

Validation of the Teamwork-in-Engineering-Education Situational Judgment Test (TEE-SJT)

Dr. Chris C. Martin, Georgia Institute of Technology

Dr. Chris C. Martin is the director of collaborative innovation in the School of Civil and Environmental Engineering at Georgia Tech. He teaches about and conducts research on wellbeing, teamwork, psychological safety, and inclusion.

Validation of the Teamwork-in-Engineering-Education Situational Judgment Test (TEE-SJT)

This is a full research paper in the empirical research category.

Teamwork is ubiquitous in engineering classes. Sometimes, students are placed in team to engage in collaborative learning, which carries distinct benefits [1], [2]. At other times, students are organized in teams to prepare them for the workplace [3], a pertinent goal considering that some industry experts view teamwork skills as more consequential than technical skills like prototyping [4]. Because teaming is now ubiquitous, engineering educators have increasingly adopted concepts from organizational psychology, such as psychological safety, and created interventions to enhance it [5], [6].

Although these intervention studies are valuable, it is hard to discern whether the interventions are truly effective. Researchers typically rely on problematic self-report techniques to evaluate effectiveness, and these face-valid techniques entail three problems. First, students can "fake good" by selecting socially desirable answers [7], [8]. Second, students may act as "good subjects" by supporting the experimental hypothesis, especially if the experimenters are well-liked faculty [9]. Third, unreflective students often rate themselves highly even if they have learned very little. [10]. The Groupwork Skills Questionnaire (GSQ) is one example of an instrument that possesses these problems [11].

This article introduces a novel instrument that mitigates these problems: a situational judgment test (SJT) that evaluates whether individuals possess skills and habits that pertain to beneficial outcomes such as psychological safety, inclusion, and justice. This instrument is broad enough for all fields of engineering because no specialized knowledge is required. Furthermore, the instrument does not explicitly ask students to evaluate themselves.

An SJT comprises realistic scenarios with questions about test-taker behavior. Consider this example: One of your team members only shows up to half of all scheduled meetings. What would you do in this situation? (a) Send them an email and ask them to attend all future meetings. (b) Meet in person to find the reason for their absences. In a typical SJT, several scenarios and questions are sequentially presented.

This article also introduces two methodological concepts to readers: first, pre-registration—a practice minimizing questionable research practices like writing hypotheses after results are known [12], [13], and second, the quantifying construct validity (QCV) procedure [14], which outputs two effect sizes representing construct validity of a scale. The QCV procedure mitigates the problematic practice of subjectively assessing convergent and discriminant validity absent specific hypotheses.

SJTs follow from the study of real-life critical incidents where effective or ineffective behavior is displayed by a candidate [15]. Adding questions and scoring rubrics to critical incidents produces SJTs [16], [17]. SJTs are sometimes termed low-fidelity simulations distinguishing them from high-fidelity techniques wherein a candidate must perform actual work. Knowledge determinant theory underlies modern SJTs, positing that SJTs predict job performance by

measuring procedural knowledge about effective and ineffective behavior [18]. This knowledge reflects personality traits and prior reinforcement by parents, teachers, and others.

Medical education is one domain where SJTs have been widely validated and adopted. A synthesis of research provides compelling evidence for their validity [18]. For instance, medical SJT performance at admission predicts internship and job performance after seven and nine years respectively [19]. SJTs are less commonly used for high-stakes assessment in engineering, likely due to different credentialing in engineering. In medicine, a small number of candidates must vie for a small number of seats in an M.D. program, which necessitates the use of tools like SJTs to differentiate between applicants with high standardized test scores. In engineering, however, a student only needs to earn a bachelor's degree to be employable, and many seats are available at the undergraduate level. Furthermore, empathy is less critical in engineering because engineers do not diagnose and treat patients in sensitive personal consultations.

SJTs are not altogether absent in engineering education. In the last six years, teams at Wake Forest University and Purdue University have developed SJTs. The first group developed a situational judgment test for ethical decision making, studying its utility for in-class use; a small-sample study validated that scores varied in the expected manner between students, early career professionals and expert engineers [20], [21]. However, there were no inferential statistics reported to support or disconfirm specific hypotheses.

The second group developed a China-based situational judgment test of intercultural knowledge, reporting that scores on the test had a modest correlation of .18 with Chinese cultural knowledge [22]. In a subsequent study, they also reported a higher mean test score among students with study abroad experience [23]. However, specific study-abroad experience in China did not explain unique variance. In a puzzling result, engineers who had traveled to China for work had lower scores than engineers who had never lived abroad.

More pertinently, an SJT to measure prosociality was developed by Motowidlo and colleagues in 2016 [24]. This SJT was tested with physicians, lawyers, community service volunteers, and human factors engineers, but it was apparently neglected by engineering educators after its publication.

One possible reason for the former neglect of SJTs in engineering education is that their development places extraordinary demands on developers' time [17]. Subject matter experts must create and review scenarios for readability, suitability, and validity. Questions and response options require expert review, followed by pilot testing and cycles of iterative refinement. Generative AI may eventually streamline this process [25].

Development of the Teamwork in Educational Engineering SJT (TEE-SJT) began with a selection of constructs to be measured. This was done by a panel of subject matter experts, including engineering professors, engineering lecturers, and postdoctoral fellows. First, the lead author presented a list of skills and habits that could be assessed, drawing on the organizational psychology literature on psychological safety, justice, and other topics. The panel expanded and edited this list to develop a long list of 28 interpersonal skills and habits that could be taught to undergraduates. They then voted on which elements to prioritize, which yielded a short list of 18

items. (Items that earned few votes were discarded.) Examples of items that earned many votes include "the habit of periodically evaluating and addressing unfairness among team members" and "the ability to differentiate between symptoms of disengagement, introversion, anxiety, and low psychological safety."

A bottom-up analysis of the retained items yielded five thematic areas based on the lead author's analysis. The first two were *psychological safety* and *justice*, which have substantial empirical evidence supporting their importance in the workplace [26], [27], [28]. Psychological safety refers to shared sense that it is safe to take interpersonal risks. The next two themes were *inclusion* and *appreciation of diversity*, intertwined topics gaining attention as engineering schools acknowledge a history of discrimination and marginalization of certain students and topics.

The final theme was supportive feedback, which less prominent than psychological safety and justice, but nevertheless of substantive importance [29], [30]. Supportive feedback conveys speaker investment in recipient growth [31], [32], making interactions feel just and inclusive without diminishing psychological safety. All five themes pertained to an NSF-supported project at our institution training undergraduates to become inclusive workforce members.

Altus Assessments a psychometrics firm specializing in SJTs, trained our team. They were selected because they created CASPER, widely used by medical schools for selection, and their consultants held doctoral degrees in psychometrics or a related field. Consultants advised our team on blueprinting—mapping assessments to constructs—covering potential hazards such as construct underrepresentation and construct-irrelevant variance.

Small working groups, containing a mix of faculty and students, then designed scenarios and items. To initiate development, seven scenarios per theme were created corresponding to seven DIAMONDS taxonomy facets [33]: Duty, Intellect, Adversity, pOsitivity, Negativity, Deception, and Sociality. For instance, the [psychological safety $\times$ duty] cell contained: "Students in a group feel they can't make bids for help because roles in a team have already been set and there is no feeling of group flexibility." Each group developed one scenario, multiple questions, and a rubric. Scenarios were reviewed and edited by the lead author and a group of teaching assistants to maximize clarity and minimize construction underrepresentation and construct-irrelevant variance.

Hypotheses Regarding Construct Validity

We adopted the Quantifying Construct Validity (QVC) procedure to evaluate construct validity with precision [14], [34]. QVC outputs two effect sizes, ranging from -1 to 1, representing how the magnitude of the match between hypothesized and observed correlations. The first, r_{alerting} , is the correlation between the set of hypothesized correlations and the respective set of z-transformed observed correlations. The second, r_{contrast} , accounts for sample size, median intercorrelation among criterion variables, and correlation magnitude. Large effect sizes jointly verify convergent and discriminant validity, supporting construct validity. However, in practice, r_{contrast} typically lags behind r_{alerting} [35], [36], [37]. Under certain situations, large values of r_{contrast} are impossible to obtain.

Table 1 displays pre-registered hypothesized correlations between the SJT and criterion-related variables. In our pre-registration, we hypothesized both r_{alerting} and r_{contrast} would surpass .60—an ambitious but reachable goal. Hypotheses were pre-registered before data collection at aspredicted.org.

Methods

Participants

The preregistered sample size was 200 because correlations stabilize at this value [38]. After preregistration, an exploratory goal comparing mean SJT scores across four subgroups was added, with power analysis indicating 206 participants needed for 80% power (effect size $f = 0.25$). Planned sample size increased to 212, though this goal was discarded due to insufficient subgroup participation. Both pilot and main studies received IRB exemptions (Protocol Numbers H20260 and H21079). Participants gave informed consent.

Participants were 215 undergraduate students at a southeastern U.S. engineering institute (216 submissions; one dropped for inattentive answering). They were recruited through mass e-mail invitations. The gender makeup was 42.8% male, 49.8% female, and 0.9% genderqueer/gender-nonconforming. The racial and ethnic makeup was 40.5% Asian, 4.2% Asian–White, 32.1% non-Hispanic White, 5.1 non-Hispanic Black, 5.6% Hispanic of any race, 0.9% Middle Eastern, and 3.7% other races. The nationality makeup was 4.7% international and 88.4% non-international. The disability-status makeup was 8.4% disabled and 86.0% not disabled. Due to missing data, proportions do not sum to 100.

There were roughly equal proportions of first-year (16.7%), second-year (19.5%), and third-year students (17.7%), and a large proportion (46%) of fourth-year (or above) students. With one exception (neuroscience), all participants were studying engineering: the most common majors were biomedical engineering (34%), aerospace engineering (28.8%), civil engineering (9.3%), materials science and engineering (7.9%), and mechanical engineering (7.0%).

Table 1: Predicted Correlations, Observed Correlations, and Construct Validity Effect Sizes

	Predicted	Observed	
Construct		Disattenuated	Raw
Big Five Facets			
Anger proneness (N)	-.10	-0.23	-0.22
Deliberation (C)	.00	0.28	0.26
Empathy (A)	.35	0.40	0.36
Frankness (E)	.10	0.05	0.05
Trust (A)	.25	0.21	0.19
Honesty-Humility	.20	0.44	0.39
Need for Power	-.25	-0.13	-0.09
Intellectual Humility	.20	0.33	0.31
Justice Sensitivity (Ben.)	.35	0.12	0.11

Justice Sensitivity (Obs.)	.30	0.12	0.11
Need for Closure	-.15	-0.08	-0.07
SDO	-.30	-0.16	-0.15
r_{alerting}		.77	.75
p		.002	.003
r_{contrast} (95% CI)		.55 [.44, .64]	.49 [.38, .59]
p		<.0001	<.0001

Measures

Situational Judgment Test. The SJT was created as described earlier. A bottom-up analysis of skills was used to generate five themes. Subteams were tasked with actual scenario and question development. In cases of scenario ambiguity, a new subteam was formed to make second-round revisions.

Pilot Test. A pilot test was conducted with six exemplary individuals to refine the test before deployment. We recruited nominations of professional engineers, who were university alumni, from faculty in the biomedical engineering programs. Faculty were asked to nominate alumni who had “a record of interpersonal excellence.” Twenty nominated individuals were invited to take the pilot test. Six individuals responded; five made full submissions and one made a partial submission.

The results were analyzed by examining a one-way tabulation of answer frequency. If at least four participants selected the same answer, that answer was confirmed as optimal. To revise cases with no consensus, we consulted with Altus Assessment psychometricians, following which we did some combination of (a) revising or pruning similar response options, (b) adding text to disambiguate situations, (c) rephrasing the optimal answer to reduce ambiguity, and (d) revising the scoring rubric by assigning partial credit to the second-best answers. Four teaching assistants from a second-year course worked with the lead researcher on these tasks. Only two questions were set up for partial credit, with all other items having a credit vs. no-credit scheme.

The SJT consisted of five blocks with one scenario per block. The appendix contains the scenarios. The number of overall questions and the number of yes/no questions (vs. multiple choice) varied by block because of discretionary decisions made by the working groups mentioned earlier.

The diversity block contained a scenario about creating teams with similar-GPA students grouped together, followed by two multiple-choice items and one yes/no item. The psychological-safety block contained a scenario about a team member hesitant about the relevance of his contribution, followed by four multiple-choice items. The inclusion-related block contained a division-of-labor scenario about a team member uncertain about their preferred task, followed by two multiple-choice questions. The supportive-feedback block contained a scenario about a team member who fails to attend meetings or execute tasks, followed by three multiple-choice items. The justice-related block contained a scenario about a team with one

overbearing member and one member who feels unheard, followed by five multiple-choice questions. All items (except yes/no items) had five response options.

Cronbach's alpha—reported for transparency—was .60. This low value should be interpreted in light of the consistency vs. validity tradeoff [39]. When researchers create a scale, they can ensure strong internal consistency by prioritizing item communalities, i.e., creating items focused on just one sub-construct. In such cases, the internal consistency is very high, but the validity is compromised. In broadening our instrument's scope, we prioritized validity over internal consistency. The TEE-SJT items should be considered part of the same network rather than indicators of a single latent factor [44]. Cronbach's alpha was also unsuited to the TEE-SJT because items were neither tau-equivalent nor congeneric.

The TEE-SJT is available on OSF at <https://osf.io/xm2yj> and the scoring guide can be obtained by emailing the author.

Criterion Variables. The 11 criterion variables were selected by the lead author because he was the only researcher with extensive knowledge of scales in personality and social psychology. These criterion variables were measured with extant validated scales. Some scales are long, e.g., 22 items, because factor analyses by the scale creators indicated that (a) there was a hierarchical factor structure, necessitating multiple items per lower-order factor, or (b) this number was necessary for sufficient precision.

Anger proneness was measured with a 10-item subscale ($\alpha = .88$) of the Faceted Inventory of the Five Factor Model (FI-FFM) [40]. Anger proneness is the dispositional tendency to be verbally aggressive, distrustful, and cynical. We hypothesized that SJT score would be negatively associated with anger proneness.

Deliberation was measured with a 9-item subscale ($\alpha = .84$) of the FI-FFM [40]. It is the tendency to think before acting. People who engage in greater deliberation do not necessarily have prosocial values. We predicted that SJT score would be uncorrelated with deliberation.

Empathy was measured with a 10-item subscale ($\alpha = .80$) of the FI-FFM [40]. It is the tendency to have greater consideration for others' needs and emotions. We predicted a moderate positive association between SJT score and empathy.

Frankness was measured with an 8-item subscale ($\alpha = .83$) of the FI-FFM [40]. It is the tendency to speak one's mind and avoid self-censorship. In cooperative work, uninhibited people may bring up uncomfortable issues but may occasionally be too blunt. We predicted a weak positive association with frankness.

Trust was measured with the 11-item subscale ($\alpha = .86$) of the FI-FFM [40]. It is the tendency to believe in others' goodness and avoid cynicism. We predicted a moderate positive association with trust.

Honesty-Humility was measured with a 16-item subscale ($\alpha = .78$) of the HEXACO-PI-R [41]. It is best construed as the absence of dispositional selfishness [42]. Low levels are associated with greed, immodesty, and unfairness [43]. We hypothesized a modest negative association with honesty-humility.

Intellectual Humility. Participants completed the Comprehensive Intellectual Humility Scale CIHS; [44], a 22-item scale that relies on a four-factor conception of intellectual humility ($\alpha = .86$). Intellectual humility refers to a disposition to view one's own cognitions and beliefs as fallible and to be less dogmatic about one's beliefs [45]. We therefore predicted that intellectual humility would be positively associated with the SJT.

Justice sensitivity (benefactor; $\alpha = .86$) and justice sensitivity (observer; $\alpha = .69$) were each measured with two-item scales [46]. Benefactor sensitivity indicates negative reactions to being unfairly over-rewarded. Observer sensitivity indicates negative reactions when observing unfairness. We predicted SJT score would be positively associated with both traits.

Need for closure was measured with the 15-item version ($\alpha = .80$) [47]. People with a strong need for closure prefer epistemic certainty and dislike ambiguity. Groups high in this need are less creative [48] and prefer task orientation over social-emotional aspects of group work [49]. Given that this trait entails a lack of egalitarianism and little intellectual generosity, we hypothesized a weak negative association with SJT score.

Need for power was measured with seven items from the Multi-Motive Grid [50], a validated hybrid motive test. Need for power is an implicit motive, an unconscious need that orients behavior and directs emotional experiences [51]. Given that it involves a preference for influencing rather than being influenced, we expected a modest negative correlation between SJT and need for power.

Social dominance orientation (SDO) was measured with the 8-item SDO7 scale ($\alpha = .80$) [52]. SDO represents the belief that some groups are intrinsically superior and deserving of higher social position. Low SDO should be associated with behavior that engenders inclusion and justice. We hypothesized that SJT score would have a negative association with SDO.

Results

For initial validation a two-parameter item-response analysis was conducted to verify that a single latent factor loaded positively on all items [53], [54]. A partial-credit model accommodated partial-credit answers [55]. The discrimination parameter was reviewed for positive values regardless of statistical significance. One item was rescored due to a negative loading. All other items performed as predicted.

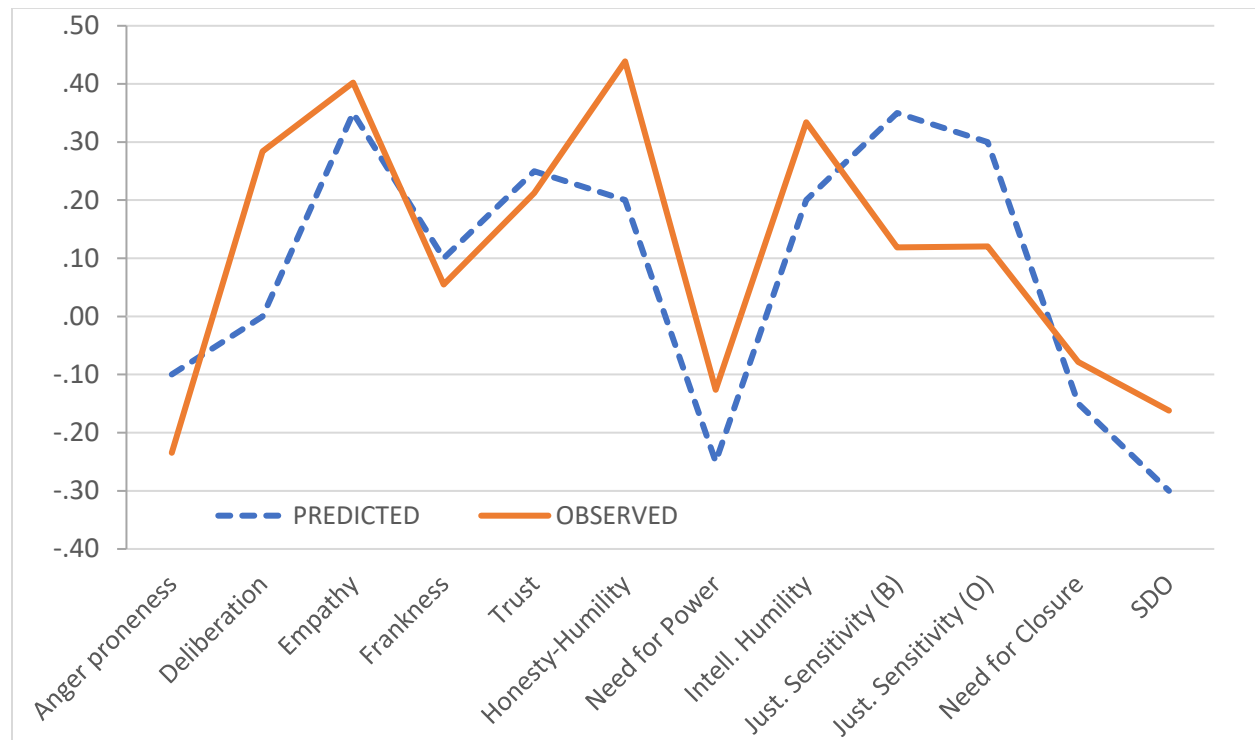
Means, standard deviations, and zero-order correlations of all variables appear in the appendix (Table A1). Most criterion-related variable inter-correlations were low, indicating distinct contributions (avoiding strong intercorrelation suggesting one underlying latent variable).

Table 1 displays hypothesized, observed correlations, and construct validity effect sizes. The disattenuated correlations are of primary concern but the raw correlations are included for reference. Overall, there was congruence between predicted correlations and observed correlations (Figure 1).

One discrepancy was that deliberation showed an observed correlation of .28 versus a predicted nil value, indicating high SJT scorers are deliberative. Similarly, honesty-humility's observed

value of .44 exceeded the predicted .20, showing high scorers lack selfishness. Justice sensitivity (benefactor and observer) showed weaker correlations than predicted ($\sim .10$).

Figure 1: Profiles of Predicted and Disattenuated Observed Correlations between SJT and Criterion Variables



Effect sizes indicated good construct validity with an r_{alerting} value of .77, surpassing the hypothesized cutoff. However, r_{contrast} fell short of the cutoff at .55, 95% CI [.44, .64], its confidence interval straddled the cutoff value. The effect size was reasonably large and unlikely to be obtained by chance. Results indicate jointly strong convergent and discriminant validity (Figure 1).

Discussion

Engineering projects can hinge on interpersonal and intra-team relationships. Industry experts have affirmed teamwork and communication sometimes outweigh specialized technical topics [4]. However, few validated instruments exist for measuring interpersonal skills in engineering programs. The TEE-SJT addresses this gap, measuring skills promoting five salutary team outcomes: justice, psychological safety, inclusion, supportive feedback, and diversity appreciation. The SJT demonstrates good construct validity via simple and advanced effect sizes.

In addition, the expected-observed discrepancies point to some strengths. The deliberation-SJT correlation of .28 exceeded the hypothesized value of zero, but this finding suggests deliberative people engage in cognitive reflection rather than impetuous decisions, possibly avoiding dismissing ideas based on superficial considerations. Thus, a positive correlation is desirable.

Propitiously, the honesty-humility correlation (.44 vs. .20) also exceeded predictions. HEXACO proponents have postulated that honesty-humility is the most consequential workplace trait because low-H-H individuals engage in deceit, exploitation, and counterproductive behavior [43], [56]. It has even been called a "nightmare trait" in leaders [57]. High SJT scores thus indicate unlikely exploitation.

Concerningly, observed justice sensitivity correlations ($\sim .10$) lagged behind predictions. The use of two-item scales may partially explain this weakness. Moreover, high SJT scorers may promote justice through high honesty-humility, deliberation, and frankness.

Unfortunately, our hypothesis of a r_{contrast} value above .60 was not strictly supported. However, published studies consistently show r_{contrast} weaker than r_{alerting} by magnitudes ranging from 0.15 to 0.50, with point estimates sometimes below .50 [35], [36], [37]. Thus, we believe this lower value does not pose validity problems.

Two other teamwork SJTs exist. The Teamwork Situational Judgment Test (Teamwork SJT), published after our study commenced [58], focuses on mission analysis, goal specification, strategy formulation, monitoring, coordination, conflict management, and motivation. With its focus on managerial efficiency, it only modestly overlaps the TEE-SJT. It may complement TEE-SJT for advanced projects such as capstone projects.

Second, a German-language SJT (SJT-TW) became available in English in 2020 [59], [60]. It covers seven team-relevant behavior categories including cooperation, planning, and communication. This SJT is less management-oriented and more relationship-oriented. It also has construct validity, with a .25 correlation ($p = .009$) between SJT score and supervisor contextual performance ratings, with no significant task-related performance association. The SJT-TW may hold promise for engineering education but it uses workplace scenarios rather than academic scenarios. Moreover, the TEE-SJT's scenarios are richer with less ambiguity, e.g., characters have names and dialogue. Nevertheless, combined administration of the SJT-TW and TEE-SJT may prove worthwhile for assessment.

A promising area for future research is to examine how well the intra-team distribution of TEE-SJT scores predicts team performance. On the one hand, if high performers (i.e., those with high TEE-SJT scores) are clustered together, the team may demonstrate greater collective intelligence in problem solving. On the other hand, the mere presence of a few high scorers may be sufficient to resolve disputes and elevate team performance. The presence of one extremely low scorer may cause exploitation and manipulation.

The chief limitation of the current work is that criterion validity was not assessed. In an ongoing project, we are assessing whether students with high TEE-SJT scores earn better ratings from peers and teachers in successive years. The predictive validity of the TEE-SJT could also be evaluated by a firm with a large employee count by administering the TEE-SJT for new hires and assessing its correlation with supervisor ratings five and ten years later. We do not have the resources to conduct such tests. Nevertheless, there is indirect evidence for external validity inasmuch as we found strong associations with honesty-humility, which is known to affect real workplace behavior.

A secondary limitation was that the blueprinting was not done by a team of experts. This was primarily due to resource limitations. It may have led to a test in which construct underrepresentation and construct-irrelevant variance were not fully managed.

The present study successfully developed and validated a situational judgment test for assessing interpersonal team skills that contribute to psychological safety, inclusion, and justice outcomes. The robust construct validity evidence, demonstrated through the QCV procedure, positions this instrument as a valuable addition to the assessment literature. Moreover, the questions are also suitable for in-class discussions about teamwork skills.

Future research should examine the test's predictive validity for actual team outcomes, explore its utility across cultural and disciplinary contexts, and use experimental methods to determine whether interventions can improve performance. Additionally, longitudinal studies examining the stability and malleability of interpersonal team skills would provide valuable insights for understanding how these competencies develop and can be enhanced over time.

Funding

This work was supported by National Science Foundation under Grant No. 1730262.

References

- [1] T. J. Nokes-Malach, J. E. Richey, and S. Gadgil, "When is it better to learn together? Insights from research on collaborative learning," *Educational Psychology Review*, vol. 27, no. 4, pp. 645–656, 2015.
- [2] J. E. Richey, T. J. Nokes-Malach, and K. Cohen, "Collaboration facilitates abstract category learning," *Mem Cogn*, vol. 46, no. 5, pp. 685–698, Jul. 2018, doi: 10.3758/s13421-018-0795-7.
- [3] D. D. Galbraith and F. L. Webb, "Teams That Work: Preparing Student Teams for the Workplace," *American Journal of Business Education*, vol. 6, no. 2, pp. 223–234, 2013.
- [4] S. Stukes, E. Bethke, M. J. O'Donnell, and J. R. Amos, "Biomedical specialized masters programs: What are they and how do they meet the needs of employers?," *Biomed Eng Education*, vol. 3, no. 2, pp. 291–301, Jul. 2023, doi: 10.1007/s43683-023-00115-5.
- [5] J. Campbell, H. M. Caruso, and L. Klotz, "Exploring an Intervention to Increase Psychological Safety on Student Engineering Design Teams," presented at the 2024 ASEE Annual Conference & Exposition, Jun. 2024. Accessed: May 30, 2025. [Online]. Available: <https://peer.asee.org/exploring-an-intervention-to-increase-psychological-safety-on-student-engineering-design-teams>
- [6] S. Scarpinella, C. Cole, S. Ritter, S. Mohammad, K. Jablokow, and S. Miller, "Can we get an intervention, please? The utility of teaming interventions on engineering design student psychological safety," *Journal of Mechanical Design*, vol. 145, no. 082301, May 2023, doi: 10.1115/1.4062395.
- [7] C. A. Pedregon, R. L. Farley, A. Davis, J. M. Wood, and R. D. Clark, "Social desirability, personality questionnaires, and the 'better than average' effect," *Personality and Individual Differences*, vol. 52, no. 2, pp. 213–217, Jan. 2012, doi: 10.1016/j.paid.2011.10.022.
- [8] M. J. Zickar and C. Robie, "Modeling faking good on personality items: An item-level analysis," *Journal of Applied Psychology*, vol. 84, no. 4, pp. 551–563, Aug. 1999, doi: 10.1037/0021-9010.84.4.551.

- [9] A. L. Nichols and J. K. Maner, "The good-subject effect: Investigating participant demand characteristics," *The Journal of General Psychology*, vol. 135, no. 2, pp. 151–166, Apr. 2008, doi: 10.3200/GENP.135.2.151-166.
- [10] G. Pennycook, R. M. Ross, D. J. Koehler, and J. A. Fugelsang, "Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence," *Psychonomic Bulletin & Review*, vol. 24, no. 6, pp. 1774–1784, Dec. 2017, doi: 10.3758/s13423-017-1242-7.
- [11] J. Cumming, C. Woodcock, S. J. Cooley, M. J. G. Holland, and V. E. Burns, "Development and validation of the groupwork skills questionnaire (GSQ) for higher education," *Assessment & Evaluation in Higher Education*, vol. 40, no. 7, pp. 988–1001, Oct. 2015, doi: 10.1080/02602938.2014.957642.
- [12] H. Gehlbach and C. D. Robinson, "Mitigating Illusory Results through Preregistration in Education," *Journal of Research on Educational Effectiveness*, vol. 11, no. 2, pp. 296–315, Apr. 2018, doi: 10.1080/19345747.2017.1387950.
- [13] B. A. Nosek, C. R. Ebersole, A. C. DeHaven, and D. T. Mellor, "The preregistration revolution," *Proceedings of the National Academy of Sciences*, vol. 115, no. 11, pp. 2600–2606, Mar. 2018, doi: 10.1073/pnas.1708274114.
- [14] D. Westen and R. Rosenthal, "Quantifying construct validity: Two simple measures.,", *Journal of Personality and Social Psychology*, vol. 84, no. 3, pp. 608–618, Mar. 2003, doi: 10.1037/0022-3514.84.3.608.
- [15] J. C. Flanagan, "The critical incident technique.,", *Psychological Bulletin*, vol. 51, no. 4, pp. 327–358, 1954, doi: 10.1037/h0061470.
- [16] M. S. Christian, B. D. Edwards, and J. C. Bradley, "Situational judgment tests: Constructs assessed and a meta-analysis of their criterion-related validities," *Personnel Psychology*, vol. 63, no. 1, pp. 83–117, Mar. 2010, doi: 10.1111/j.1744-6570.2009.01163.x.
- [17] J. Corstjens, F. Lievens, and S. Krumm, "Situational judgment tests for selection," in *The Wiley Blackwell Handbook of the Psychology of Recruitment, Selection and Employee Retention*, Hoboken, NJ: Wiley-Blackwell, 2017, pp. 226–246.
- [18] F. Lievens and S. J. Motowidlo, "Situational Judgment Tests: From Measures of Situational Judgment to Measures of General Domain Knowledge," *Industrial and Organizational Psychology*, vol. 9, no. 1, pp. 3–22, Mar. 2016, doi: 10.1017/iop.2015.71.
- [19] F. Lievens and P. R. Sackett, "The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance," *Journal of Applied Psychology*, vol. 97, no. 2, pp. 460–468, 2012, doi: 10.1037/a0025741.
- [20] J. Koehler *et al.*, "Development of Situational Judgment Tests to Educate the Whole Engineer and Assess Complex Character-Based Ethical Decision-Making," in *2025 IEEE Frontiers in Education Conference (FIE)*, Nov. 2025, pp. 1–8. doi: 10.1109/FIE63693.2025.11328311.
- [21] O. Pierrakos *et al.*, "An Exploratory Investigation on the Use of Situational Judgement Tests to Cultivate Character and Evaluate Character in Engineering Learning Environments," in *2025 IEEE Frontiers in Education Conference (FIE)*, Nov. 2025, pp. 1–9. doi: 10.1109/FIE63693.2025.11328159.
- [22] B. K. Jesiek, S. E. Woo, S. Parrigon, and C. M. Porter, "Development of a situational judgment test for global engineering competency," *Journal of Engineering Education*, vol. 109, no. 3, pp. 470–490, 2020, doi: 10.1002/jee.20325.

- [23] K. A. Davis, B. K. Jesiek, and D. B. Knight, “Exploring scenario-based assessment of students’ global engineering competency: Building evidence of validity of a China-based situational judgment test,” *Journal of Engineering Education*, vol. 112, no. 4, pp. 1032–1055, 2023, doi: 10.1002/jee.20552.
- [24] S. J. Motowidlo, K. Ghosh, A. M. Mendoza, A. E. Buchanan, and M. N. Lerma, “A context-independent situational judgment test to measure prosocial implicit trait policy,” *Human Performance*, vol. 29, no. 4, pp. 331–346, Aug. 2016, doi: 10.1080/08959285.2016.1165227.
- [25] S. Krumm, A. M. Thiel, N. Reznik, J.-P. Freudenstein, P. Schäpers, and P. Mussel, “Creating a psychological test in a few seconds: Can ChatGPT develop a psychometrically sound situational judgment test?,” *European Journal of Psychological Assessment*, 2024, doi: 10.1027/1015-5759/a000878.
- [26] M. Adamovic, “Organizational justice research: A review, synthesis, and research agenda,” *European Management Review*, vol. 20, no. 4, pp. 762–782, 2023, doi: 10.1111/emre.12564.
- [27] J. A. Colquitt *et al.*, “Justice at the millennium, a decade later: A meta-analytic test of social exchange and affect-based perspectives,” *Journal of Applied Psychology*, vol. 98, no. 2, pp. 199–236, 2013, doi: 10.1037/a0031757.
- [28] A. C. Edmondson and D. P. Bransby, “Psychological Safety Comes of Age: Observed Themes in an Established Literature,” *Annual Review of Organizational Psychology and Organizational Behavior*, vol. 10, no. Volume 10, 2023, pp. 55–78, Jan. 2023, doi: 10.1146/annurev-orgpsych-120920-055217.
- [29] F. Anseel, A. S. Beatty, W. Shen, F. Lievens, and P. R. Sackett, “How Are We Doing After 30 Years? A Meta-Analytic Review of the Antecedents and Outcomes of Feedback-Seeking Behavior,” *Journal of Management*, vol. 41, no. 1, pp. 318–348, Jan. 2015, doi: 10.1177/0149206313484521.
- [30] S. J. Ashford and L. L. Cummings, “Feedback as an individual resource: Personal strategies of creating information,” *Organizational Behavior and Human Performance*, vol. 32, no. 3, pp. 370–398, Dec. 1983, doi: 10.1016/0030-5073(83)90156-3.
- [31] P. Marciano, *Carrots and stick don’t work: Build a culture of employee engagement with the principles of respect*. New York: McGraw-Hill, 2010.
- [32] D. S. Yeager *et al.*, “Breaking the cycle of mistrust: Wise interventions to provide critical feedback across the racial divide,” *Journal of Experimental Psychology: General*, vol. 143, no. 2, pp. 804–824, 2014, doi: 10.1037/a0033906.
- [33] J. F. Rauthmann *et al.*, “The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics,” *Journal of Personality and Social Psychology*, vol. 107, no. 4, pp. 677–718, 2014, doi: 10.1037/a0037250.
- [34] R. M. Furr and S. Heuckeroth, “The ‘Quantifying Construct Validity’ Procedure: Its Role, Value, Interpretations, and Computation,” *Assessment*, vol. 26, no. 4, pp. 555–566, Jun. 2019, doi: 10.1177/1073191118820638.
- [35] G. Bombel, J. L. Mihura, and G. J. Meyer, “An examination of the construct validity of the Rorschach Mutuality of Autonomy (MOA) scale,” *Journal of Personality Assessment*, vol. 91, no. 3, pp. 227–237, May 2009, doi: 10.1080/00223890902794267.
- [36] S. Griffin, D. Lynam, and D. Samuel, “Dimensional Conceptualizations of Impulsivity,” *Personality Disorders: Theory, Research, and Treatment*, vol. 9, May 2017, doi: 10.1037/per0000253.

- [37] J. D. Miller, D. R. Lynam, and W. K. Campbell, "Measures of narcissism and their relations to DSM-5 pathological traits: A critical reappraisal," *Assessment*, vol. 23, no. 1, pp. 3–9, Feb. 2016, doi: 10.1177/1073191114522909.
- [38] F. D. Schönbrodt and M. Perugini, "At what sample size do correlations stabilize?," *Journal of Research in Personality*, vol. 47, no. 5, pp. 609–612, Oct. 2013, doi: 10.1016/j.jrp.2013.05.009.
- [39] J. D. W. Clifton, "Managing validity versus reliability trade-offs in scale-building decisions," *Psychological Methods*, vol. 25, no. 3, pp. 259–270, 2020, doi: 10.1037/met0000236.
- [40] D. Watson, E. Nus, and K. D. Wu, "Development and validation of the Faceted Inventory of the Five-Factor Model (FI-FFM)," *Assessment*, vol. 26, no. 1, pp. 17–44, 2019, doi: 10.1177/1073191117711022.
- [41] K. Lee and M. C. Ashton, "Psychometric properties of the HEXACO-100," *Assessment*, vol. 25, no. 5, pp. 543–556, Jul. 2018, doi: 10.1177/1073191116659134.
- [42] K. J. Diebels, M. R. Leary, and D. Chon, "Individual differences in selfishness as a major dimension of personality: A reinterpretation of the sixth personality factor," *Review of General Psychology*, vol. 22, no. 4, pp. 367–376, Dec. 2018, doi: 10.1037/gpr0000155.
- [43] K. Lee and M. C. Ashton, *The H Factor of Personality: Why Some People are Manipulative, Self-Entitled, Materialistic, and Exploitive—And Why It Matters for Everyone*. Wilfrid Laurier Univ. Press, 2013.
- [44] E. J. Krumrei-Mancuso and S. V. Rouse, "The Development and Validation of the Comprehensive Intellectual Humility Scale," *Journal of Personality Assessment*, vol. 98, no. 2, pp. 209–221, Mar. 2016, doi: 10.1080/00223891.2015.1068174.
- [45] M. R. Leary *et al.*, "Cognitive and Interpersonal Features of Intellectual Humility," *Pers Soc Psychol Bull*, vol. 43, no. 6, pp. 793–813, Jun. 2017, doi: 10.1177/0146167217697695.
- [46] A. Baumert *et al.*, "Measuring four perspectives of justice sensitivity with two items each," *Journal of Personality Assessment*, vol. 96, no. 3, pp. 380–390, May 2014, doi: 10.1080/00223891.2013.836526.
- [47] A. Roets and A. Van Hiel, "Item selection and validation of a brief, 15-item version of the Need for Closure Scale," *Personality and Individual Differences*, vol. 50, no. 1, pp. 90–94, Jan. 2011, doi: 10.1016/j.paid.2010.09.004.
- [48] A. Chirumbolo, S. Livi, L. Mannetti, A. Pierro, and A. W. Kruglanski, "Effects of need for closure on creativity in small group interactions," *European Journal of Personality*, vol. 18, no. 4, pp. 265–278, 2004, doi: 10.1002/per.518.
- [49] E. De Grada, A. W. Kruglanski, L. Mannetti, and A. Pierro, "Motivated Cognition and Group Interaction: Need for Closure Affects the Contents and Processes of Collective Negotiations," *Journal of Experimental Social Psychology*, vol. 35, no. 4, pp. 346–365, Jul. 1999, doi: 10.1006/jesp.1999.1376.
- [50] K. Sokolowski, H. D. Schmalt, T. A. Langens, and R. M. Puca, "Assessing achievement, affiliation, and power motives all at once: the Multi-Motive Grid (MMG)," *Journal of personality assessment*, vol. 74, no. 1, pp. 126–45, Feb. 2000, doi: 10.1207/S15327752JPA740109.
- [51] D. C. McClelland, R. Koestner, and J. Weinberger, "How do self-attributed and implicit motives differ?," *Psychological Review*, vol. 96, no. 4, pp. 690–702, 1989, doi: 10.1037/0033-295X.96.4.690.

- [52] A. K. Ho *et al.*, “The nature of social dominance orientation: Theorizing and measuring preferences for intergroup inequality using the new SDO₇ scale,” *J Pers Soc Psychol*, vol. 109, no. 6, pp. 1003–1028, Dec. 2015, doi: 10.1037/pspi0000033.
- [53] F. M. Lord, *Applications of Item Response Theory to Practical Testing Problems*. Mahwah, NJ: Lawrence Erlbaum Associates, 1980.
- [54] T. Raykov, “Item Response Theory and Modeling with Stata,” *Measurement: Interdisciplinary Research and Perspectives*, vol. 21, no. 2, pp. 117–128, Apr. 2023, doi: 10.1080/15366367.2022.2133528.
- [55] E. Muraki, “A generalized partial credit model: Application of an EM algorithm,” *Applied Psychological Measurement*, vol. 16, no. 2, pp. 159–176, Jun. 1992, doi: 10.1177/014662169201600206.
- [56] G. Hodson, A. Book, B. A. Visser, A. A. Volk, M. C. Ashton, and K. Lee, “Is the Dark Triad common factor distinct from low Honesty-Humility?,” *Journal of Research in Personality*, vol. 73, pp. 123–129, Apr. 2018, doi: 10.1016/j.jrp.2017.11.012.
- [57] R. E. de Vries, “Three Nightmare Traits in Leaders,” *Front. Psychol.*, vol. 9, Jun. 2018, doi: 10.3389/fpsyg.2018.00871.
- [58] G. E. Littlepage, D. S. Jr. Steffensen, J. Yi, and H. Hembree, “Validation of the Teamwork Situational Judgment Test.,” *Group Dynamics: Theory, Research, and Practice*, vol. 26, no. 4, pp. 322–340, Dec. 2022, doi: 10.1037/gdn0000157.
- [59] J.-P. Freudenstein, N. Remmert, N. Reznik, and S. Krumm, “English Translation of the SJT-TW,” *ZIS - The Collection of Items and Scales for the Social Sciences*, 2020, doi: 10.6102/ZIS285.
- [60] T. Gatzka and J. Volmer, “Situational Judgment Test für Teamarbeit (SJT-TA),” *Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)*, 2017, doi: 10.6102/ZIS249.

Appendix

Scenarios Presented to Test Takers in the TEE-SJT

Category	Scenario
Diversity	You are one of the TAs in a class where students work on semester-long design projects. The instructor has asked you and two other TAs, Grace and Robert, to divide students into four-person teams. The professor says, “Your job as TAs is to optimize the teams so that students collaborate successfully and learn interpersonal skills.” You are now in a meeting with Grace and Robert without the professor present. Grace says, “I think people with similar GPAs should be grouped together—I have a 3.5 and I want to keep my GPA up by working with similar students. I think other students who have high GPAs feel the same way.”
Psychological Safety	You are on a team of six second-year students: In week 3 of the semester, Peter brings a scientific paper to a team meeting and says, “I think this could be useful for us because it’s on our project topic.” Tremayne asks, “What does the paper say?” Peter is quiet for a moment and then says to the team, “This is embarrassing because I read the paper and still don’t understand it.” Elizabeth says, “Well, we don’t have a lot of time on our hands so I’m not sure spending time on it is worthwhile.”
Inclusion	You’re a student on a seven-person project team. In a recent lecture, the professor announced that project grades will be higher if all members of project teams—not just some—have contributed ideas during team discussions and worked on the final product. TAs will monitor who is contributing. Later, the TA hands out project instructions to people in your team, and everyone starts volunteering for tasks. Eventually, all but one person has volunteered for something; this person seems to be listening to the team discussion but looks confused like they’re struggling somehow. Another student turns to this student who hasn’t spoken and asks, “What can you do?” The student pauses and then replies, “I don’t know.”
Supportive Feedback	You and a teammate Aditya are discussing your class project after a team meeting. Recently, your team noted a lack of communication about work progress. At that meeting, all members agreed to use a shared Google sheet to log accomplishments and to-do items. You both noted that Sam, another teammate, hasn’t been contributing to the shared Google sheet. You and Aditya are frustrated, and you’re pretty sure the rest of the team is also frustrated. While Sam has been showing up to the meetings, he hasn’t edited the Google sheet in six weeks and hasn’t reported any individual progress at the group meetings for at least three weeks.
Justice	You are a second-year student on a project team. Your team is being overseen by a teaching assistant. All seven team members talk at least a

little bit during meetings. However, since week 2, Jordan has talked much more than other members of the team during each meeting and often talks over others. Jordan is an upbeat, positive person; you and Jordan are friends and hang out together outside of meetings.

In week 6, Jordan did work beforehand that was supposed to have been done by the whole team together. It is now week 7. Jordan again did more work beforehand, and just talked over Samir who was trying to say something.

Samir says, "I feel like my ideas aren't being heard as much as other people on the team. I think everyone should be able to do their own work instead of somebody taking another person's work."

Jordan replies, "I feel like I took a lot of initiative, which is a good thing. Like we're all overachievers here, and I think everyone could step up."

Samir says, "Well, it's a team project so everyone's ideas should be considered equally."

Jordan says, "Yes, but we're not doing the level of work that would get us an A."

Full Situational Judgment Test

Starting on the next page, you'll be shown scenarios with you playing a role in them.

Please imagine that you're in the scenario and answer the questions based on what you would do.

Section I

You are one of the teaching assistants (TAs) in a class where students work on semester-long design projects. The instructor has asked you and two other TAs, Grace and Robert, to divide students into four-person teams. The professor says, “Your job as TAs is to optimize the teams so that students collaborate successfully and learn interpersonal skills.”

You are now in a meeting with Grace and Robert without the professor present. Grace says, “I think people with similar GPAs should be grouped together—I have a 3.5 and I want to keep my GPA up by working with similar students. I think other students who have high GPAs feel the same way.”

1. Assume that you respond to Grace. Given these possible responses, what would you say?
 - o I’m not sure this system would work. People want everyone to put in equal effort and GPA doesn’t always align with effort. (1)
 - o I’m not sure this system would work. The professor wants students to collaborate. GPA doesn’t indicate that you’re good at collaborating. (2)
 - o I think this system could work. People want to work with people with higher GPAs so those students can take on most of the workload. This system will prevent that from happening. (3)
 - o I think this system could work. When I’m on teams, I want to work only with the best people too. I don’t care who they are, but I don’t want dead weight on the team. (4)
 - o I’m not sure this system would work. Everyone would like at least one A student on their team. (5)
2. After you respond, Robert says, “Well, I think there should be varying GPA within a team because that will help. People at each level of GPA have unique skills and perspectives that can contribute to groupwork.”
 - o I also think a mix of GPAs is better for groupwork. (1)
 - o I think we really need to focus on diversity of experience rather than some configuration based on GPA. This way each person brings unique knowledge to the group. (2)

- o GPAs don't show how well people will work together in a team, so I also don't think we'll gain anything by paying attention to GPA. (3)
- o Students who have the very highest GPAs often work very hard and try to outdo people. It would be bad to put them all on the same team because the team members won't collaborate. (4)
- o Having a mix of strong and weak students in a team is good, because the strong students get to explain content to weaker students, which helps everyone. (5)

3. Privately, do you feel that organizing students according to GPA is a good idea?

- o Yes (1)
- o No (2)

Section II

You are on a team of six second-year students: In week 3 of the semester, Peter brings a scientific paper to a team meeting and says, "I think this could be useful for us because it's on our project topic."

Tremayne asks, "What does the paper say?"

Peter is quiet for a moment and then says to the team, "This is embarrassing because I read the paper and still don't understand it."

Elizabeth says, "Well, we don't have a lot of time on our hands so I'm not sure spending time on it is worthwhile."

4. What would you do?

- o You say to the whole team, "Yeah, if we don't really know what it says, it's probably useless." (1)
- o You say to Elizabeth, "Well, this paper could tell us if we're heading in the wrong direction, which is important, so let's look at the beginning of the paper." (2)
- o You privately say to Peter afterwards, "Thanks for looking up that paper. Can you email it to me so I can read it?" (3)
- o You wait for others to respond first and then decide what to say. (4)

5. In the next few weeks, you hear privately from three teammates. They think the project is going to fail given its current direction but they're afraid of speaking up in team meetings. They

say that they're silent partially because they have only some background knowledge and they're not confident about their opinions.

At the next meeting, the facilitator asks everyone to talk about team dynamics and selects you to speak first. What would you say?

- o We have a problem of work being split up unevenly. Going forward, I think everyone should look at papers that might be useful. (1)
- o Some of us are naturally assertive and others are naturally introverted. Going forward, I think we should less hard on shy people who are taking risks by speaking up. (2)
- o We have a lack of unity in the team and we're splitting into factions. Going forward, we need to think of ourselves as one team with a common goal. (3)
- o In this project, we're working under uncertainty. So, in meetings, it's going to be helpful for each of us to openly express opinions even when we're unsure that those opinions are valid. (4)
- o We're working on a project for a grade, so we need to know about all the mistakes we could make. Going forward, each of us should be free to express criticism of any idea at any point. (5)

6. Assume that instead of being a team member, you are the teaching assistant that supervises this team. You know about everything that has occurred thus far. You plan to attend the next team meeting. As the TA, which of these things would you plan to do?

- o You'll look for an instance where someone's comment is ignored, and then say, "Hold on. What about that idea?" (1)
- o You'll listen attentively and decide whether to meet with individual students after the meeting. (2)
- o You'll say, "I've heard there have been issues with the group. Let's keep this in mind at this meeting." (3)
- o You'll say, "In the future, you need to split up the work based on strengths and weaknesses." (4)

7. You and the rest of the team later learn from one team member considers himself to be extremely introverted. Which of these things would you suggest as a team policy?

- o Create a Google sheet that's open during and shortly after meetings so people can comment on work in progress during the meeting without having to talk. (1)

- o Create a check-in at the beginning and end of each meeting, where each person on the team states their opinion on the state of the project and their individual goals. (2)
- o Meet as a team without that member and discuss how to evaluate his contributions. (3)
- o You personally meet with that team member and ask him what system would work well for him. (4)

Section III

You're a student on a seven-person project team. In a recent lecture, the professor announced that project grades will be higher if all members of project teams—not just some—have contributed ideas during team discussions and worked on the final product. TAs will monitor who is contributing.

Later, the TA hands out project instructions to people in your team, and everyone starts volunteering for tasks. Eventually, all but one person has volunteered for something; this person seems to be listening to the team discussion but looks confused like they're struggling somehow. Another student turns to this student who hasn't spoken and asks, "What can you do?" The student pauses and then replies, "I don't know."

8. Which of these would you do?

- o You say to this student: "What are you interested in?" (1)
- o You say to this student: "Do you think you could do this task?" naming a task that hasn't been taken (2)
- o You say to this student: "Could you help a teammate with a task that they volunteered for?" (3)
- o You say to this student, "Well, you need to do something." (4)
- o You wait for another teammate or the TA to do something. (5)

9. You have a dinner out for the team in the second week of the semester. The student who was quiet doesn't show up even though you know his schedule was open. What would you guess as the reason for their absence?

- o He's a shy person and probably doesn't enjoy being in large groups unless he knows everyone in the group well. (1)
- o He is probably catching up on work in other classes in order to stay on top of his workload. (2)

- o He feels like he has to pretend to be interested in this task that someone else gave him, so he probably doesn't enjoy being in the team. (3)
- o He probably forgot to put this dinner on his calendar. (4)
- o He only wants to have a professional relationship with the team (5)

Section IV

You and a teammate Aditya are discussing your class project after a team meeting. Recently, your team noted a lack of communication about work progress. At that meeting, all members agreed to use a shared Google sheet to log accomplishments and to-do items.

You both noted that Sam, another teammate, hasn't been contributing to the shared Google sheet. You and Aditya are frustrated, and you're pretty sure the rest of the team is also frustrated. While Sam has been showing up to the meetings, he hasn't edited the Google sheet in six weeks and hasn't reported any individual progress at the group meetings for at least three weeks.

10. What would you do at this point?

- o Wait and see if the behavior corrects itself (1)
- o Approach Sam about the situation (2)
- o Approach the team, including Sam, about the situation (3)
- o Approach other members of the team about the situation (4)
- o Approach the TA or facilitator about the situation (5)

11. Assume that you decide to give feedback to Sam. When and where would you talk to Sam?

- o You talk to Sam in person as soon as you can (1)
- o You talk to Sam through online chat as soon as you can (2)
- o You talk to Sam at the end of the semester (3)
- o You write information on Sam's end-of-semester peer evaluation, which he will receive (4)

12. If you were to talk to Sam in person, which of these things would you say to Sam?

- o I'm frustrated with the amount of work I have had to do because you don't follow through on your contributions to the project. Because you don't update the Google sheet, I feel like you aren't doing anything at all. (1)
- o Thanks for talking with me. You seem comfortable with the group and talking in meetings, but I noticed that you haven't updated the Google sheet in 6 weeks. Can you start updating it again like the team agreed? (2)
- o Look, I'm concerned you might hurt the team's grade. It feels like you don't care about the group because you don't update the Google sheet. We don't know what you're doing so it's hard to trust you. Can you get your act together? (3)
- o You're a great person and I enjoy working with you but you're not updating the spreadsheet like we agreed we would, which means we don't know what you are doing. I'm sure you're trying your best, but be honest, are you interested in this class? (4)

Section V

You are a second-year student on a project team. Your team is being overseen by a teaching assistant. All seven team members talk at least a little bit during meetings. However, since week 2, Jordan has talked much more than other members of the team during each meeting and often talks over others. Jordan is an upbeat, positive person; you and Jordan are friends and hang out together outside of meetings.

In week 6, Jordan did work beforehand that was supposed to have been done by the whole team together. It is now week 7. Jordan again did more work beforehand, and just talked over Samir who was trying to say something.

Samir says, "I feel like my ideas aren't being heard as much as other people on the team. I think everyone should be able to do their own work instead of somebody taking another person's work." Jordan replies, "I feel like I took a lot of initiative, which is a good thing. Like we're all overachievers here, and I think everyone could step up."

Samir says, "Well, it's a team project so everyone's ideas should be considered equally."

Jordan says, "Yes, but we're not doing the level of work that would get us an A."

13. Assume your goal is to resolve this conflict. What would you do?

- o You say nothing, and let Jordan and Samir keep talking for a while. You expect them to reach a resolution or the teaching assistant to intervene. (1)
- o You say to Jordan and Samir, "I think we can prepare and share work on a Google Doc in the future before meeting. If Jordan think it's not enough for an A, she can suggest edits during the meeting." (2)
- o You say to the team, "I do think Jordan has been doing more work. She has more work to present—that's why she talks more during team meetings, which is fair." (3)

- o You say to the team, “Samir has a point—others might feel like they’re not able to be heard. Maybe we can delegate time more equally so everyone has a chance to talk.” (4)

14. The meeting is over and you’re just talking to Jordan. Everyone else has left. What would you say to Jordan?

- o Since we’re friends, I can be honest—I think sometimes you do other people’s work and you do talk more. It can make us feel like we’re not contributing to the project. (1)
- o Yes, it’s stressful but the semester will be over soon. We should be planning what to do on spring break. (2)
- o I don’t think Samir realizes how much you do for the team. I think you’re doing fine. (3)
- o I want to stay out of this. There’s definitely some tension, and you and Samir need to talk this out. (4)
- o I think you and Samir need to have a conversation and I’d be happy to come and help work it out. (5)

15. Assume that instead of making a statement, you ask Jordan a question. You are just talking to Jordan now and everyone else has left. What would you ask Jordan?

- o What point do think Samir was trying to make? (1)
- o Why do you care what Samir thinks? (2)
- o How do you think the rest of the group feels about the situation? (3)
- o Why don’t you do less work for the next module to even things out and see what happens? (4)

16. You know the professor well and often talk to her during office hours. The professor says, “Someone on your team complained to me about how people on the team are slacking. Do you feel like people are slacking?” Assume you agree with this assessment. What would you say to the professor?

- o I think the team is fine overall. We are getting the work done. (1)
- o People are putting in unequal amounts of work: some are putting in more effort and some are putting in less, which is unfair. (2)

- o Some people are talking over others and doing too much work, so other people feel like they can't contribute, which is unfair. (3)
- o Jordan is taking on too much and gets upset when others don't do the same amount of work. But they can't because Jordan already did it. (4)
- o We're having some issues. Could you give us some more guidance? (5)

17. It is now week 10. There are five weeks left in the semester. Jordan is still being overbearing. The professor considers penalizing overbearing students with a low grade for teamwork and asks for your opinion privately. What would you say?

- o I think that would be fair because your rubric for the class specifies that teamwork is part of the grade. (1)
- o I think that would be fair if you're consistently using the same criteria across all teams. (2)
- o I think that would be unfair because engineering students should be graded on the quality of their technical work alone. (3)
- o I think that would be fair as long as you communicated to people earlier that you might do this if there was no improvement. (4)
- o I think that would be unfair because people would accuse you of being biased. (5)

Table A1

Descriptive Statistics and Zero-Order Correlations Between Key Variables

	M	SD	N	SJT	Ang	Del	Emp	Fra	Tru	H-H	Pow	Int	JSB	JSO	Clos
SJT	20.8	5.6	215												
Anger Proneness	2.3	0.7	211	-.22											
Deliberation	3.7	0.6	211	.26	-.35										
Empathy	3.8	0.5	212	.36	-.31	.39									
Frankness	3.2	0.7	212	.05	.13	-.24	-.12								
Trust	3.3	0.6	211	.19	-.40	.09	.29	-.01							
Honesty-Humility	3.5	0.5	213	.39	-.27	.22	.37	-.02	.27						
Need for Power	2.8	0.6	209	-.09	-.01	-.04	.05	.02	-.12	-.05					
Intel. Humility	3.7	0.5	208	.31	-.40	.20	.33	.11	.24	.39	-.05				
Justice Sens. (B)	3.1	1.1	206	.11	.03	.10	.28	-.21	.04	.14	.03	-.12			
Justice Sens. (O)	3.6	0.9	206	.10	.08	.07	.19	.01	.08	-.05	-.00	-.04	.45		
Need for Closure	4.5	0.8	206	-.07	.30	.12	-.02	-.16	-.32	-.18	-.10	-.30	.16	.17	
SDO	2.6	1.0	207	-.15	.25	-.14	-.28	.05	-.38	-.25	.04	-.25	-.21	-.21	.14

Note. Correlations with absolute values $> .14$ and $.22$ are significant at the $.05$ and $.001$ levels respectively. No correction for attenuation due to unreliability was applied to the correlation matrix