Approaches to Evidencing Intra-Team Equity in Student Collaborative Design Decision-Making Interactions

Dr. Andrew David Moffat, University of Michigan

Andrew Moffat is a Postdoctoral Research Fellow at the University of Michigan, working with the Engineering Education Research Unit and Center for Academic Innovation on an NSF-funded project to assess the effectiveness of Tandem, an in-house software platform designed to support and nurture teamwork skills in undergraduate engineering students. Andrew has a background in education research and evaluation, having previously worked on a project at the University of Leeds, UK, evaluating an institution-wide curriculum transformation initiative. He holds a PhD in Applied Linguistics from the University of Nottingham, UK, prior to the undertaking of which he spent a decade teaching English as a foreign language.

Dr. Robin Fowler, University of Michigan

Robin Fowler is a Technical Communication lecturer and a Engineering Education researcher at the University of Michigan. Her teaching is primarily in team-based engineering courses, and her research focuses on equity in communication and collaboration as well as in group design decision making (judgment) under uncertainty. She is especially interested in how power relationships and rhetorical strategies affect group judgment in engineering design; one goal of this work is to to understand factors that inhibit full participation of students who identify with historically marginalized groups and investigate evidence-based strategies for mitigating these inequities. In addition, she is interested in technology and how specific affordances can change the ways we collaborate, learn, read, and write. Teaching engineering communication allows her to apply this work as she coaches students through collaboration, design thinking, and design communication. She is part of a team of faculty innovators who originated Tandem (tandem.ai.umich.edu), a tool designed to help facilitate equitable and inclusive teamwork environments.

Rebecca L. Matz, University of Michigan

Becky Matz is a Research Scientist on the Research and Analytics team at the Center for Academic Innovation at the University of Michigan. She directs and supports research projects across Academic Innovation's portfolio of educational technologies and online learning experiences. Prior to joining Academic Innovation, she focused on STEM education assessment and research, connecting faculty with data, and developing interdisciplinary activities for introductory chemistry and biology courses at Michigan State University. Becky earned her Ph.D. in Chemistry and M.S. in Educational Studies from the University of Michigan.

Spencer JaQuay, University of California, Irvine

Work in Progress: Approaches to Evidencing Intra-Team Equity in Student Collaborative Design Decision-Making Interactions

Introduction

This work-in-progress paper reports progress on our goal to find a means of identifying and evidencing behavior change in undergraduate engineering student teamwork following an equity-focused intervention. The intervention in question is implemented by Tandem, an in-house software platform that provides students with a space to give frequent feedback regarding their teamwork experience, and then offers customized instructional material to the team to change behaviors and improve the team dynamic [1]. The project on which this paper reports is part of a wider initiative to assess the effectiveness of Tandem.

Major goals driving the development of Tandem include promoting equity in student teamwork and providing future engineering professionals with expertise working in diverse 21st-century team-based environments. Equity is therefore the focus of this evaluation project, which seeks methods of identifying equity, or the lack of it, in team interactions. This paper sets out a working definition of equitable interactions before exploring the ways in which prior work has operationalized and analyzed the manifestation of equity in interaction. It then explores potential metrics that might enable the analysis of interactional equity on a scale large enough to provide empirical evidence of behavior change following a Tandem intervention.

Defining Equity

Equity is frequently differentiated from the semantically-similar notion of equality by the incorporation of a moral aspect: while equality is equated with 'sameness', equity argues that identical treatment regardless of individual differences may reproduce and perpetuate social disadvantage, and therefore prioritizes 'fairness' (e.g., [2]). Equity nevertheless anticipates equality, of some kind of potentiality further down the line. Equity in an education policy context implies the potentially uneven apportionment of resources, with priority given to individuals or groups facing greater social barriers to educational attainment, in order to achieve equality of access to the benefits of education and, more broadly, equality of the potential for societal participation that education facilitates ([3], [4]).

In the micro-level context of interpersonal group interactions, this redressing of an imbalance of a key resource can be analyzed in terms of each team member's social and cultural capital: the social networks to which they belong or can claim affiliation, and the societally-valued practices and forms of knowledge that they can demonstrate [5]. These forms of capital, and their physical manifestations (known as 'status characteristics', [6], which include demographic factors such as gender and race), create power differentials within the group, reproducing wider social hierarchies and constraining the interactional roles available to less capital-rich team members. Such roles are dynamically negotiated in the team interaction through the process of 'positioning' [7]: interactional, often linguistic, moves that position the speaker or other team members as more or less valuable and authoritative, irrespective of actual levels of expertise. Equity in team interactions may therefore be conceptualized as a process of compensating for the uneven distribution of social and cultural capital, in order to level the playing field and provide

all team members with an equal right and ability to assume positions of value to the team and the task [8].

Measuring Interactional Equity

This project therefore seeks a means of quantitatively identifying indicators of positioning in team members' language use during the interaction, and observing the distribution of these indicators among team members. Cases in which distribution of positioning is strongly imbalanced (where one team member is consistently positioned as 'expert', for example), will be considered candidates for inequitable interactions, with the caveat that although equality may be considered "a rough proxy for equity" [9], this method is not directly observing whether team members had the opportunity and ability to assume positions, but rather whether they did. This paper reports on a comparison across qualitative and quantitative methods of capturing conversational influence, with a focus on assessing the quantitative methods for usefulness as potential scalable metrics.

Method

Analysis is performed on transcript data from an in-person team discussion task in a first-year engineering class at a large midwestern university in the United States during the Spring 2022 semester. The teams generated and evaluated possible solutions for clearing snow from a residential driveway during a 20-minute in-class task practicing the implementation of a decision matrix, and they submitted an audio recording of their conversation at the end of the class meeting. For this analysis, the three best quality audio recordings were chosen for transcription. All three teams had four students. Transcription was initially done using the automated transcription tool Otter.ai and followed by extensive manual correction, as the tool had difficulty with the high level of background noise in the recordings.

Four analyses were performed on the conversations, two qualitative and two quantitative. To develop our understanding of the interactions, two members of the research team produced descriptive analyses of the interactions based on listening to the recordings and reading the transcripts. A third member of the research team coded the transcripts using a coding scheme for interactional positioning [10]. The transcripts were then analyzed using two tools for automated text analysis, each of which is briefly described below. The four analyses were compared for their potential in highlighting aspects of interactional equity. This project was considered exempt by our Institutional Review Board.

Coding Procedure

Transcripts were coded following [10] to summarize how team members positioned themselves and each other as more or less valuable with regard to the task. The authors of [10] identify four ordinally-ranked positions, from expert to novice with two intermediate levels, and a fifth 'facilitator' position. These five categories and examples of their linguistic realizations are summarized in Table 1. In [10], data are segmented at 15-second intervals, the authors applying a single code per segment per team member. However, the rationale for this is unclear, and we chose to apply codes utterance-by-utterance as this resulted in a finer-grained analysis. Brief utterances of agreement such as 'Ok', and 'Yeah', were disregarded, again following the source material. Coding was undertaken by one researcher, reflecting the exploratory nature of this work: this methodological choice is discussed further at the end of this paper.

Table 1 - Descriptions and examples of interactional positioning codes, taken from [10].

Positional move (code)	Description	Example from data
Expert (C1)	Firm statements of fact or firm or strong disagreement	"The least amount of time is gonna be the kid [shoveling]".
Intermediate expert (C2)	Softened statements or softened disagreement, using hedging, question tags etc.	"Safety should probably be first"
Intermediate novice (C3)	Questions that demonstrate understanding and make constructive contributions to the discussion	"Do we need 'effective' on the list?
Novice (C4)	Questions or statements that convey helplessness or general confusion	"Yeah I've never actually lived with snow"
Facilitator (C5)	Metalevel statements or questions that facilitate the discussion	"Should we move on to the matrix thing?"

Linguistic Inquiry and Word Count (LIWC)

Linguistic Inquiry Word Count (LIWC) is an automated text analysis tool for categorizing words and making psychological inferences. The analysis focuses on the use of function words like pronouns and articles to infer cognitive, emotional, and biological processes. It returns a broad range of measures based on the proportions of an individual's words falling within various dictionaries, as well as four *summary variables* calculated using proprietary, non-transparent algorithms that produce a standardized score [11]. LIWC is rigorously tested and validated, and has been extensively used in the scientific community, referenced in over 20,000 publications [12]. Three constructs of LIWC-22 were trialed as potential indicators of positioning among team members: *power* (i.e., need for power or awareness of status), *clout* (i.e., displaying status, confidence, or leadership), and *tentativeness* (i.e., uncertain language). At the group level, the LIWC construct *prosocial behavior* (i.e., signally helping or caring about others), was used to gauge the supportive climate of the team at the group level.

Group Communication Analysis

Group Communication Analysis (GCA) is a tool for analyzing social dynamics in group interactions, and identifying social roles occupied by individual team members [13]. It employs natural language processing techniques to generate scores for team members in six categories. In large datasets, clustering techniques are then used to identify emergent socio-cognitive roles; this step was unavailable to us with our small data set. Three of the six measures generated by the tool were trialed for the potential to highlight inequity: *participation*, or "group-relative mean participation" is a measure of member contributions "above or below what we might expect from perfectly equal participation"; *overall responsiveness* is a measure of how each team member takes up the semantic content of others' language; and *social impact* is in a sense the opposite: the degree to which an individual's contributions are taken up by others. These latter two measures should identify equity-related team characteristics, such as a team member's ideas being ignored by their teammates.

Results

The following sections present a discussion of the four analyses organized by transcript. The three teams that provided data for analysis are identified as F2, F14, and F22. Summaries of the informal descriptive analysis are presented with discussion of how the other forms of analysis support or contradict these impressions. Figure 1 presents the LIWC construct *prosocial behavior*, measured for our purposes on a team level, while Figures 2-4 show the results of the

coding, individual-level LIWC measures, and GCA analyses, respectively. The researchers conducting the descriptive analysis are designated R1 and R2, while speakers in each team interaction are labeled S1-S4.

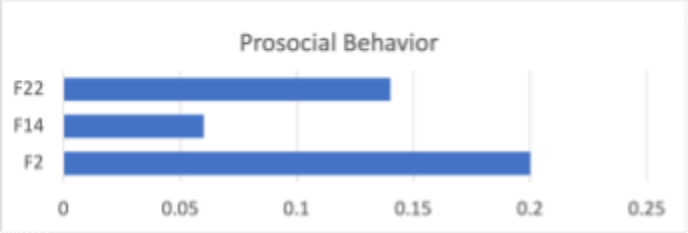

Figure 1 - LIWC team-level prosocial behavior score.

F2: Both researchers (R1 and R2) consider F2 to be a broadly equitable group (R1: "seems pretty collaborative"; R2: everyone "offering ideas pretty equally"). Both agreed that S2 and S4 demonstrate greater "inchargeness," especially in the second half of the discussion, but that this is not oppressive or unreasonable. There is also a common impression that S3 is somewhat overlooked and talked over at times, and that S4 facilitated and directed more than offered substantive ideas. Despite these elements of imbalance, the overall impression is that this is a positive and amicable team environment, and this is supported by the highest *prosocial behavior* score of any of the teams (see Fig. 1).

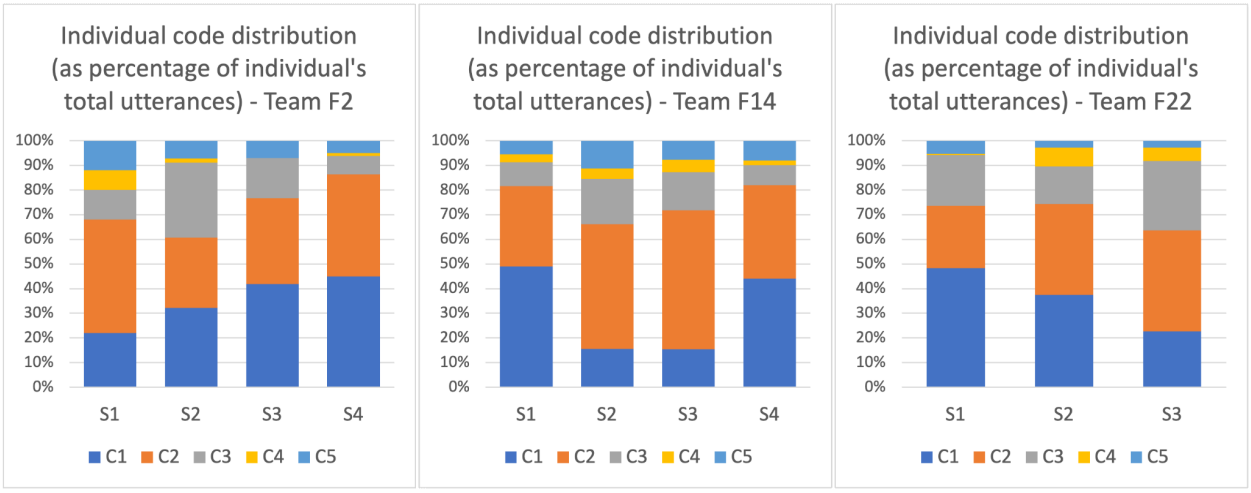

Figure 2 - Distribution of positioning moves (C1-C5; see Table 1) among members (S1-S4) in each team from manual coding. S4 in team F22 had no coded utterances.

The coding and quantitative analyses present a mixed, and somewhat contradictory, complement to these impressions. The coding (Fig. 2) agrees that S4 is most strongly positioned as expert, although with the lowest proportion of facilitator codes, contradicting that part of the qualitative description. S4 scores highly on *clout* in the LIWC analysis (Fig. 3), but is surpassed in this regard by S3, who is regarded as lacking in influence in the descriptive analysis. The coding is more closely aligned with the LIWC analysis of S3. S2, who is described as showing greater "inchargeness," displayed greater *power* awareness according to the LIWC analysis, in line with the descriptive analysis. S2 and S4, described as being more in control of the interaction, also have the highest level of *tentativeness* in the LIWC analysis, suggesting that their leadership is expressed in a way that invites others' input. The GCA analysis (Fig. 4) is somewhat at variance with the others, suggesting that S1 and S4 are the greater participators. The *overall responsiveness* scores are very similar for all team members, but the *social impact* scores corroborate the observation that S3 seems disempowered.

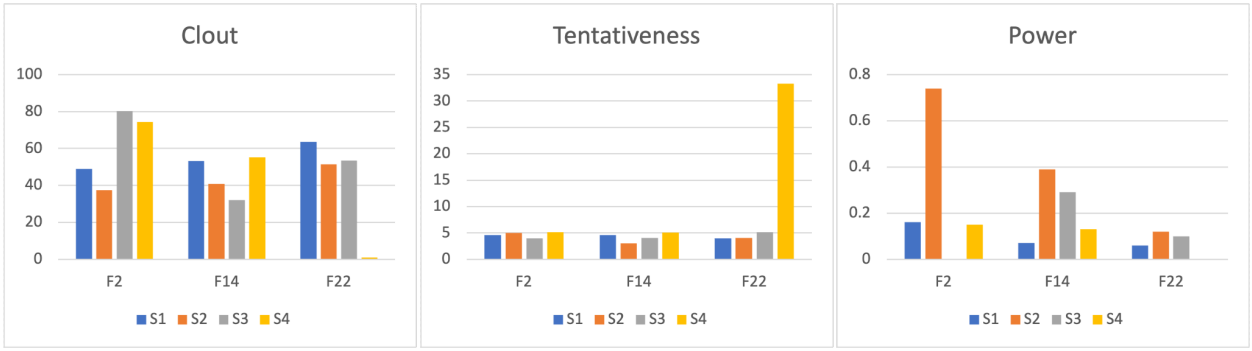

Figure 3 - Scores for each member (S1-S4) of each team for each of the three LIWC constructs. The results for team F22 are skewed by S4's very small number of utterances.

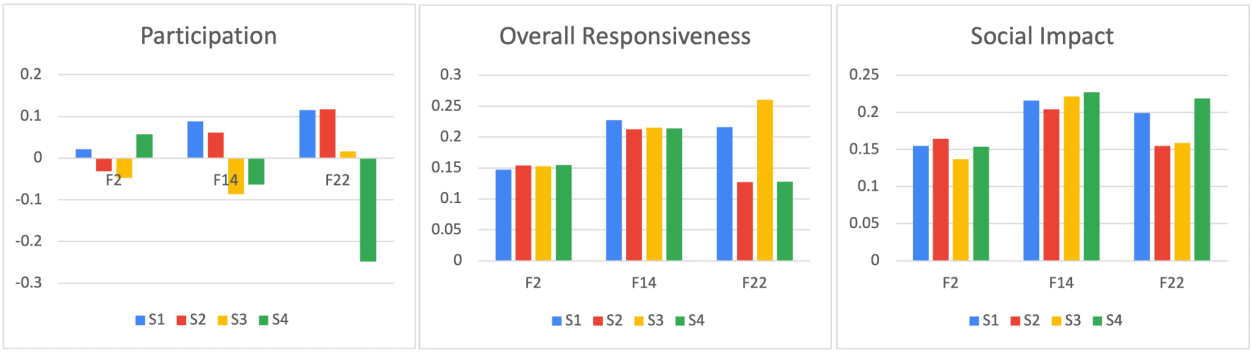

Figure 4 - Scores for each member (S1-S4) of each team for each of the three GCA constructs. The results for team F22 are skewed by S4's very small number of utterances.

F14: The consensus is that this conversation is less equitable than F2, with S1 and S2 dominating the discussion (R2: "more in conversation with each other"), and making unilateral changes to the team product. The impression of a less constructive team environment is supported by the lowest *prosocial behavior* score among these teams (Fig. 1). An interesting example of positioning by reference to personal experience occurred when debating the best weight to give to identified solutions: S2: "A two? Okay. Yeah, I used to live in D.C. We never got more than like three inches of snow there." S1: "I'm from northern Vermont so we get a lot". And later, S1: "I think six, having used a snowblower a lot". S2: "Oh, okay. I've never used one before". This is supported in the coding analysis (Fig. 2) with S1 showing the highest proportion of expert positioning moves. The similar profiles of S2 and S3 in Figure 1 are less congruent, however, as is S4's apparent similarity with S1. It may be that if the coding proportions were normalized by number of contributions the result would be closer to the descriptive analysis. The GCA analysis (Fig. 4) clearly demonstrates the difference in *participation* between S1/S2 and S3/S4. However, the LIWC *clout* scores are congruent with the coding analysis of S1 and S4, as, to a lesser extent, are the GCA *social impact* scores.

F22: This team is somewhat ambiguous with regard to equity. R2 considers it to be the "most egalitarian," and R1 observes that the team document doesn't get updated or modified without consensus. Both researchers note that S3 tries to contribute an idea early on that is quickly dismissed by the rest of the team, but that the idea is provisionally retained until being replaced

later by a new idea that is mutually agreed to be better. There is an incident in which S1 goes to talk to the instructor and S2 and S3 come up with an alternative idea in his absence. They wait for S1 to return to consult before making changes. This may be interpreted as equitable (waiting for consensus) or inequitable (S2 and S3 deferring to S1's authority). The coding leans toward the latter interpretation, with a higher proportion of expert positioning moves shown by S1, followed by S2 and S3. The LIWC *clout* scores also favor S1, although the range of these scores (discounting S4) is the least of all the teams, implying greater equity. S1 has the highest *social impact* score in the GCA analysis (again discounting S4), while S3 has the highest *overall responsiveness* score. This result perhaps suggests that S3 was more likely to go along with the ideas of others, although this conclusion is somewhat questioned by S1's also-prominent *responsiveness* score. The GCA *participation* scores clearly support the notion that S3 did not, or was unable to, contribute as much as S1 and S2, corroborating the inequitable interpretation.

Discussion and Next Steps

There were aspects of congruence between the four analyses, and some areas of contrast. In some cases, the quantitative analyses provided an useful alternate perspective to help differentiate competing interpretations in the subjective description. The qualitative coding provided a useful mediator between the descriptive analysis and the quantitative tools. However, the manual process of coding limits the scope of its application, especially as it should be performed more rigorously using multiple coders and assessed for intercoder reliability. While this methodological choice is by no means mandatory in all qualitative research, this deductive coding analysis would benefit from multiple coders as "an important check on selective perception" [14] in the application of the codes. The LIWC *clout* measure provided useful differentiation between team members, but the *tentativeness* measure demonstrated less variance, and the *power* measure was more challenging to interpret. The group-level *prosocial behaviors* measure aligned intuitively with the descriptive analysis of group climate. The GCA *participation* measure provided a useful indication of the balance of individual contributions to the interaction, but doesn't help us differentiate between team members who don't wish to contribute and those who do but are constrained from being able to. Comparing the analyses for evidence of equity is hampered by the lack of a baseline, and future work will incorporate other sources of data, including students' own feedback on the interactions.

A potential "wish list" of characteristics for a sophisticated equity metric would include the ability to analyze dynamic change over the course of an interaction, as well as the ability to detect idea uptake and attribution. This goal is to an extent accomplished with GCA's *social impact* metric, but a yet more ideal solution would enable the tracking of a specific idea through the interaction. Additionally, further natural language processing solutions may enable an automated approach to the coding implemented manually in this paper, where positioning moves can be reliably mapped on to syntactic forms, and these forms identified automatically. All these are directions which the project will explore in its next phases, alongside self-reported student data collected by Tandem itself attesting to students' perceptions and experiences of teamwork. Greater attention will also be paid to data collection and processing, with transcripts tailored more carefully to the needs of the tools. GCA, for example, calculates *participation* based on utterances, and not enough consideration was given during transcription to exactly where utterance boundaries were, and how overlapping utterances (a very common feature) were represented. Using tried and tested transcription conventions from Conversation Analysis

literature may prove fruitful in this regard. Greater care will also be taken to isolate teams during data collection, to produce cleaner audio data; noisy data created significant difficulties in the transcription process.

References

1. R. Fowler, L. K. Alford, S. Sheffield, C. Hayward, T. S. Henderson, and R. L. Matz, "Supporting Equitable Team Experiences Using Tandem, an Online Assessment and Learning Tool," in *ASEE Annual Conference*, 2021, https://peer.asee.org/37787
2. P. Braveman, "Defining Equity in Health", *Health Policy and Development Journal*, col. 2, no. 3, pp. 180-185, December 2004, https://www.jstor.org/stable/25569979
3. W. J. Jordan, "Defining Equity, Multiple Perspectives to Analyzing the Performance of Diverse Learners", *Review of Research in Education*, vol. 34, pp. 142-178, March 2010, http://doi.org/10.3102/0091732X09352898
4. S. Marginson, "Equity, status and freedom: a note on higher education", *Cambridge Journal of Education*, vol. 41, no. 1, pp. 23-36, March 2011, http://doi.org/10.1080/0305764X.2010.549456
5. T. Pham and L. Pham, "An analysis of small group interactions of Vietnamese students under the Bourdieusian theoretical lens," *Higher Education Research and Development*, vol. 37, no. 2, pp. 359–372, 2018, https://doi.org/10.1080/07294360.2017.1370437
6. J. Berger, B. P. Cohen, and M. Zelditch, "Status Characteristics and Social Interaction" *American Sociological Review*, vol. 37, no. 3, pp. 241–255, 1972, https://www-jstor-org.ezp2.lib.umn.edu/stable/2093465
7. R. Harré, F. M. Moghaddam, T. P. Cairnie, D. Rothbart, and S. R. Sabat, "Recent Advances in Positioning Theory," *Theory & Psychology*, vol. 19, no. 1, pp. 5–31, 2009, https://doi.org/10.1177/0959354308101417
8. I. Esmonde, "Mathematics learning in groups: Analyzing equity in two cooperative activity structures," *Journal of the Learning Sciences*, vol. 18, no. 2, pp. 247–284, 2009, https://doi.org/10.1080/10508400902797958
9. M. Slattery and P. Hutchison, "Assessing equity in collaborative learning situations: A comparison of methods," *Proceedings of the International Conference of the Learning Sciences, ICLS*, pp. 1149–1152, June 2018.
10. D. T. Brookes and Y. Yang, "Social positioning in small group interactions in an investigative science learning environment physics class," *Physical Review Physics Education Research*, vol. 17, no. 1, p. 10103-1 - 10103-13, 2021, https://doi.org/10.1103/PhysRevPhysEducRes.17.010103
11. R. L. Boyd, A. Ashokkumar, S. Seraj and J. W. Pennebaker, *The development and psychometric properties of LIWC-22.* Austin, TX: University of Texas at Austin, 2022. https://www.liwc.app/
12. Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods", *Journal of Language and Social Psychology,* vol. 29, pp. 24–54, 2010, https://doi.org/10.1177/0261927X09351676
13. N. M. M. Dowell, T. M. Nixon, and A. C. Graesser, "Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions," *Behavior Research Methods*, vol. 51, no. 3, pp. 1007–1041, 2019. http://doi.org/10.3758/s13428-018-1102-z

14. R. Cole, “Inter-Rater Reliability Methods in Qualitative Case Study Research,” *Sociological Methods and Research*, pp. 1–39, 2023. http://doi.org/10.1177/00491241231156971