RESEARCH REPORT

Equity during peer conferences: A linguistic comparison by race and gender

Daniel L. Reinholz[1]

San Diego State University, USA

Abstract: *This paper provides an analysis of students' peer assessment conversations in introductory college calculus. Prior research shows that this type of activity can support meaningful student learning. However, previous studies have suggested that students from different groups (e.g., by race or gender) may have different opportunities to participate in such discussion-based activities. Accordingly, this paper explores the participation of students in peer assessment conversations, by focusing on the types of feedback and word choices used by different groups of students, by race and gender. Based on computer-aided textual analysis, this paper provides insights into the types of words used by different students in the class. While there was evidence of inequities in participation between men and women, the results for race were inconclusive. These results suggest that peer conferences have some potential for producing more equitable participation in calculus.*

Keywords: Equity, peer assessment, participation

Introduction

Imagine walking through a large-enrollment introductory STEM classroom with hundreds of students trading papers and providing constructive feedback on one another's problem solving. As you walk around the classroom, you overhear a student conferencing with their peer:

Part three of your solution is clearly wrong. You didn't even show units and your answer doesn't make sense when interpreted physically.

As you continue walking, you overhear another student:

I'm not sure about your response to number two. I wonder if you could show more work, and could you explain more to me about what you were thinking? I have a feeling that you may have computed the distances using an incorrect model of the physical setup.

What could you say about the relative status of the speaker and listener in each example based on the words they used? Which words would help you draw inferences?

In the first excerpt, the speaker strongly asserts that their peer's work is incorrect. This type of feedback discusses the product of the work but provides no insight into why an incorrect answer may have been reached. In contrast, the second speaker describes what they think about their peer's work, with a strong focus on the problem-solving process. That is, the speaker is concerned with why the listener may have had an incorrect result, not just that it was an incorrect result. While feedback is known to promote learning, not all feedback is

equally useful, with feedback that focuses on the why being more helpful than that focusing on the what (Hattie & Timperley, 2007; Reinholz, 2017b). Thus, by analyzing the types of feedback across a variety of speaker-listener pairs, it may be possible to infer something about the value of these interactions to promote learning.

What else could you infer? In the first excerpt, few first-person pronouns are used. Rather than saying what they think about their peer's solution, the speaker simply makes statements about what the listener's solution should be. In contrast, the second excerpt features a large number of 'I' statements. The feedback is clearly coming from the perspective of the speaker. Such pronoun usage can reveal status hierarchies (Pennebaker, 2011). The first speaker, who uses many more second-person pronouns is likely of higher status than the second speaker who uses primarily first-person pronouns. Thus, simply by looking at the types of pronouns students use in the conversations, it may be possible to uncover subtle status hierarchies in the classroom.

This paper focuses on the use of words as they relate to status hierarchies. Research shows that students with less status have fewer opportunities to learn, which is a source of inequity (e.g., Cohen & Lotan, 1997). This paper addresses the following research question:

> *What quantitative patterns exist in the language used by students in different racial and gender groups in peer feedback conferences?*

In light of this question, the implications of using peer feedback as a tool to create more equitable learning opportunities are discussed. This paper focuses on mathematics, but the results should generalize to other disciplines, especially given that the focal activity, Peer-Assisted Reflection (PAR) (Reinholz, 2015a), has been used across STEM disciplines.

Theoretical Framing

Classroom discourse is a key part of learning (Bransford, Brown, & Cocking, 2000; Lampert, 1990; Sfard, 2008). However, students simply speaking during class through low level contributions is insufficient (cf. Cazden, 2001; Mehan, 1979). Rather, students need to engage with disciplinary content in ways that require higher-order thinking (cf. Stein, Grover, & Henningsen, 1996). This means that students should not merely provide answers to short computations, but rather, they should spend time explaining disciplinary concepts.

Discourse is a powerful learning vehicle, but it is also a means of potential inequity (Esmonde & Langer-Osuna, 2013; e.g., Herbel-Eisenmann, Choppin, Wagner, & Pimm, 2012). For instance, students from certain groups tend to receive lower-level participation opportunities, based on their gender (Sadker, Sadker, & Zittleman, 2009), race (McAfee, 2014), and immigration status (Planas & Gorgorió, 2004). These students tend to be those who are already underrepresented in the STEM disciplines (i.e. women, and students of color). While the actions leading to such patterns are unintentional, they have negative consequences for many learners, and must be addressed.

Equity is a matter of justice (Secada, 1989). An equitable classroom is one in which students receive instruction in accordance with what they need as learners. This means that equity is subjective and depends on the feelings and experiences of individual learners. As such, it is difficult for an outsider to make strong claims about what is an equitable classroom, because they cannot share the same lived experiences as individual students in the classroom. Nevertheless, it is clear that if students from historically-marginalized groups receive fewer opportunities to learn (i.e. less than equal) than their historically-dominant peers, that would signify inequity. In this way, equality is taken as a necessary but insufficient baseline for equity (Secada, 1989). In other words, if all students receive at least equal opportunities to participate, it is a positive (yet insufficient) step in the right direction.

Educators have worked for decades to reduce inequity, but it still remains prevalent in a wide variety of mathematics settings (Adiredja & Andrews-Larson, 2017; Gutiérrez, 2008; Leyva, 2017; Martin, 2003; President's Council of Advisors on Science and Technology, 2012). Through these efforts, some powerful techniques have

been developed for supporting equity. One well-known example is the set of techniques associated with Complex Instruction. These instructional moves (e.g., framing tasks as requiring multiple, diverse skills to be completed; assigning competence to low-status students) help mitigate status hierarchies in heterogeneous classrooms, leading to more equitable outcomes for all students (Cohen & Lotan, 1997; Nasir, Cabana, Shreve, Woodbury, & Louie, 2014). In other words, power imbalances (e.g., who is perceived as a high-status) lead to less equitable outcomes (Engle, Langer-Osuna, & Royston, 2014; Langer-Osuna, 2016), but when these imbalances can be addressed, learning becomes more equitable (Cohen & Lotan, 1997).

Against this backdrop, the present study focuses on issues of equity and status in peer feedback conferences, particularly with respect to race and gender. Peer conferences are an important feature of peer assessment (cf. Falchikov & Goldfinch, 2000; Topping, 2009) and offer unique opportunities for addressing issues of inequity in discourse. In particular, peer conferences generally involve only two students, so the complexities of promoting participation from all students in a small group or a whole class are reduced. Moreover, peer conferences position students as competent authorities, because they must critically judge the work of their peers, which provides them with space in the classroom to act as experts (cf. Engle & Conant, 2002; Reinholz, 2015a).

Finally, it is worth noting that race and gender are complex, socially-constructed identities, which organize the social positioning of individuals (Davies & Harré, 1990). As such, putting students into these discrete categories can be problematic, because it obscures this complexity; gender is not a binary, and the reality of students' lives is much more complicated and their identity might change under different circumstances. Nevertheless, such essentialization can be 'strategic,' as a tool to highlight or address inequities (Gutierrez, 2002). The value of using static categories to promote equity can be seen in stereotype threat research.

Stereotype threat occurs when high-pressure circumstances (e.g., taking an exam) cue negative stereotypes related to academic performance in particular domains (e.g., Black and Latinx students as bad at math). This could happen, for instance, when there are only five women taking an exam in an introductory course with over 100 men, because their identities as women would be highly salient in this context. When comparisons are cued in such social contexts, it negatively impacts the performance of students from certain groups, thereby further marginalizing them (Spencer, Steele, & Quinn, 1999; Steele, 1997).

To see how stereotype threat operates, it is important to temporarily flatten the complex categories of race and gender, so we can understand how racialized and gendered stereotypes marginalize particular students. Thus, this strategic essentialism makes it possible to illuminate subtle patterns of inequity (e.g., men speaking more than women), as a step towards greater equity. Still, complementary approaches that treat social markers more fluidly are necessary too (e.g., Nasir, McLaughlin, & Jones, 2009).

Method

Context

The present study took place in Calculus I at a large, Hispanic-serving, PhD-granting institution with high research activity. There are approximately 30,000 undergraduate students (45% women). The racial demographics of undergraduates are roughly: 4% African American, 0.3% Native American, 6% Filipino, 7.3% Asian, 0.3% Pacific Islander, 8.3% International, 27.6% Hispanic/Latinx, 6.2% multiple ethnicities, 35% White, and 5% Other/Unknown. Data for the mathematics department and the particular calculus section were not available, but as shown below, the demographics of students in the course does not represent the full diversity of the university, as is typical in introductory calculus.

The course consisted of a combination of large lectures (100-200 students) taught by full-time instructors and smaller breakout recitation sections (30-40 students) taught by Graduate Teaching Assistants. This paper focuses on a single large-lecture section (N=124) taught by the author, which met three times weekly

for 50 minutes at a time (Reinholz, 2017a). In addition, students in the lecture met twice weekly for 50 minutes for their recitation sessions, but those sessions are not a focus of this paper.

Design

Each week students engaged in a peer assessment learning activity called Peer-Assisted Reflection, or PAR (Reinholz, 2015a; Reinholz & Dounas-Frazer, 2016), for a total of 14 assignments. The goal of PAR is for students to develop self-assessment skills as they assess the work of their peers (Black, Harrison, & Lee, 2003; Reinholz, 2015b). Specifically, PAR consists of a four-part cycle through which students: (1) complete a draft solution to a conceptual mathematics problem for homework, (2) reflect on their solution by identifying which aspects of their solution they would like to receive feedback on, (3) trade papers with a peer in class and exchange peer feedback, and (4) revise their work before turning in their solution. Students receive homework credit both for the correctness of their solution and for completing the PAR process, which encourages students to revise their work (in practice nearly all students do so). Prior studies show that PAR has a significant positive impact on student learning (Reinholz, 2015a, 2016), but the learning impact of PAR is not the focus of the present study.

This implementation of PAR differed from prior iterations (in nearly 20 classroom contexts, across STEM disciplines), because it took place in a large-lecture course, which imposed different logistical constraints. In terms of the actual PAR process, students were able to engage productively during their large lecture sessions: they simply turned to a peer, traded papers, and conferenced about their work. The course structure also meant that students received limited feedback on their PAR solutions and engagement in the PAR process, however, the instructor was constrained by the logistics of teaching a large course.

In this implementation of PAR, students chose their partners. Because students chose their partners, they often picked students with whom they were friends or more comfortable. This resulted in a large number of same-gender pairings. For instance, of the conversations that included women, 42 of them were entirely comprised of women, and 29 of them were mixed gender. This indicates that women were mostly talking with other women. As such, the results of this study may differ if random assignment of partners was used, as in some prior iterations of PAR.

During the feedback exchange component of PAR (step 3), students read each other's work silently for five minutes and write comments, and then have five minutes to discuss their feedback. Forcing students to engage silently with each other's work before the discussion helped ensure that students actually talked about their peers' solutions, not just the problem. Moreover, PAR positions both students as competent, as they both provide feedback to one another, rather than creating an asymmetric relationship in which only one student provides feedback to the other. This was a feature designed to promote student authority (cf. Engle & Conant, 2002). In the context of a large-lecture course, this was intended to provide all students with opportunities to engage in meaningful talk about mathematics, which can otherwise be difficult to facilitate in whole-class conversations.

Participants and Data

A total of 84 students participated in the study (in a class of N=124). For all of these students, demographic information was collected from the university's office of Institutional Research (see Table 1).

Table 1.

Participant demographics (N = 84)

	Women	Men	Total
African American	0	3	**3**
Asian/Pacific Islander	4	11	**15**
Hispanic	8	16	**24**
International	0	1	**1**
Multiple Ethnicities	1	4	**5**
White	9	19	**28**
Unknown	4	4	**8**
Total	**26**	**58**	**84**

The primary data source for this article was students' peer conversations. During their PAR conferences students recorded their conversations (the second part of their feedback exchange) using audio recorders on their cellular phones. For students who did not have a cellular phone, they were offered the use of a dedicated audio recorder, but no students elected to use this option. A total of 172 conversations were recorded. While the majority of conversations consisted of student dyads, some of the conversations involved three students at a time. To account for a variable number of students in certain conversations, a unit of 'participant-conversations' was used for analysis, which represents how many times some student from a particular group participated in some conversation. The demographics of students by participant-conversations is given in Table 2. Because of the small number of contributions from the international students they were dropped from racial group analyses, as were the students of unknown race.

Table 2.

Breakdown of demographics for participant-conversations (N = 370)

	Women	Men	Total
African American	0	12	**12**
Asian/Pacific Islander	37	36	**73**
Hispanic	22	83	**105**
International	0	2	**2**
Multiple Ethnicities	1	16	**17**
White	45	97	**142**
Unknown	11	8	**19**
Total	**116**	**254**	**370**

To support data analysis, all conversations were transcribed and linked to student names. This allowed for demographic information to be attached to each individual student contribution. Cleaning of the dataset and data analysis was completed in R statistics, using a variety of text processing packages (e.g., stringi, lsr, lexicon). This cleaning process involved removing punctuation and converting all text to lowercase so that it could be processed using the appropriate packages.

Analytic Methods

Feedback Types. Student conferences were analyzed for the type of feedback provided using a prior coding scheme that focused on process, product, and person feedback (Reinholz, 2017b). By looking at the different types of feedback provided by students, it was possible to explore if there were differences in feedback style based on race or gender. Process-focused feedback describes how to reach a mathematical solution or why the solution works. Product-focused feedback describes whether or not answers given are correct. Person-

focused feedback consists of praise or negative personal comments; in practice, negative personal comments have almost never been encountered in empirical studies of PAR. Further information about these categories can be found elsewhere (Reinholz, 2017b).

Linguistic Inquiry. To analyze student word usage, Linguistic Inquiry and Word Count 2015 (LIWC) software was used (Pennebaker, Boyd, Jordan, & Blackburn, 2015). LIWC provides a standard for automated text processing that allows results to be compared across numerous sources. LIWC provides output on numerous dimensions, including four summary scales (ranging from 0 to 100):

1. Analytical Thinking – higher numbers indicate more formal logical thinking, while lower numbers suggest more informal, personal text;

2. Clout – higher numbers signify status and confidence, while lower numbers imply humility or a tentative, hesitant style;

3. Authentic – high numbers represent honest and personal text, while lower numbers are guarded and distant text;

4. Emotion – higher numbers are more positive and upbeat, while lower numbers are anxious, sad, or hostile.

In addition, I considered the use of first-person pronouns to indicate lower status, and second person pronouns to indicate higher status (Pennebaker, 2011). I also considered positive and negative emotion words, informal language, and cognitive language. All of these variables are standard outputs from LIWC. Table 3 provides examples of words from these different categories (Pennebaker et al., 2015).

Table 3.

Example words used in LIWC analysis

Category	Example
Positive emotion	Love, nice, sweet
Negative emotion	Hurt, ugly, nasty
Informal language	Damn, OK, hmm
Cognitive language	Cause, know, ought

While language analyses can provide insight into peer conferences, they must be interpreted with caution. A wealth of literature highlights differences in word usage based on gender (Argamon, Koppel, Fine, & Shimoni, 2006), task characteristics (Newman, Groom, Handelman, & Pennebaker, 2008), topic (Bamman, Eisenstein, & Schnoebelen, 2014), and age (Huffaker & Calvert, 2005). Still, some commonalities exist across settings. For instance, women tend to use first-person singular, cognitive, and social words more, while men use more articles, and there are no meaningful differences in the use of first-person plural or positive emotion words (Pennebaker, 2011). While less attention has been given to racialized differences in text analysis (e.g., pronoun usage), a vast literature connects issues of race, culture, identity, power, and discourse (e.g., Holland, Lachicotte, Skinner, & Cain, 1998; Leonardo, 2009; Nasir, Hand, & Taylor, 2008).

In sum, one can expect differences in word usage by students from different groups in the peer assessment process, simply by virtue of their membership in particular gender, racial, or other demographic groups. Yet, it will be difficult to predict in advance what these differences will be. Nevertheless, this paper will provide a reference as others continue to look at such patterns of word usage in other educational contexts.

Results

The amount of words and type of feedback was analyzed and are presented overall as well as by gender and by ethnicity. Results from the Linguistic Inquiry using LIWC (such as amount of clout that) are also presented.

Overall amount of talk and type of words used

Table 4 provides a summary of word usage for the class as a whole. To contextualize these results, they are compared to two prior iterations of PAR (Reinholz, 2017b). These prior iterations of PAR also took place in introductory college calculus, but in smaller class sizes. Like Phase II of the prior study, students in the present calculus class also had whole-class discussions about how to give feedback to their peers. These conversations were not part of the Phase I study. Table 4 shows that in the present study each student contributed an average of 149.09 words to each conversation. The standard deviation of 92.80, however, reveals a large variance in how individual students participated.

Table 4.

Classroom-level word usage (by participant conversation)

	Present (N=370)	Phase I (N=116)	Phase II (N=184)
Total Words (Mean)	149.1 (SD=92.8)	163.2 (SD=80.5)	295.2 (SD=117.8)
Process Words (Mean)	6.0	4.9	9.00
Product Words (Mean)	1.3	0.7	1.1
Person Words (Mean)	3.5	2.4	2.5

Table 4 also shows that conversations during the present study were shorter than those in prior iterations. The Phase II conversations contained more feedback in these three categories than did the current study. Yet, when looking at density of feedback--the amount of feedback based on how many words were spoken--it is highest in the current study. In other words, it seems that students were saying more with fewer words, and likely with less off-topic talk. Given variation in the implementation of PAR and student populations, it is difficult to identify the cause of these differences.

Amount of talk by gender and race

Figure 1 shows a large spread in the amount of talk by gender. The spread for amount of talk by men was much larger than women (many words and few words). Thus, at least at the individual level, gender differences appear in amount of talk, indicating both lower-status and higher-status men. Figure 2 provides the same data for race. Here we see that the spread was greatest for White students, and White students spoke more on average than their peers in other racial groups.
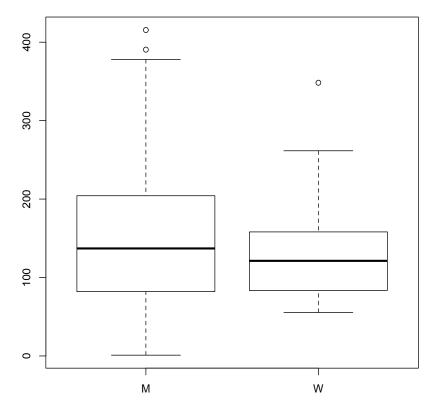
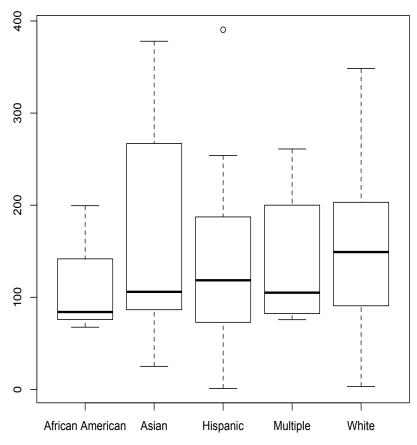*Figure 1. Average number of words per participant-conversation by gender*

Figure 2. Average number of words per participant-conversation by race

Type of feedback by gender and race

Figure 3 shows feedback types by gender. The data are expressed as an "equity ratio" (Reinholz & Shah, 2018). Actual participation is the count of coded items through classroom observation using EQUIP. Expected participation is based on a group's demographic representation in the classroom. The equity ratio can fall into three categories: greater than one, less than one, or equal to one. To illustrate, imagine a classroom where 40% of students are women. If actual participation from women is 60% of total participation, then the equity factor would be 1.5—this indicates disproportionately greater participation from women relative to their demographic representation. If actual participation from women turns out to be 30%, then the equity factor would be 0.75—this indicates disproportionately less participation. If actual participation is 40%, the equity factor would be 1.0, which indicates proportional participation. Overall, this metric makes it possible to make claims about equity while accounting for differences in the raw numbers of students from different social marker groups in a given classroom.

Here we see that men contributed more total words than one would expect $\chi^2(1, N = 55284) = 102.67$, $p = 3.9 * 10^{-24}$, Cramer's V = 0.53 (large effect size). Women used more person-focused feedback (i.e. praise), $\chi^2(1, N = 1387) = 39.18$, $p = 3.85 * 10^{-10}$, Cramer's V = 0.33 (medium effect size). Men used more product feedback (i.e. describing right or wrong), $\chi^2(1, N = 484) = 19.2$, $p = 1.2 * 10^{-5}$, Cramer's V = 0.22. There were no significant differences for process words.
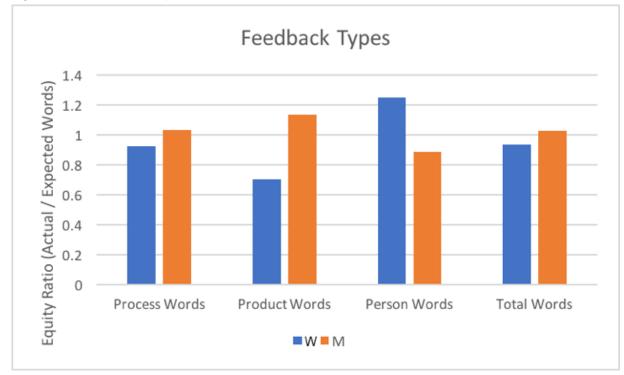

Figure 3. Feedback types by gender

What can be inferred from these results? Figure 3 indicates that men and women behaved the same when it came to giving process-focused feedback, which is the most valuable type for learning. Yet, there were also stylistic differences with men focusing more on correctness of the solution and women offering more praise. This may also speak to possible status issues, with men feeling more confident to make assertions about correctness. On the whole, men did talk more, but the equity ratio for total words was near one. Thus, this statistically significant difference may have less practical significance.

Figure 4 shows feedback type by race. Here we see that Black and White students contributed more total words than one would expect $\chi^2(4, N = 51955) = 400.59$, $p = 2 * 10^{-85}$, Cramer's V = 0.54 (large effect size). Differences were also significant for person-focused feedback $\chi^2(4, N = 1314) = 15.44$, $p = 0.0038$, Cramer's

V = 0.11 (small effect size). There were no significant differences for process words or product words, after correcting for multiple comparisons. Again, there were no differences for process words, the most important type of feedback, but there were some potential differences in status indicated by total number of words.
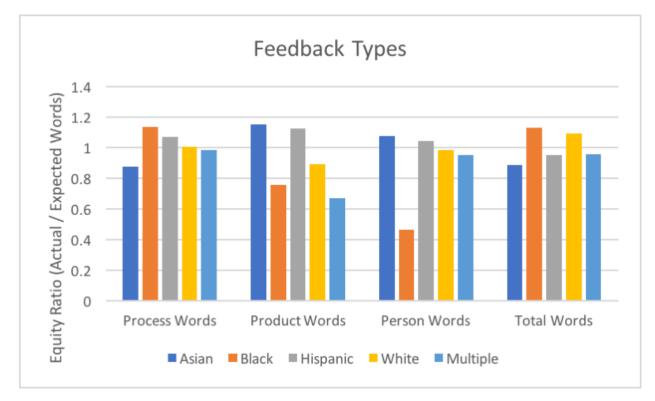

Figure 4. Feedback types by race.

Linguistic Inquiry

Table 5 provides a summary of LIWC output. For reference, the results for natural speech have been included in the right column (Pennebaker et al., 2015).

Table 5.

LIWC results

		Women	Men	Asian	Black	Hispanic	White	Multiple	(Ref.)
Analytic		16.6	21.2	21.2	33.3	16.4	19.2	17.2	18.4
Clout		49.4	52.9	47.4	48.1	60.5	48.8	46.8	56.3
Authentic		46.3	49.4	48.0	56.3	45.6	50.4	58.0	61.3
Tone		87.8	79.2	84.3	69.4	77.6	83.6	85.6	79.3
Selected Results									
Pronouns		18.6	19.1	17.8	17.7	20.0	19.1	20.3	13.6
	I	4.1	4.6	4.1	5.2	4.4	4.6	5.8	4.8
	You	3.5	3.8	3.1	2.9	4.7	3.5	3.7	4.0
	Impersonal	10.0	9.4	9.6	8.4	9.7	9.8	9.5	7.5
Positive Emotion		4.2	3.6	4.0	3.1	3.6	3.8	3.9	5.3
Negative Emotion		0.6	0.7	0.7	0.8	0.8	0.6	0.5	1.2
Cognitive		17.4	17.1	17.1	18.3	17.8	16.9	16.7	12.3
Informal		4.7	3.7	5.3	5.9	3.2	4.0	3.1	7.1

The results show modest gender differences in each category, with women using fewer analytic words, having lower clout, speaking more distantly, and being more upbeat in the feedback. The women tended to use more positive emotion words, and a more informal style. These results indicate that the men were likely of higher status, consistent with the sociometric surveys and total talk length.

There were also differences by race. For instance, Hispanic students had the highest clout. We also see the Black students in the sample had high levels of analytic speech and clout (although not as high as Hispanic students). This shows that Hispanic and Black students were able to take on positions of relatively high-status in the way that they discussed mathematics in these conferences. More notably, we do not see a bias toward White and Asian students, as research in other settings might suggest (Stinson, 2008).

Beyond this descriptive analysis, Table 6 provides the ANOVA results for the four summary variables for gender and Table 7 is for race. When these tests for statistical significance were computed, no significant differences were found. Thus, at least for this dataset, it was not possible to detect statistically-significant differences in these key LIWC variables. Thus, it may be the case that there were far more similarities than differences in how students engaged in their conferences.

Table 6.

ANOVA results for LIWC summary variables by gender

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Analytic	1	1019	1019.0	3.027	0.0856*
Residuals	82	27602	336.6		
Clout	1	9	8.5	0.016	0.899
Residuals	82	43292	527.9		
Authentic	1	365	364.9	0.65	0.423
Residuals	82	46058	561.7		
Tone	1	2330	1330.2	2.431	0.123
Residuals	82	44868	547.2		

Table 7.

ANOVA results for LIWC summary variables by race

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Analytic	6	467	77.9	0.213	0.972
Residuals	77	40537	526.5		
Clout	6	2763	460.5	0.875	0.518
Residuals	77	44606	579.3		
Authentic	6	1816	302.7	1.121	0.358
Residuals	77	44606	579.3		
Tone	7	3710	618.4	1.121	0.358
Residuals	77	42488	551.8		

Discussion

Promoting equity in classroom interactions is a challenge and an ongoing concern for undergraduate mathematics (e.g., Adiredja & Andrews-Larson, 2017). This paper focuses on quantitative analyses of student speech as it relates to status. The results showed some areas of inequity in peer conferences (e.g., some

individuals speaking more than others), but on the whole the results were promising for equitable opportunities to participate in this activity. There were no statistically significant differences in LIWC categories for race or gender, and for analytic speech and clout, numerical differences actually favored Black and Latinx students. It is possible that such differences would be statistically significant in a larger dataset. Still, men did have a greater total talk time, which may be an indicator of higher status. Moreover, since opportunities to talk are opportunities to learn, the men in the class may have taken up proportionally more of these opportunities to learn.

Does this actually represent equity? It is unclear. Equity is a value judgment, and one that is difficult to make as an outside observer. Moreover, the present study is limited because it did not include student interviews or any other mechanism to provide further insight into students' subjective experiences. Still, it appears that peer conferences provided opportunities for students of various groups to engage meaningfully, which is a step in the direction of equity.

As a tool to promote equitable engagement, peer conferences have a great deal of potential. Because students work in a one-on-one setting, it is less likely that students will not have an opportunity to participate at all. More than just participating, structured peer conferences help students learn exactly what is expected of them and the types of feedback they should provide, which helps all students engage with mathematics in more meaningful ways (i.e. focusing on processes over products or praise). The structured nature of the activity makes it more likely that students from different groups will have an equal opportunity to contribute, rather than allowing historically dominant students to dominate. While this study took place in mathematics, peer assessment (and in particular, PAR), has been used in a variety of STEM settings (e.g., physics, biology), so the results likely generalize across domains, at least to some extent. Future research is required, however, to more fully determine the extent to which peer conferences can produce more equitable participation in calculus.

References

Adiredja, A. P., & Andrews-Larson, C. (2017). Taking the sociopolitical turn in postsecondary mathematics education research. *International Journal of Research in Undergraduate Mathematics Education*, 1–22.

Argamon, S., Koppel, M., Fine, J., & Shimoni, A. R. (2006). Gender, genre, and writing style in formal written texts. *Text & Talk: An Interdisciplinary Journal for the Study of Discourse*, *23*(3), 321–346.

Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, *18*(2), 135–160.

Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice*. Berkshire, England: Open University Press.

Bransford, J., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Academies Press.

Cazden, C. B. (2001). *Classroom discourse: The language of teaching and learning*. Portsmouth, NH: Heinemann.

Cohen, E. G., & Lotan, R. A. (1997). *Working for equity in heterogeneous classrooms: Sociological theory into practice*. New York, NY: Teachers College Press.

Davies, B., & Harré, R. (1990). Positioning: The discursive production of selves. *Journal for the Theory of Social Behaviour*, *20*(1), 43–63.

Engle, R. A., & Conant, F. R. (2002). Guiding principles for fostering productive disciplinary engagement: Explaining an emergent argument in a community of learners classroom. *Cognition and Instruction*, *20*(4), 399–483.

Engle, R. A., Langer-Osuna, J. M., & Royston, M. M. de. (2014). Toward a model of influence in persuasive discussions: negotiating quality, authority, privilege, and access within a student-led argument. *Journal of the Learning Sciences*, *23*(2), 245–268.

Esmonde, I., & Langer-Osuna, J. M. (2013). Power in numbers: Student participation in mathematical discussions in heterogeneous spaces. *Journal for Research in Mathematics Education*, *44*(1), 288–315.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*(3), 287–322.

Gutierrez, R. (2002). Enabling the Practice of mathematics teachers in context: Toward a new equity research agenda. *Mathematical Thinking and Learning*, *4*(2&3), 145–187.

Gutiérrez, R. (2008). A "gap-gazing" fetish in mathematics education? problematizing research on the achievement gap. *Journal for Research in Mathematics Education*, *39*(4), 357–364.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112.

Herbel-Eisenmann, B., Choppin, J., Wagner, D., & Pimm, D. (2012). *Equity in discourse for mathematics education: Theories, practices and policies*. New York, NY: Springer.

Holland, D., Lachicotte, W., Skinner, D., & Cain, C. (1998). *Identity and agency in cultural worlds*. Cambridge, MA: Harvard University Press.

Huffaker, D. A., & Calvert, S. L. (2005). Gender, identity, and language use in teenage blogs. *Journal of Computer-Mediated Communication*, *10*(2), 00–00.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal*, *27*(1), 29–63.

Langer-Osuna, J. M. (2016). The Social construction of authority among peers and its implications for collaborative mathematics problem solving. *Mathematical Thinking and Learning*, *18*(2), 107–124.

Leonardo, Z. (2009). *Race, whiteness, and education*. New York, NY: Routledge.

Leyva, L. A. (2017). Unpacking the male superiority myth and masculinization of mathematics at the intersections: A review of research on gender in mathematics education. *Journal for Research in Mathematics Education*, *48*(4), 397–433.

Martin, D. B. (2003). Hidden assumptions and unaddressed questions in mathematics for all rhetoric. *The Mathematics Educator*, *13*(2), 7–21.

McAfee, M. (2014). The kinesiology of race. *Harvard Educational Review*, *84*(4), 468–491.

Mehan, H. (1979). "What time is it, Denise?": Asking known information questions in classroom discourse. *Theory into Practice*, *18*(4), 285–294.

Nasir, N. S., Cabana, C., Shreve, B., Woodbury, E., & Louie, N. (2014). *Mathematics for equity: A framework for successful practice*. New York, NY: Teachers College Press.

Nasir, N. S., Hand, V., & Taylor, E. V. (2008). Culture and mathematics in school: Boundaries between "cultural" and "domain" knowledge in the mathematics classroom and beyond. *Review of Research in Education*, *32*(1), 187–240.

Nasir, N. S., McLaughlin, M. W., & Jones, A. (2009). What does it mean to be african american? constructions of race and academic identity in an urban public high school. *American Educational Research Journal*, *46*(1), 73–114.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, *45*(3), 211–236.

Pennebaker, J. W. (2011). The secret life of pronouns. *New Scientist*, *211*(2828), 42–45.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin. Retrieved from https://repositories.lib.utexas.edu/handle/2152/31333

Planas, N., & Gorgorió, N. (2004). Are different students expected to learn norms differently in the mathematics classroom? *Mathematics Education Research Journal*, *16*(1), 19–40.

President's Council of Advisors on Science and Technology. (2012). *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*. Washington, D.C.: Executive Office of the President.

Reinholz, D. L. (2015a). Peer-Assisted Reflection: A design-based intervention for improving success in calculus. *International Journal of Research in Undergraduate Mathematics Education*, *1*(2), 234–267.

Reinholz, D. L. (2015b). The assessment cycle: A model for learning through peer assessment. *Assessment & Evaluation in Higher Education*, 1–15.

Reinholz, D. L. (2016). Improving calculus explanations through peer review. *The Journal of Mathematical Behavior*, 44, 34–49.

Reinholz, D. L. (2017a). Co-Calculus: Integrating the academic and the social. *International Journal of Research in Education and Science*, *3*(2), 521–542.

Reinholz, D. L. (2017b). Peer conferences in calculus: The impact of systematic training. *Assessment & Evaluation in Higher Education*, *42*(1), 1–17.

Reinholz, D. L., & Dounas-Frazer, D. R. (2016). Using peer feedback to promote reflection on open-ended problems. *The Physics Teacher*, *54*(6), 364–368.

Reinholz, D. L., & Shah, N. (2018). Equity analytics: A methodological approach for quantifying participation patterns in mathematics classroom discourse. *Journal for Research in Mathematics Education*, *49*(2), 140–177.

Sadker, D., Sadker, M., & Zittleman, K. R. (2009). *Still failing at fairness: How gender bias cheats girls and boys in school and what we can do about it*. New York, NY: Simon and Schuster.

Secada, W. G. (1989). Educational equity versus equality of education: An alternative conception. In W. G. Secada (Ed.), *Equity in education* (pp. 68–88). New York, NY: Falmer.

Sfard, A. (2008). *Thinking as communicating: Human development, the growth of discourses, and mathematizing*. New York, NY: Cambridge University Press.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613–629.

Stein, M. K., Grover, B. W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*(2), 455–488.

Stinson, D. W. (2008). Negotiating sociocultural discourses: The counter-storytelling of academically (and mathematically) successful African American male students. *American Educational Research Journal*, *45*(4), 975–1010.

Topping, K. J. (2009). Peer Assessment. *Theory Into Practice*, *48*(1), 20–27.