Influences on Variability of Perceptions of Behavior on Student Engineering Project Teams

Emily Miller, University of Virginia

Emily Miller is a graduate student in Systems and Information Engineering at the University of Virginia. She has previously worked for the National Integrated Cyber Education and Research Center and as a researcher at the University of Virginia, Olin College of Engineering and Ohio State. Her research interests include motivation, expertise recognition, and teamwork.

Prof. Reid Bailey, University of Virginia

Reid Bailey is an Associate Professor in the Department of Systems and Information Engineering at the University of Virginia. He holds a BSE from Duke University and an MSME and PhD from Georgia Tech, all in mechanical engineering. His professional interests include engineering design, engineering education, and the environment.

Influences of Variability on Perceptions of Behavior on Student Engineering Project Teams

Introduction

Within an engineering student project team, there are many interpersonal dynamics at play. Clashes between group members, even on a small scale, seem inevitable in nearly every group. Perceptions held by individuals about other team member's contributions to the project are not universally held and can be driven by attributes of the judger, the individual being judged, the relationship between the two, and/or the group at large. Furthermore, behaviors considered unacceptable to students come in a myriad of varieties – previous work lists eleven distinct types. [1] This study uses a newly-created survey called the Team Behaviors and Attitudes Survey to assess the subjective perceptions of teammates' behaviors using a Social Relations Model to explore the potential underlying sources of variance of these perceptions.

Background

Educators' care toward their students is exhibited not only by supporting a student's learning and performance, but also by optimizing their experience within the classroom. In their seminal work, Erickson and Schultz argue for increased research efforts surrounding the student experience in this way:

> "Why would any educator need to know about student experience of the curriculum? We can test students to find out if they learned or not. If there is sufficient money and we are well organized enough we can reteach those who didn't learn the first time around and catch those who slipped through the cracks and somehow didn't get taught at all. But if we are concerned that a wide range of students learn judgment and reasoning, as well as facts, perhaps a clearer understanding by educators of students' subjectivity in school is required." [2]

How students perceive their educational experience is of critical importance. Their interpretation of the role they are asked to play on a team environment, the attitudes of their teammates, and whether they are ultimately successful can flavor their attitudes about engineering and their ultimate persistence in the discipline. Thus, the student experience as viewed from the student perspective deserves attention and care from researchers.

Teams and Assessments

In engineering classrooms, teamwork experiences are commonly employed to (1) help students thrive in the team environment they are likely to enter post-graduation, (2) support students' development of communication, innovation/creativity, and design skills [3] and (3) support the ABET accreditation criteria that states students must develop "an ability to function on multidisciplinary teams". [4]

Team environments in educational settings are unique in that the performance of a team (and thus, of an individual) is impacted by the behavior and idiosyncrasies of different team members. These behaviors may aggregate by way of composition, in which all team

contributions are weighted equally and the team outcome is a linear combination of team member inputs, or compilation, in which more complex relationships emerge. [5] Either way, team outcomes depend on the contributions of its members; and, accordingly, one's learning experience and grade are impacted by the work of others on the team. [5], [6] For an individual participant on a dysfunctional team, this impact can be negative despite the hard work of an individual or positive despite a failure to contribute.

To mitigate this risk, many educators have adopted peer assessments. These assessments typically ask students to rate one another's performance on a Likert scale [7], [8]. When adopted at intervals during the project, these assessments can be used formatively to help diagnose problems and correct errant behaviors. They can also be used as a motivational tool to encourage active participation by using teammate perceptions of an individual as a factor in one's project grade. [9] Teamwork often takes place away from the eyes of the instructor; thus, these tools encourage accountability and discourage "social loafing" – a phenomenon in which individuals tend to exert less effort when working collectively than individually. [3] Furthermore, these peer assessments have been found to promote cooperation, higher levels of performance, and team member satisfaction. [10]

A taxonomy of behaviors typically evaluated by instructors in these assessments was developed by Baker in a meta-analysis of the literature on this subject. She identifies eight components [11]:

1. Attended group meetings; was available and on time
2. Was dependable, kept his or her word
3. Submitted quality work
4. Exerted effort and took an active role
5. Cooperated and communicated with others
6. Managed group conflict
7. Made cognitive contributions; possessed and applied necessary knowledge and skills
8. Provided structure for goal achievement

However, most evaluation tools are developed by instructors. As such, the desired behaviors as listed are top-down rather than bottom-up. How the students themselves are perceiving their own learning environment is vitally important to their persistence in engineering [12][14]. A second study suggests that, though many behaviors overlap, some aspects of teammate behavior viewed as important to students are not reflected in most instructor-created peer assessments. This study lists eleven behavior components important to teammates in engineering education settings. The more unexpected components of poor team behavior include expecting teammates to contribute beyond their "fair share", being unwilling to take on tasks beyond clearly articulated expectations of teammates, and directly or indirectly inhibiting the group from completing its work in a timely manner. [1] Other studies, more quantitative in scope, have identified dysfunctional teams using a "dysfunctional index" of test scores, analogous to a signal-noise ratio. [15]

Expertise Recognition

Thomas-Hunt and Phillips note that perceptions do not always match reality by saying, "the complexity of most organizational tasks makes it difficult for expert members to demonstrate

the correctness of their perspective prior to the completion of the group's task and the receipt of feedback from sources external to the group. Consequently, teams often have difficult assessing the veracity of members' claims of expertise" [16]. Evaluations of expertise within a group may differ between individuals, and thus may be colored by the idiosyncrasies of those individuals. These evaluations also may or may not reflect the true nature of the object of evaluation. Some studies have found that among women, educational status as objectively measured by education level attained does not significantly predict expertise evaluations [16], [17]

This may be explained in part by social identity theory [18] which posits that easily accessible differences between people, like gender, form the basis for in-group/out-group distinctions. People tend to prefer the in-group in a context-specific way. For example, a male job applicant may be preferred for a traditionally "male" position such as police chief whereas the female job applicant may be preferred for a traditionally "female" position such as nurse. These preferences are not always explicitly known to the evaluator. One study found that male and female applicants for the same position were not found to actually possess different strengths, but rather that the evaluator redefined the criteria deemed necessary for job success to favor the specific credentials that the in-group candidate happened to possess. The evaluators in this case were not aware of their own implicit biases. In fact, perceiving one's own judgments as objective actually predicted greater bias. [19] As engineering is widely known to be a male-dominated field, women are often perceived to be members of the "out-group" in this context. Thus, expertise judgments made about peers may be influenced by the implicit biases associated with the in-group/out-group distinction.

The greatest variance in expertise evaluations is exhibited, not through the attributes of the person being evaluated (such as gender or education level completed) as one might expect, but rather through the attributes of the person providing the evaluation and the relationship between the two individuals. [17] Van Der Vegt similarly suggests that levels of interpersonal commitment felt between individuals on a team were predicted by the relationship between those two group members and not by the makeup of the group as a whole. [20]

This study hypothesized that, in the same way, perceptions of the severity and frequency of negative team behaviors differ between individuals and that those perceptions may be influenced by factors not associated with actual performance (such as gender). In particular, it hypothesized that the discrepancies in perceptions of an individual can be explained by the attributes of the relationship between the dyad (one person judging a second person) and/or attributes of the individual judging others.

Methods

Participants in this study were recruited from undergraduate engineering project classes at the University of Virginia. Examples include Synthesis Design II (ENGR 1420) and Civil Design (CE 4991). With the consent of class instructors, a sign-up email was circulated including the link to the survey itself. In some cases, a short presentation was also given in the relevant class to recruit participants. Participants must have been currently working on a team for a time period of at least 4-6 weeks at the time of the survey, and they could take the set of two surveys (Basic Psychological Need Satisfaction Scale and Team Behaviors and Attitudes

Survey) at their own leisure on a computer via Qualtrics. Both surveys were completed at the same time. In this way, about 350 students were contacted, yielding 89 respondents. (response rate: 25.4%)

Team Behaviors and Attitudes Survey

In conjunction with this established survey, a second survey was constructed for the purposes of this study. The Team Behaviors and Attitudes Survey (TBAS) explores teammate perceptions of one another's behavior over the course of the class project using a round-robin format. In particular, it examines eleven distinct types of negative teaming behaviors identified in previous work [1], including expecting too much from others, failing to advance toward project's completion, failing to prioritize project, inconsistency of contribution, inconsistency with an engineering identity, lack of communication, lack of competence, experience, or skills, lack of initiative, procrastination, restricting others' work, and unreliability. It additionally asks for overall impressions of each teammate's contributions to the project with three quantitative questions and demographic information, such as sex, age, and year in school.

Ideally, each individual on a team would complete the TBAS survey to provide a more complete picture of how each individual was widely perceived to behave. However, not every member from each team participated. Therefore, the incentive structures encouraged volunteering in conjunction with other teammates. The payment associated with participation in this study increased for each individual on a team if the entire team participated.

The survey was built in a branched format. First, respondents answered demographic questions about themselves such as their year in school, major, and age. They also indicated their team's size and their teammates' first names. They answered a series of 3 questions about their overall impression of contributions made to the project by each teammate. Then, they read a list of 11 vignettes that corresponded to the 11 identified negative team behavior types [1]. To reduce fatigue and order effects, all 11 behavior vignettes were presented at once. Listing previously named teammates one at a time, respondents indicated whether each teammate exhibited each behavior binarily (yes/no) in one long form. Answering "no" constituted a score of "0" for that particular behavior and teammate. Answering "yes" led to a series of four quantitative and three qualitative questions about that teammate and behavior. The four quantitative scores assessed frequency, effect on personal experience, effect on team's work quality, and effect on team overall. The last three were averaged, yielding a score between 1 and 7 for that teammate. In total, behaviors were scored from 0 to 7, so that each teammate had a score for each behavior. For the full TBAS text, see the appendix. This branched format provided multiple benefits. First, it allowed participants to compare the behavior vignettes to one another and see the differences more clearly. Second, if respondents knew that answering "yes" required more work beforehand, they might have been less honest about which behaviors did and did not occur.

Sample

In total, 89 unique respondents completed both surveys at a fairly large public university. Their demographics are detailed in Figure 1. These students had an average age of 21.07

(with five students not giving their age).

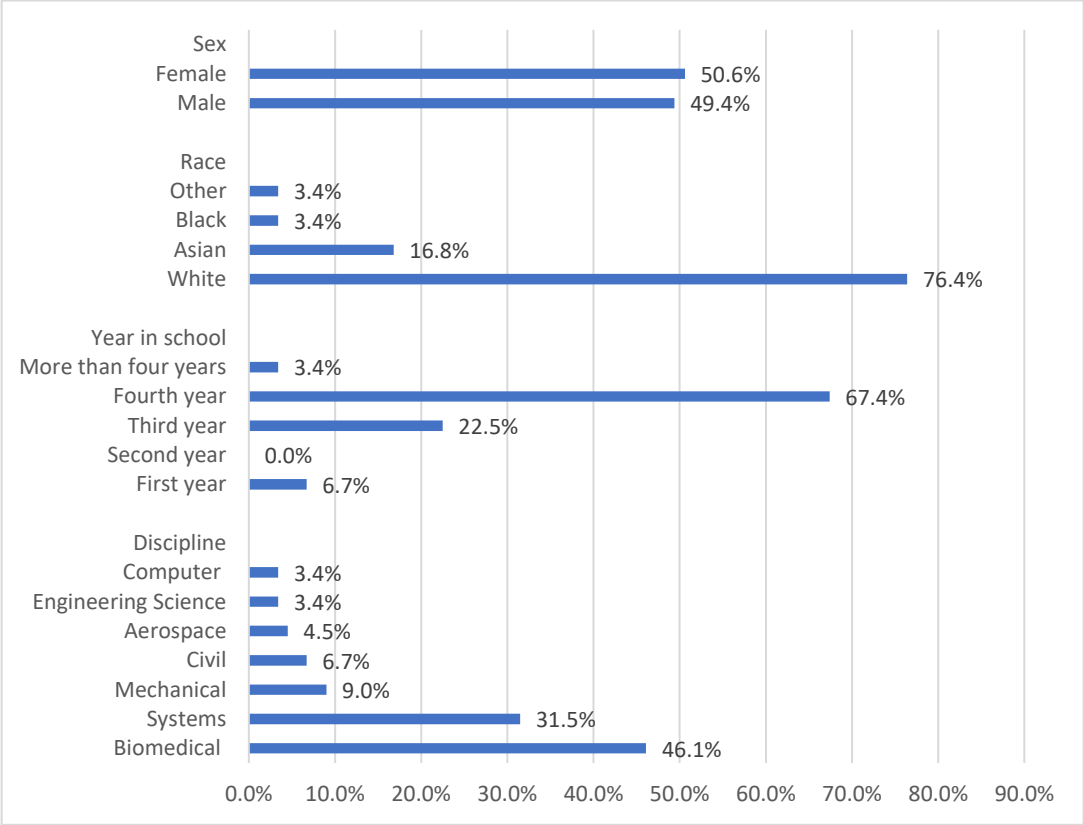

Figure 1: Participant demographics

These participants represented 47 unique teams that ranged in size from 2 to 7 people total (including the rater). This dataset provided behavior information on 178 unique "targets" (or "ratees"), 66 of whom were also raters. The TBAS asks participants to provide demographic information about themselves but not about their teammates (other than first names). To reconstruct likely genders for "targets", the following procedure was used. First, some targets were also raters and thus had a listed gender. Second, qualitative data provided about a target sometimes used gender-specific pronouns, which were taken to be the correct gender. Third, the 1996 census list (average respondent age was 21.07) of most popular boys and girls names was searched and the higher listing was chosen as the correct gender. Fourth, if the name was not on the census list (which were most often international names) a google-search was completed of the gender commonly associated with that name. In total, 93 (46%) were previously indicated as a rater, 32 (16%) were taken from qualitative data pronouns, 66 (33%) were taken from the 1996 census list, and 10 (5%) from a google search. 61% of unique targets were male, reflecting team makeup more than anything else. Out of these 329 pairings, each behavior was said to exist with the following frequencies (Table 1):

Table 1: Behavior frequencies by pairings

Expecting too much from others	13.4% (44 pairings)
Failure to advance project toward completion	13.7% (45)
Failure to prioritize project	16.4% (54)
Inconsistency of contribution	17.0% (56)
Inconsistency with an engineering identity	7.0% (23)
Lack of communication	13.7% (45)
Lack of competence, experience, or skills	9.4% (31)
Lack of initiative	19.1% (63)
Procrastination	9.4% (31)
Restricting others' work	6.4% (21)
Unreliability	9.4% (31)

Results

Based on the literature in expertise recognition [17], [20], which suggests that variance in the perceptions of expertise among team members (as assessed by educational level attained) is based less on the actual education level of the target and more on the attributes of the actor (rater) and the relationship between the pair, the following hypothesis is made. It is evaluated using a Social Relations Model.

Hypothesis: Variance of behavior perceptions will be based more on the actor than the target.

The Social Relations Model (SRM) is most often used with round-robin data in which individuals rate one another within groups. Snijders and Kenny first developed the SRM model when dealing with relationships between family members. These relationships are certainly not independent of one another – the relationship between two individuals within a family may depend on person #1, person #2, the dynamics between the pair, and the dynamics of the family as a whole. [21] Similarly, on a project team, one's perceptions of a teammate's exhibited behavior may depend on the individual making a judgment, the individual being judged, the relationship between the pair, and the dynamics of the group as a whole. For the purposes of this study, these individuals will be referred to as the actor, target, dyad, and group, respectively.

The social relations model for multiple groups is based on the following equation:

$$Y_{ijk} = \mu + F_k + A_{ik} + B_{jk} + R_{(ij)k} + E_{ijk}$$

in which the response Y_{ijk} is the sum of an overall mean μ, F_k is the random effect of family k, A_{ik} is the effect of actor i on group k, B_{jk} is the effect of target j on group k, $R_{(ij)k}$ is the dyad effect, and E_{ijk} and E_{jik} are assumed to be uncorrelated. [21] Unlike more traditional multilevel models, the levels of analysis are crossed rather than nested, meaning that each unit of the lower level is not contained in exactly one unit of the higher level. [22] In this

case, level one includes the actor and target, level two is the dyad, and level three is the group. The SRM model takes account of these complex and interrelated levels of analysis by using random effects to allow for a complicated correlational structure within a group, and the fixed effects allow the effects of covariates within each level. [21]

In this study, SAS software was used to estimate the SRM using conventional multilevel modeling (PROC MIXED COVTEST) as opposed to the method initially proposed by Snijders and Kenny that uses dummy variables and an obscure software package. Seven parameters exist in an SRM model: the mean, actor variance, target variance, group variance, error variance, actor-target covariance, and error covariance. The weakness of using conventional multilevel modeling is that the actor-target covariance is assumed to be zero. Because the group variance depends on the actor-target covariance, the method also does not give a group covariance estimate. However, the method has been shown to estimate actor variance, target variance, and error covariance accurately in comparison to the original method. Error covariance refers to the two members' correlated relationship effects, while the actor-target covariance refers to the correlated actor effect and target effect [23]

The full TBAS dataset includes 329 unique pairings (individual #1 perceiving individual #2) from 47 different teams. However, for this round-robin model only groups in which every member participated in the study were included. This smaller dataset includes only 132 unique pairings from the 11 full teams represented. These teams ranged in size from two to seven people, which was accounted for by the model.

Contrary to what is suggested in expertise recognition literature, the results suggest that the target's attributes contribute more to the variance than the actor's attributes. The actor variance estimate for this model is 0.126 ($p=0.003$), while the target variance estimate is 0.180 ($p=0.001$). The error co-variance is -0.009 and the standard error of the model is 0.093. Contrary to what is suggested in expertise recognition literature [17], the results suggest that the target's attributes contribute more to the variance of the overall scores than the actor's attributes in this case. The error covariance estimate, which speaks to the effect of the relationship between the dyads, was close to zero. Furthermore, it is still currently unclear to what extent the group may contribute to the variance.

Discussion

In her work on expertise recognition, Joshi found that when the various levels of groups were examined (actor, target, dyad, and group), the greatest source of variance was not the "attributes (such as gender or education level completed) of the person being evaluated (i.e., the target) as one might expect, but rather through the attributes of the person providing the evaluation (i.e., the actor) and the relationship between the two individuals (i.e., the dyad)". Joshi considered all four levels of cross-level groups in conjunction with one another, without directly comparing one level to another. [17] Indirectly, this can be interpreted as saying that when evaluating others' expertise, actors contribute more to the variance than targets. Through evaluating overall assessments of teammates in this study, however, the opposite was found – that targets contribute more to the variance than actors. At least one notable difference exists between the two studies. The overall teammate evaluation questions in the TBAS survey encouraged students to take a critical eye to their teammates' behavior in terms of the success of the team, personal experience, and frequency of enjoyment. Joshi's study,

on the other hand, was framed as identifying a teammate's highest education level achieved. Evaluations of teammate behaviors are experienced on a personal level over the course of the project, whereas education level must generally be inferred from observed behaviors. It may be the case that more concrete concepts (like behavior observation) yield perceptions that are based more heavily on the target while more higher-level or abstract concepts (like an inference of education level) may be influenced more easily by the actor's attributes and preconceptions of that individual.

Validity and Reliability

The TBAS used eleven vignettes based on, but not identical to, behavior definitions from past work [1] to communicate the same ideas about types of behavior to the participants. How can we know whether participants understood the same concepts that the vignettes were intended to portray? Each time a participant indicated that a particular individual on their team exhibited a particular behavior based on their reading of the vignette, one of the open-ended questions asked was: "In what way do you feel _____'s behavior during the project is/was similar to the passage above?" These responses were randomized and de-identified. The web-based computer software *Dedoose* was used to code all 366 excerpts based on the 11 original behavior definitions independently of which behavior the participant had intended to indicate. A given excerpt could be coded as more than one behavior according to the official definitions. Agreement between the intended behavior and the coded definition would indicate that the participants did understand the concept the vignette was intended to portray. In Table 2, the descriptor rows show the behavior each vignette was intended to portray by the participant, and the code columns show how the researcher coded each vignette according to the given definitions. For most behaviors, the most frequently appearing behavior definition code was in fact the intended behavior. The exceptions to this are *Inconsistency of contribution*, which was more often coded as *Failure to prioritize project*, and *Unreliability* which was equally frequently coded as *Failure to advance project toward completion.* By examining the qualitative excerpts, it seems that *Inconsistency of contribution,* a behavior meant to indicate long-term variability ("My teammate contributed at the beginning of the semester, but after they landed a job offer they lost interest in our project") was often misunderstood as short-term irregularities in performance ("Sometimes they didn't seem as responsive to the group text and sometimes they would"). Furthermore, *Unreliability* was a behavior meant to showcase lack of trust between teammates ("I didn't know if I could trust them to get their part done or whether I should just do it myself") that instead was often interpreted as simply not contributing the way they ought ("My teammate did little for the project").

Table 2: Intended behavior vs. code counts

Intended behavior / **Codes of responses**	Expecting too much from others	Failure to advance project toward completion	Failure to prioritize project	Inconsistency of contribution	Inconsistency with an engineering identity	Lack of communication	Lack of competence, experience, or skills	Lack of initiative	Procrastination	Restricting others'work	Unreliability
Expecting too much from others	15	8	2				2	7		1	1
Failure to advance project toward completion	1	27	5		1	2	4	6	1	1	2
Failure to prioritize project			41								
Inconsistency of contribution	1	9	16	11	2	6	1	2	6		
Inconsistency with an engineering identity		5		1	8		3	1			
Lack of communication		4	8			29		2	3	4	
Lack of competence, experience, or skills	1	4	3		1	2	20	4		5	
Lack of initiative		4	2	2	1		3	39	3	1	2
Procrastination		6	2	1		1			15	4	6
Restricting others' work		5	3			2	5		1	9	
Unreliability	3	10	6			1	7	1	3	4	10

Because each team is composed of different individuals, typically reliability measures are inappropriate in the context of this study. Agreement was ascertained by comparing ratings of

an individual target by multiple actors for each of the eleven behaviors. Teammates were considered in agreement if there is unanimity for whether an individual possessed a behavior or not – severity of scores was not taken into account. Only teams in which all members participated in the study were considered. Of targets with two teammates, 93.9% of behaviors were unanimous. Of targets with three teammates, 78.8% were unanimously rated. Of those with four teammates, 58.2% of behaviors were unanimously rated. Of those with five teammates, 28.8% of behaviors were unanimous. Reaching unanimity is considerably more difficult as more voices are added.

Limitations

This study uses conventional multilevel modeling to estimate differences in variance for actors, targets, dyads, and groups. While multilevel modeling is effective at estimating certain parameters (actor variance, target variance, and error covariance), it is restricted in the sense that the actor-target covariance is fixed to zero. Using this limited model, one cannot know the true actor-target covariance or group variance. Thus, this paper largely aims to compare actor variance and target variance.

When data were originally collected for this study, participants were incentivized to volunteer in conjunction with their other teammates. Though members representing 47 teams took part in the study, only 11 of these had every single team member participating. Of these 11 teams, two are two-person teams. The social relations model does not require that each team be the same size, but this limited the sample size of participants representing full teams used in this work. In a smaller sample, the occurrence of each of eleven behaviors is low (ranging from 7 times for *Unreliability* to 33 times for *Lack of initiative*). A larger overall sample size including more full teams is desirable for future analysis.

Additionally, this work focuses on undergraduate students on long-term engineering project teams at a mid-sized university. Certain particularities are inherent to this context. For example, in small schools in which students know one another apart from the classroom setting, students are apt to carry in their own judgments about their teammates that are not based on their performance within the group. This effect may impact the results. In this study, students occasionally but not often knew each other prior to the project. Still, results may not necessarily be extended to other contexts.

Future Work

This study attempts to compare the actor and target's influence on eleven specific behaviors despite the fact that because the four levels are crossed rather than nested, complicated relationships between them arise. When this multilevel model is applied to evaluations of individual behaviors, it is expected that different behaviors will yield different patterns of how the gender of the actor and target influenced them. The behaviors may be categorized into three different groups: those where actor effects outweigh target effects, those where target effects outweigh actor effects, and those where the two effects are roughly equal. Behaviors belonging to each of these groups may share characteristics that help explain the results.

For some behaviors, the actor effects contribute more to the variance than the target effects. Behaviors like *Inconsistency with an engineering identity* is likely to fit in this category. In general, this category seems to be particularly subject to the attitudes and interpretation of the actor, which could be the result of a second variable such as gender. In the literature regarding expertise recognition, Joshi found that "when evaluating the expertise of male targets, male actors did not differentiate among targets based on educational level. When male actors evaluated female targets, however, they did differentiate between targets based on educational status, rating highly educated women lower than less educated women". [17] In the qualitative data of past work [1], both of the behaviors with actor effects outweighing target effects were said to occur sometimes on the basis of the actor's attitudes about gender. For example, the actor believed the target was not a "real engineer" for some reason or that they didn't possess the skills necessary to complete engineering tasks, both of which are attitudes that frequently reflect gender attitudes. Because these judgments are generally based on perceptions of attitudes, it makes sense that they are based on the actor's characteristics and not the target's – there are a myriad of ways to "be an engineer", regardless of whether this fits into the actor's mental model. Further analysis is required to investigate this possible relationship.

In other cases, the actor effects may contribute less to the variance than the target effects. One likely behavior in this category is *Lack of communication.* Perceptions of these negative behaviors are more likely to be unambiguous, observable, warranted, and ultimately less subject to the attitudes of the actor.

For cases in which the actor and target effects are roughly equal, co-construction by the actor and target of the reality of the relationship seems likely. Future work that identifies the actor-target covariance and group variance may shed further light on this area.

Future work would also include examining how gender plays into the actor-target dynamic and how it impacts perceptions of negative behaviors will be explored. Do actors rate targets more severely or more frequently when the dyad pattern is "Male -> Female", for example? Social identity theory, [18] which posits that easily-accessible differences between people, like gender, form the basis for in-group/out-group distinctions, may suggest that "Male -> Male" and "Female -> Female" dyads may rate less severely than "Female -> Male" and "Male -> Female" dyads, for example.

Furthermore, expansion of the material in this paper using the PROC MIXED COVTEST dummy variable method [23] of estimating the Social Relations Model to obtain accurate actor-target covariance and group variance measures of the dataset is recommended. Round-robin data is somewhat unique in that perceptions of individuals exist with cross-classified variables (individuals, dyads, and groups).

References

[1] E. Miller, L. Hirshfield and D. Chachra, "Classifying dissatisfaction: Student perspectives on teammate performance," in *American Society for Engineering Education*, New Orleans, LA, 2016, .

[2] F. Erickson and J. Shultz, "Students' experience of the curriculum," in *The Handbook of Research on Curriculum: A Project of the American Research Association* Anonymous New York: Macmillan., 1992, pp. 465-485.

[3] M. Borrego *et al*, "Team Effectiveness Theory from Industrial and Organizational Psychology Applied to Engineering Student Project Teams: A Research Review," *J Eng Educ*, vol. 102, *(4)*, pp. 472-512, 2013.

[4] (). *Accreditation*.

[5] G. L. Stewart, I. S. Fulmer and M. R. Barrick, "An Exploration of Member Roles as a Multilevel Linking Mechanism for Individual Traits and Outcomes," *Person. Psychol.*, vol. 58, *(2)*, pp. 343-365, 2005.

[6] S. Sonnentag and J. Volmer, "Individual-Level Predictors of Task-Related Teamwork Processes: The Role of Expertise and Self-Efficacy in Team Meetings," *Group & Organization Management*, vol. 34, *(1)*, pp. 37-66, 2009.

[7] A. Zhang, "Peer Assessment of Soft Skills and Hard Skills," *Journal of Information Technology Education*, vol. 11, *(1)*, 2012.

[8] L. E. Gueldenzoph and G. L. May, "Collaborative Peer Evaluation: Best Practices for Group Member Assessments," *Business Communication Quarterly*, vol. 65, *(1)*, pp. 9-20, 2002.

[9] S. Fallows and B. Chandramohan, "Multiple Approaches to Assessment: Reflections on use of tutor, peer and self-assessment," *Teaching in Higher Education*, vol. 6, *(2)*, pp. 229246, 2001.

[10] A. Erez, J. A. Lepine and H. Elms, "Effects of Rotated Leadership and Peer Evaluation on the Functioning and Effectiveness of Self-Managed Teams: A QuasiExperiment," *Person. Psychol.*, vol. 55,*(4)*, pp. 929-948, 2002.

[11] D. F. Baker, "Peer Assessment in Small Groups: A Comparison of Methods," *Journal of Management Education*, vol. 32, *(2)*, pp. 183-209, 2008.

[12] C. Mcdowell *et al*, "The impact of pair programming on student performance, perception and persistence," in *25th International Conference on Software Engineering, 2003. Proceedings.* 2003.

[13] S. E. Cross and N. V. Vick, "The Interdependent Self-Construal and Social Support: The Case of Persistence in Engineering," *Person. Soc. Psychol Bull.*, vol. 27, *(7)*, pp. 820-832, 2001.

[14] R. Suresh, "The Relationship between Barrier Courses and Persistence in Engineering," *Journal of College Student Retention: Research, Theory & Practice*, vol. 8, *(2)*, pp. 215-239, 2006.

[15] C. Hsiung, "Identification of Dysfunctional Cooperative Learning Teams Based on Students' Academic Achievement," *J Eng Educ*, vol. 99, *(1)*, pp. 45-54, 2010.

[16] M. C. Thomas and K. W. Phillips, "When What You Know Is Not Enough: Expertise and Gender Dynamics in Task Groups," *Person. Soc. Psychol Bull.*, vol. 30, *(12)*, pp. 15851598, 2004.

[17] A. Joshi, "By Whom and When Is Women's Expertise Recognized? The Interactive Effects of Gender and Education in Science and Engineering Teams," *Adm. Sci. Q.*, vol. 59, *(2)*, pp. 202-239, 2014.

[18] H. Tajfel and J. C. Turner, *The Social Identity Theory of Intergroup Behavior*. 2004.

[19] L. U. Eric and G. L. Cohen, "Constructed Criteria: Redefining Merit to Justify Discrimination," *Psychol. Sci.*, vol. 16, *(6)*, pp. 474-480, 2005.

[20] D. V. Van, J. S. Bunderson and A. Oosterhof, "Expertness diversity and interpersonal helping in teams: Why those who need the most help end up getting the least," *Academy of Management Journal*, vol. 49, *(5)*, pp. 877-893, 2006.

[21] T. A. B. Snijders and D. A. Kenny, "The social relations model for family data: A multilevel approach," *Personal Relationships*, vol. 6, *(4)*, pp. 471-486, 1999.

[22] T. A. B. Snijders and R. J. Bosker, *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. Sage Publishers, 2012.

[23] D. A. Kenny, "Estimation of the SRM Using Specialized Software," unpublished.

Appendix: Team Behaviors and Attitudes Survey

How do you believe _____'s behavior impacted the overall success of the team?

1. It strongly hindered the team's success
2. It somewhat hindered the team's success
3. It somewhat supported the team's success
4. It strongly supported the team's success

How do you believe _____'s behavior affected your overall personal experience working on the project?

1. Strong negative impact
2. Slight negative impact
3. Slight positive impact
4. Strong positive impact

How frequently did you enjoy working with _____?

1. Never
2. Less than half the time
3. More than half the time
4. Always

For each teammate (computer auto-fills each team member's name)

For each member of your team, indicate whether you recognize this type of behavior in the way they conducted themselves throughout the course of the project. If the behavior occurred at least **once**, please indicate "yes".

"This person expected others to pick up their slack, as they wanted to avoid responsibilities themselves. I felt that they took advantage of other teammates by asking them to take on more than their "fair share"." ***Yes No***

"Generally, my teammate just didn't seem to move the project closer to completion or add any value. They were not actively engaged - you could often find them watching others work, sitting idly, or simply listening to group discussions. They seemed unwilling to participate in the work itself." ***Yes No***

"To my teammate, other commitments always seemed to take precedence over our project. They were unwilling to devote much of their time to making the project the best it could be. They may have been frequently absent in team meetings." ***Yes No***

"Sometimes my teammate was available and an active contributor, and sometimes they were not. Sometimes they seemed committed to the project's success, and sometimes they seemed to

fall off the grid. Perhaps they were complacent until a deadline loomed, or perhaps they helped until deciding to accept a job offer. Either way, my attitude toward them changed drastically over the course of the project." ***Yes No***

"Their attitude toward the project did not align with how I feel an engineer should behave or feel. I identify as an engineer more than I believe this person does. For example, it bothered me that they took on tasks that were not technical in nature. I may have felt that while I was truly enthusiastic about the engineering work, my teammate was mostly concerned about getting an A." ***Yes No***

"This person was not an effective communicator, either within the team or with others. They didn't work to make sure we were all on the same page. They may have failed to distribute information important for everyone to have, gotten easily off track, or generally failed to connect with the rest of the team." ***Yes No***

"My teammate lacked the requisite competence, experience, and/or skills to effectively contribute to the project. They sometimes seemed to be "behind" the rest of the group members in terms of content understanding. Sometimes the work they'd present to the group was incorrect or needed to be reworked. I was constantly spending group work time explaining (what I considered to be) obvious questions about content, rather than focusing on important issues." ***Yes No***

"This person was willing to complete tasks explicitly assigned to them, but never seemed to take the next step. I can't point to one aspect of the project over which my teammate seemed to take ownership." ***Yes No***

"My teammate seemed to always wait until the last possible moment to get their work done. Even if they thrived under the pressure and did manage to pull off a successful finished product, it was done at the last minute. This habit put a strain on my schedule, especially when I depended on them to finish their part before I could do mine." ***Yes No***

"My teammate directly or indirectly prevented the group from making effective progress on the project. They may have demanded work that I felt was sufficient needed to be redone, or simply redoing it themselves. They were active team members, but I felt that they detracted from the work's quality." ***Yes No***

"I couldn't really trust my teammate to follow through on what they promised to do. I was always worrying and wondering whether they were spending their time getting their work done or if I should just do it myself. I felt like I needed to keep reminding them about their obligations." ***Yes No***

For each member and behavior they respond "yes" to…

> In what way do you feel _____'s behavior during the project is/was similar to the passage above? Please limit your response to 1-2 sentences.

How frequently did this behavior bother you over the course of the project?

1. Almost never
2. Rarely
3. Sometimes
4. About half the time
5. Often
6. Usually
7. Almost always

To what extent did this behavior affect your overall personal experience working on the project?

1. The behavior had a strong negative impact
2. The behavior had a moderate negative impact
3. The behavior had a slight negative impact
4. No effect
5. The behavior had a slight positive impact
6. The behavior had a moderate positive impact
7. The behavior had a strong positive impact

How specifically do you feel this behavior impacted your own experience working on the project?

To what extent did this behavior affect the quality of work your team was able to achieve?

1 - The behavior strongly hindered the quality of work
2- The behavior moderately hindered the quality of work
3 - The behavior slightly hindered the quality of work
4 - No effect
5 - The behavior slightly supported the quality of work
6 - The behavior moderately supported the quality of work
7 - The behavior strongly supported the quality of work

To what extent did this behavior affect the team's overall success?

1 - The behavior strongly hindered the team's success
2- The behavior moderately hindered the team's success
3 - The behavior slightly hindered the team's success

4 - No effect
5 - The behavior slightly supported the team's success
6 - The behavior moderately supported the team's success
7 - The behavior strongly supported the team's success

How specifically do you feel this behavior impacted the team's overall success?