



Interviewing Peer Leaders: Inter-Reliability of a Growth-Oriented Selection Rubric

JOURNAL ARTICLE

JILLIAN P. RAGER

JAIME S. MILLER 

JOSEPH M. KUSH 

DEIDRE LESLIE

ASHLEY C. CARR

**Author affiliations can be found in the back matter of this article*

VIRGINIA TECH.
PUBLISHING

ABSTRACT

Peer leaders are student employees selected by program administrators in higher education to support their fellow classmates and peers in college. This study evaluated the inter-reliability of a selection rubric used by raters during group interviews for a peer leader program, Supplemental Instruction (SI), that maintains a team of over 60 peer leaders each semester. Results confirm that the rubric possesses a substantial degree of reliability, with an overall average agreement percentage of 88%. Researchers discuss reasons for this statistically significant result, including effective and theory-based processes, thorough training, and openness to change and growth. Implications include more emphasis on educators' process in creating assessment tools, including rubrics, and validation of other selection tools presently used in the field of higher education to determine students' eligibility for peer leadership roles.

CORRESPONDING AUTHOR:

Jaime S. Miller

Academic Support, Auburn
University, US

jmiller@auburn.edu

TO CITE THIS ARTICLE:

Rager, J. P., Miller, J. S., Kush, J. M., Leslie, D., & Carr, A. C. (2026). Interviewing Peer Leaders: Inter-Reliability of a Growth-Oriented Selection Rubric. *Journal of Peer Learning*, 17(1): 4, 1–13. <https://doi.org/10.21061/jopl.151>

Peer leaders' roles and use in higher education are ever-growing within various institutional contexts (Cuseo, 2010a; Ender & Kay, 2001; Ganser & Kennedy, 2012; Skalicky et al., 2024). Keup and Skipper (2010) define a peer leader as "an undergraduate student who has been selected to serve as a mentor or peer educator to other students through a position with a school-run organization" (Slide 9). As this definition suggests, selection of peer leaders is an essential step in peer leader programs, even considered a "building block" of program administration (Hunter & Heath, 2001, pp. 42–43). As such, interviewing becomes a common pathway for selecting students to serve as peer leaders. Hunter and Heath (2001) suggested the use of rubrics during the interview process to promote fairness and equity while screening applicants. However, the use of rubrics for the evaluation of peer leaders in higher education specifically remains an understudied area with important implications.

The remainder of the article is outlined as follows. We begin with an overview of peer leadership within the field of higher education. Next, we consider the role of rubrics in education, focusing on their use during interviews and for hiring decisions. We then introduce the Supplemental Instruction Interview Rubric (SIIR) and its developmental history. Finally, we outline the steps for using the rubric in practice, noting contextual considerations in the selection process.

PEER LEADERSHIP IN HIGHER EDUCATION

Benefits of peer leadership are well discussed in the field of higher education (Shook & Keup, 2012; van der Meer, et al., 2022). Recognizing peer leaders' role in supporting students' experiences in college (Ender & Kay, 2001), including their work to connect their peers to resources (Colvin & Ashman, 2010; Cuseo, 2010a; Cuseo, 2010c), model and coach positive and productive habits (Colvin & Ashman, 2010; Cuseo, 2010a), serve as a support and champion (Colvin, & Ashman, 2010; Cuseo, 2010a; Ender & Kay, 2001), and facilitate socialization in college (Irwin et al., 2023; Jarratt et al., 2024; Linley, 2015; 2017; 2018), it is critical to select students who demonstrate capacity and interest in helping their peers.

Supplemental Instruction (SI) is one such example of peer leadership in higher education that requires a selection process. SI was created through the work of Dr. Deanna Martin in 1973 at the University of Missouri–Kansas City and is now an international program model at many institutions of higher education (The International Center for Supplemental Instruction [ICSI], 2022). SI often targets courses with high rates of students receiving a grade of D, F, or withdrawing from the course (ICSI, 2022) and has been shown to increase course grades and retention in degree programs for the students that regularly attend (Bowman et al., 2023). Within SI programs, SI Leaders are charged with facilitating sessions among peers, attending class lectures, incorporating collaborative learning techniques, and facilitating active learning strategies (ICSI, 2022). Generally, SI programs consider the selection of SI Leaders based on the applicants' skills and characteristics related to social aptitude, pedagogical knowledge, and logistical or administrative responsibility (McDaniel, 2008). There has been limited guidance and information from the International Center on selection processes, thus there are numerous methods and processes for SI Leader selection across programs. Because we intend for this rubric to be broadly applicable and relevant across the field, we will refer to SI Leaders as peer leaders throughout this manuscript in alignment with the implied universality of the term (Cuseo, 2010a).

Scholars note the importance of interviewing as a major administrative element of peer leader programs (Hunter & Heath, 2001; Keller & Porter, 2020) and consider a variety of approaches to recruit and select peer leaders (Hamid & VanHook, 2001). Empowering current peer leaders to recommend their peers as part of the selection process is a recommended practice (Breslin et al., 2018) and aligns with approaches that support students' competency and confidence (Gannouni & Ramboarison-Lalao, 2018). Hunter & Heath (2001) encourage program staff to be thoughtful about the goals, structure, and overall design of interviews. Without appropriate training or guidance, raters may drastically differ in which aspects of traits they consider most important in peer leader selection.

Researchers have proposed incorporating rubrics during the interview and selection processes. Specifically, rubrics set expectations by identifying and listing levels or scales of criteria (Panadero & Romero, 2014, as cited in Taylor et al., 2024). While there are many types of rubrics, all have three features: assessment criteria, a grading strategy, and standards or quality definitions (Jonsson & Svingby, 2007; Popham, 1997, as cited in Reddy & Andrade, 2010; Sambamurthy & Cox, 2016). Within education, rubrics have historically been used in classroom learning experiences or peer evaluations of student work (Reddy & Andrade, 2010; Taylor et al., 2024), including providing feedback on work, grading, and identifying areas of improvement (Reddy & Andrade, 2010). Scholars have noted that rubrics with strong fundamental properties (e.g., standardization, concrete expectations at each performance level, transparency) lead to increased consistency of scoring, clarity of academic expectations, and even increased academic performance (Chan & Ho, 2019; Taylor et al., 2024). Yet, critiques of rubrics include stifling creativity, undermining critical thinking, and omitting student input (Chan & Ho, 2019; Taylor et al., 2024). Further, in their literature review of rubrics in higher education, Reddy and Andrade (2010) note limited insight into the reliability and validity of the rubrics being used.

Within interview and selection processes, rubrics are useful performance evaluation tools, as they align organizations' values and responsibilities to positions in fair and consistent ways (Culpepper et al., 2023; Jonsson & Svingby, 2007; Leyzberg et al., 2017; Sambamurthy & Cox, 2016). Comparable to their use in education, rubrics for selection processes can help alleviate issues of ambiguity, social biases, and lack of reliability and validity evidence (Culpepper et al., 2023; Sambamurthy & Cox, 2016). In general, the effectiveness of rubrics in interview and selection processes as a means of evaluating performance can be strengthened by (a) ensuring an iterative and thoughtful design process, (b) calibrating with other search committee members, (c) utilizing potentially different rubrics for different phases of the search, and (d) defining a decision-making process for evaluators (Culpepper et al., 2023; Leyzberg et al., 2017; Sambamurthy & Cox, 2016).

Although the use of rubrics for interviewing and hiring processes has been well documented, there has been limited insight into the use of rubrics for interviews and selection of peer leaders, specifically. For example, Berg and Stoner (2016) discussed the process used for selection of resident assistants (RAs), wherein applicants participated in a group process interview and were scored on a standardized rubric that spoke to the skills of communication, attitude, teamwork, leadership, problem-solving, and overall recommendation. The rubric scores, totaled with the scores from the application and individual interviews, were compared to the RAs' performance evaluation, and they did not yield significant results (Berg & Stoner, 2016). The authors speak to the necessity of evaluating the rubric components more thoroughly, such as strengthening the criteria and aligning interview rubrics with evaluation rubrics (Berg & Stoner, 2016). Similarly, Leyzberg et al. (2017) outlined the creation and implementation of an interview rubric for graduate teaching assistants (TAs). The authors found that implementation of the hiring rubric resulted in improved ratings of TAs by their students. While there is minimal literature on using rubrics for interviews and selection of peer leaders, it appears that the guiding principles from more generalized rubric use are relevant.

One additional consideration concerns the evaluation of applicants through interviews that are conducted by peers. Within education, there is evidence that peer rating can be a helpful practice for reducing teacher workload and promoting learning, growth, and collaboration (Cheung-Blunden & Khan, 2018; Falchikov & Goldfinch, 2000; Moore & Teather, 2013; Zhang et al., 2008). Although peer ratings generally demonstrate strong reliability and validity between teacher and student grades, critics of peer rating have noted variability in rating skills and students' distrust of or disinterest in others' ratings (Ashenafi, 2017; Cheung-Blunden & Khan, 2018; Zhang et al., 2008).

Overall, the efficacy of peer ratings may be improved by (a) using rubrics and other supporting materials (e.g., scenario-based assessments following scripts), (b) allocating adequate time for rater training and development, and (c) providing constructive feedback focused on growth (Moore & Teather, 2013). Ultimately, peer rating is a valuable way to support peer leader programs, but careful consideration of tools, training, and standardization is necessary.

Recognizing the robust and varied status of rubrics in peer leader interview and selection processes, this study tested the inter-reliability of the rubric presently used during group interviews among SI Leader applicants in Academic Support, named the Supplemental Instruction Interview Rubric (SIIR). The SIIR was originally developed by Carr (2018) and revised by Rager & Leslie (2023). This rubric aims to promote growth for peer leaders, mitigate bias and subjectivity within peer leader interview processes, and proactively avoid any discriminatory practices in peer leader selection. These efforts within the office have resulted in an interview and selection rubric that is theory-informed, reflective of our office culture, and considers the growth potential among applicants in the context of the positional responsibilities of the peer leader role. The published rubric (see Table 1) is presently implemented by program staff, including both program practitioners and current SI Leaders (referred to in this article as raters).

The SIIR rubric in Academic Support reflects staff considerations of the anti-deficit achievement framework (Harper, 2010; 2012), the dignity-building culture of the office (Ansley et al., 2025), person-centered theory (Gladding, 2012; Rogers, 1961; Witty & Adomaitis, 2014), and the applicants' potential for growth within the peer leader position duties. These core principles work in tandem to promote strong processes for peer leader selection.

DIMENSION	INSUFFICIENT (1)	DEVELOPING (2)	PROFICIENT (3)	DISTINGUISHED (4)
Familiarity with SI (knows basic elements)	Misinformed about or unaware of SI elements	Mentions elements with some possible misconceptions	Describes or uses elements mostly accurately	Describes or uses elements accurately
Familiarity with Academic Support	Misinformed about or may not acknowledge other AS programs or mission	Mentions other AS programs or mission	Describes other programs or AS mission mostly accurately	Describes other programs or AS mission accurately
Communication (verbal)	May be difficult to understand, unclear response, speaks off topic	Speaks nonspecifically, may demonstrate minimal processing of ideas	Speaks to specific actions or ideas, positive tone, understandable in the group	Volume, enunciation, and demonstration of ideas are clear, concise, and effective
Collaboration	Dominates or unwilling to contribute to group work	Willing to contribute, but may repeat or restate others	Effectively adds value to conversation or provides new ideas	Synthesizes discussion, adds new ideas or evaluates ideas, may lead group effectively, may use SI facilitation skills
Decision-making	Rigid or binary thinking	Considers new ideas, but does not apply	Demonstrates flexibility in thinking, teachable	Demonstrates flexibility and application in thinking
Communication (nonverbal)	Misaligned body language and speech, negative affect	Generally engaged, may not be actively participating, instances of negative affect	Actively listening to facilitators, positive affect, is approachable and friendly, focused	Actively listening to facilitators AND group, positive affect with aligned body language and speech, engaged
Mindset of an SI Leader (understands why SI works)	Uses undesired terminology such as "teaching," "lecturing," "kids" (as an SI Leader), inaccurately describes elements or responsibilities	References experience with SI, describes fundamental aspects of SI accurately	Mentions self-directed learning, student leading, active learning, learning styles, group work, skill integration (may not use the exact word)	Describes self-directed learning, learning styles, active learning, group work, student leading, skill integration

Anti-Deficit Achievement Framework (ADAF)

Deficit approaches in research involve research questions that amplify student failure instead of achievement, especially with students of color (Harper, 2010; Irizarry, 2009, as cited in Perez et al., 2017). Further, deficit approaches blame individuals for not meeting certain academic and social standards and are unlikely to examine systemic barriers that contribute (Reyes & Duran, 2021). As a result, deficit approaches often result in implicit biases for educators and feelings of inadequacy for students (Harper, 2010). Work by Harper (2010) ultimately led to the creation of the Anti-Deficit Achievement Framework (ADAF). Harper's (2010) analysis of the experiences of

Table 1 Supplemental Instruction Interview Rubric (SIIR).

Note: Created by Carr (2018) and adapted by Rager & Leslie (2023).

Black male students in STEM degree programs at various institutions across the United States revealed that research questions primarily focused on why college students of color in STEM were performing more poorly and achieving less than White peers (i.e., deficits) as opposed to focusing on assets and attributes. These deficit questions were reframed into the ADAF, and the framework was further refined and applied to a broad, nationwide study of Black male undergraduates at 42 different colleges and universities (Harper, 2012). This report on the study concluded with the necessity of a “paradigm shift in the way achievement is studied” (Harper, 2012, p. 25). The Anti-Deficit Achievement Framework (Harper, 2010) was used to create a selection tool that ensures students are evaluated on their attributes and potential as peer leaders and promotes their long-term employment with the program.

The framework reframed how our peer leader selection was approached by avoiding focus on what the applicant is *not* doing and instead focusing on what the applicant *is* doing. For example, raters were trained to look for how the applicants explicitly demonstrated the dimensions in the rubric within the interview as opposed to scoring based on how the applicants did not meet their own ideas or expectations. Further, the dimensions of the rubric were shared with applicants during the interview to address the cultural capital assumptions discussed within the ADAF. Finally, the rubric was designed to avoid both subjective judgments and emphasizing raters’ own beliefs and assumptions about applicants’ eligibility for the peer leader position. As a result of the influence of the ADAF, the rater, upon selecting a certain score, would ideally be able to reference specific behaviors from the interview that align with the criteria rather than intangible judgements or feelings.

Workplace Culture

Academic Support staff have fostered a work culture informed by dignity: the work culture in the office is further described by Ansley and scholars (2025). Inspired by the work of Donna Hicks (2021), the authors illustrated the application of dignity to the workplace through the telling of a student employees’ experience within the office (Ansley et al., 2025). Recognizing this culture, we honored that not all applicants will perform their best during interviews and appreciated that the applicants are undergraduate students with various levels of work experience and knowledge. For example, raters considered the ways applicants can learn and grow, their dispositions for performing on the team, and focused on behaviors during the interview.

Person-Centered Theory

To ensure alignment among the theoretical frameworks, workplace culture, and a reflective and rigorous program practice, we utilized person-centered principles within the rubric. Person-centered counseling theory involves three core conditions of unconditional positive regard, empathy, and congruence (Rogers, 1961; Witty & Adomaitis, 2014). Unconditional positive regard refers to the facilitator being nonjudgmental, empathy refers to the facilitator seeking understanding of the client’s point of view, and congruence refers to the facilitator acting with authenticity and genuineness (Witty & Adomaitis, 2014). Person-centered principles have been applied and utilized in group settings; facilitators lead groups using the core conditions and active listening (Gladding, 2012). Facilitators also have the responsibility to “create a climate that makes a difference in the growth of group members” (Gladding, 2012, p. 347). In this SI program’s context, the raters were facilitators who implemented these principles within the peer leader group interview experience. The raters (as facilitators of the group interview experience) were trained with these principles as a mindset; i.e., they should facilitate the interview and rate the applicants with authenticity and genuineness, hold positive regard for the applicants, actively listen to each applicant, and seek understanding and meaning in their responses. By positioning the group interview experience and training facilitators to embrace these principles, the applicants became more collaborative and productive within the interview activities, thereby providing them the opportunity to demonstrate their knowledge, skills, and potential for growth in the role. The three core conditions and group facilitation principles were necessary for interview facilitators to lead the group effectively and align with the anti-deficit framework and workplace culture that informed the selection rubric.

Peer Leader Position Responsibilities and Duties

This rubric is presently used for SI Leader interviews. SI Leaders are a type of peer leader (Cuseo, 2010b) and as part of their training, an SI Leader will learn the core tenets of SI: content mastery, peer connections, and skill integration (ICSI, 2019). SI Leaders must be strong referral agents for campus resources, collaborate well, demonstrate leadership, and be confident decision-makers. The work requires thinking of themselves simultaneously as a peer, a helper, a learner, and a leader. Further, the SI Leader is responsible for attending class, serving as a model student, planning active learning review for group study sessions, and building relationships with the faculty member and students. Woven into these responsibilities are the skills of communication, decision-making, and collaboration. The rubric was designed with these responsibilities in mind.

Application

The rubric has been refined and rewritten since 2018 to honor these core principles and frameworks and to address key wonderings from the researcher-practitioner team authoring this paper:

- How can we empower current peer leaders to select future peers in the same role?
- In what ways can we approach a peer-led selection process that is ethical and equitable?
- How do we select peer leaders who can grow, learn, and serve well in this role?

As the rubric is intended for use by peer leaders evaluating their peer applicants, the selection criteria and process became demystified and streamlined for raters through using the principles of Anti-Deficit Achievement Framework (Harper, 2010), the dignity-building culture of the office (Ansley et al., 2025), person-centered theory (Gladding, 2012; Rogers, 1961; Witty & Adomaitis, 2014), and the applicants' potential for growth within the peer leader position duties. See Table 1 for the current rubric.

Each criterion and dimension of the rubric was written with intentionality, including positive language that encourages growth over time while simultaneously scaffolding the potential ranges of an applicant. For example, consider the description of the Communication (verbal) dimension scored at the Developing (2) level: "speaks nonspecifically, may demonstrate minimal processing of ideas." This phrasing measures the level of what the applicant did in actuality rather than the level they did not achieve. In addition, the phrase "spoke nonspecifically" invokes more concrete conceptualization than something such as "spoke poorly." By removing such ambiguous or subjective language, raters better understood what is meant by each criterion description and could appropriately focus on applicants' demonstrated skills.

THE PRESENT STUDY

This article aims to address the understudied use of rubrics for evaluating peer leaders in part by assessing the reliability of our peer leader selection rubric. Using rating data from $N = 12$ raters assessing $N = 132$ undergraduate student applicants during the 2024–25 academic year, various inter-rater agreement coefficients were calculated. By assessing the reliability of a promising interview selection rubric, results from the current study offer strong evidence with meaningful implications for peer interviews in higher education more generally.

Although the SIIR rubric had recently undergone extensive revisions for fairness, clarity, and job responsibility alignment, the overall reliability of the rubric remained largely unknown. The current study investigated the inter-rater agreement and reliability of the SIIR rubric to better understand the degree to which different raters selected the same or similar scores for applicants across a range of criteria and dimensions. By providing evidence of the reliability of the SIIR rubric, researchers and practitioners alike can assess the effectiveness of the interview process, from training raters to selection; ensure consistency amongst raters; and promote standardization of practice for peer leaders.

SAMPLE

Ratings of students were obtained from a large-sized public university in the Southern U.S. from the Fall 2024 and Spring 2025 semesters. Institutional Review Board (IRB) approval was attained for this study. The raters digitally signed an IRB-approved consent form before submitting their ratings. All IRB-approved procedures were followed in the interview and rating process. The Fall 2024 subsample consisted of $n = 35$ undergraduate student applicants while the Spring 2025 subsample consisted of $n = 89$. Due to removing one student who received ratings from only a single rater, the final analytic sample consisted of $N = 124$ students who received ratings from two to three raters each (mean raters per subject = 2.69). A total of $N = 12$ raters provided scores across the fall and spring semesters. The raters included 1 professional staff member, 1 paraprofessional staff member, and 10 current SI Leaders who were undergraduate students.

The Group Interview Structure

During this study, the group interview included up to five applicants and three raters. Importantly, group interviews demonstrated how applicants interacted with, listened to, and collaborated with others (Hunter & Heath, 2001). This was especially relevant to the Supplemental Instruction program model in which applicants must be able to interact effectively with others as part of the job responsibilities.

Based on the quality of their application, undergraduate student applicants were invited to interview by self-selecting their interview date and time using an online sign-up platform where they could not see the names of other applicants or raters. Group interviews were then facilitated by a team of undergraduate peer raters who had prior experience working as SI Leaders. In advance of the interviews, all peer raters completed training where they reviewed the facilitation guide and rubric and practiced with a case interview. After raters reviewed the materials, the SI Coordinator further explained procedures and checked for raters' understanding of the rubric. Raters then practiced using the rubric via a case study activity. The case study was a narrative, fictional depiction of the interview and what could occur based on the raters' genuine lived experiences from previous interviews. After reading the case study, raters assigned scores to the three fictional applicants using the rubric. Finally, the selection process was thoroughly discussed with an overall goal of increasing raters' knowledge of and comfortability with the rubric and facilitation expectations.

Group interviews were approximately 75 minutes in length. Each interview was divided into three sections, including (1) an opener, (2) three main activities, and (3) a closer. The opener involved each applicant stating their name and responding to an icebreaker question. In the first main activity, applicants shared a prepared mock introduction speech that briefly described themselves, the fundamentals of SI, and the benefits of SI. The second main activity was a mock planning activity, in which each applicant reviewed a predetermined list of active learning strategies and selected one that they were most interested in using during a SI session. The applicants were asked to explain the reasoning for their choice. The final main activity was a scenario activity in which the entire group was presented with a challenging scenario and tasked with discussing the context of the issue, identifying the main challenges, and proposing a resolution. The applicants collaborated on their responses and presented them to the raters. The interview concluded with a dice-rolling activity: applicants rolled dice and responded to associated questions regarding the interview, the SI program, or their personality. The interview concluded with the facilitator providing next steps (e.g. timeline, communication) and an opportunity for applicants to ask questions. Raters then independently submitted their ratings for each applicant using the rubric.

RUBRIC DIMENSIONS AND SCORING

The rubric consisted of seven distinct criteria or dimensions, including (1) familiarity with SI, (2) familiarity with academic support, (3) verbal communication, (4) collaboration, (5) decision-making, (6) nonverbal communication, and (7) mindset of an SI Leader. Applicants received ratings in each dimension across the following four scoring categories: 1 = insufficient, 2 = developing, 3

= proficient, and 4 = distinguished. Scores across the seven dimensions were summed, resulting in a total score for each applicant ranging from 7–28. It is important to note that in practice, total scores are used for hiring decisions.

DATA ANALYSES

Given the ordinal nature of the scoring categories, an unweighted Cohen's kappa coefficient would likely yield inefficient, misleading results (Cohen, 1960; 1968). Extensions by Janson and Olsson (2001; 2004) proposed a weighted kappa statistic for interval- and ratio-level data, with Gwet (2014) further introducing ordinal weights. Considering two scoring categories k and l , an unweighted kappa coefficient would have the weighting scheme:

$$w_{kl} = \begin{cases} 0 & \text{if } k \neq l \\ 1 & \text{if } k = l \end{cases}$$

It can be seen that a weight of 1 indicates perfect agreement, while a weight of 0 implies complete disagreement. However, an unweighted kappa coefficient would treat all disagreements as equal weight. For example, consider three raters assigning scores from the four categories described above. If rater a provided a score of 4 = distinguished, rater b 's score of 3 = proficient would be given a weight of zero, as would rater c 's score of 1 = insufficient. While raters b and c would both receive weights equal to zero because their scores were not exactly equal to rater a 's score, the amount of disagreement between raters a and b (i.e., the 1-point difference) is less than the amount of disagreement between raters a and c (i.e., the 3-point difference). Given q total scoring categories, categories could instead be ranked ordinally, such that adjacent categories imply a less serious disagreement than nonadjacent categories. Thus, defining a set of weights $w_{kl} \in [0,1]$ suggests that any weight $0 < w_{kl} < 1$ indicates partial agreement. Continuing with the example, such ordinal weights would imply that raters a and b , though not equal in their scores, were in more (partial) agreement than raters a and c . Specifically, ordinal weights may be constructed as (Klein, 2018):

$$w_{kl} = \begin{cases} 1 - \frac{\binom{|k-l|+1}{2}}{\binom{q}{2}} & \text{if } k \neq l \\ 1 & \text{if } k = l \end{cases}$$

In the current study, given $q = 4$ total scoring categories, the resulting ordinal weights were 0.83 for any 1-category disagreement, 0.50 for any 2-category disagreement, and 0 for a 3-category disagreement (in addition to a weight of 1 for agreement). In addition to calculating reliability coefficients for each dimension, a reliability coefficient was calculated for total scores using ratio weights (i.e., total scores having the potential to range from 7–28). These weights were then used to construct various kappa coefficients, with suggested interpretations as (Landis & Koch, 1977): 0.00–0.20 = poor, 0.21–0.40 = slight, 0.41–0.60 = fair, 0.61–0.80 = moderate, 0.81–1.00 = almost perfect. Both exact percent agreement and Gwet's AC (Gwet, 2014) coefficients were calculated for each dimension as well as for total scores.

RESULTS

In general, results indicated that the rubric possessed a substantial degree of inter-rater reliability, with agreement coefficients across the various dimensions ranging from 0.71 to 0.97 (see Table 2). As expected, exact percent agreement coefficients were always the highest coefficients, while the chance-adjusted Gwet's AC coefficients were always the lowest (comparatively within a given dimension). The two dimensions with the lowest inter-rater agreement coefficients were SI familiarity ($\bar{\kappa} = 0.80$) and mindset of an SI Leader ($\bar{\kappa} = 0.80$), while the dimension with the highest values was collaboration ($\bar{\kappa} = 0.88$; see Figure 1). Additionally, total scores demonstrated extremely strong percent agreement (0.97) and Gwet's AC (0.86) coefficients.

DIMENSION	AGREEMENT COEFFICIENT	
	PERCENT AGREEMENT	GWET'S AC
SI familiarity	0.88	0.72
Academic Support familiarity	0.87	0.76
Verbal communication	0.89	0.75
Collaboration	0.92	0.83
Decision-making	0.89	0.77
Nonverbal communication	0.89	0.74
Mindset of an SI Leader	0.88	0.71
Total score	0.97	0.86

Table 2 Inter-Rater Reliability Statistics Across Rubric Dimensions.

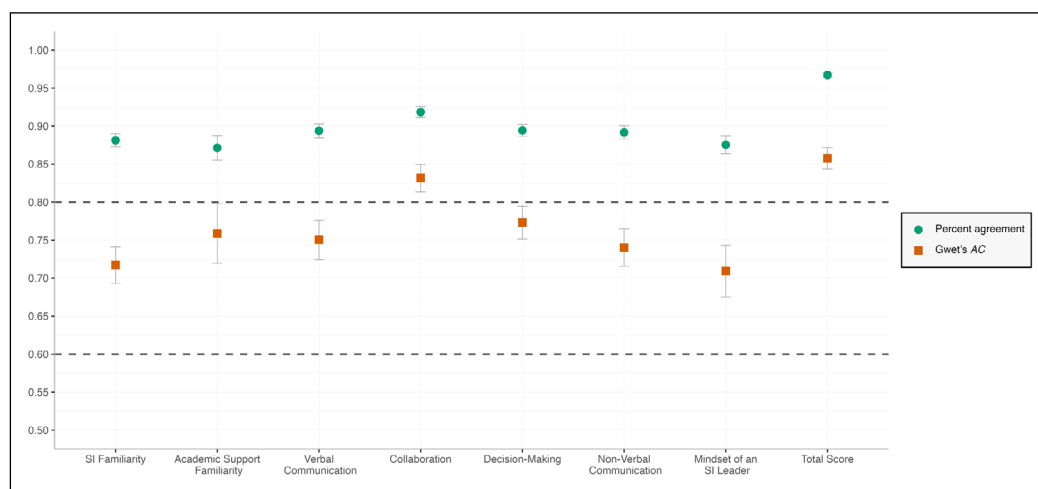


Figure 1 Various Agreement Coefficients Across Each Dimension.

Note: Kappa coefficients may be interpreted as 0.61–0.80 = substantial, 0.81–1.00 = almost perfect.

DISCUSSION

The current study examined the inter-rater reliability of a rubric developed for assessing applicants' performance when interviewing for a peer leader position. Using a variety of inter-rater coefficients, results largely demonstrated that the rubric possessed a substantial degree of reliability, with an overall average agreement percentage of 88%. However, the dimensions of the rubric varied in their reliability coefficients, with collaboration showing the highest degree of reliability, and SI familiarity and mindset of an SI Leader showing the lowest. Results from the present study led the team to reflect upon current practice with the rubric. Consequently, three findings emerged that provide insight into effective program administration for selecting peer leaders.

RELIABLE RUBRICS ARE BUILT ON EFFECTIVE WORK PROCESSES

The results of this study led us to believe that the process we took to create an inter-reliable rubric worked. The rubric proved to be inter-reliable, and this outcome was the result of bringing together practitioners' knowledge and experiences to build something collaboratively over time. The rubric incorporated theoretical understanding, and changes to the rubric were made since 2018. Our process to build this rubric honored the knowledge and praxes of the people who do this work (e.g. practitioners and raters). To this end, iterative feedback loops were built into the process so raters could provide their insights.

RELIABLE RUBRICS REQUIRE RESPONSIVE TRAINING

The results of this study led us to believe that the training raters experienced was sufficiently helpful and useful. As outlined earlier, this training approach created space for raters to learn about the rubric, practice with it, and reflect upon the scores as a group. The opportunity to engage with the materials and practice enabled raters to feel more confident in the evaluation task. Prior to

this study, the authors were unsure if the training was sufficient. Yet, the study confirmed that the training generated reliable scoring among raters and successfully met the training goals.

RELIABLE RUBRICS BENEFIT FROM DYNAMIC FEEDBACK LOOPS AND CHANGE

This rubric has evolved and changed since 2018 to be responsive to the theoretical framings and workplace culture of the office. As research-practitioners, we embraced the opportunity for change to honor the knowledge and lived experiences of colleagues, raters, and applicants. This commitment required dynamic yet straightforward work processes for collecting and responding effectively to feedback each semester. For example, in addition to more formalized surveying methods, the lead author of this paper routinely asked the evaluators questions to build understanding of raters' experiences with the group interviews:

What needs to be reworded? What language is unclear? How do you define this word or phrase? What does it mean to you? How does the facilitation guide flow for you? What instructions remain unclear, complicated, or confusing? In what ways did the training prepare you for the interview? What was missing from the training from your perspective?

The responses were noted, reflected upon in preparation for the next semester of interviews, and implemented when updating materials.

LIMITATIONS

There were several limitations in this study. First, selection bias among student applicants and raters existed within the study. For example, selection bias existed among those interested in serving as peer leaders while in college as well as among those willing to be raters for SI program interviews. Second, the rubric could only evaluate what was covered during the group interview. While a comprehensive facilitator guide existed for the raters' use, the rubric could be used effectively only if all elements of the interview existed. A third limitation of this study was the overall utility of the rubric outside of peer leader selection, as well as the rubric's generalizability. The results of this study could vary in future studies based on the raters, the program type, or institutional context. A final limitation of this study is that demographic information was not collected on the raters nor applicants. Therefore, the influence of social identities is not known.

FUTURE DIRECTIONS

This study builds opportunities for future research and practice. With regard to research, we hope this work inspires practitioners and researchers alike toward generating new scholarship related to peer leader program practices. In particular, we recommend replication studies using this same rubric. We invite other institutions and peer leader programs to partner with us in this effort. We also recommend a future study that disaggregates the results by the social identities of both the raters and the applicants. While total scores are used in practice for hiring decisions, we are hesitant to provide a singular cutoff for all practitioners (e.g., must score above 22/28). Future research may consider the implications of such scoring decisions and their predictive utility.

We hope that an emphasis on process can better support and prepare practitioners to build selection processes for peer leaders in the field of higher education. To that end, we recommend more studies about peer leader selection processes including the validation of interview tools (e.g. rubrics, rating scales) that are used among peer leader programs. We further recommend discussion among practitioners about decision-making in developing and implementing selection processes within peer leader programs.

While we invite readers to use or adapt this rubric (see [Table 1](#)), especially for SI programs, we acknowledge that inter-rater reliability of this rubric within our institution's work may not translate to others' contexts. With this in mind, we encourage adaptation to other institutional or programmatic contexts. We ask that (1) any replication or adaption be shared back with the corresponding author to build a network of practitioners doing this work, and (2) appropriate

citations are included in your own work (e.g. Retrieved and Adapted from the Auburn University Academic Support Supplemental Instruction Program, 2026). Ultimately, we hope this work encourages practitioners to create and test their own rubrics for inter-reliability and other measures of validity. Finally, more conversation about how raters are prepared (e.g. trained, supported) to use rubrics for peer leader selection will help elevate practitioners' practices in the field.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Jillian P. Rager

Academic Support, Auburn University, US

Jaime S. Miller  orcid.org/0000-0002-6981-2889

Academic Support, Auburn University, US

Joseph M. Kush  orcid.org/0000-0003-0183-494X

IXL Learning, US

Deidre Leslie

Academic Support, Auburn University, US

Ashley C. Carr

Academic Support, Auburn University, US

REFERENCES

- Ansley, S., Miller, J. S., Rager, J., Brode, C., & Chaplin, C.** (2025). Dignity by design. In J. Tippetts & E. Kirby (Eds.), *ASCEND to higher retention rates: A collection of best practices and practical strategies from higher education experts for increasing student retention, completion, and student success* (Vol. 3). Tippetts & Kirby Consulting.
- Ashenafi, M. M.** (2017). Peer-assessment in higher education: Twenty-first century practices, challenges, and the way forward. *Assessment & Evaluation in Higher Education*, 42(2), 226–251. <https://doi.org/10.1080/02602938.2015.1100711>
- Berg, S. A., & Stoner, J. C.** (2016). Selecting resident assistants: The relationship between candidate assessment and job performance. *The Journal of College and University Student Housing*, 42, 26–36.
- Bowman, N. A., Preschel, S., & Martinez, D.** (2023). Does Supplemental Instruction improve grades and retention? A propensity score analysis approach. *Journal of Experimental Education*, 91(2), 205–229. <https://doi.org/10.1080/00220973.2021.1891010>
- Breslin, J. D., Kope, M. H., O'Hatnick, J. L., & Sharpe, A. G.** (2018). Students as colleagues: A paradigm for understanding student leaders in academic support. *The Learning Assistance Review*, 23(2).
- Carr, A. C.** (2018). *Supplemental Instruction Interview Rubric (SIIR)*. [Unpublished work].
- Chan, Z., & Ho, S.** (2019). Good and bad practices in rubrics: The perspectives of students and educators. *Assessment & Evaluation in Higher Education*, 44(4), 533–545. <https://doi.org/10.1080/02602938.2018.1522528>
- Cheung-Blunden, V., & Khan, S. R.** (2018). A modified peer rating system to recognise rating skill as a learning outcome. *Assessment & Evaluation in Higher Education*, 43(1), 58–67. <https://doi.org/10.1080/02602938.2017.1280721>
- Cohen, J.** (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46. <https://doi.org/10.1177/001316446002000104>
- Cohen, J.** (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. <https://doi.org/10.1037/h0026256>
- Colvin, J. W. & Ashman, M.** (2010). Roles, risks, and benefits of peer mentoring relationships in higher education. *Mentoring & Tutoring: Partnership in Learning*, 18(2), 121–134. <https://doi.org/10.1080/13611261003678879>
- Culpepper, D., White-Lewis, D., O'Meara, K., Templeton, L., & Anderson, J.** (2023). Do rubrics live up to their promise? Examining how rubrics mitigate bias in faculty hiring. *The Journal of Higher Education*, 94(7), 823–850. <https://doi.org/10.1080/00221546.2023.2168411>
- Cuseo, J.** (2010a). Peer leadership: Definition, description, and classification. *E-Source for College Transitions*, 7(5), 3–5. https://sc.edu/about/offices_and_divisions/national_resource_center/publications/search/details.php?id=1545

- Cuseo, J. (2010b). Peer leadership: Situation-specific support roles. *E-Source for College Transitions*, 7(6), 4–5. https://sc.edu/about/offices_and_divisions/national_resource_center/publications/search/details.php?id=1541
- Cuseo, J. (2010c). Peer power: Empirical evidence for the positive impact of peer interaction, support, and leadership. *E-Source for College Transitions*, 7(4), 4–6. https://sc.edu/about/offices_and_divisions/national_resource_center/publications/search/details.php?id=1552
- Ender, S. C., & Kay, K. (2001). Peer leadership programs: A rationale and review of the literature. In S. L. Hamid (Ed.), *Peer leadership: A primer on program essentials* (Monograph Series No. 32, pp. 1–11). The National Resource Center for The First-Year Experience and Students in Transition.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287–322. <https://www.jstor.org/stable/1170785?seq=1>
- Gannouni, K., & Ramboarison-Lalao, L. (2018). Leadership and students' academic success: Mediating effects of self-efficacy and self-determination. *International Journal of Leadership in Education*, 21(1), 66–79. <https://doi.org/10.1080/13603124.2015.1123300>
- Ganser, S. R., & Kennedy, T. L. (2012). Where it all began: Peer education and leadership in student services. *New Directions for Higher Education*, 157, 17–29. <https://doi.org/10.1002/he.20003>
- Gladding, S. (2012). Transactional analysis, reality therapy, Adlerian, and person-centered groups. In S. Gladding (Ed.), *Groups: A counseling specialty* (6th ed., pp. 321–351). Pearson.
- Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (4th ed.). Advanced Analytics.
- Hamid, S. L., & VanHook, J. (2001). First-year seminar peer leaders, programs, and profiles. In S. L. Hamid (Ed.), *Peer leadership: A primer on program essentials* (Monograph No. 32, pp. 37–52). University of South Carolina, The National Resource Center for The First-Year Experience and Students in Transition.
- Harper, S. R. (2010). An anti-deficit achievement framework for research on students of color in STEM. *New Directions for Institutional Research*, 2010(148), 63–74. <https://doi.org/10.1002/ir.362>
- Harper, S. R. (2012). *Black male student success in higher education: A report from the National Black Male College Achievement Study*. Center for the Study of Race and Equity in Education, University of Pennsylvania. [https://rossierapps.usc.edu/facultydirectory/publications/231/Harper%20\(2012\)%20Black%20Male%20Success.pdf](https://rossierapps.usc.edu/facultydirectory/publications/231/Harper%20(2012)%20Black%20Male%20Success.pdf)
- Hicks, D. (2021). *Dignity: Its essential role in resolving conflict*. Yale University Press. <https://doi.org/10.12987/9780300263503>
- Hunter, M. S., & Heath, M. M. (2001). The building blocks of the peer leader programs: Recruitment, selection, and training. In S. L. Hamid (Ed.), *Peer leadership: A primer on program essentials* (Monograph No. 32, pp. 37–52). The National Resource Center for The First-Year Experience and Students in Transition.
- Irwin, L. N., Miller, J. S., Morgan, K., & Linley, J. (2023). A critical ecological exploration of peer socialization agents' sense of belonging. *Journal of College Student Development*, 64(5), 509–523. <https://doi.org/10.1353/csd.2023.0911787>
- Janson, H., & Olsson, U. (2001). A measure of agreement for interval or nominal multivariate observations. *Educational and Psychological Measurement*, 61(2), 277–289. <https://doi.org/10.1177/00131640121971239>
- Janson, H., & Olsson, U. (2004). A measure of agreement for interval or nominal multivariate observations by different sets of judges. *Educational and Psychological Measurement*, 64(1), 62–70. <https://doi.org/10.1177/0013164403260195>
- Jarratt, L., Martin-Stanley, C. R. II, Linley, J. L., Mollet, A. L., Anderson, D. L., & Mohebbi, M. (2024). Underlife: Peer socialization agents' resistance to higher education diversity narratives. *Journal of Diversity in Higher Education*, 17(3), 430–442. <https://doi.org/10.1037/dhe0000431>
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144. <https://doi.org/10.1016/j.edurev.2007.05.002>
- Keller, P., & Porter, H. D. (2020). A terminological study of peer education in higher education [White paper]. College Reading and Learning Association. https://cdn.ymaws.com/crla.net/resource/resmgr/white_papers/crla_2020_whitepaper_peer_ed.pdf
- Keup, J., & Skipper, T. (2010, March 21–24). Findings from the 2009 National Survey of Peer Leaders. [Conference session]. American College Personnel Association Annual Convention, Boston, MA, United States.
- Klein, D. (2018). Implementing a general framework for assessing interrater agreement in Stata. *The Stata Journal: Promoting Communications on Statistics and Stata*, 18(4), 871–901. <https://doi.org/10.1177/1536867X1801800408>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>

- Leyzberg, D., Lumbroso, J., & Moretti, C.** (2017). Nailing the TA interview: Using a rubric to hire teaching assistants. *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education*, 128–133. <https://doi.org/10.1145/3059009.3059057>
- Linley, J. L.** (2015). *We are (not) all bulldogs: Minoritized peer socialization agents' critical sensemaking about collegiate contexts* [Doctoral dissertation, Michigan State University]. MSU Libraries Digital Repository. <https://d.lib.msu.edu/etd/3464>
- Linley, J. L.** (2017). We are (not) all bulldogs: Minoritized peer socialization agents' meaning-making about collegiate contexts. *Journal of College Student Development*, 58(5), 643–656. <https://doi.org/10.1353/csd.2017.0051>
- Linley, J. L.** (2018). Racism here, racism there, racism everywhere: The racial realities of minoritized peer socialization agents at a historically white institution. *Journal of College Student Development*, 59(1), 21–36. <https://doi.org/10.1353/csd.2018.0002>
- McDaniel, A.** (2008). Recruiting and training Supplemental Instruction Leaders. In M.E. Stone & G. Jacobs (Eds.), *Supplemental Instruction: Improving First-Year Student Success in High-Risk Courses* (pp. 39–56). University of South Carolina.
- Moore, C., & Teather, S.** (2013). Engaging students in peer review: Feedback as learning. *Issues in Educational Research*, 23(2), 196–211.
- Panadero, E., & Romero, M.** (2014). To Rubric or Not to Rubric? The Effects of Self-Assessment on Self-Regulation, Performance and Self-Efficacy. *Assessment in Education: Principles, Policy & Practice*, 21(2), 133–148. <https://doi.org/10.1080/0969594X.2013.877872>
- Perez, D., Ashlee, K., Do, V., Karikari, S., & Sim, C.** (2017). Re-conceptualizing student success in higher education: Reflections from graduate student affairs educators using anti-deficit achievement framework. *Journal on Excellence in College Teaching*, 28(3), 5–28.
- Popham, W. J.** (1997). What's wrong – and what's right – with rubrics. *Educational Leadership*, 55(2), 72–75. http://skidmore.edu/assessment/handbook/Popham_1997_Whats-Wrong_and-Whats-Right_With-Rubrics.pdf
- Rager, J. P., & Leslie, D.** (2023). *Adapted Supplemental Instruction Interview Rubric (SIIR)*. [Unpublished work].
- Reddy, Y. M., & Andrade, H.** (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435–448. <https://doi.org/10.1080/02602930902862859>
- Reyes, H. L., & Duran, A.** (2021). Higher education scholars challenging deficit thinking: An analysis of research informed by community cultural wealth. *Journal of Critical Scholarship on Higher Education and Student Affairs*, 6(1), 7–20.
- Rogers, C. R.** (1961). *On becoming a person: A therapist's view of psychotherapy*. Houghton Mifflin.
- Sambamurthy, N., & Cox, M. F.** (2016, June). *Developing a Behavior-Based Rubric for Assessing Professional Skills During Hiring*. [Paper presentation]. 2016 ASEE Annual Conference & Exposition, New Orleans, LA, United States. <https://doi.org/10.18260/p.27288>
- Shook, J., & Keup, J.** (2012). The benefits of peer leader programs: An overview from the literature. *New Directions for Higher Education*, 142(157), 5–16. <https://doi.org/10.1002/he.20002>
- Skalicky, J., Speed, H., van der Meer, J., & Young, D. G.** (2024). An international snapshot of peer leadership in higher education. *Journal of Leadership Education*, 24(1), 84–101. <https://doi.org/10.1108/JOLE-02-2024-0031>
- Taylor, B., Kisby, F., & Reedy, A.** (2024). Rubrics in higher education: An exploration of undergraduate students' understanding and perspectives. *Assessment & Evaluation in Higher Education*, 49(6), 799–809. <https://doi.org/10.1080/02602938.2023.2299330>
- The International Center for Supplemental Instruction.** (2019). *SI Supervisor manual*.
- The International Center for Supplemental Instruction.** (2022). *What is Supplemental Instruction?* <https://info.umkc.edu/si/>
- van der Meer, J., Skalicky, J., Speed, H., & Young, D. G.** (2022). Focusing on the development of the whole student: An international comparative study of the perceived benefits of peer leadership in higher education. *Open Journal of Social Sciences*, 10, 14–35. <https://doi.org/10.4236/jss.2022.103002>
- Witty, M. C., & Adomaitis, R.** (2014). Carl Rogers and client-centered counseling. In R. D. Parsons & N. Zhang (Eds.), *Counseling theory: Guiding reflective practice* (pp. 171–199). Sage. <https://doi.org/10.4135/9781483399621.n7>
- Zhang, B., Johnston, L., & Kilic, G. B.** (2008). Assessing the reliability of self- and peer rating in student group work. *Assessment & Evaluation in Higher Education*, 33(3), 329–340. <https://doi.org/10.1080/02602930701293181>

TO CITE THIS ARTICLE:

Rager, J. P., Miller, J. S., Kush, J. M., Leslie, D., & Carr, A. C. (2026). Interviewing Peer Leaders: Inter-Reliability of a Growth-Oriented Selection Rubric. *Journal of Peer Learning*, 17(1): 4, 1–13. <https://doi.org/10.21061/jopl.151>

Submitted: 31 July 2025

Accepted: 05 January 2026

Published: 11 August 2026

COPYRIGHT:

© 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <https://creativecommons.org/licenses/by/4.0/>.

Journal of Peer Learning is a peer-reviewed open access journal published by VT Publishing.