Graduate Student Self and Adviser Ratings on Professional Competencies

Mr. Bret Austin Arnold, University of Tulsa

Bret is a doctoral student of Industrial and Organizational Psychology at the University of Tulsa. His recent projects concern how personality shapes team-related behaviors and the degree to which antecedents of workplace burnout differ across cultures. Most recently, Bret has joined the University of Tulsa's cross-disciplinary STEM ProDev team. The team has recently designed and piloted a training program that develops the professional soft-skills of graduate engineering students.

Alison J. Kerr, University of Tulsa

Alison Kerr is a graduate student at The University of Tulsa. She is pursuing a doctoral degree in Industrial-Organizational Psychology. Her research interests include training development and evaluation as explored across a variety of academic disciplines and organizational settings. She is currently assisting on a number of training projects aimed at developing engineering students on relevant non-technical professional skills including ethical practice and presentation.

Dr. Bradley J. Brummel, University of Tulsa

Dr. Brummel is an Associate Professor of Industrial/Organizational Psychology at The University of Tulsa. He received his PhD from the University of Illinois at Urbana-Champaign. He conducts research on training and development with a specific focus on professional development, ethics, and coaching.

Dr. Michael W. Keller, University of Tulsa

Michael Keller is an associate professor of mechanical engineering at the university of tulsa. His research and teaching interests are in solid mechanics, both experimental and theoretical, and materials science.

Graduate Student Self and Advisor Ratings on Professional Competencies

Introduction

Professional societies, government agencies, and other organizations have issued numerous calls to strengthen the engineering workforce. However, this declaration of need does not answer the question of how. One possible direction focuses on improving the field relevant, but non-technical skills that help engineers as professionals. The authors of this paper have recently completed the first round of a pilot program that professionally develops graduate engineering students [1]. As a part of the program, students compare self-given competency ratings to those from select peers and their academic advisor. This multi-source feedback (MSF) approach to development gives participants a glimpse of their professional reputation from different angles. After all rater's submit their feedback, ratees meet with their advisors and create a development plan based on these results. Equipped with multi-source feedback, professional grade training tools, and a specially designed development plan, these engineers begin practicing the competencies possessed by high-performing, professional engineers.

This paper is comprised of two main components. First is a description of the program and the rating format was specially designed to reduce common sources of rater-error in subjective measurement. We discuss how behaviorally anchored rating scales (BARS) can provide a shared frame of reference among raters and how we constructed BARS for each of our nine competencies. Next, our discussion presents the initial feedback results in two lights. The first consults the data as a needs assessment of the piloted cohort. By looking at the competencies of highest and lowest ratings, we can infer their areas of strength and weakness. The second addresses differences in ratings based on rater type to illustrate the value of the MSF approach. Raters have different opportunities to observe different behaviors, and receiving feedback from multiple sources gives the ratee a glimpse of their professional reputation in each of these contexts. We conclude this paper with brief insights into how the multi-source feedback can guide their professional development goals.

Identifying Competencies

Soft skills like communication and teamwork are a necessary complement to the technical skills acquired in graduate training [2]. However, a survey of industry and business leaders of STEM fields recognized a gap between technical and professional competence [3]. In response, professional societies, government agencies, and other organizations have issued calls to strengthen the engineering workforce by developing these competencies. But while the recognition of this need has growing consensus, the best method for developing these skills remains unclear. To help address this gap in our own students, we designed a development program targeting these non-technical, soft skills.

The first step in program development was identifying which competencies to target. To articulate the soft skills valued by engineering professionals, we first created a competency model. Competencies describe one's knowledge, skills, abilities, or advantageous traits within a certain domain [4]. A competency model describes a desired set of employee attributes that help

the organization achieve strategic goals [5]. Once created, this model can be used to select, train, and develop employees. To create the model, we consulted the accreditation framework of the Accreditation Board of Engineering Technology (ABET), work assessment centers for managers, other competency frameworks [6], and faculty members at the Department of Engineering and Natural Sciences at the University of Tulsa.

After consulting with advisors of the engineering department, definitions of each competency were drafted and disseminated to the remaining engineering faculty for comment. Reactions were uniformly positive, and while some definitions were modified based on faculty input no competencies were added at this point. The final list is organized into three new categories as displayed in Table 1: communication-based, project-based, and interaction-based competencies. The full list of competencies with their final definitions can be found in the Table 2.

Table 1. Final set of professional competencies.

Communication	**Project**	**Interaction**
Oral Communication	Planning	Cultural Adaptability
Written Communication	Problem Solving	Teamwork
	Creativity	Leadership
		Conflict Management

Table 2. Competency list and definitions

Competency	**Definition**
Conflict Management	• Uses effective strategies for dealing with conflict • Recognizes and openly addresses conflict appropriately • Arrives at constructive solutions while maintaining positive working relationships
Creativity	• Develops and encourages novel ideas or solutions to problems • Acquires information from multiple sources and develops a clear perspective on an issue or topic • Anticipates future trends and assesses the likelihood and feasibility of possible responses
Cultural Adaptability	• Maintains a consistent standard of treatment toward all individuals • Values interaction with people from diverse backgrounds • Displays sensitivity to the needs, feelings, and viewpoints of others and expresses courtesy, neutrality, and respect
Leadership	• Guides, directs, and motivates others using regular, specific, and constructive feedback • Balances the interests, abilities, goals, and priorities of self and others with the needs of the group • Commands attention and respect while working toward goal achievement

Oral Communication	• Clearly conveys information with appropriate purpose and detail • Matches communication style with audience • Listens effectively and responds to input
Planning	• Prioritizes information and uses that information to set short and long-term goals • Monitors tasks and activities of self and others to ensure objectives are met and goals are accomplished • Accomplishes goals and completes work in one area without neglecting other projects
Problem Solving	• Recognizes problems and potential challenges in their work • Identifies solutions and evaluates costs and benefits of each • Makes timely decisions, plans course of action, and carries out action accordingly
Teamwork	• Values the contributions of all team members toward meeting the team objectives • Shares information and encourage others to do the same • Remains flexible within the dynamics of a group context and ~~can~~ works effectively with almost anyone
Written Communication	• Expresses thoughts clearly and succinctly across all written formats • Uses proper grammar and spelling • Follows a logical flow and has a developed sense of style

Measuring Competence and Reducing Error

After identifying the our target competencies, we determined our method of measurement. Because our rating process is analogous to performance appraisal, we turned to the research of Industrial and Organizational (I-O) Psychology to guide our approach. When applied at work, these ratings can be used for administrative purposes, like how to best allocate bonuses, or for developmental purposes, like drawing attention to strengths and weakness to improve performance [7]. An effective development program applies these ratings to create a low-pressure environment where ratees can identify their needs for improvement, set developmental goals in those areas, and achieve those goals with available resources. This captures the goal of our professional development program. Therefore, we draw on research from this approach to inform later decisions about interpretation and application of MSF ratings.

Another important decision about competency measurement is whether measurement should be subjective or objective. Subjective measurements are those provided by raters while objective measures are collected without a rater. Examples of objective measurements include the number of widgets produced, number of items sold, and the number of errors made. Unfortunately, many aspects professionalism elude quantitative measurement—consider cooperation with management or maintaining ethical standards. Consequently, objective measures can be quite restricted in scope. In contrast, subjective ratings allow raters to consider a broad range of reference points before making their assessment. This requires, however, a careful consideration to sources of rater error that contaminate subjective ratings. After considering both approaches, we determined subjective measures were most fitting and carefully considered the sources of rater-error detailed below.

Sources of Rater Error

If rater measurements were perfect, the scores provided by each rater would reflect only the ratee's degree of competence. In reality, ratings are subject to rater biases and distortions [8]. In extreme cases, subjective ratings are better measures of a rater's bias than the performance of the target [9]. We anticipated and identified these biases in order to improve the accuracy of the MSF ratings.

Leniency and halo bias are two common sources of error to consider before interpreting our pilot results. Leniency bias occurs when raters provide inflated ratings to their ratees. For example, a manager subject to leniency bias may rate over 90% of their employees as being "above average" [10]. A common motivation for leniency in MSF systems is to avert consequences of providing negative feedback to a superior [11], so we should note that higher scores provided by subordinates may indicate leniency. Furthermore, peers may be socially motivated to distort their reports to improve or maintain their relationships with the ratee [12], [10]. The second bias mentioned, halo bias, describes a situation in which a ratee's high score on one dimension leads to increased scores on other, unrelated dimensions. This occurs when a rater is unable to identify and differentiate between performance dimensions [13] or when high performance in one area (e.g. likability) skews the ratings of all other competencies. Our competency model and behaviorally anchored rating scales (BARS) carefully delineate the dimensions of each soft-skill to minimize the impact of halo bias. Furthermore, emphasizing the developmental nature of the feedback further reduces effects of halo or leniency bias [14]. This means our program is at least somewhat resilient to these biases simply by its nature.

Another source of error involves the opportunity of the raters to observe the rated behavior. Differences in opportunity to observe a behavior will manifest as differences in scores between rater types and is detectable using the MSF system. Identifying these difference in perception can be helpful for identifying the absolute competence of participants, but variation in perceptions is also of interest. For example, if a participant received high teamwork ratings from their supervisor, but low teamwork ratings from their peers, then this suggests that the ratee has a different reputation for teamwork among their peers than their advisor.

Reducing Error Through BARS

Behaviorally anchored rating scales (BARS) are instruments used to improve the accuracy of subjective performance measurements. Generally, raters are presented with a performance scale with points ranging from *Very Low* to *Very High*. Raters then select the point on the scale that best reflects the target's performance. To improve the accuracy of the rating, BARS include descriptions of each behavior that are consistent with a given rating. For example, rather than the rating of 1 being described as *Very Low* performance, BARS would describe in concrete terms what this level of job performance would look like. In the beginning of performance appraisal research, rating employee performance on an ordinal scale (e.g. 5-point scale) was not uncommon. Later research identified two considerable problems with this approach: raters must infer what behaviors are relevant and should therefore influence their rating, and raters must infer what each value on the scale represents [15]. . In short, BARS reduce rater error by reducing the number of inferences required of the rater [13].

The first inference is about what constitutes "performance". Performance is multidimensional in any role, but raters must consolidate a large amount of information in order to provide a single, unidimensional score. For example, simply asking raters to provide ratings on a single construct like professionalism or non-technical competence would result in a wide array of responses. Raters would likely have many different behaviors in mind when they provided their ratings. BARS address this by delineating performance dimensions (e.g. the nine competencies) and allowing raters to provide ratings on each. This allows reviewers to find in what areas the ratee is most and least effective.

The second problem is determining what constitutes "effective" performance. Even after performance is decomposed into smaller sub facets, the scales of each dimension may have unclear values. For example, rater's may be asked to creativity scores on a 7-item scale where 1 indicates low performance and 7 indicates high creativity. Determining what constitutes the difference between a 4 or 5, per se, falls to the rater. BARS resolve this problem by linking clearly observable behaviors to each value on the scale. An example of BARS designed for our program are displayed in Table 3.

In summary, a challenge of subjective ratings concerns the variability of interpretation made by each rater. BARS were introduced as a way of helping ratee's establish a shared frame of reference when rating employees by distinguishing different dimensions of performance and providing behavioral examples of each score. This was the rationale for introducing BARS for our multi-source feedback tool. However, like the competencies themselves, the BARS for our program must be individually constructed to match each competency. This process is detailed in previous ASEE paper [16] but the following two paragraphs summarize the process.

Creating our BARS

To determine the behavioral anchors of each rating scale, we collected information about the experiences of engineering faculty members. Researchers consolidated their anecdotal experiences into six behaviors for each competency [16]. These six behaviors were intended to capture various points on a 9-item scale.

The six items of each scale were then presented to a focus group of faculty members and graduate students from multiple departments (n=12). The focus group was asked to label the competency these behaviors represent and then order the behaviors based on the degree to which they reflect competence in that area. An example of the Oral Communication BARS is presented in Table 3, which displays the ratings for Oral Communication as they were presented to raters for the final project. We repeated this process to generate BARS for each competency.

Table 3. Final BARS for Oral Communication

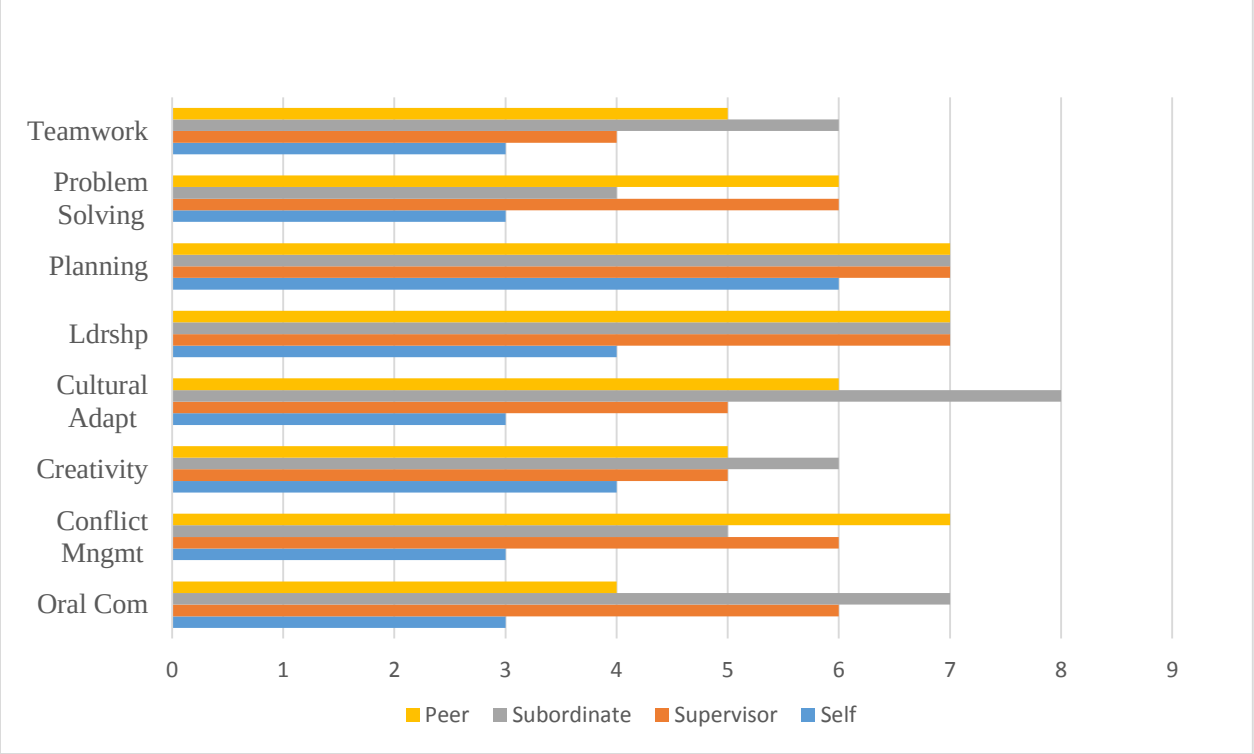
Rating	Anchor
1	•Is not at all comfortable speaking to or in front of others • Is unable to convey information clearly or coherently
2	
3	•Is somewhat able to convey information, especially in less formal settings, but may not be particularly comfortable doing so • Has difficulty expressing things concisely, may get caught up in details, but is able to get main ideas across
4	•Can convey information to others so that the audience understands the gist, but tends to ramble or struggle to get to the point in a concise way • Tends to use uncomfortable language or mannerisms and struggles to adjust communication style to suit the audience
5	
6	• Can convey information concisely and the content of the message is strong but does not consistently connect with the audience • May dominate conversations or meeting dialogue, but speaks in a clear and compelling manner in doing so
7	•Verbally conveys information with ease and in an engaging tone
8	
9	•Speaks clearly and concisely and is exceptionally engaging

Procedure

Engineering and Psychology faculty recruited graduate engineering students for the professional development program from the University of Tulsa. The creators of the program offered presentations and a question and answer session to explain the structure and mission of the project. The first pilot of the program involved 12 graduate students of engineering at the University of Tulsa. Participants were asked to invite a subordinate, a peer, and an advisor to participate in the multisource feedback report. Respondents varied in their number of raters (Max = 6, Min = 2, M = 3.29). Two students were removed from these results because they had only one rater. Every student was assessed by their academic advisor. In addition, participants selected several classmates to provide competency ratings. They were asked to select peers who had an opinion the ratees' respect, taken several classes with the ratee, listened to ratee presentations, read ratee writing, been involved with ratee lab research, and worked with the ratee on any form of the professional project.

An organization called Reliant Talent Management Solutions provided a platform for survey survey distribution. Raters provided scores for each ratee using the previously constructed BARS. In an open response section below each BARS, raters were asked to provide concrete examples of the ratee behavior to justify their ratings. The Reliant software then generated a report for each of the ten students that compared self, supervisor (advisor), peer, and subordinate ratings side by side. An example of the score comparison is presented in Figure 1.

Figure 1. Example of Participant Profile



Results

By organizing the competencies based on their mean scores, we gather a general sense of greatest strengths and weaknesses of the cohort. These are presented in Table 4. As a cohort, the participants received highest ratings on the competencies of Teamwork (*M* = 6.79, *SD* = 1.05) and Cultural Adaptability (*M* = 6.67, *SD* = 1.25) while the lowest ratings described the cohort's Creativity (*M* = 4.81, *SD* = 1.03) and Oral Communication (*M* = 5.11, *SD* = .87) skills.

Table 4. Means and SDs based on rater type

Rater Type	*M*	*SD*
Supervisor	5.77	1.18
Subordinate	5.70	0.97
Self	5.62	1.49
Peer	5.28	1.58
Total	5.68	1.10
Corrected Total	5.73	1.18

Note. The corrected total is the mean of all ratings provided by raters other than the ratee.

Seeing the trends among rater types allows us to find whether a particular type of rater is providing lenient or stringent scores relative to other rater types. Disproportionately high scores given by subordinates, for example, could indicate subordinates were lenient. However, this does not appear to be the case. The means and standard deviations of each rater type are presented in Table 5.

Table 5. Means and SDs based on rater type

Rater Type	*M*	*SD*
Supervisor	5.77	1.18
Subordinate	5.70	0.97
Self	5.62	1.49
Peer	5.28	1.58
Total	5.68	1.10
Corrected Total	5.73	1.18

Note. The corrected total is the mean of all ratings provided by raters other than the ratee.

The overall scores of each rater type are organized in descending order to aid comparison. On average, the supervisors gave higher ratings than peers. The corrected total average shows that when self-ratings are removed from the mean of scores, participant's self-ratings were generally in line with those provided by others.

Earlier we mentioned that some raters might have a greater opportunity to observe competency-related behaviors. To evaluate these potential discrepancies among raters, we computed the means and standard deviations between rater types within individual competencies. This data is presented in Tables 6 to 14. To bring attention to patterns of difference, each row of data is organized by the highest to lowest mean score given on the competency.

Discussion

The data presented here are entirely descriptive, the sample size precludes the use of inferential statistics to determine rating differences. However, the data from this pilot are valuable in at least two ways: they provide a glimpse of the cohort's overall reputation and how that reputation varies among different raters.

Table 6. Average Conflict Management scores by rater type

Rater	*M*	*SD*
Supervisor	6.11	0.99
Subordinate	6.00	1.00
Self	5.40	1.69
Peer	4.85	1.76

Table 7. Average Oral Communication scores by rater type

Rater	M	SD
Subordinate	6.00	0.00
Self	5.30	1.19
Peer	5.30	1.47
Supervisor	4.56	1.17

Table 8. Average Creativity scores by rater type

Rater	*M*	*SD*
Subordinate	5.50	0.50
Supervisor	5.11	1.29
Peer	4.95	2.05
Self	4.30	1.49

Table 9. Average Cultural Adaptability scores by rater type

Rater	*M*	*SD*
Supervisor	7.11	1.20
Self	6.80	1.78
Peer	5.55	1.62
Subordinate	5.00	1.00

Table 10. Average Leadership scores by rater type

Rater	*M*	*SD*
Self	5.70	1.49
Supervisor	5.67	1.33
Subordinate	5.50	2.50
Peer	4.65	1.55

Table 11. Average Planning scores by rater type

Rater	*M*	*SD*
Subordinate	7.00	0.00
Peer	6.00	1.55
Supervisor	5.67	1.15
Self	5.60	1.69

Table 12. Average Problem Solving scores by rater type

Rater	*M*	*SD*
Supervisor	6.22	1.03
Self	6.20	1.25
Peer	6.10	1.62
Subordinate	5.33	1.70

Table 13. Average Teamwork scores by rater type

Rater	*M*	*SD*
Supervisor	7.11	0.87
Self	6.00	1.79
Peer	5.70	1.40
Subordinate	5.50	1.50

Table 14. Average Written Communication scores by rater type

Rater	*M*	*SD*
Subordinate	5.50	0.50
Self	5.30	1.10
Peer	4.40	1.20
Supervisor	4.33	1.56

By looking at the overall mean scores on each competency (Table 4), we find that our cohort scored highest in the competencies of teamwork and cultural adaptability and the lowest scores were in oral communication and creativity. Taken as a whole, we might suggest to the cohort that oral communication and creativity are valuable targets for well-rounded development. However, despite the fact that these scores are on the same 1 to 9 point rating scale, the same score on two competencies does not necessarily indicate equal skill in both competencies. For example, a six in oral communication is not necessarily equivalent to a six in teamwork. Therefore, we caution against conclusions about relative competence. Another lesson to learn from overall competency scores is in the variance of overall scores. The standard deviation of scores within each competency indicated that students tended to be within one and one-half points from the mean of each competency—extreme scores were uncommon in our pilot sample.

Second, our descriptive statistics showed that raters of different types gave different scores. This can be especially helpful for professional development because these differences in scores among raters capture differences in professional reputation between their advisor, peers, and subordinates. For example, Table 5 shows the means and standard deviations of overall scores based on rater type. When the ratings for all competencies were averaged together, supervisors tended to give the highest score (M = 5.77, SD = 1.18) and the lowest scores were given by peers (M = 5.28, SD = 1.58). This is a notable finding because we might otherwise expect peers to be more socially motivated to provide lenient ratings of the participants but these averages suggest peer raters did not inflate their scores.

Perhaps more interestingly, trends can be observed within each competency by examining Tables 6 through 14. This organization of scores shows differences by rater type for each of the nine competencies. This helps us discern whether raters of certain types provided higher or lower scores than other types on particular competencies. This is especially helpful for the professional reputation if the cohort differs across rater types. For example, we can see whether participants have a different reputation for leadership among their subordinates compared to their advisor. Several notable patterns emerge in the three competencies with greatest mean differences: Teamwork, Cultural Adaptability, and Oral Communication. Recall that Teamwork and Cultural Adaptability were the competencies with highest mean scores and were considered overall strengths of our cohort. Now we see that this perspective was not shared among all raters.

First, when providing ratings for Teamwork, participants' self-ratings tended to give responses similar to those of their peers (M = 6.00, 5.70 respectively) but advisors provided higher ratings than both (M = 7.11). Put simply, ratee's appeared to have a better reputation for teamwork among their supervisors than among their peers. One explanation is that peers more frequently observe Teamwork-related behaviors than do supervisors.

A similar trend can be observed in the competency of Cultural Adaptability in Table 9. Supervisors tended to give the highest ratings (M =7.11) but subordinates (M = 5.00) and peers (M = 5.5) tended to provide lower ratings. Like before, this may reflect the advisor's opportunity to observe the ratee's culturally adaptive behaviors beyond the classroom. Peers and subordinates may provide ratings more consistent with their extracurricular behavior. Indeed, anchors for this scale include the likelihood to "seek out" cultural experience, a behavior unlikely to occur within the classroom.

Finally, when providing ratings for Oral Communication, subordinates (M = 6.00) provided higher scores than supervisors (M = 4.56). Clearly, the oral communication directed to advisors and subordinates will vary in complexity. For example, advisors may be referring to the participant's felicity in conference presentations while the subordinate is referring to emails or business writing which contain directives.

Conclusion

Several calls have been issued to develop professional soft-skills as a complement to technical graduate engineering work [2], [3]. Last year we responded to those calls by designing a professional development program [1], [16]. This paper provided a brief overview of the program's construction and content. We then discussed how our preliminary results showed the cohort's overall strengths and weaknesses, and also that their professional reputation for some competencies varies between subordinates, peers, and advisors. Insights like these will help participants interpret their reports, and design the best development goal for well-rounded growth.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant Number 1545211.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] Streets, V. N., Brummel, B. J., Keller, M. W., & Younis, R. M. (2017). Fostering graduate student professionalism using developmental coaching techniques. Proceedings of the American Society of Engineering Education Conference.

[2] Wendler, C., B. Bridgeman, F. Cline, C. Millett, J. Rock, N. Bell, & P. McAllister. 2010. *The Path Forward: The Future of Graduate Education in the United States.* Princeton, NJ: Educational Testing Service.

[3] Meier, R. L., M. R. Williams, & M. A. Humphreys. 2000. Refocusing our efforts: Assessing non-technical competency gaps. *Journal of Engineering Education, 89,* 377-385.

[4] Vazirani, N. 2010. Competencies and competency model: A brief overview of its development and application. *SIES Journal of Management, 7,* 121-131.

[5] Lucia, A. D. & R. Lepsinger. 1999. *The Art and Science of Competency Models,* San Francisco, CA: Jossey-Bass.

[6] Wherry, R. J., Sr., & Bartlett, C. J. (1982). The control of bias in ratings: A theory of rating. *Personnel Psychology, 35,* 521–551. doi:10.1111/j.1744-6570.1982.tb02208.x

[7] Dorfman, P., Stephan, W., & Loveland, J. (1986). Performance appraisal behaviors: Supervisor perceptions and subordinate reactions. *Personnel Psychology, 39*(3), 579–597.

[8] G. C. Thornton III, R. Hanson, and D. E. Rupp, Developing organizational simulations: A guide for practitioners and students: Psychology Press, 2003.
[9] Lance, C. E. (1994). Test of a latent structure of performance ratings derived from Wherry's (1952) theory of ratings. *Journal of Management, 20*, 751–771.
[10] Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives.* Thousand Oaks, CA: Sage.
[11] Spence, J. R., & Keeping, L. M. (2010). The impact of non-performance information on ratings of job performance: A policy-capturing approach. *Journal of Organizational Behavior, 31*, 587–608.
[12] Murphy, K. R., Cleveland, J. N., Skattebo, A. L., & Kinney, T. B. (2004). Raters who pursue different goals give different ratings. *Journal of Applied Psychology, 89*, 158–164.
[13] Saal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin, 88*, 413–428.
[14] Wong, K. F. E., & Kwong, J. Y. Y. (2007). Effects of rater goals on rating patterns: Evidence from an experimental field study. *Journal of Applied Psychology, 92*, 577–585.
[15] Schwab, D. P., G. Heneman III, & T. A. DeCotiis. 1975. Behaviorally anchored rating scales: A review of the literature. *Personnel Psychology, 28*, 549-562
[16] Keller, M. W., Brummel, B. J., Streets, V. N., Kerr, A. J., Younis, R. M., Tecle, L., & Crunkleton, D. W. (2017). Professional competencies with behaviorally anchored ratings for graduate students. Proceedings of the American Society of Engineering Education Conference.