Plato’s *Republic* and Elizabeth Bishop’s poetry as well as works by a scientist (Einstein), and a composer (Messiaen). Levels of student participation, inclusiveness, peer-to-peer interaction, and critical thinking were used as measures of discussion performance. Diverse classes had significantly higher rates of *participation* as measured by the mean number of words spoken by students and the proportion of student to discussion leader word counts. Diverse classes had greater *inclusiveness* as measured by the greater number of different students participating in the upper quartile of participation and the mean number of different students participating in discussion threads. Diverse classes had much higher percentages of *student-to-student interactions* in the discussions.

The non-diverse classes had higher percentages of *critical thinking propositions* as measured by the proportional frequency of high-level arguments and text-based claims in individual turns of discussion, while diverse classes had higher percentages of opinions and claims based on personal experiences. However, students in diverse classes also joined formal claims to their opinions in 25% of turns, whereas the non-diverse classes only did so in less than 10%. While significantly different critical responses were found in both diverse and non-diverse classes, diverse classes had higher proportions of high-level questions, re-phrasings and greater numbers of elaborations/clarifications of communications responding to previous turns. Two patterns of class discussion emerged from the research: 1) participation, inclusiveness, student interaction, and critical responsiveness may be supported by a discussion climate in which there is permission to express opinions and claims based on personal experience that afforded students relatively easy entry into discussions and 2) a climate in which text-based claims and arguments are the standard may result in a relatively small hierarchy of students capable of making such statements. A three-stage model was proposed that encouraged discussion leaders to allow students to voice opinions as a means of affording wide participation, then follow-up by encouraging students to make text-based claims and full arguments.

As measured by several criteria in the diverse groups: wide and inclusive participation, high peer-to-peer interaction, greater understanding by a larger proportion of students that opinions need to be backed up with evidence, and exposure to more high-level claims and responses; versus instructor domination, lower student participation, active critical thinking in only a limited number of students, and lower peer to peer responsiveness in the non-diverse groups, we conclude that the diverse classes provided more value - - the “greatest good to the greatest number” - -to their students than the non-diverse classes.

Introduction

In Justice Sandra Day O'Connor's majority opinion in *Grutter v. Bollinger*, 2003, the Supreme Court affirmative action case, she argued that diversity leads to educational benefits for all because of a "robust exchange of ideas" (U. S. 539, 17). These benefits are "important and laudable," she said, because "classroom discussion is livelier, more spirited, and simply more enlightening and interesting" when the students have "the greatest possible variety of backgrounds" (U.S. 539, 17). Although "backgrounds" are not specified here the fact that this case was concerned with affirmative action leads us to select racial and ethnic backgrounds as critical in this study. What makes for a good discussion? A "robust exchange" suggests that participants make claims and offer opinions that provoke responses by others. We take "livelier" to mean that student participation is high and that a relatively high proportion of students are involved. Certainly, an instructor-dominated discussion is neither lively nor spirited. "More spirited" is an evocative dimension, but difficult to measure. Perhaps more salient to classroom discussion are "enlightening" and "interesting". These terms suggests that students are exchanging well-formulated ideas that have potential for being informative and instructive, *i.e.*, that students might learn something and can be operationally defined as encouraging critical thinking. Other outcomes related to good discussions might be "well-debated" and mutually supportive and elaborative communications that lead to a range of ideas. As a freshman program, the discussions also serve to help students at the start of their college careers to find their courage to speak up in group settings.

The present study analyzed transcribed student communications in discussion sections of a freshman "Great Works" program in a liberal arts college. The pedagogic intent of this program is for students to learn to interpret texts and, hence, these are "interpretive discussions" (Haroutunian-Gordon, 2009). In an experimental design, two randomly assigned classes with a critical mass of ethnic and racial diversity were compared with two randomly assigned classes that were composed entirely of white students, *i.e.*, non-diverse. Each set of classes discussed the same four works in science, music, philosophy and poetry during an eight-week period. According to O'Connor's majority opinion, the diverse classes should have displayed more robust and enlightened exchanges of ideas during this period.

Theoretical and Research Background

Why does the Associate Justice, her supporting colleagues, a good proportion of the public, and many researchers believe that ethnic and racial diversity leads to a “robust exchange of ideas” in classroom discussions? Both Gurin (2003) and Milem (2003) identified classroom discussion as an important setting for facilitating diverse students to contribute to academic benefits for all participants. Using self-report measures, researchers found positive faculty assessments (Milem, 2003 (HERI/UCLA) and positive student assessments (Chang, 1996) that diversity contributed to richer classroom interactions among students. Other studies also showed that diversity improved cross-racial understanding and friendships, provided long-term social and economic benefits, and high student academic achievement was associated with those who had diversity training and social associations with minorities (Bowen & Bok, 1998; Orfield & Kurlaender, Eds. 2001; Chang, Witt, Jones, & Hakuta, Eds. 2003; Gurin, 2003).

In Gurin’s (2003) expert testimony in *Grutter v. Bollinger*, she formulated a learning model to explain the beneficial impacts of diversity in student groups. Combining major theoretical sources (Erikson, Piaget, Langer), Gurin argued that diversity supports self-conscious, active thinking processes because it is associated with novel, dis-equilibrating, and unpredictable communications in peer group settings. These interactions are likely to produce contradictory expectations that support complex thought and engaged learning. Referring to studies by King and colleagues (1996, 1998), Gurin further articulated active thinking as academic communications involved in the assessment of multiple truth claims according to conceptual soundness, evidence, and meaningfulness.

While Page (2007) argued strongly on theoretical grounds for the benefits of identity and cognitive diversity on the productivity of groups, his review of the relevant literature was more ambivalent:

The evidence for identity diverse groups, though, is far from unequivocal. Some identity diverse groups perform well. Others do not. The same is true of identity diverse cities and countries. This makes sense, given the conditional nature of the logic. First the linkages between identity diversity and cognitive diversity may not be strong in all cases. Second, many identity diverse groups have differences in their fundamental preferences, which, as we have seen cause problems. Third, people who differ often have trouble communicating and getting along. Nevertheless, if we look

closely at the evidence across the scales - groups, cities, and countries - - we find that well managed identity diversity does produce benefits (p. 314).

Opponents of affirmative action, however, insist that race does not carry any special capabilities that contribute to educational benefits (Hopwood vs. Texas 861 F. Supp. 551 (W. D. Tex. 1994)), and doubts linger about the influence of diversity on benefits to all students. Even earnest supporters of affirmative action concede that empirical studies of academic benefits have lagged and that the existing research has largely employed self-report, not objective, methodologies (Williams & O'Reilly, 1998).

While largely limited to the analysis of organizational work groups and laboratory studies and already a decade old, Williams and O'Reilly's analysis of the effects of demography and diversity in organizations remains the most definitive review of empirical research. The model employed by these authors is that variations in social categories of participants have effects on group processes, which, in turn are associated with effects on group performance. Identity diversity, that is, racial and ethnic diversity for example, has been found to have two kinds of effects on group processes: a) negative effects involving conflict, communications problems, and factionalism; and b) positive effects involving more diverse kinds of information based on background and experience.

With respect to beneficial performance outcomes, laboratory studies have shown that diverse groups produce higher quality ideas, range of perspectives considered, and number of alternatives generated (McCleod & Lobel, 1992; Watson, Kumar & Michaelson, 1993). In field studies, O'Reilly, Williams & Barsade (1997) found both creativity and implementation ability in the diverse groups they studied. Williams and O'Reilly added, however, that one of the key factors contributing positively to beneficial outcomes, or reducing the potential for negative outcomes, is the *proportion* of diverse group members. They suggest there may be upper limits to the extent that minority representation, including both sex and ethnic and racial diversity, may be optimal in successful integration into majority groups so as to minimize conflict, factionalism and communications problems. Among the theories used to explain possible negative effects are similarity/attraction: that members of a group are more attracted to, and hence more comfortable with those who are similar; and, social categorization: that dissimilar

members are subject to stereotypes and this leads to lower cohesiveness, polarization and anxiety.

In Page's (2007) model of how diversity contributes to group productivity, he argued that identity diversity needed to be linked to cognitive diversity to confer benefits to the group. Thus, identity diversity had to be associated with diverse cognitive perspectives (ways of representing situations and problems), diverse interpretations (ways of categorizing or partitioning perspectives), diverse heuristics (ways of generating solutions to problems) and diverse predictive models (ways of inferring cause and effect) to provide benefits to groups. Moreover, the net positive effect of diversity is dependent on overcoming potential communications problems associated with identity differences. Page's model is more relevant, nuanced and appropriate to a study of undergraduate discussions of great works, because the primary purpose of such discussions is to encourage students to make interpretations based on individual cognitive perspectives and diversity presumably endows the group with a greater range of perspectives and potential interpretations. Wortham (personal communication) proposed, "Only by contributing partial ideas together a set of people might reach a new conclusion, which none of them could have reached alone." In this setting, college discussants of great works, it was neither likely that race and ethnicity *per se* would produce conflict nor was it probable that race-based factionalism would be the case.

Discussions of "Great Works": Group composition, social processes, and performance

How do these diversity models apply to college discussions of great works? To begin with, such discussions are relatively distinctive when compared with organizational work groups or laboratory studies that have had a long history in the sociology and social psychology of human groups. The purpose of college discussions is to provide a forum for students to express ideas and to learn to collaborate with others, including both discussion leaders and peers in building interpretive knowledge. Typically, the social composition of these groups is not designed. In contrast to work groups, university discussions treat a larger range of content, moving from reading to reading periodically, compared with the former who are engrossed with solving particular problems. And, there is no expressed performance outcome as such except that there should be wide

participation and that the group makes sense of the particular reading assigned. In general, the goals of academic discussions are weakly defined, if at all. Another critical difference is that in the work place most studies involved leaderless groups, because the research objectives concerned the emergence of leaders and/or hierarchies. In university (and K-12 classroom) discussions, the instructors, hereafter discussion leaders (DLs) play an important role. Therefore, university discussions of great works differ considerably from organizational work groups who have more specific objectives and more specific outcomes. Consequently, we cannot refer to studies in the literature that are precisely comparable to the present investigation.

Effects of critical mass on group performance

As Williams and O'Reilly suggested, proportion or critical mass of diverse participants may be a key factor in group cohesion and have implications for performance. Mannix and Neale (2005) found that both token diversity and parity were unproductive, the former leading to neglect and isolation and the latter to too much conflict. Yet an adequate number of diverse students, who formed collaborative sub-groups, did facilitate productivity. Williams and O'Reilly concluded that "proportions of [diversity] are important in priming salient categories" (p. 111) and they advised that researchers "consider proportional measures within groups" (p. 116). Normally, the group composition of university discussions is random and would generally reflect the overall demographic distribution of the university. Therefore, any controlled study of diversity would have to assign diverse students to particular sections and constitute all non-diverse classes as comparison groups.

Effects of discussion leader pedagogy on diversity and discussion

Researchers suggested that active learning pedagogies, in which students have opportunities and explicit strategies to discuss and exchange ideas and worldviews in peer groups, are needed to maximize the potential of diversity to enable cognitive and academic benefits for all students (Hurtado, 2001; Milem, 2001). Nunn (1996) found that oral discussion strategies that support peer-to-peer interactions were more effective than typical instructor-dominated discussions. In our own research on family and teacher educational discourse processes we found that guided discussions in which discussion leaders had explicit strategies for managing discussions, whether in natural or

experimental conditions, resulted in high quality questioning, elaborative response and feedback (Beck, Fitzgerald, & Pauksztat, 2003; Beck & Clarke-Stewart, 1998; Beck & Wood, 1993). Because discussion group leaders vary in pedagogic philosophy, communications processes may also vary potentially masking the effects of group composition. Discussion leaders may differ in the degree to which they use high-level initiations, *e.g.*, questions, to stimulate discussion or how they help sustain threads (themes) of sub-discussions. They also may vary in their own level of participation and how much they encourage peer-to-peer interaction. Other dimensions on which leaders show differences in best practices are the extent to which they: use “scaffolding” (indirect prompts and hints) in helping students to construct their own ideas; respond to all students personally; support formulation of students high-level claims and arguments; and periodically weave together and summarize the communications of group members. Therefore, any controlled study of diversity would need to minimize variance contributed by discussion leader approaches. Below, we report on a discussion leader guide used to control variance that highlighted such techniques as encouragement of wide student participation and peer-to-peer communication; the guide also recommended that discussion leaders use of high-level initiations (*e.g.*, explanatory questions), scaffolding, and summaries and that they minimize their own participation.

Discussion group performance outcome measures

We measured discussion performance by levels of student participation, inclusiveness, peer interaction, and critical thinking:

Participation

It was assumed that the aggregate level of student participation could be used as a measure of productive discussions that benefit all students. In college discussions there is more or less continuous talk and both the instructor and students assume responsibility for sustaining the discussion. Anytime students are not speaking, the discussion leader is speaking and vice versa, *i.e.*, a zero-sum game. Within limits, therefore, teacher discussion leader speech is inversely proportional to class productivity. We think that productive discussions are characterized by relatively high student participation, whereas faculty discussion leaders dominate unproductive discussions. In face-to-face studies of

experimental groups, the measure of participation has been based on the amount of time any member speaks, either to the group at large or to specific others (Bales, 1950; Bales, Strodtbeck, Mills, & Roseborough, 1951; Bonito, 2002; Kalma, 1991; Lamb, 1980; Mullen, Salas, & Driskell, 1989; Shelley & Troyer, 2001a, 2001b; Stasser & Taylor, 1991; Stein & Heller, 1979). Speaking time, as a behavioral style, may help to establish a more advantageous position in a group. Schmid Mast (2001) used speaking time as a measure of behavioral dominance in a study of female and male differences in small group discussions about child rearing practices. Group members also rated each other on a sociometric instrument. It was found that speaking time was highly positively correlated with the group members' dominance ratings (how much members estimated others spoke) of each other, thus providing external validity for speaking time as a dominance behavioral measure. We will assume that word count (the length of the message) may be used as a proxy for speaking time. In a study of online discussions, Beck, Fitzgerald, & Pauksztat (2003) demonstrated that student mean word count was an effective measure of individual student performance. In face-to-face discussions, relatively high word count is related to meaningful and complex contributions (Pilot Study findings in present investigation). Other students or the discussion leader seldom tolerate irrelevant or oblique speech and nearly all speech by our reading was on task. Therefore, *word count* in transcribed discussions will be used as an equivalent measure of speaking time and, therefore, participation.

Inclusiveness

Inclusiveness may be operationally defined as the number of different speakers who participated. Several studies have demonstrated that, in unassigned, initially leaderless groups, a stable influence hierarchy and role system will emerge and persist after initial instability (Homans, 1950; Bales, 1950). What discussant attributes contribute to the development of status hierarchies? Strangers initially can use observed status characteristics such as age, gender, and race to establish leadership and influence hierarchies and, subsequently, these may be reinforced or altered through other processes. Burt (1999) contrasted the human capital explanation (people do better and hence attain higher positions in a status hierarchy because of their personal attributes such as knowledge, skill, and charisma) with the social capital explanation (people do better

because they are better socially connected). Gould (2002) used a comparable dichotomy in summarizing groups of factors that led to interaction hierarchies. He specified an individualist or market framework in which outcomes are unequal because individuals vary in qualities that have locally meaningful importance, such as talkativeness or confidence. In this case differentiation occurs because people make different contributions. This was contrasted with a social structural framework in which outcomes are unequal because of the quality of social positions one occupies, largely independent of personal qualities (for example, membership in an in-group).

While the explanations for the formation of hierarchies remain contested, the criterion of inclusiveness in great works discussions implies that more numerous dominance hierarchies are preferable to smaller ones. The models proposed above indicate that speakers with some favorable attributes will be addressed more frequently, but other attributes such as talkativeness implies that such speakers will participate more because of intrinsic characteristics. For either quality, inclusiveness may be examined at two scales of participation: discussion as a whole; and, discussion in threads (themes). A relatively inclusive discussion group would have a greater number of different speakers collaborating in threads and a greater number of students who engage at the highest levels of participation.

Peer-to-Peer Interaction (PTP)

Discussion researchers have emphasized that in productive discussions students should interact with each other and not always with the instructor. This measure will be dependent on how active the DL is and whether s/he encourages PTP interaction. The leader should participate enough to frame inquiry and be responsive, but should not participate too much: the more the leader has the floor, the fewer opportunities exist for student participation. Once the overall discussion and threads are initiated, the leader lessens his or her contributions so that students may get a turn. As threads develop, a good discussion leader may assume the role of “just another discussant”. Souter and Rudge (2007), in particular, stressed the value of the teacher as a co-learner. In productive discussions, the discourse is connected: "the discussion is characterized by multiple, interactive, connected turns; succeeding utterances build on and extend previous ones" (Saunders and Goldenberg (2007, 224-225). Connectivity implies that a threaded

line of inquiry is being followed – the overall question and important sub-questions are being addressed – and that participants are responding to, following up on each other’s communications. Therefore, a high PTP group would be one in which there were a high percentage of conversational sequences in which students were responding to each other.

Critical thinking

We propose that levels of critical thinking, provided that the construct may be adequately defined, could function as qualitative performance measures. In Derek Bok’s book, *Our Underachieving Colleges* (2006), he made several points about the importance of structured discussion in teaching students to think. He cited educational researchers, McKeachie *et al.* (1986), Pascarella and Terenzini, (2005) and Prince (2004) in concluding that whether using “measures of transfer of knowledge to new situations, or measures of problem-solving, attitude change, or motivation for further learning, the results show differences favoring discussion methods over lecture” (p. 117). Bok argued that instructors who do best at teaching critical thinking structure group problem solving processes, including discussions: “They encourage their students to think for themselves by challenging them with interesting questions and using class discussions, collaborative projects, and other forms of active learning to develop habits of critical thinking” (p. 119). Bok contrasted the discussion-based approach of college instructors with those in professional schools. He reflects on the discussion-based methods of the Harvard Business School and most law schools: “Why, for example, do so many college instructors continue to lecture long after most professional schools have drastically curtailed such methods in favor of problem-based discussion?” (p. 122). Bok concluded that there were several reasons for this neglect (p. 125): older professors were taught by lectures and they perpetuate the practice; Socratic teaching takes more time and effort and instructors cannot prepare for discussions in the same way; it is difficult to arouse and direct a group of unresponsive students; and, finally, “instructors may not be adequately trained in teaching by discussion” (p.117).

Nevertheless, claims Bok, “the advantages of Socratic discussion, small-group instruction and other forms of active learning are becoming widely enough recognized that even skeptical professors may have to change their ways eventually” (p. 126). He argues that discussion-based teaching enables students to “test their cognitive skills and

receive prompt feedback on the results” (pp. 118-19). During active learning students employ their reasoning skills, have their curiosity awakened, and have their common misconceptions confronted.

In the present study, students are engaged in interpretive discussions (Haroutunian-Gordon, 2009) that make sense of the assigned texts. Having read and been presented with interpretation-seeking questions on major works, posed by DLs principally but also by peers, students participate in class discussion either by offering opinions and/or trying out their interpretations and making claims supported by evidence drawn from the text and personal experience. Instructors teach indirectly and socratically, principally by asking follow-up questions and offering elaborations and clarifications to help other students with their developing propositions. During the discussion, students collaborate cognitively by helping each other through questions and elaborations that clarify each other’s interpretations and may engage in the exchange of opinions, claims and extended arguments. In the present interpretive discussions there is practice of critical thinking in a relatively unconstrained context in which there is both a competition and collaboration of ideas. When students receive feedback from peers and instructors on their arguments, they experience their arguments from other vantage points as formulated by other speakers and they need to adapt. Presumably, diverse students occupy relatively more diverse vantage points from which to formulate and criticize claims.

We find it persuasive, therefore, that discussion-based teaching and learning, such as the discussion of great works in this study, contribute to students’ critical thinking because:

- In discussions students engage in active learning because they “do the work”. In order to participate students are required to present their own independent interpretations of texts. The benefits of diversity have been linked to differences in interpretations made by participants of varying backgrounds and experience.
- Discussions offer students opportunities to make claims and offer extended arguments about their positions. In arguing they practice critical thinking by engaging in reasoning about texts: making claims, offering supportive evidence, and drawing conclusions. The levels of claims and

arguments may serve, therefore, as performance measures in discussion groups.

We have argued that in rich discussions students need to make claims about what they are reading and in high-level participation they support their claims and opinions with evidence from the text or other sources. In this regard discussion leaders may encourage students, who have not substantiated their claims, to follow up and extend their own communications in subsequent turns so that they achieve greater depth. Equally important is that other students in the group support individuals by rephrasing and restating their propositions. This is important because not only are discussions opportunities for individuals to try out their ideas, but because discussions provide opportunities for cognitive collaboration. Also important are student turns in which they agree or disagree with previous ideas. Carried to their conclusion, agreements and disagreements may lead to healthy debates. Naturally, discussion leaders may support all these forms of collaboration, but when provided by peers they may engender the kinds of discussion one esteems outside of class as well.

Measures of critical thinking, therefore, will be determined by an extensive set of codes used to derive qualitative differences of student and DL turns of conversation. The codes, examples which are found in the analyses following, will be used to assess two forms of critical thinking: 1) propositions, including claims and arguments referring to the text and/or personal experience; and 2) critical responses, including questions, elaborations/clarification, re-phrasings, and agreements/disagreements.

At issue in this study is the impact of identity diversity on all students' educational development in college classrooms. Does classroom racial and ethnic diversity contribute to participation, inclusiveness, peer interactive, and critical discussions that benefit all students? How are participative, inclusive, peer interactive, and critical thinking discussions defined and measured? If there are differences between diverse and non-diverse classes, what are the explanatory mechanisms?

Methodology

Sample

The study investigated 16 plenary, whole-group discussions in four classes in a Freshman Studies great works program at a Midwestern liberal arts college. Discussion

leaders in the program were recruited, randomly, from an initial pool of 25 instructors. Using random assignment from freshman class rosters, two classes were experimentally composed of a critical mass of racial and ethnic diversity (ca. 25%) and using random assignment two classes were experimentally composed of 100% majority students. Diverse students (N = 7: 1 student withdrew in the first week) consisted of Posse Program students - - selected from inner city settings and typically of “minority” race (African-American and Hispanic) -- as well as other domestic and international students of color. Eight diverse classes that discussed 4 different works in different weeks (Science-Einstein, Music-Messaien, Philosophy-Plato, Poetry-Bishop) were compared with eight non-diverse classes that discussed the same 4 works in the same weeks of the program.

Discussion leader instruction prior to experiment

Prior to the course the four participating instructors were provided with a one-hour consultation based on a review of best practices (see Appendix 1). The objective was to control for potential differences in leadership style and, hence, to minimize systematic variations in discussion leader (DL) contributions to group performance. Of course, DL approaches are formed over a long period of development and we cannot guarantee that our goal was achieved although consultants reported that the criteria were met with agreement by participating faculty.

Procedure: Collection and transcription of classroom discussions.

A reliable method was employed for collecting digital audio recordings and transcriptions in a highly complex environment. Using live observers, placards with student ID numbers, and software enabling the production of tones in the digital sound track, student observers successfully identified the 15-17 speakers (including DLs) in each discussion. We found a reliable and economic service for transcribing MP3 audio files into word processing documents. After raw transcripts of class discussions were prepared, observers entered student and DL identifications for each turn of conversation. Researchers were capable of assigning turns of conversation to particular speakers with more than 99% accuracy. A programmer created a utility to transfer the electronic transcripts to SPSS for statistical analysis. The SPSS files contained verbatim discourse and word counts for each individual turn of discussion.

Measuring participation, inclusiveness, PTP interaction, and critical thinking in discussions

Measures of discussion *performance* were developed at three levels of grain: a) Turns, b) Threads, and c) Whole discussion. A turn is a speech by a particular individual in the sequence of speeches in a discussion. Threads are themes contributed to by multiple speakers. Threads have unique start and stop points in turns, although some themes ended and started in the same turn. In such cases turns were broken into A and B parts. In Appendix 2 find an example of a thread. The whole discussion refers to all of the speech (words) in a transcript.

Participation was measured by quantifying student participation in the whole discussion. Two measures were used: 1) Student whole discussion word count – the higher the word count the greater the participation and 2) Student to DL word count ratio in whole discussions -- the higher the proportion of student to DL speech, the greater the participation.

Inclusiveness was measured by two indices: 1) Number of different student speakers who contributed to different threads – the greater the average number of different participants, the greater the inclusiveness; and, 2) Size of the student dominance hierarchy as indicated by number of students in the upper quartile of students' cumulative word counts across all discussions - the greater the size of the hierarchy (number of different individuals) the greater the inclusiveness.

Peer-to-Peer Interaction was measured by the percentage of turn sequences in which students responded to each other and not to the DL.

Critical thinking was measured by applying codes to individual turns of discussion. Examples of non-obvious codes are found in analyses of each variable in the text below. The reliability of the codes was high. Pairs of researchers independently coded whole transcripts. The author resolved discrepancies. Reliability was re-assessed in additional studies of individual diverse students' interactions with the DLs and other students (data not reported in this study). We conclude that overall the coding reliability was greater than .90.

High-level critical thinking propositions:

1. Claims based on the text and cultural knowledge sources.

2. Arguments are defined as consisting of two or more sentences that are connected by reasoning. Arguments consist of claims and at least one of the following -- principles, evidence and conclusions.

Low-level critical thinking propositions:

1. Opinions
2. Claims based on personal experience
3. Procedural information
4. Lecture information
5. Text information
6. Cultural information

Critical responses to previous communications:

High-level critical responses:

1. Elaborations and clarifications
2. High-level questions: explanations; hypothetical; rhetorical; conditional
3. Agreements and disagreements
4. Re-phrasings
5. Scaffolding that pointed to ways to improve ideas

Low-level critical responses

1. Questions asking for procedural information
2. Questions eliciting opinions

Hypotheses

All seven hypotheses, following, proposed that diverse classes would exhibit higher levels of performance than non-diverse classes.

H1. Participation. Diverse class discussions exhibit higher aggregated levels of student participation in four whole class discussions than non-diverse classes.

H2. Participation. In diverse class whole discussions students exhibit proportionally higher levels of student participation compared with discussion leaders participation than in non-diverse classes.

H3. Inclusiveness. Diverse class discussions have greater numbers of different students contributing to threads.

H4. Inclusiveness. Diverse class discussions have a more inclusive dominance hierarchy that contained a greater number of students.

H5. Peer-to-Peer Interaction. Diverse class discussions have a higher percentage of turns in which students responded to students.

H6. Critical thinking propositions. Diverse class discussions demonstrate greater percentages of high-level critical thinking propositions than non-diverse classes.

H7. Critical thinking responses. Diverse class discussions demonstrate greater percentages of high-level critical responses to previous communications than non-diverse classes.

Results

Overall statistics of the study

A total of 158,185 words were spoken in the study of which students spoke 94,049 (59.5%) and DLs, 64,135 (41.5%). Comparable results were found in comparing the number of turns spoken by students, 4002 (60.7%) and DLs, 2583 (39.3%). The total number of threads was 387 with almost equal numbers in the diverse classes (196) and the non-diverse classes (191). Moreover, the threads were nearly the same average length: 17.69 turns/thread in diverse classes and 16.32 turns/thread in non-diverse classes.

H1. Diverse classes exhibit higher levels of student participation than non-diverse classes

Participation: Mean student word count in all diverse vs. non-diverse class discussions.

Across all four works, the *diverse classes* spoke a mean of 2026.3 (SD = 1523.5) words per student, while the *non-diverse classes* only spoke a mean of 1103.9 (SD = 985.4) words per student. This difference is significant: $t = 2.777$, $p = .008$. Thus, students in diverse classes spoke nearly 84% more words than in non-diverse classes. In Table 1 the results demonstrate that not only do diverse classes speak significantly more than non-diverse classes but also there is much greater variation in the diverse classes.

Student Participation in Diverse and Non-Diverse Classes

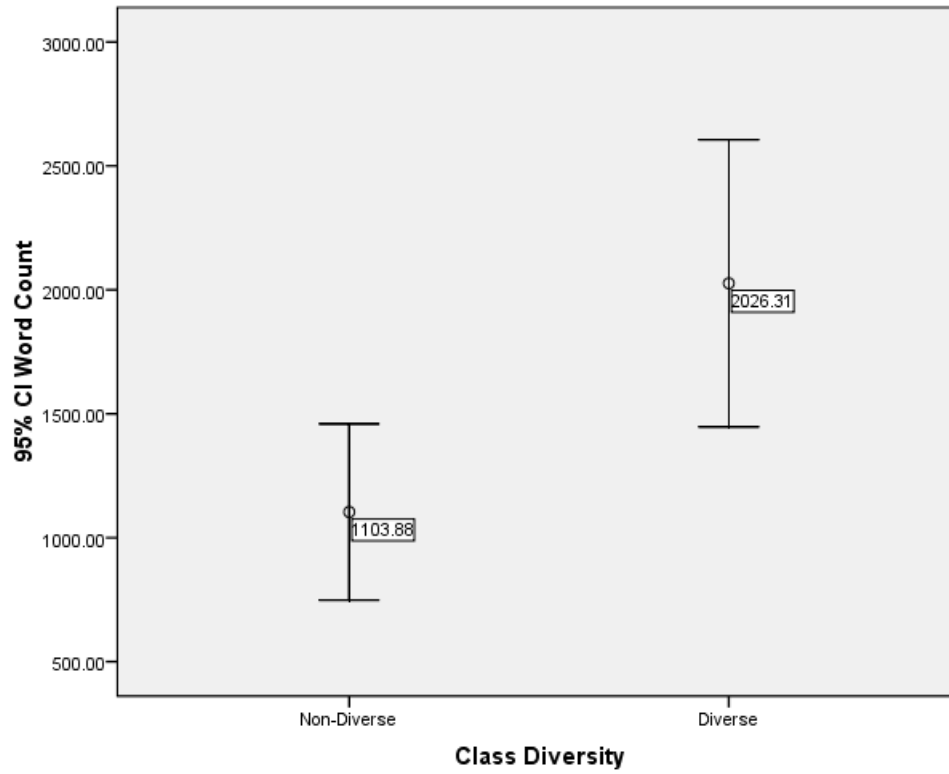


Table 1

There were no statistical differences in mean word count across all works or in discussions of any of the four works in diverse classes *between diverse and non-diverse students*, $t = .266$, $p = .792$ (NS). Table 2 also reveals that there is much greater variation in diverse classes students' word counts than in non-diverse classes.

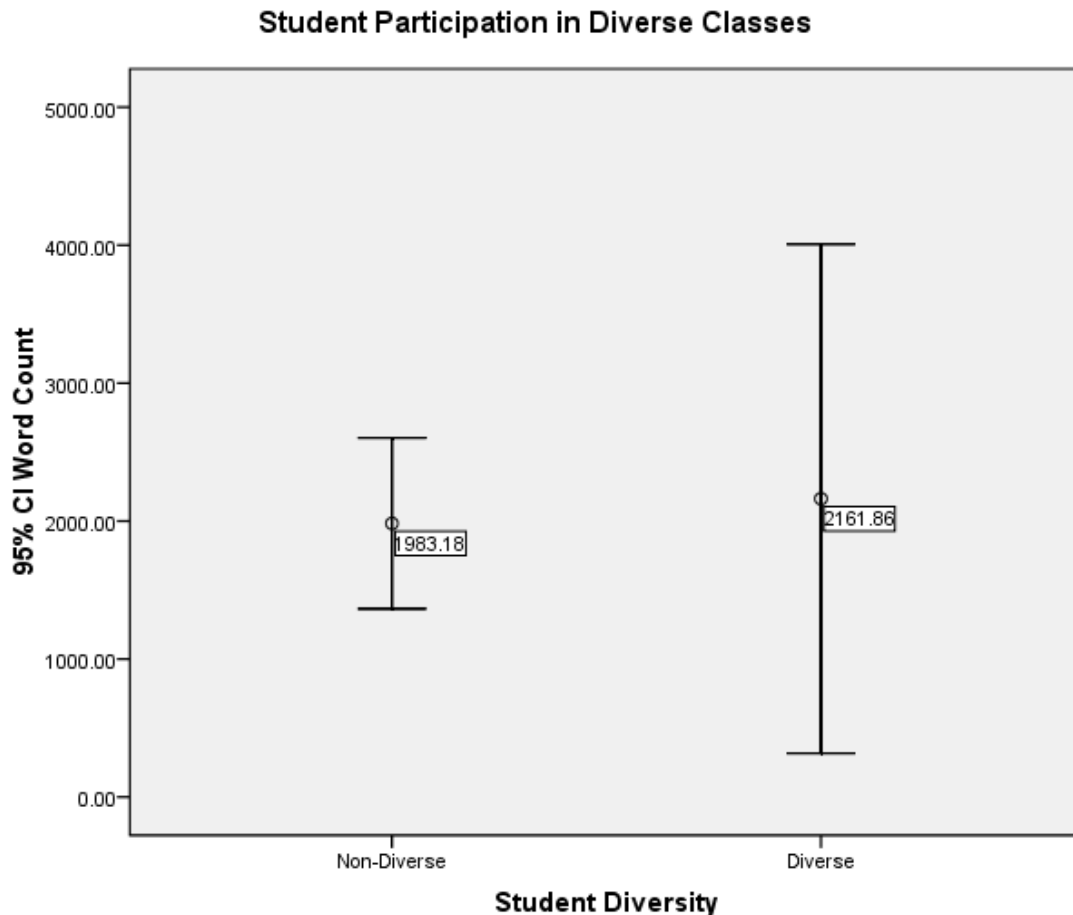


Table 2

This finding is important because the higher participation in diverse classes is not due to higher word counts by diverse students or by individual majority students but by greater participation in the diverse classes as a whole. Further, when the 7 diverse student word count scores were analyzed as frequencies in the total student population ($N = 61$), it was found that they participated normatively: 2 diverse students had frequencies in the upper 75th quartile; 2 diverse students had frequencies in the 50th quartile; 3 students in the 25th

quartile; no diverse students had frequencies in the lowest 0 quartile. While the mean participation scores for diverse and non-diverse students within diverse classes are not significantly different, the graphic and the much higher standard deviation score indicates that diverse student participate more diversely, that is, with a significantly greater range of scores.

H2. In diverse class whole discussions students exhibit proportionally higher levels of student participation compared with discussion leaders participation than in non-diverse classes.

Balance of participation between DLs and students

The diverse classes spoke 85,125 words and the non-diverse classes 73,041 words. In the diverse classes the students spoke 58,766 words and the DLs 26,359 words. In the non-diverse classes, students spoke only 35,264 words while the DLs spoke *more* words, 37,777. We found that 69% of the total words spoken in diverse classes came from student participation (DL = 31%) compared to 48% in homogeneous classes (DL = 52%). The percentage difference between student participation is marginally significant ($z = 1.66$, $n_1+n_2 = 61$, $p = .097$). Although the small sample size limits us to a marginal claim that the diverse classes had a more favorable student to DL balance of participation, it should be noted that in these diverse classes students spoke 38% more than the instructors and in the non-diverse classes, the students spoke 4% less than the instructors.

Summary: Participation

In both measures students in diverse classes had higher levels of participation than in non-diverse classes. The diverse classes had significantly higher mean word counts and participated in a much greater proportion of the discussion relative to the DLs. Therefore, Hypotheses 1 and 2 were supported by the findings.

H3. Diverse class discussions have greater numbers of different students contributing to threads.

Threads: Number of different student participants

The findings for number of different students participating in threads are shown in Table 3. In diverse classes an average of 4.50 (SD = 1.98) students collaborated in threads, whereas in the non-diverse classes only 3.71 (SD = 1.84) students collaborated. Recall that there was no significant difference concerning the length of threads between diverse and non-diverse classes. A t-test revealed that this difference is significant, $t = 4.006$, $p = .000$. Thus, there was greater participation in threads by more students in diverse classes than in non-diverse classes.

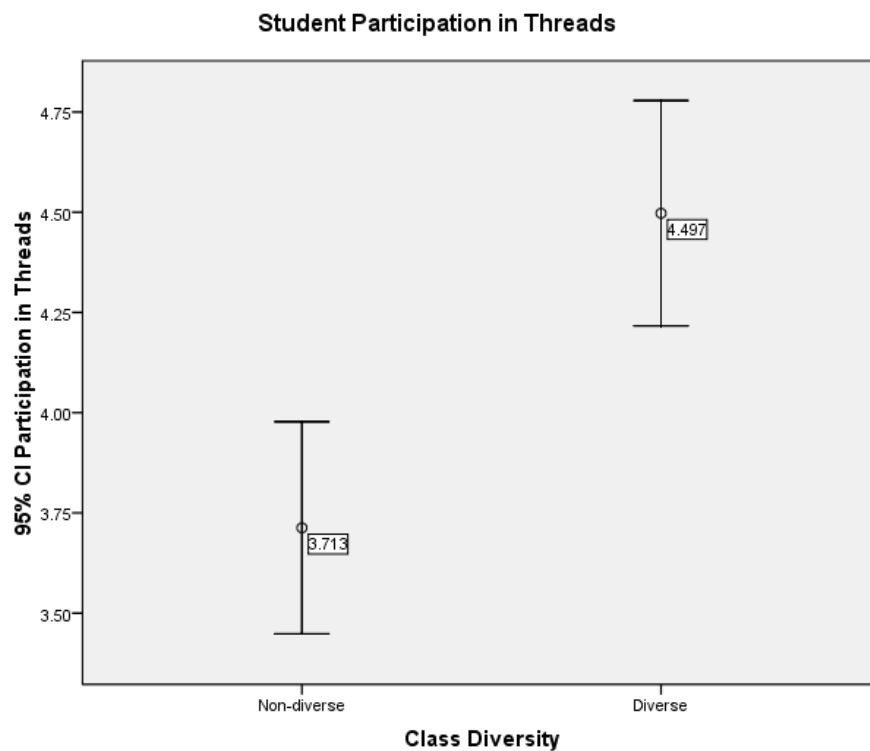


Table 3

H4. Diverse class discussions have a more inclusive dominance hierarchy that contained a greater number of students.

Size of dominance hierarchy

We computed the size of the dominance hierarchy by pooling all student word count scores from both diverse and non-diverse classes and ranked the scores into

quartiles. The numbers of students from diverse and non-diverse classes were compared in the upper quartile. It was found that 11 students from the diverse classes and 4 students from the non-diverse classes were represented in the upper quartile. Thus, the size of the dominance hierarchy in diverse classes has nearly 3 times as many members as the non-diverse groups.

Summary: Inclusiveness

In both measures diverse classes were found to be more inclusive than non-diverse classes. A greater number of different speakers in diverse classes collaborated in threads and a greater range of students spoke at high levels in diverse classes. Therefore, Hypotheses 3 and 4 were supported by the findings.

H5. Diverse class discussions have a higher percentage of turns in which students responded to students (PTP).

There were differences between diverse and non-diverse classes in the percentage of PTP turn sequences. These differences were not the result of any in class communications by the DLs that specifically encouraged PTP interactions. The diverse classes had higher percentages of PTP interactions. Nearly one-third (32.4%) of all turns in diverse classes were those in which students responded to other students. By contrast, only 11.3% of the turns in non-diverse classes were those in which there were PTP interactions. In part, this might have been predicted by the results of previous hypotheses that showed greater participation by DLs in the non-diverse classes and a greater number of different student speakers in the threads of diverse classes. We conclude, therefore, that in diverse classes there was greater PTP than non-diverse classes.

Critical thinking: Mean percentages of critical thinking indicators in diverse and non-diverse groups in individual turns

The analyses following use percentages of various categories of critical thinking propositions and critical thinking responses to make comparisons between diverse and non-diverse classes. The percentages are based on the frequencies of the coded variables in the total number of turns spoken by each of the two diverse classes and the two non-diverse classes across four discussions. All participants spoke a total of 6585 turns across

all discussions, of which student turns comprised 4002 and DL turns were 2583. Only the student turns are analyzed here.

The diverse classes spoke a total of 2307 turns and the non-diverse classes spoke 1695. Because the analyses are based on such a large sample of cases, many indicators were statistically significantly higher in diverse or non-diverse groups when compared with each other. Yet certain indicators had very low frequencies. For example, student discussion turns that referred to works other than the text under consideration were mentioned only .5% by all groups. These extremely low frequency indicators, less than 1% of all turns, have been eliminated from the findings. Note that because the total number of turns varied between the diverse and non-diverse classes it may occur that while there are significant percentage differences between the groups, the total number of instances of each code may actually be higher in a comparison group that had a significantly lower percentage.

H6. Diverse class discussions demonstrate greater percentages of high-level critical thinking propositions than non-diverse classes.

The following codes that met the frequency and statistical significance tests were used to indicate *high-level critical thinking propositions*:

Text-based Claims

Text-based claims were one of the most frequently employed critical propositions. These propositions meet the requirement of “making sense of the text” in discussions of great works. Students refer to ideas or exact passages in the text when making these claims. Following is an example of statements that were coded as text-based claims:

Maybe this is reading way too much into it, but I feel like if you go down to the fifth stanza where the child puts in a man with buttons like tears, that maybe it is like the grandfather who died on a certain day that is coming up. And the grandmother knows when the date is. But the kid maybe doesn't, so she knows when it's coming in the almanac. (Bishop discussion)

The students in the non-diverse classes made more claims based on the text (14.0%) than students in the diverse classes (11.7%). This difference is statistically significant: Chi-square = 5.007, $p = .025$. However, the diverse group actually uttered a greater number of

such claims (269) compared to the non-diverse (238). This raises the question of whether the percentage or the absolute number is more relevant. As an indicator of overall group critical thinking we feel the proportion is more important. And yet members of groups with a higher absolute number would have been exposed to a greater number of examples. The latter, of course, is also a consequence, in part, of the greater number of words spoken in the diverse classes.

Arguments

Arguments are defined as consisting of two or more sentences that are connected by reasoning. Arguments were coded if the speaker made a claim that was backed by warrants and/or evidence and/or drew a conclusion. The words “because” or “so” may frequently appear in arguments. Following is an example of a statement coded as an argument:

I think that he's right in saying that you are not, this eliminates all accountability to everyone else, but it doesn't eliminate accountability to yourself, because you'll always know that you've done these things. Which, I think, that a just man wouldn't be able to live with his own accountability and so, I think that's what separates the just man from the unjust man. It's not that the unjust man is just better at getting away with it. I think it's just that the unjust man is all right with living with the consequences himself. (Plato Discussion)

The students in the non-diverse classes made more arguments (12.6%) than students in the diverse classes (5.7%). The non-diverse classes made 214 arguments, while the diverse classes only had 131. This difference is statistically highly significant: Chi-square = 59.861, $p = .000$. This measure is, certainly, the most important measure of critical thinking quality and the non-diverse group uttered proportionally more than twice as many arguments as the diverse groups.

Summary: High-level critical thinking

Two important measures of high-level critical thinking were analyzed. In both cases the non-diverse classes exhibited significantly higher percentages of these propositions. However, diverse classes actually uttered more text-based claims.

Low-level Critical Thinking

Opinions

Opinions, in which a student uses words like “I think” or “I believe” without any other supporting justification was a frequently employed form of speech during the discussions. An example of an opinion proposition: “I thought it was interesting, but it's kind of hard to follow.” (Plato discussion)

Students in the diverse classes provided significantly more (369) unsubstantiated opinions (16.0%) than in the non-diverse groups, 144 (8.5%). Chi-square = 49.171, $p = .000$.

Claims based on personal experience

Experience-based claims were relatively infrequent but met our criteria for inclusion. Students refer to personal experiences in support of claims. Experience-based claims are higher level than opinions, but still fail to refer to the text for evidence.

I was going to say, maybe the horror was like she was almost ashamed that she was looking at it [a woman's breasts in National Geographic], because she was...Like, you get the idea that she was so young. That maybe she thought the people around her wouldn't want her looking at that. You know, like when you are watching something on TV that you don't want your parents to know that you are watching, whether it's bad or not. Like you feel ashamed when they catch you, so maybe she thought something like that. The people around her would judge her. (Bishop discussion)

The students in the diverse classes made more claims based on personal experience (1.8%) than students in the non-diverse classes (.8%). This difference is statistically significant: Chi-square = 7.492, $p = .006$. However, the diverse groups only uttered 41 of these claims compared with 13 for the non-diverse groups.

Procedural Information

In this category the speakers simply offer simple but helpful procedural information, such as: “it's on pg. 82.” Students in the diverse classes provided more procedural information (5.2%) than in the non-diverse groups (2.8%). Chi-square = 14.847, $p = .000$.

Lecture Information

Each of the four works was accompanied by a lecture. These lectures occurred prior to study discussions. This information is sometimes helpful in a discussion but does not include why the information pertains. Students in the diverse classes provided significantly more information from the lectures (2.3%) than in the non-diverse groups (1.1%). Chi-square = 8.558, $p = .003$.

Text Information

Text information consisted of literal citations from the various texts used in the curriculum. Students in the non-diverse classes provided significantly more information from the texts (26.0%) than in the non-diverse groups (7.8%). Chi-square = 248.80, $p = .000$.

Cultural Information

Statements in this category refer to information drawn from general cultural information and/or personal experience. Students in the non-diverse classes provided more information from the external cultural sources (9.3%) than in the diverse groups (6.5%). Chi-square = 10.156, $p = .001$.

Summary: Low-level critical thinking

The diverse groups had a much greater frequency of utterances in which relatively low-level critical thinking was exhibited: unsubstantiated opinions; claims based on personal experience; procedural information; and information from lectures. Only opinion-type and personal-experience-based statements, however, represented knowledge claims, while all the information categories were offered without explanation as to relevance.

Of a total of 513 opinions uttered by all students, 369 (71.9%) were made by students in diverse classes.

Within some of the low-level categories, the non-diverse classes also provided significantly more low-level information from the texts and cultural sources than the diverse classes. Other than the clear-cut difference on *opinions*, we are inclined to discount any differences between the groups concerning low-level critical thinking.

In any case, given that the non-diverse classes had significantly higher percentages of high-level critical thinking, and lower percentages of opinions, we may conclude that Hypothesis 6 was not supported.

H7. Diverse class discussions demonstrate greater percentages of high-level critical responses to previous communications than non-diverse classes.

High-level critical responses

As for critical thinking propositions, the comparisons are based on the percentages for each of the indicators that were explicitly critically responsive across all turns in four discussions for the two diverse classes and the two non-diverse classes.

Elaborations and clarifications

In elaboration/clarification critical responses, students either add to or seek clarification of previous statements uttered by other students or DLs:

T18. DL He is mending his net. Yeah. By the way, what's a gloaming?

T19. S4. Gloaming is a period of twilight. It is an Irish term. As the sun is sort of set, but it is still light out. Even though there is no sun. (Bishop discussion)

The non-diverse classes contributed elaborations/clarifications in 13.9% of all turns, while the diverse classes had 11.6% of their turns on this indicator. This difference is significant: Chi-square = 4.91, $p = .027$. However, the diverse classes actually uttered more turns with this indicator (267) than the non-diverse classes (236).

Questions

Students in the diverse classes consistently asked more high-level questions than in the non-diverse classes. In no case other than procedural information questioning did question levels reach a higher level of usage in the non-diverse classes, although the percentages were equivalent in the use of conditional questions. To simplify the statistical analysis, the following high-level question types were combined: explanation; clarification; hypothetical; and rhetorical. The diverse classes had these questions in 10.4% of the turns while the non-diverse classes only had 4.4%. This difference is highly significant: Chi-square= 49.821, $p = .000$.

Re-phrasing

The diverse classes had a higher percentage of statements in which the speaker rephrased (2.9%) the previous speaker's utterance when compared with the non-diverse classes (1.6%).

S10: I think, we finished it.

S16: We finished Thrasymachus (from Plato discussion)

This difference is significant: Chi-square = 6.921, $p = .009$.

Summary: High-level critical responses

The findings for critical responses were mixed. Whereas, non-diverse classes exhibited significantly higher percentages of high-level elaboration/clarification responses, diverse classes displayed significantly higher percentages of high-level questions and re-phrasings. Moreover, given that there were actually a greater number of elaboration/clarification statements in the diverse classes, the results suggest that these groups were more critically responsive than the non-diverse groups

Low-level Critical responses

Procedural information questions

Students in the non-diverse classes asked more low-level procedural information questions (4.6%) than students in the diverse classes (2.8%). Chi-square= 9.536, $p = .002$.

Agreement

The diverse classes had a higher percentage of statements in which the speaker simply agreed (12.7%) with the previous speaker ("o.k.", "Right") when compared with the non-diverse classes (7.7%). This difference is significant: Chi-square = 26.162, $p = .000$. We note that the levels of *disagreement* were of insufficient frequency to use in making comparisons between diverse and non-diverse classes.

No other measure of critical responses met the frequency test or was statistically significant.

Summary: Low-level critical responses

The non-diverse groups asked more low-level procedural information questions and the diverse classes uttered more statements of agreement with other members of the groups. On balance, the agreement statements, while low-level, are compatible with our

conclusion that the diverse classes were somewhat more critically responsive than the non-diverse classes

Summary: Critical Responses

Given these results for both high-level and low-level critical responses, we conclude that *Hypothesis 7 was moderately supported.*

Discussion

Participation, Inclusiveness, Peer-to-Peer Interaction, and Critical Thinking in Diverse and Non-Diverse Groups

The results showed that the two principal goals of interpretive discussions, to afford students a context to try out their ideas and to construct high-level propositions using critical thinking might be met in different ways. In this study, high levels of participation, inclusiveness, and PTP were encouraged, apparently, in an environment characterized by a high proportion of students who had good interpretive abilities, while strong aggregate amounts of critical thinking were present in environments where a small proportion of students had excellent interpretive skills. The Supreme Court decision that inspired this study claimed that diversity would benefit all students. If participation, inclusiveness, and PTP are the criteria then that claim is satisfied. However, the decision also emphasized that diversity leads to more enlightening exchanges. If that is the criterion, then we cannot confirm that diversity is associated with this outcome.

The findings may reveal a trade-off between participation, inclusiveness, and PTP on the one hand and level of critical thinking on the other. The data suggest that the diverse groups benefited in that they had much higher levels of participation, inclusiveness, PTP, and critical responsiveness than the non-diverse groups. Yet, there may have been a cost in terms of discussion quality. While the non-diverse groups had much weaker participation, inclusiveness, and PTP, they made more arguments and text-based claims, while diverse groups had higher levels of opinions and claims based on personal experiences. We add, however, that there were no significant differences between the diverse and non-diverse students within the diverse classes on critical thinking variables. It was not the case that minority students spoke more opinions, for example, than other majority students in the diverse classes.

Opinions and claims based on personal experience, which have been identified as weaker forms of knowledge propositions than arguments, may have a lower threshold for communication in a group discussion. Did the diverse classes establish a climate in which opinions and experience-based claims became normative? Sprott (1958), reviewing dozens of studies of small group interaction, concluded that normativeness is a prerequisite for continuous interaction:

The interacting parties must develop a mutually agreed system of expectation in accordance with which the action of the individual is geared to the reaction of others, and theirs, in turn, to his responses. What expectations, what rules, will be precipitated will clearly depend partly on the general culture within which the participants have been brought up...(p. 144)

In the present discussion groups, such normativeness was not explicitly formed at the outset. Perhaps, in the diverse groups, the presence of diverse students helped to create a classroom climate in which students were free to express their opinions and make claims based on personal experience. If this were so, then a larger number of students might have been afforded entry into the discussion. It is tempting to speculate that the diverse groups may have tacitly established norms or a classroom climate in which student speakers introduced themselves to the group through statements that revealed their personal takes on the works in the courses. Perhaps, diverse groups encourage such personal revealing as a way for the group to get to know each other as a pre-condition of further interaction. In diverse groups students might need to appreciate different backgrounds as information to speak to each other comfortably. To work together we need to know more about each other. Alternatively, the diverse groups in this study may not have been as well prepared to formulate high-level claims and arguments as the non-diverse groups.

How would such norms get established in the diverse groups? We examined DL statements for critical propositions and responses. While there were few differences between the instructors – a comforting fact on the one hand that DLs were not contributing unduly to the variance that we have explained as differences in diversity – the DLs in the diverse classes did make a statistically significant higher percentage of *opinion* statements (3.5%) than the DLs in the non-diverse classes (2.0%). Moreover,

they asked a higher percentage of questions *eliciting opinions* (8.6%) than DLs in the non-diverse classes (4.7%). Therefore, the DLs in the diverse classes may have played an instrumental role in establishing a climate in which sharing personal opinions and experiences were a valid means of participating in the group's discussion.

On the other hand, in the non-diverse classes only a small number of students contributed high-level critical propositions, such as arguments. In fact, only four students in the two non-diverse classes contributed 1/3 of the arguments. A similar pattern held for critical responses in which only a few students spoke the elaborations/clarifications: four students produced almost 2/3 of elaboration/clarification responses. Did another set of tacit norms, therefore, regulate interaction in the non-diverse groups such that higher-level critical propositions were the gold standard? Only a few students may have been able to meet that standard. Whereas the diverse groups featured an easier pathway into the discussion thus maximizing participation, the non-diverse groups, by setting very high standards, not only facilitated very high-level performance by a small hierarchy, but also, perhaps, excluded the majority of students who, as freshmen, were not yet prepared to jump in at that level.

While our approach to data analysis led to clear-cut differences in the kinds of propositions used in diverse and non-diverse groups, it may have hidden aspects of the complexity of student speech. When students offer opinions, for example, in the same turn they may also put forward other kinds of propositions. Further examination of the data revealed that in 22.5% of the turns in which students communicated opinions they also made claims, including text-based claims (17.9%) and claims based on external sources of evidence, such as cultural knowledge, news or historical sources (4.6%). For example, here is a turn in which a student offers an opinion, but then makes a text-based claim and an argument.

All right. Just going off what (*Speaker 7*) said. I, I think that our educational system is more about like presenting facts, or information, or however we're going to describe it, and then leaving it open for you to decide whether or not you think it's right or not. But I feel like Socrates' system, there is one right answer, and any other answer is just simply not allowed. And so we can all have different opinions and neither of them, or none of them are, like, right or wrong. They're just different, but not for Socrates. (Plato discussion)

Moreover, in diverse classes opinions were accompanied by arguments in 11.7% of the cases. Thus, in a considerable percentage of cases students in the diverse classes were not inclined to let their opinions remain unalloyed. Rather, they joined their opinions with more formal claims and to a lesser extent with arguments. It is interesting that the non-diverse groups, who had proportionally only half as many opinion statements, joined their opinions with claims in only 9.4% of their turns. This does not diminish the conclusion that non-diverse groups had stronger text-based claims and arguments but it mediates the finding to suggest that some students in diverse classes may have used simpler propositions as an initial means for formulating, subsequently, more complex ones.

Follow-up questions for the study instructors

Do high levels of participation, inclusiveness and student interaction in a discussion group imply lower levels of critical thinking? Are there two patterns in discussion groups: a more egalitarian pattern such as in the diverse groups with high participation and lower critical thinking; and a more hierarchical, less participatory pattern with higher critical thinking as in the non-diverse groups? At this juncture of the study we decided to approach the four DLs to ask them for their views. Our intent was to supplement the current investigation with a method to assess our interpretations of the “climate” and other factors that may have been operative in the discussion environments as well as to inform our recommendations for future DL training and research. We asked the four instructors to respond in writing to the conclusions reached above and a proposed model for moving from simpler to more complex propositions. The instructors’ responses are in italics. Instructors 1 and 2 led diverse classes, while 3 and 4 led non-diverse classes. The author’s interpretive comments of the instructor’s criticism are included in plain text after each instructor’s response.

Q1. “Based on your experience of this class can you explain why you think the diverse classes had higher rates of participation and inclusiveness than the non-diverse classes?”

Instructor 1 (diverse class)

I am a little inclined to believe that the lower critical thinking threshold may have been important here. I imagine that I had a tendency to try to build a welcoming discussion environment, perhaps at the cost of creating a less

rigorous discussion environment. I would expect that opinion level responses were high in my class because I sought inclusion as a high priority. (Instructor 1)

In this statement we find some evidence for identifying the DL as a mediating factor in the effects of diversity on group discussion. We hasten to add that we provided neither advice nor constraints in advance of the study as to how to manage a diverse class. Such interference would have defeated the experimental design, indeed the very objectives of the study. We suspect that this very forthcoming statement by the instructor is a clue as to how expectation norms (see Sprott above) get established in a class discussion.

I agree that frequent opinion level contributions limit the sophistication of discussions. The thin line to balance on is that one on which you try to make the discussion an inviting place for those who don't tend to speak and the requirement that their contributions...well...contribute. Pedagogically I am unsure on whether the inviting or the demanding environment fosters better student discussion. Having any statement be acceptable allows all students to be part of the group and encourages more statements by all. I would say lowering the level of critical inquiry is necessary to participate in the discussion does lower the overall quality of the discussion, but as an instructor I am sure that I let that often happen because it is easier, and less stressful to myself and the students. A bit more stress in the discussion may be better for all of us. (Instructor 1)

This is specific support for our claim that opinions, for example, may be used by a greater number of students to gain entry to the group discussion. The instructor raises an issue not previously considered: that lowering the level of critical inquiry lowers the stress level in the group. On the one hand, wide participation indicates that the group is unstressed from this perspective. But clearly, the instructor is ambivalent about this, as some stress - - waiting out the group after a difficult inquiry or demanding higher levels of critical thinking - - may be needed to raise levels of critical thinking.

I will say that students often bring opinion and personal experience into discussion as a means of establishing credentials on which to make their claims. "Well when I was a kid I had to earn money for my shoes, and I don't think Plato understands human nature and how people want to make money" or whatever. The idea is that those opinion statements are, as you say, vehicles for accessing textual references. These vary greatly however. Opinion and experience are often used as trump card to say why a textual reference is inane. "I have seen this personally, therefore author x doesn't get it." The reference to their own lives can act as a great vehicle for bringing the esoteric and different into the "real world"

of the student's life, but the students tend to counter pose those things (the "real world" and the world in the text) and their experiences are generally given precedence.

Here we hit on one of the big problems that this whole line of inquiry examines. How do we get students who are not good at engaging these ideas from their starting point, unable to readily engage the material on its own level and dependant on their own experiences to interact meaningfully with the texts, to the point that we want them to reach, able to deal with the text and the ideas presented on their own, able to speak about justice in the terms that Plato uses. What path will move some students up to the level of other students without overburdening one group and boring and frustrating the other? Diversity in the classroom, if managed, might be able to help in this process if the opinion oriented students can be brought up to higher level critical thinking, and the high level critical thinking students can be informed by the opinions and experiences of other with backgrounds different than their own. (Instructor 1)

The instructor argues that all students may use opinion and personal experience to make their claims. Indeed, the students in the non-diverse classes also voiced a considerable number of opinions. Of extraordinary interest is this instructor's explanation about how students personalize what to them, in many cases, are their first encounters with extremely unfamiliar and difficult texts. Is there a case, therefore, for the voicing of opinions and personal experiences as a positive contribution to discussions? This comment implies an important role for opinions and personal experiences and lends support to our stage model for managing discussions, of which more below.

Instructor 2 (diverse class)

A difference that I noted to myself was that this [course] section seemed to be intellectually consistent. What I mean is that they all (for the most part) appeared to begin at a similar level. Usually my Freshman Studies sections seem to have 3 or 4 students on the most accomplished side of the spectrum and then 3 or 4 at the opposite end. This class, for whatever reason had no one that I would identify as walking in as a super-student. While not a one of them would have verbalized this they all seemed to understand that they were intellectual equals and the lack of hubris led the group to want to help each other understand the works more fully. These young people were extraordinarily generous with one another. By that I mean they would allow each other time and space during discussions to wander, ponder, speculate and yes, fail. Even when one of them might be headed in the wrong direction they seemed to allow their classmates to be wrong without any embarrassing consequences. Certainly, they would point out inaccuracies and disagree with passion but no one seemed to be scared to

fail in front of the class; thus I think I witnessed (participated in?) a great deal of intellectual “risk-taking”. (Instructor 2)

This instructor’s response leads us to the conclusion that no matter how rigorous our randomized design, in which both the instructors, in a first step, and the students, in a second step, were both selected randomly, it was possible that the resulting diverse class in this case contained many individuals “at the same level”. While this may be true, according to our data, there was still variation within the diverse classes relative to the non-diverse classes. As reported above, the upper quartile of speakers, using word count as a measure, contained 11 speakers of 30 total in both diverse sections (compared with 4 in the non-diverse sections). In fact, there were performance hierarchies in the diverse classes but they were relatively large and may have given the instructor, and the students, the impression that all the students in the group were at equivalent levels. There are two implications: the randomized design produced a relatively equal class in this diverse section just as it led to a small dominance hierarchy in the non-diverse classes; given the complexity of the discussion, referred to in our response to Instructor 1, it is difficult for DLs to be fully aware of the range of performance in discussion groups.

Please allow me to clarify a final point: These traits and behaviors that I have been discussing were true for all the students in this class. These comments are not strictly and solely about students who may be considered “diverse”.

This comment supports our claim that the diverse students were apparently no different than majority students within the diverse classes. Concerning the high use of opinions and personal experiences:

*I think I believe this might be accurate. I’m hedging somewhat here simply because analyzing this sort of data is not my area. That said, I did allow what, in hindsight, seems like a tad more personal experience statements to occur. To begin with, it is my experience that a great many college-age people will use this discussion tool (crutch perhaps?) I felt that as long the point being made was relevant and that we always returned to the text to prove our point (or at least offer evidence) this was a permissible (and perhaps valuable) tool. My thinking was that it’s difficult to tell students they can’t use analogies from their own experiences when Plato so often uses analogy and simile to make his point, when Einstein talks so much about his thought experiments and Elizabeth Bishop writes poetry like *The Moose* or *In the Waiting Room* (“I am an Elizabeth...”).*

I feel that despite the implication that utilizing personal experience, as a tool or metaphor or analogy is “bad” that maybe it’s simply a tool like any other and can be used for good or misused for ill.

This instructor supports the use of opinions and personal experiences provided that there is eventual “return to the text to prove the point”.

Instructor 3 (non-diverse class)

All in all, I can say that my section of Freshman Studies this past year was much less talkative than the one I taught a year ago. I would say that this is because I had a higher number of students who seemed very intimidated to speak up in class and also because I had at least three students who mastered the material well and spoke up often (and perhaps further intimidated the students who were uncomfortable with speaking up. For my specific section, I had at least two students who were more skilled at critical thinking than most Freshmen I have encountered.

In a separate question made only to non-diverse instructors, Instructor 3 also commented that s/he interacted more frequently with students who made text-based claims or arguments. Thus, this DL chose, or could see no other way than to reinforce the hierarchy.

This accords with my personal experience teaching a non-diverse section. However, I will share that in the fall of 2007, I taught a non-diverse group that perfectly fit the above description for a diverse group (that is, much more participation from everyone in the class, but in general a lower level of critical discourse.

Instructor 4

With respect to greater critical thinking and making text-based claims, I believe that these two are very closely and directly linked. To answer one of the specific questions, I probably did respond with a special enthusiasm when students made such claims or contributions. I think that the conclusion which states “very high standards based on shared academic norms not only facilitated very high-level performance by a small hierarchy, but perhaps excluded the majority of students who as freshmen were not yet prepared to jump in at that level” does seem plausible.

The non-diverse instructors, therefore, offered support for the inverse relationship between participation and critical thinking. Both of these instructors interacted more frequently with students in the hierarchy. Although such communications would tend to

reinforce and probably stabilize the hierarchy, it seems a natural course of action. After all, the discussion leaders are there to discuss and they will respond to those who are communicating frequently and to those who are exhibiting high levels of critical thinking. However, now and then it would also seem plausible for them to encourage low or moderate participation speakers to join in so as to expand the hierarchy.

Proposal for a stage model of discussion

Our initial findings, prior to eliciting DL commentary, had led us to suggest a possible stage model for encouraging both the widest participation and highest quality in a diverse setting:

- 1) To be part of the discussion, actively, requires each student to join the group by making statements that satisfy the overall purpose, making sense of the text, but each in his or her own way. Students utter opinions and personal experiences that afford them entry into the discussion and such communications might let others know “where they are coming from”, *i.e.*, their background.
- 2) In a subsequent stage, students could move on to more difficult claims based on the text or other sources.
- 3) Finally, students attempt full arguments by marshaling claims, warrants and evidence in longer turns of discussion.

The model assumes that it may be more facilitative to move from simpler to more complex propositions. In non-diverse classes such a ladder pathway may not have been available. Following our model, the implications for DLs would seem to be clear: if students do not join evidence-based claims to their opinions, then contingent questioning and scaffolding should be employed to help them to achieve this goal. This would seem to be part of the general advice for DLs as indicated in points in the DL Guide (in Appendix 1), 4. “Respond to students personally”, and 6. “Support students production of high-level claims through questions and alternative arguments.” In settings where DLs establish tacit norms of high-level critical thinking, they may need to encourage more inclusive participation by students on the periphery.

Q. 2 “This leads us to suggest a possible stage model for encouraging both the widest participation and highest quality in a diverse setting. Please respond.”

Instructor 1

Those would seem to be logical steps. In my class I tend to start off with very straightforward questions to "lubricate their jaws." I work on the assumption that speaking has its own inertia and once they get going it is easier to demand more of them. I do agree that this staged model of discussion could be useful. They could begin essentially by engaging the room, becoming participants, and then move on to engage more specific and sophisticated ideas, thus engaging the text and higher-level course aims (Instructor 1).

Instructor 2

Stage 1. I agree with Stage 1 but take some umbrage with the following: “Students utter opinions and personal experiences that let others know where they are coming from.” My sense of this is that while it is certainly helpful for the students’ opinions or analogies to give each other a sense of background, my primary pedagogical reason for “allowing” this discussion technique is to encourage insight into the text via the students’ opinions or experiences. In other words, it seems useful in assisting students in expressing themselves – particularly if they haven’t quite developed fully the skill-set necessary for advanced college-level articulation yet. Using illustrations from life experiences can be useful in helping the group understand one another.

Stage 2. I believe Stages 2 and 3 can be combined: students should be encouraged to back up claims and opinions and justify their experiential anecdotes by returning to the text. They (or at least, in my experience, most of our Freshman Studies freshmen) can acquire this skill/utilize this technique almost immediately.

Stage 3. “I concur! In a subsequent stage, students could move on to more difficult claims based on the text or other sources.”

Instructor 3

This non-diverse class instructor agreed with the model, but suggested that Stage 1 should include the following: *Or by asking comprehension-level questions about the text.*

Instructor 4

Of course I believe that “that discussion is a form of active learning leading to cognitive benefits,” but I do not believe that the students who participate actively in a specific discussion are the only ones who benefit & learn. Allowing a smaller higher-level hierarchy to provide a good model (with “very high standards”) for

the rest might be a very good rung on the “ladder” to high-level participation for the others.

Summary: Instructor responses to the model

Three of the four instructors accepted the stage model with minor corrections. One instructor suggested that experience-based claims were a valid source of interpretations as well as being useful for students to obtain entry into the discussions. Instructor 4, however, indicated that a small hierarchy with high quality could provide a useful model for the non-participating students who might participate subsequently. This instructor seemed unwilling to trade critical thinking for active participation by all students.

Conclusion: Does diversity lead to benefits - - the greatest good for the greatest number - for all students in college discussions?

We conclude that with respect to participation, inclusiveness, and peer interaction, many more diverse class students benefited compared with non-diverse classes. While the average level of critical thinking propositions was higher in non-diverse classes only a few of those students actively benefited. The fact that instructors in these sections tended to converse with the hierarchy meant that the other students were relegated to audience roles. Most of the students in the non-diverse classes were only passively exposed to the text-based claims and arguments used to measure high-level critical thinking. One non-diverse class instructor stated that the very high standards of this smaller group provided a good model for the other students. If this is so, then the diverse class performance with respect to text-based claims might be reconsidered. Recall that although the non-diverse classes had a small but significant edge in the percentage of text-based claims among their students, there were a higher absolute number of text-based claims in the diverse sections, no doubt associated with the greater participation of the diverse sections. Thus, all the diverse class students were exposed to more text-based claims than were the non-diverse classes. The same was true of critical responses. In the diverse classes, the students voiced a greater number of elaborations/clarifications, high-level questions, and re-phrased their peers' previous communications. When we add to this the fact that diverse classes had more PTP interaction, it is compelling to conclude that diverse class students had much greater exposure to a core principle of discussions: that discussants

follow-up on each others' communications; responsiveness makes for a rich, sustainable discussion.

Another aspect of the discussions in the diverse classes might be re-examined in answering the question as to which groups benefited more. This concerns the percentage of diverse class students who may have voiced opinions, but also added claims in the same turns of conversation. This measure is important because some students who voice opinions know that better evidence is required to make interpretations of works. While less than 10% of the non-diverse groups combined opinions with text-based claims, 22.5% of the diverse classes did so. Perhaps, the weaker participants in the non-diverse classes did not learn as well from the model hierarchy as supposed?

Finally, what about the status and role of opinions and personal experiences in interpretive discussions? When students' arguments contain evidence based on personal experience this raises issues of relevance, aptness and application just as with factual quotes, such as those based on text. Such knowledge is that gathered from past events, whether by conscious observation or by consideration and reflection (Scott, 1994). These experiences are likely to have been transformed through the student speaker's cultural background. Diverse students, and perhaps other students may believe that their experiences and opinions based on their experiences are valid and relevant in informing their interpretations. In speaking of the advantages of student diversity, we often refer to their unique backgrounds and, hence, experiences, do we not? The diverse class instructors argued that these kinds of contributions played a useful role both in affording easy entry into the discussion and in making use of students' backgrounds to make interpretations and, perhaps, letting others know where they were coming from. Three of the four instructors also subscribed to the stage model we proposed that found utility in opinions/personal experiences as a way of getting started, provided that stronger evidence-based claims were used in subsequent stages.

In summing up, let us weigh the evidence. As measured by several criteria in the diverse classes, there was wide and inclusive participation, high peer-to-peer interaction, greater understanding by a larger proportion of students that opinions need to be backed up with evidence, and exposure to more high-level claims and responses. In the non-diverse classes there was greater instructor domination, lower student participation, active

critical thinking in only a limited number of students, and lower peer-to-peer responsiveness. We conclude that the diverse classes provided more value -- the “greatest good to the greatest number” -- to their students than the non-diverse classes.

Implications for Future Research

The results of the preliminary study are encouraging but have need of replication in a larger sample of liberal arts colleges and other higher education settings. Sample size is always important but particularly problematic here. Because the sample was experimentally divided into diverse and non-diverse groups (without diverse members), we could not merge the groups to conduct regression analyses with sufficient numbers that might have shed light on the relative contributions of diverse students to the participation, inclusiveness and critical thinking outcomes. *Because every test of diverse vs. non-diverse members within diverse groups failed to detect any differences in the productivity measures employed*, this study failed to determine the precise role, if any, that diverse students played in the success of diverse groups. Thus, we had to formulate an alternative explanation for the results, to wit that the diverse classes had different norms and expectations from the non-diverse groups, norms that tacitly or with some DL priming encouraged the voicing of propositions like opinions and claims based on personal experience that afforded easier entry into the discussion. Another explanation was possible: the diverse groups and their DLs may have had lower expectations for themselves and believed themselves less capable of uttering higher-level critical thinking propositions. For either explanation, we need to call attention to the possibility that either the DL and/or the students are tacitly operating under what has been termed the “deficit model” for the students in their classes. Deficit models assume that diverse or different students are not as well prepared as majority students to meet requirements. The present study methods cannot measure such expectations for diverse students if, in fact, they existed. Future research designs would need to conduct post-observation debriefings if these expectations are suspected.

Furthermore, the research needs an experimental design to control threats to validity imposed by natural variations in DL and student participation. We can build from our current DL training program to alert instructors to the presence of student dominance hierarchies, which are a small sub-group of students who engage in frequent discussion.

The more we can use the DL Guide to encourage practices to raise the participation rates of all students, *i.e.*, the weaker the hierarchy, the more valid will be the test that racial and ethnic diversity contributes to aggregate benefits, such as those used in this research, to group performance. Naturally, this would satisfy the equally important goal of increasing learning in a larger distribution of students if we accept that discussion is a form of active learning leading to cognitive benefits. The two instructors in diverse classes added comments about DL training:

I think the expanded training of DLs could help to add greater control for the study. More importantly it would be a helpful pedagogical tool. It is not always easy to recognize the variety of participation patterns developing in the classroom. Some of the terminology in the study was entirely new to me. Thus I was not keyed in to observing them. Furthermore there is no standard rubric for what constitutes a good discussion, hence we tend to seek what we each find satisfying in the classroom. Clearly enunciating these models of class participation (hierarchies, turns, opinion statements....) could help DLs recognize point at which their discussions are changing in positive or negative ways. (Instructor 1).

The DL training could be used to control threats to validity imposed by natural variations in DLs (Instructor 2)

Instructor 1 comments that the very measures used to assess discussions, such as “hierarchies, turns, etc.” should be included in the DL Guide as ways to sensitize leaders to how discussions are going. Given the fact that each discussion has thousands of words, hundreds of turns, and dozens of threads it is difficult for DLs to cope with the complexity. Any heuristics, therefore, might serve DLs to get a handle on the discussions. Instructor 2 welcomes additional training, but also points out an intrinsically difficult issue in conducting studies of diversity: that DLs vary. Although the instructor left this unsaid, it is also the case that DLs have been leading discussions without any training for many years and have probably constructed a distinctive approach that may be resistant to change in some cases.

The points about using the methodology - - hierarchies, turns, opinions, etc. - - to cognize the complexity of the discussion are a constructive addition to the DL Guide. It also reminds us that a significant outcome of the study is methodological. We have

developed a set of quantitative indices that discriminated performance in diverse and non-diverse groups. These measures should be useful in going forward to conduct more definitive studies using larger samples of classes. Future studies could now benefit from theoretical advances made in this investigation. The primary question, does diversity benefit all students, has been refined. The new questions concern the management of discussion groups under different conditions: how a critical mass of diversity affects the climate, expectations and the formation of hierarchies in groups; how discussion leaders manage hierarchies and how they respond variably to diversity in terms of how they facilitate entry of all students into discussions; and, how they mediate the students' stages of critical thinking.

References

- Bales, R.F. (1950). A set of categories for the analysis of small group interaction. *American Sociological Review*, 15, 257-263.
- Bales, R. F., Strodtbeck, F. L., Mills, T. M., & Roseborough, M. E. (1951). Channels of communication in small groups. *American Sociological Review*, 16, 461-468.
- Beck, R. J. & K. A. Clarke-Stewart, K. A. (1998). Improving 5 year-old children's narrative recall and comprehension. *Journal of Applied Developmental Psychology*, 19(4), 543-569.
- Beck, R. J. & Wood, D. (1993). The dialogic socialization of aggression in a family's court of reason and inquiry. *Discourse Processes*, 16(3), 341-362.
- Beck, R. J., Fitzgerald, W. J., & Pauksztat, B. (2003). Individual behaviors and social structure in the development of communication networks of self-organizing online discussion groups. In B. Wasson, S. Ludvigson, & U. Hoppe (Eds.), *Designing for Change in Networked Learning Environments* (pp. 313-322). Dordrecht: Kluwer.
- Beck, I.L. and McKeown, M. G. (2007). How teachers can support productive classroom talk: Move the thinking to the students. In R. Horowitz (Ed.), *Talking Texts: How speech and writing interact in school learning* (pp. 207-220). Mahwah, NJ: Erlbaum.
- Bok, D. (2006). *Our underachieving colleges*. Princeton, NJ: Princeton University Press.
- Bonito, J. A. (2002). The analysis of participation in small groups: Methodological and conceptual issues related to interdependence. *Small Group Research*, 33(4), 412-438.
- Burt, R. S. (1999). The social capital of opinion leaders. *Annals of the American Academy of Political and Social Science*, 556, 37-54.
- Chang, M. J., Witt, D., Jones, J., & Hakuta, K. Eds (2003). *Compelling interest: Examining the evidence on racial dynamics in colleges and universities*. Palo Alto, CA: Stanford University Press.

- Costa, P.T., & McCrae, R.R. (1992). Normal personality assessment in clinical practice: The NEO Personality Inventory. *Psychological Assessment*, 4(1), 5-13.
- Furnham, A., Christopher, A., Garwood, J., & Martin, N.G. (2008). Ability, demography, learning style, and personality trait correlates of student preference for assessment method. *Educational Psychology*, 28(1), 15-27.
- Gould, R. V. (2002). The origin of status hierarchies: A formal theory and empirical test. *American Journal of Sociology*, 107(5), 1143-1178.
- Griebler, S. (2009). *A study of personality and student participation in college discussion classes*. Honor's Thesis, Lawrence University, Appleton, Wisconsin.
- Grutter v. Bollinger et. al. (2003). Certiorari to the U.S. Court of Appeals for the 6th Circuit, No. 020241. Argued April 1, 2003. Decided June 23, 2003.
- Gurin, P., Dey, E. L., Hurtado, S., & Gurin, G. (2002). Diversity and higher education: Theory and impact on educational outcomes. *Harvard Educational Review*, 72 (3), 330-366.
- Hacker, D.J. & Graesser, A. C. (2007). The role of dialogue in reciprocal teaching and naturalistic tutoring. . In R. Horowitz (Ed.), *Talking Texts: How speech and writing interact in school learning* (pp. 253-270). Mahwah, NJ: Erlbaum.
- Haroutunian-Gordon, S. (2009). *Learning to teach through discussion: The art of turning the soul*. New Haven, CT: Yale.
- Kalma, A. (1991). Hierarchisation and dominance assessment at first glance. *European Journal of Social Psychology*, 21, 165-181.
- Lamb, T. A. (1980). Paralanguage hierarchies in dyads and triads: Talking first and talking the most. *Social Behavior and Personality*, 8, 221-224.
- McKeachie, W. J., Patrick, P., Lin, Y. G., and Smith, D. A. F. (1986), *Teaching and learning in the college classroom: A review of the research literature*.
- McLeod, P. & Lobel, S. (1992). The effects of ethnbiic diversity on idea generation in small groups. Paper presented at the Annual Academy of Management Meeting, Las Vegas, Nevada.
- Mannix, E. & Neale, M. (Eds.), *Research in the management of groups and teams* (Vol.1). Greenwich, CT: JAI Press.

- Mullen, B., Salas, E., & Driskell, J. E. (1989). Salience, motivation, and artifact as contributions to the relation between participation rate and leadership. *Journal of Experimental Social Psychology*, 25, 545-559.
- Nathan, M. J., & Kim, S. (2007). Regulation of Teacher Elicitations and the Impact on Student Participation and Cognition. Working Paper No. 2007-4. Wisconsin Center for Education Research. Madison: UW Madison.
- Nunn, C. (1996). Discussion in the college classroom: Triangulating observational and survey results. *The Journal of Higher Education*, 67 (3), 243-266.
- O'Reilly, C., Williams, K., & Barsade, S. (1997). Group demography and innovation: Does diversity help? In E. Mannix & M. Neale (Eds.), *Research in the management of groups and teams* (Vol. 1). Greenwich, CT: JAI Press.
- Orfield, G. , & Kurlaender, M. (2001). *Diversity challenged: Evidence on the impact of affirmative action*.
- Pascarella, E. & Terenzini, P. (2005), *How college affects students*, V.2: A third decade of research.
- Page, S. E. (2007). *The Difference: How the power of diversity creates better groups, firms, schools, and societies*. Princeton, NJ: Princeton.
- Prince, M. (2004). "Does active learning work? A review of the research," *Journal of Engineering Education*.
- Saunders, W. M. & Goldenberg, C. (2007). The effects of an instructional conversation on English Language Learners' concepts of friendship and story comprehension. In R. Horowitz (Ed.), *Talking Texts: How speech and writing interact in school learning* (pp. 221-252). Mahwah, NJ: Erlbaum.
- Schmid Mast, M. (2001). Gender differences and similarities in dominance hierarchies in same-gender groups based on speaking time. *Sex Roles*, 44(9/10), 537-556.
- Scott, J. W. (1994). The evidence of experience. In J. Chandler, A. I. Davidson, & H. Harootunian, Eds. *Questions of Evidence: Proof, practice and persuasion across the disciplines* (pp. 363-387). Chicago: University of Chicago.

- Shelly, R. K., & Troyer, L. (2001a). Emergence and completion of structure in initially undefined and partially defined groups. *Social Psychology Quarterly*, 64(4), 318-332.
- Shelly, R. K., & Troyer, L. (2001b). Speech duration and dependencies in initially structured and unstructured task groups. *Sociological Perspectives*, 44(4), 419-444.
- Sprott, W. J. H. (1958). *Human groups*. London: Penguin.
- Stasser, G., & Taylor, L. A. (1991). Speaking turns in face-to-face discussions. *Journal of Personality and Social Psychology*, 60, 675-684.
- Souter, A. O., & Rudge, L. (2005). What the discourse tells us: Talk and indicators of high-level comprehension. Unpublished manuscript. Presented at the American Education Research Association Annual Meeting. Montreal, Quebec, Canada.
- Stein, R. T., & Heller, T. (1979). An empirical analysis of the correlations between leadership status and participation rates reported in the literature. *Journal of Personality and Social Psychology*, 37, 1993-2002.
- Watson, W., Kumar, K., & Michaelson, L. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogenous and diverse task groups. *Academy of Management Journal*, 36, 590-602.
- Williams, K. Y., & O'Reilly, C. A. (1998). Demography and diversity in organizations: A review of 40 years of research. In B. M. Staw & L. L. Cummings (Eds.), *Research in Organizational Behavior* (pp. 77-140). Greenwich, CT: JAI Press.

Appendix 1

Discussion Leader Guide: a Review of the Literature

To control for discussion leader variance all instructors received instruction in active discussion pedagogies. An extended treatment of this factor is needed because differences that are attributed to diversity among discussions groups may be due to differences in instructor leadership.

The research literature on discussions in academic settings has largely addressed discussion leader strategies and practices associated with productive discussions. In a rare higher education study, Nunn (1996) found that oral discussion strategies that support peer interactions were more effective than typical instructor-dominated discussions. Beck and McKeown's (2007) study of middle-school students suggests that good teachers help students construct meaning and help them build meanings, whereas poor teachers do the cognitive work themselves and have the goal of simply getting the facts out. Good discussion leaders are responsive to discussants, not simply providers of information or interpretations. Beck and McKeown offered that ill-formed student meanings should be scaffolded by teachers, that is, they respond to the provisional meanings of students with indirect calls for clarification and elaboration. Also investigating middle-school settings, Haroutunian-Gordon (1998) proposed that discussion leaders of high level interpretive discussions structure their guidance through basic and follow-up questions they have developed about the text. Basic Questions refer to the "deepest point of doubt" or the "question the leader wants to resolve". The discussion leader's follow-up questions are aimed at developing a rich interpretive discussion of issues that form a part of the Basic Question.

Saunders and Goldberg (2007) studied the effects of "instructional conversations" (IC) in middle school English Language Learner students compared with controls. The controls received traditional recitation style discussion: teachers ask questions that have known answers; students respond; and, teachers offer evaluation or feedback. In contrast, the experimental IC group was taught to use open-ended, explanation-seeking

questioning (e.g., why-questions) to elicit student ideas and a variety of techniques to expand ideas and inquire into the bases for why students are arguing as they are. A recitation approach is more likely to produce relatively brief student turns of conversation communicating *facts* about the text, whereas IC leads to *longer student turns in which they communicate deeper meanings*. In several studies they have conducted, these researchers and colleagues found that in ICs teachers talk relatively less, students more, and students learn, progressively, to "follow up on each other's comments and *maintain topics for longer stretches of conversation*" (p. 226). Thus, productive discussions are characterized by collaborative discourse that is interactive and sustained, and individual student contributions that communicate deeper meanings.

In their recent experimental study of the effects of instructional conversation on young English Language Students, Saunders and Goldenberg (2007, 224-225) and Souter and Rudge (2005) listed some elements of teacher practices for supporting good discussions: establish non-threatening atmosphere; encourage wide participation; ask high-level questions; maintain thematic focus; provide background knowledge when needed; promote complex language; elicit bases for positions or arguments; encourage mutual responsiveness and connected discourse among discussants; support affective, inter-textual comparisons, shared knowledge and elaborated explanation; engage in exploratory talk; be a co-learner when appropriate; and conduct academic "revoicing" (restatement of students' comments in academic language). The following guidelines summarize empirical findings in greater detail.

1. Leaders use high-level initiations, such as explanatory questions, to initiate threads (coherent discussion sub-topics). *The leader encourages general participation using high-level initiation techniques, such as, questions.*

Elicitations function to initiate particular threads. Elicitation questions serve to establish the point of inquiry of a thread and all communications are made in response to the question and to other points made by discussants in response to the question. We can distinguish several kinds of questions in discussions: informational questions refer to who, what, where, when questions that have known answers and are usually considered low-level. Nevertheless, informational questions and responses may play a beneficial role in giving contextual meaning

to a thread and may help in sustaining a thread. Interpretive questions, which call on discussants, specifically, to address the text, are very important in stimulating interpretations that refer to particular passages. High-level explanation or why-type questions, of course, are critical in provoking comments and opinions as part of knowledge claims and arguments. Other types of questions - - clarification, hypothetical and rhetorical - - appear more frequently as follow up communications and are not used as initiations. Although questions are the primary method for setting topics, Nathan (2007) found other elicitation formats in his study of discourse processes in a middle-school math classroom: provocative statements; prolonged utterances; use of student's name; and specific requests for students to do something. Clearly, high-level elicitation techniques do not guarantee high-level responses, but we propose that such elicitations, if taken up, at least mean that discussants are attempting to respond to questions about important topics. In college-level discussions, both instructors and students may initiate discussions.

2. Leaders encourage and help *sustain extended threads* by performing as “just another discussant”. Productive discussions have a *thematic, conceptual focus* in which students are expected to pursue a theme suggested by the text. A thread is an extended series of turns on a consistent topic or sub-topic of an overall theme. "A thread is a characterization of consecutive segments of discourse" (Stanton Wortham, personal communication). The coherency of a thread is characterized by contributions that consistently develop answers or responses to the question or claim that initiated the thread. A new thread is distinguished from a preceding thread by treating different content of the theme. For example, in one of our sample discussions of Martin Luther King’s “I have a dream” speech there is a general thread that deals with segregation. Subsequent threads may treat more differentiated aspects: segregation in the North and South; segregation of restroom facilities; integration through basketball. Another general thread focuses on economic inequalities. Subsequent threads that segue from economic inequalities explore such subtopics as urban ghettos, geographic economic differences and affirmative action in sports. Still another kind of thread in this

transcript deals with the relation of racial freedom to economic opportunity. This latter thread seems to integrate previous threads dealing with segregation and economic inequalities.

3. Leaders foster group responsivity/connectivity by encouraging students to respond succinctly to each other's communications. In productive discussions, *the discourse is connected: "the discussion is characterized by multiple, interactive, connected turns; succeeding utterances build on and extend previous ones"* Saunders and Goldenberg (2007, 224-225). Connectivity implies that a thread or thematic line of inquiry is being followed – the overall question and important sub-questions are being addressed – and that participants are responding to, following up on each other's communications.
4. Leaders respond to students' communications personally. *The leader is highly responsive to student contributions*, as well as supporting student-to-student interaction. In addition to follow-up questions, other kinds of leader responses serve to provide feedback to students, such as, agreement/disagreement, providing information statements, and scaffolding that may help them elaborate and improve their knowledge claims. For example, by restating, marking (calling attention to), and annotating student remarks (see Beck & McKeown, 2007) leaders help students better understand their own messages and provide them with a more substantial basis for making further communications. Concurrently, these same leader communications help other students in the group understand the previous student's communication and, hence, provide them, potentially, with a greater understanding for facilitating their further responses. Such elaborations of meaning also serve the goal of sustaining the discussion.
5. Leaders minimize their own participation and foster wide student participation. The leader should participate enough to frame inquiry and to be responsive, but should not participate too much. Discussion is a zero-sum game to a great extent: the more the leader has the floor, the fewer opportunities exist for student participation. Once the overall discussion and threads are initiated, the leader lessens his or her own contributions so that students may get a turn. As threads develop, a good discussion leader may assume the role of "just another

discussant”. Souter and Rudge (2007), in particular, stressed the value of the teacher as a co-learner.

6. Leaders support individual students’ construction of high-level claims and arguments and peer-to-peer collaborative construction of high-level claims and arguments through follow up questions and alternative claims and arguments. In rich discussions students need to make claims about what they are reading and in high-level participation they support their claims with evidence from the text or other sources. In this regard discussion leaders may encourage students to follow up and extend their own communications in subsequent turns so that they achieve greater depth. Depth of individual communications may be assessed by the extent to which students create academic arguments with claims, principles, evidence and conclusions. This is important not only because discussions are opportunities to try out their ideas, but also because substantive communications provide greater opportunities for other students to respond. Research on discussions suggests that knowledge building is induced through questions probing what students think about the text. In effect, the questions are search terms used by the discussion leader for the most part to address and elicit the group’s knowledge and understanding of the text. However, when individual students respond to such questions the potential knowledge they communicate is only a *knowledge claim* in an academic argument. According to Toulmin (1963), the basic layout of arguments in jurisprudential logic (appropriate to the kinds of arguments used with reference to textual meanings) consists of data that lead to a conclusion under the authorization of warrants. Warrants, which are generally unexpressed, are "logical bridges that allow one to argue a step between data and conclusions" (p. 98). Warrants may assume various forms: premises, assumptions, justifications, reasons, beliefs or any form of pre-suppositional understanding to authorize a conclusion. Knowledge claims in arguments about readings may be supported by various kinds of evidence, principle or theory that is offered by a speaker as an interpretation of the text. *Evidence* refers to references in readings. There are other weaker forms of evidence supporting claims: opinions; commentary; background information, etc. *Principle* refers to a warrant, such as

justice, which when combined with evidence might be used to draw a *conclusion* or theory.

7. Leaders use scaffolding - - statements that indirectly point to or hint at productive lines of inquiry - - to support individual and group knowledge building. Recently, Hacker and Graesser (2007) reported successful use of scaffolding in tutorial dialogues. Scaffolding is an adult or expert-facilitated process that enables a novice to solve a problem that would be beyond his or her unassisted efforts. A successful scaffolder changes instructions on the basis of earlier responses and estimates of current abilities (Wood, Bruner, and Ross, 1976).
8. Leaders periodically “weave” or summarize the group’s knowledge building. *The leader “manages to keep everyone engaged in a substantive and extended conversation, weaving individual participants’ comments into a larger tapestry of meaning.”* The Discussion Leaders accomplish this by summarizing and synthesizing communications made by student participants and, perhaps, their own previous utterances. In so doing, the leader responds to and connects up student contributions. Such syntheses also serve to punctuate themes with concluding understandings and these kinds of communications imply that the ultimate goal of the discussion is to create shared meanings. While these have likely been accumulating throughout the discussion, the “tapestry” presents them in a coherent, connected whole.

Appendix 2

Thread Example from Discussion of Plato's Republic:
"Teaching students how or what to think in college compared to The Republic"

T = Turn number.

S = Student

DL = Discussion Leader

T100 S8 Umm, I was... I agree with what (*Speaker 13*) said, but I also was thinking about what you said about how our society teaches us how to think, while this society teaches them what to think. And I was actually thinking it was, what, where we're kind of... We're taught certain things. And we can think however we want, as long as we're thinking the way they want us to think. Does that... Do you know what I mean? Like, well, I just feel like, if you go through like a public school process these days, public high school process, whatever, you're being taught facts that you have to spit out about history and math and English and all these different things. And you can say whatever you want, as long as you're doing what they want you to do. And you can't really out-step any of those boundaries. Which, in here, you're given all these ideas and you're given all these things that you can live by, but you have the decision to, to formulate new ways to, to do things. Um, I feel like I'm not making myself very clear. But...

T101 DL No!

T102 S10 What about what we're doing right here? Isn't this, isn't one of the, like, key things of Lawrence is teaching you how to think, not what to think?

T103 S8 Well...

T104 S10 I think that's where I picked it up from. Because this, we're not spouting facts, we're not, like memorizing...

T105 S8 But we are spouting facts. I mean, we're taking a book and we're just spouting facts that we read about. We're not talking...

T106 S10 No, we're responding our interpretations of what we thought ...

T107 S8 No, we're not. We're not responding to interpretations. We're responding to things that we've seen in a book. If we wanted to say our own ideas, we would have to think of brand new things, and then talk about them. And we're not doing that. We're talking about ideas that Socrates and Plato formulated. We're not talking about ideas that, you know, like, (*Speaker 8*) formulated. We're just taking these ideas and comparing them to other things. We're not coming up with completely brand new ways to think.

T108 S5 I don't think we're being taught what to think. I mean, we're just kind of taking our interpretations of what we're told. And rather than when something in our own

way, it's still our interpretation. I mean, it's not, we're not sitting in a lecture. No one's telling us this is right, this is wrong. We're entitled to our own interpretation to make life, as we're doing it now.

T109 DL OK. (*Speaker 1*), then (*Speaker 11*).

T110 S1 I think it's like really tricky. Because, I mean, there are so many different types of education in the US, like, and I mean, you can't really compare, really, I mean, high school, generally, like you said, is a public education. But here we have a private education, and I think that there's going to be a difference between the two. And I don't know, it's, so when they're talking about in their public education, they're talking about, basically, the state training, or educating the guardians. Like, is that a public education, or is it privatized? I, man, that's where I was a little confused about this.

T111 DL I'll go out on a limb and say that I think his plans are sort of this is a public state-run process...

T112 S1 Oh.

T113 DL ...throughout.

T114 S11 What my belief is, is in "The Republic," they're teaching you how to think, and here they're actually teaching you what to think. Surprisingly, here they give it, they're teaching you what to think, but under the guise of how to think. Here, they're forcing most people to believe a certain way. I don't agree with that, so my beliefs are opposite of that. I'm trying to explain something that's very, very hard.

T115 S12 In so-called Lawrence...?

T116 S10 Which one is forced on you?

T117 S11 I felt like in Socrates' and in Plato's world, they're teaching you how to think, because, as we all know, Socrates formed his arguments through questions. But here they're teaching you what to think under the guise of how to think.

T118 S12 What do you mean by that?

Acknowledgments:

We gratefully acknowledge the support of the Spencer Foundation and the Freshman Studies Program, Lawrence University in the conduct of this study. Dr. Bill Skinner, Director of Institutional Research managed the database and computed the statistics. Professor Stanton Wortham, University of Pennsylvania, consulted on the use of theories and methods in the study of discourse processes.