

The Context of “Confidence”: Analyzing the Term Confidence in Resident Evaluations



Janae K. Heath, MD, MS¹ , Meagan E. Alvarado, MD², Caitlin B. Clancy, MD¹, Todd D. Barton, MD¹, Jennifer R. Kogan, MD¹, and C. Jessica Dine, MD, MS¹

¹Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA; ²Yale School of Medicine, New Haven, CT, USA.

BACKGROUND: Despite similar performance metrics, women medical trainees routinely self-assess their own skills lower than men. The phenomenon of a “confidence gap” between genders, where women report lower self-confidence independent of actual ability or competency, may have an important interaction with gender differences in assessment. Identifying whether there are gender-based differences in how confidence is mentioned in written evaluations is a necessary step to understand the interaction between evaluation and the gender-based confidence gap.

OBJECTIVE: To analyze faculty evaluations of internal medicine (IM) residents for gender-based patterns in the use of iterations of “confidence.”

DESIGN: We performed a retrospective cohort study of all inpatient faculty evaluations of University of Pennsylvania IM residents from 2018 to 2021. We performed n-gram text-mining to identify evaluations containing the terms “confident,” “confidence,” or “confidently.” We performed univariable and multivariable logistic regression to determine the association between resident gender and references to confidence (including comments reflecting too little confidence), adjusting for faculty gender, post-graduate year (PGY), numeric rating, and service.

SUBJECTS: University of Pennsylvania IM residents from 2018 to 2021.

KEY RESULTS: There were 5416 evaluations of IM residents (165 women [51%], 156 men [49%]) submitted by 356 faculty members (149 women [51%]), of which 7.1 % ($n=356$) contained references to confidence. There was a significant positive association between the mention of confidence and women resident gender (OR 1.54, CI 1.23–1.92; $p<0.001$), which persisted after adjustment for faculty gender, numeric rating, and PGY level. Eighty evaluations of the cohort explicitly mentioned the resident having “too little confidence,” which was also associated with women resident gender (OR 1.66, CI 1.05–2.62; $p=0.031$).

CONCLUSION: Narrative evaluations of women residents were more likely to contain references to confidence, after adjustment for numerical score, PGY level, and faculty gender, which may perpetuate the gender-based

confidence gap, introduce bias, and ultimately impact professional identity development.

KEY WORDS: assessment; gender bias; confidence gap; narrative evaluation.

J Gen Intern Med 37(9):2187–93

DOI: 10.1007/s11606-022-07535-z

© The Author(s), under exclusive licence to Society of General Internal Medicine 2022

Despite similar performance metrics, women medical trainees routinely self-assess their own skills lower than men.^{1–5} Compared to men, women medical students underestimate their abilities² and report more anxiety about performance during standardized patient encounters.¹ Women surgical trainees report lower operative confidence and perceived skill^{3–5} despite equal performance on objective assessments.⁶ These differences between performance and self-assessment and self-concept (or how an individual perceives themselves) have been referred to as a gender-based “confidence gap” and are not unique to medicine. In fact, women in diverse fields such as science, engineering, economics, athletics, and academia report lower self-esteem and self-confidence, independent of actual ability or competency.^{7–19} In academic medicine, the gendered confidence gap has been associated with differences in important career achievements including choice of jobs, negotiations, and applications for leadership positions.^{20–24} Given the association with longitudinal career achievements, an improved understanding of the factors contributing to the development and propagation of the gender-based confidence gap in medical trainees could inform efforts towards gender equity in medicine.

In addition to gender-based differences in self-assessment, gender-based differences have also been repeatedly demonstrated in faculty assessments of medical trainees, with women residents receiving lower competency ratings and different content in evaluations,^{25–29} despite no differences in actual skill.^{30–37} In medical training, there is a complicated but integral interaction between self-assessment and confidence with faculty evaluation of learner competence. For example, resident confidence can influence faculty perceptions of

Prior Presentations Results have previously been presented at the University of Pennsylvania Department of Medicine Research Day on June 10th, 2021 and at the Alliance for Academic Internal Medicine (AAIM) APDIM Online 2021 conference on September 24th, 2021.

Received November 30, 2021

Accepted March 29, 2022

Published online June 16, 2022

resident competence and ultimately shape feedback to and evaluations of learners.³⁸ When women receive verbal feedback about confidence (in the context of both too much confidence as well as too little confidence), they subsequently self-censor and devalue attributes stereotypically perceived as feminine.¹¹ This creates a vicious cycle negatively impacting self-concept, potentially further widening the confidence gap, perceived or real (Fig. 1).^{39–42} While verbal feedback about confidence has a known negative impact on a medical trainee's self-concept,¹¹ it is unknown whether subtle gender-based differences in narrative evaluation (such as differences in commenting on confidence) may also negatively impact self-concept. This, in turn, could further widen the "confidence gap" and may represent a previously unrecognized source of bias.

In order to address the ongoing gender disparities in resident assessment, and the potential impact on professional development and competency achievement, it is crucial to explore the relationship between competency, confidence (both actual and perceived), and assessment. It remains unclear whether word use in narrative evaluations, specifically the use of iterations of the term "confidence," plays a role in the development of a gendered confidence gap. Prior to identifying the perceived impact of these language differences on a medical trainee's self-assessment, it is necessary to understand if there is a differential presence of the word confidence in evaluations by gender. To our knowledge, there is no research exploring the differential use of iterations of the term confidence in faculty evaluations of residents. Therefore, we aimed to analyze end-of-rotation narrative evaluations of internal medicine (IM) residents for gender-based patterns in the use of the word "confidence" and its iterations.

METHODS

Setting and Participants

We performed a retrospective cohort analysis of all submitted narrative faculty evaluations of Hospital of the University of Pennsylvania internal medicine (IM) residents, medicine-pediatric residents, and medicine-dermatology residents from 2018 to 2021.

Due to differences in evaluation forms and processes between ambulatory and inpatient rotation evaluations, we

limited our analysis to inpatient evaluations. Inpatient faculty evaluations of residents consisted of 11 items rated on a 9-point entrustment scale (Supplemental Table 1), as well as a single mandatory free-text comment. Faculty completed evaluations at the end of each inpatient clinical rotation, usually representing one to 2 weeks of exposure to the resident, via an online evaluation platform. All evaluations were de-identified and all pronouns were removed from the narrative comments prior to analysis.

Data Collection and Processing

To analyze the free-text comments, we performed n-gram text-mining to convert each comment into discrete one-word phrases (unigrams). We counted the presence of the terms "confidence," "confident," or "confidently" (hereafter referred to as mentions of confidence) in each free-text comment, creating a matrix of binary variables for analysis (labeled "0" or "1" to denote the absence or presence of confidence in the comment, respectively).

Following the identification of unigrams, two authors (JH, CC) manually reviewed each evaluation to identify any mentions of confidence that did not refer to the resident. For example, "I am confident the resident is on track..." was recoded to identify mentions of confidence that only referenced the resident's attributes.

To further analyze the data, we used structured pile sorting,^{43,44} a qualitative research method used to group qualitative findings into "piles." Using this approach, we manually reviewed each comment for themes reflecting too little confidence as a negative attribute, too much confidence as a negative attribute, the growth or trajectory of confidence (neither negative nor positive, e.g., "growing confidence"), and confidence mentioned as a positive attribute (e.g., "excellent level of confidence"). The use of pile sorting enabled a manual count of the qualitative data, allowing for additional analysis. A single comment was generally sorted into only one of the themes but could be categorized into multiple themes. Any discrepancies were discussed as a group to ensure agreement in coding.

Data Analysis

We performed univariable analyses to determine the association between references to confidence (in all variations) and

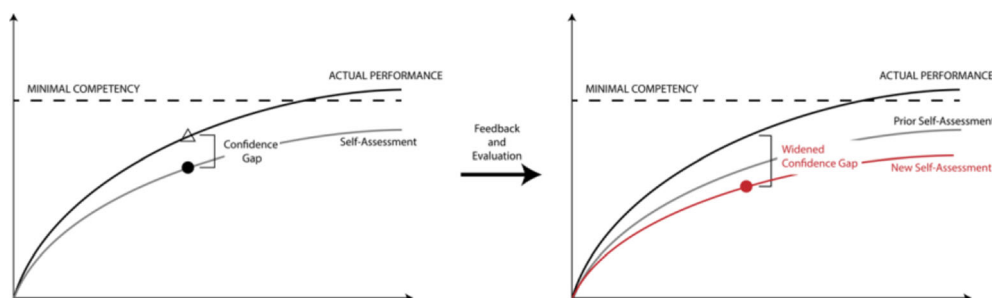


Figure 1. Conceptual framework of evaluator feedback on confidence in an individual's self-concept, and the resultant impact on the confidence gap.

resident gender, post-graduate year (PGY), numeric rating, and faculty gender. We also performed univariable analyses to determine if service acuity level impacted the presence of references of confidence. Service type is stratified into a binary variable of either ward rotation (which included all non-ICU-based rotations) or critical care rotation (which included all ICU-based rotations). For the purpose of this analysis, ratings of the 11 evaluation questions were averaged into a mean overall numeric score (creating an average of the evaluation subscales as a single score for use in the regression analysis). We used stepwise variable selection (inclusion threshold $p=0.2$)⁴⁵ and multivariable logistic regression analysis to determine associations between resident gender and references to confidence. We also analyzed the impact of faculty gender (and the interaction with learner gender) due to the potential for gender concordance of faculty and trainee to influence overall assessments.⁴⁶

We performed a subgroup descriptive analysis and multivariable logistic regression to identify associations with negative comments of confidence (both too little confidence and too much confidence), positive comments about confidence, and comments about the trajectory of confidence (each identified via pile sorting). We used logistic regression to determine associations between the theme of the comment and resident gender, PGY level, numeric rating, faculty gender, and rotation.

Given the a priori hypothesis that there may be gender differences in the perception of “quiet confidence” as a positive attribute, we used n-gram text mining to identify the presence of the terms “quietly confident,” and “quiet confidence.” We then performed an additional analysis looking for associations of “quiet confidence” with resident gender, adjusting for PGY level, numeric rating, faculty gender, and rotation.

Statistical analysis was completed using STATA version 15.1 (StataCorp, College Station, TX).

The institutional review board at the University of Pennsylvania approved the study.

RESULTS

Between 2018 and 2021, there were 5416 submitted evaluations of IM, medicine-pediatric, and medicine-dermatology residents (consisting of 165 women [51%] and 156 men [49%]). This included 295 residents (50% women) in the internal medicine program, 24 residents (63% women) in the medicine-pediatric program, and 2 residents (50% women) in the medicine-dermatology program. The evaluations were submitted by 347 unique faculty members (149 women [43%]). In total, there were 2138 evaluations [40%] submitted by women faculty in the cohort (Table 1). The mean numeric rating of the evaluations was 7.8 (SD 1.1) which did not differ by resident gender (7.81 versus 7.78, $p=0.30$). Supplemental

Table 1 Demographics of Cohort

	Total (n, %)
Completed evaluations	5416 (100%)
Mentions of confidence	356 (7.1%)
Mentions of too little confidence	80 (1.5%)
Mentions of too much confidence	7 (0.1%)
Faculty gender	
Written by men faculty	3278 (60%)
Written by women faculty	2138 (39%)
Resident gender	
Written about men residents	2619 (48%)
Written about women residents	2797 (52%)
Academic year of evaluation	
2018–2019	1853 (34%)
2019–2020	1509 (28%)
2020–2021	2054 (38%)
Post-graduate year (PGY)	
PGY1	2262 (42%)
PGY2	1671 (31%)
PGY3+	1483 (28%)
Mean rating on evaluation	7.8 (+/- 1.1)
Service	
ICU-based rotation	1722 (32%)
Non-ICU-based rotation	3694 (68%)

Figure 1 provides a graphic representation of the mean values of the ratings for the cohort.

In the cohort, 356 (7.1%) resident evaluations contained a reference to confidence. Of these, 219 (62%) occurred in evaluations of women residents. Evaluations containing references to confidence were associated with 16 unique women and 9 unique men. The majority of evaluations mentioning confidence were in internal medicine residents (95% [$n=340$]), with only 4% [$n=15$] in medicine-pediatric residents and less than 1% in medicine-dermatology residents [$n=1$].

There was a significant association between the mention of confidence in the narrative evaluation with resident gender, where confidence was mentioned more frequently in evaluations of women (OR 1.54, CI 1.23–1.92; $p<0.001$). A mention of confidence was also associated with a lower mean numerical rating (OR 0.88, CI 0.80–0.96; $p=0.005$) and an earlier PGY level (OR 0.81, CI 0.71–0.92; $p=0.002$).

The mention of confidence was not significantly associated with faculty gender (OR 1.18, CI 0.95–1.47; $p=0.131$) or ward versus ICU rotation (OR 0.83 CI 0.66–1.05; $p=0.121$). However, these associations were below the pre-specified p value for variable inclusion in the multivariable model and were included in the overall regression model (Table 2).

Table 2 Associations in Mentions of Confidence in Evaluations

Univariable regression	Odds ratio (CI)	p value
Resident gender (women)	1.54 (1.23–1.92)	<0.001
PGY level	0.81 (0.71–0.92)	0.002
Faculty gender (women)	1.18 (0.95–1.47)	0.131
Service (ICU)	0.83 (0.66–1.05)	0.121
Mean rating on evaluation	0.88 (0.80–0.96)	0.005
Multivariable regression	Odds ratio (CI)	p value
Resident gender (women)*	1.52 (1.33–1.90)	<0.001

*adjusted for PGY level, faculty gender, service, and mean rating on evaluation

After adjustment for mean numeric rating, PGY level, faculty gender, and service type, the association between resident gender and mention of confidence persisted (OR 1.52, CI 1.33–1.90; $p < 0.001$). There was no significant interaction effect between resident gender and PGY level, nor between faculty and resident gender.

Table 3 describes the pile sorting results with example comments. Using pile sorting, we identified 80 evaluations (22% of all mentions of confidence) that explicitly mention a lack of confidence as a negative attribute (Table 3). Evaluations of women residents were more likely to contain mentions of a lack of confidence (OR 1.66, CI 1.05–2.62; $p = 0.031$). This finding persisted after adjustment for PGY level, faculty gender, service, and mean numeric rating. Of the 56% of evaluations mentioning confidence as a positive attribute ($n = 198$), there was a significant association with resident gender (OR 1.69, CI 1.26–2.28, $p < 0.001$), with evaluations of women residents more likely to contain positive references

to confidence. Among these positive comments, there were no gender differences in the “quietly confident” phenotype ($n = 14$ [7% of positive comments], $p = 0.58$).

Seven evaluations (2% of all evaluations mentioning confidence) referenced overconfidence as a negative attribute, which was not significantly associated with resident gender (OR 0.70, CI 0.16–3.14; $p = 0.643$). The remainder of the comments discussed the perceived improvement of the resident’s confidence and overall growth trajectory ($n = 71$, 20% of all evaluations mentioning confidence), which was not significantly associated with resident gender (OR 1.04 CI 0.65–1.67, $p = 0.86$).

DISCUSSION

References to confidence were more likely to occur in narrative evaluations of women residents compared to men. This

Table 3 Pile sorting Results of Comments on Confidence with Representative Comments

Pile sorting theme	Evaluations	Representative comments
Mention of too little confidence as a negative attribute	Total evaluations 22% ($n = 80$) Evals of women 64% ($n = 51$) Evals of men 36% ($n = 29$)	“Curious and conscientious. Knowledgeable, but lacks in confidence.” “Solid JAR. Excellent empathy, very efficient. Constructive Comments: Speak up more on rounds, adding your ideas to the presentations of others. Try to show more confidence in your own differentials and plans.” “We worked together early in her intern year. She has really grown and developed. ... She needs to reframe how she is articulating her thoughts on rounds from: I’m not sure but I think it is X to My Ddx for this problem is A,B,C in this order. The “i’m not sure” sends a message to the listeners that she is not confident.”
Mention about trajectory and growth of confidence	Total evaluations 20% ($n = 71$) Evals of women 54% ($n = 38$) Evals of men 46% ($n = 33$)	“...worked very hard during her first inpatient medicine rotation. Over the course of the 2 weeks, she became more efficient and confident. She was thoughtful in asking the right questions and considering relevant data for decision making. She developed excellent rapport with her patients and kept them regularly updated on progress.” “... builds excellent rapport with patients and has a calm, yet clear bedside manner, particularly skilled in explaining complex medical care to a wide variety of patients. ... He is building his skills in developing independent assessments and plans for patients and I think his confidence will grow with more experience and pushing himself to be the primary decision maker whenever possible.” “...a great team member during our recent MICU sojourn. Over the course of the week, she grew in her confidence and ability to communicate the plan for very complex critically ill patients. Great attitude, super smart, great work ethic. All the ingredients to continue to thrive during residency.”
Mention of too much confidence as a negative attribute	Total evaluations 2% ($n = 7$) Evals of women 43% ($n = 3$) Evals of men 57% ($n = 4$)	“... is an excellent intern. She has a very solid fund of knowledge. She develops very good relationships with her patients and families. ... My suggestion for feedback would be that she be more open minded to constructive feedback and alternative views. She is often very confident in her assessments which although a strong suit, can be detrimental when she is not willing to accept alternative diagnostic possibilities.” “Performed like he was more experienced than he actually is. Very competent. Excellent presentations. Strong fund of knowledge. No deficiencies that I noted at all. A point of constructive feedback would be that he comes across as perhaps a bit too confident.” “... is a strong resident in terms of clinical skills. She was always engaged and motivated. I do encourage her to recognize that limits to those abilities are appropriate, even for senior residents. I do not want to convey the message that she should not be confident. Rather, I hope to cultivate that humility is desirable and can promote patient safety and be an effective leadership style.”
Mention of confidence as a positive attribute	Total evaluations 56% ($n = 198$) Evals of women 64% ($n = 127$) Evals of men 36% ($n = 71$)	“... is performing at or above expected level for a mid-year intern. He has excellent bedside manner and excels at building therapeutic physician patient relationships. He is very effective at counseling his patients on their daily management plans using patient friendly language and answers questions with confidence and understanding.” “Confident, intelligent, and caring. ... performed at the level of a junior resident. Was thorough, efficient and hard working.” “... was such a pleasure to work with. He is hardworking, eager to learn and grow, humble, but confident. He had many highly complicated patients with multi-system organ failure which he handled with aplomb.”

finding persisted after adjustment for mean numerical rating, PGY level, hospital service, and faculty gender. Women residents were more likely to be described as being underconfident compared to male residents.

While recent qualitative work highlights women perceive differences in feedback on confidence,¹¹ this is the first study to evaluate differences in the use of iterations of confidence within narrative evaluations of residents. This mirrors work in non-medical fields, describing a gender-based confidence gap (in the absence of a performance gap),^{7,10,39,41} and also aligns with findings of gender-based assessment differences within medical education. Specifically, prior work has identified linguistic differences in multiple settings along the medical education continuum, demonstrating qualitative gender-based differences in word use in letters of recommendation, narrative end-of-rotation evaluations of residents, and narrative evaluations of faculty.^{27,47–49}

Interestingly, we found that evaluations of women residents were more likely to contain references to confidence overall, including positive (as well as negative) mentions of confidence. This increased mention of confidence as a positive attribute may be surprising at face value, but we posit several potential explanations. Increased positive comments about confidence for women could represent evaluator efforts to balance stereotypical gendered attributes (i.e., praising women for displays of confidence). Alternatively, women residents may be more likely to self-identify confidence as an area they are actively working on based on prior feedback, thus shaping the evaluator's assessment (Fig. 1). Finally, evaluators may be more likely to mention confidence in evaluations of women because they perceive that trait as atypical and/or noteworthy in women, whereas confidence may be considered the expectation or norm for men. Overall, more work is needed to understand how gender informs resident interpretations of references to confidence in evaluations, and how this impacts self-concept, even when described as a positive finding.

It is worth noting that evaluations containing references to confidence were more likely to have lower associated mean numerical rating and be associated with a resident at an earlier stage of training. While this could indicate that references to confidence reflect lower perceived competence, it is important to note that overall we found no gender-based differences in performance ratings. The more frequent mentions of confidence (including low confidence) in women persisted despite PGY level and mean numeric rating, supporting that this was not a reflection of perceived competence alone.

Gender-based differences in references to confidence have significant implications, particularly on residents' self-concept.¹¹ Self-concept has been associated with ultimate career achievement, salary, and job performance^{7,8,23} and has consistently been found to be lower in women.^{22–24} Identifying subtle gender-based differences in word use in narrative evaluation in medical training could identify a previously unrecognized source of bias that potentially drives the gender-based "confidence gap." Importantly, this unrecognized source of

bias could ultimately impact the professional identity formation of women in academic medicine and their associated long-term career trajectories.

Overall, this study offers new insights and identifies potential bias within narrative evaluations of medical residents. How differential references to confidence in narrative evaluation are interpreted and incorporated into overall professional identity formation is an important future step. Additionally, the association between mentions of confidence with an ultimate career path, including choice of subspecialty training and leadership positions, would be an interesting area of further study. There is also a potential role for faculty development, aiming to minimize rater inferences to address these observed gender-based differences.

Although our study yielded a large sample size of evaluations, generalizability may be limited given this is a single institution study in a single specialty (internal medicine). Further evaluation in additional specialties, especially including fields with different amounts of gender diversity, would be particularly useful. Our cohort did not contain any individuals who identified as non-binary, gender-queer, or gender-fluid, but we recognize that the experience of these individuals is unique and an important area for further study. Additionally, in our analysis, we opted to focus on inpatient evaluations only given the differences in evaluation forms between settings, which further limits generalizability. We did not evaluate the association between mentions of confidence in narrative evaluation with ultimate self-perception of the individual, nor the perceptions or intent of the evaluator.

Finally, it is worth noting that the perception of confidence is an inference on the part of the evaluator, as confidence is an internal state. We are unable to differentiate whether comments about confidence were purely inferences on the part of the evaluator or whether trainees shared their own self-assessment of confidence level with their evaluator. Furthermore, we cannot know whether women were more likely to raise confidence in self-assessments or as an area for feedback from evaluators. Regardless, noting a difference in references to confidence in narrative evaluations is an important first step in this work. Further studies are needed to determine whether and how these gender-based differences in mentions of confidence in narrative evaluations ultimately impact learners and their trajectory.

CONCLUSIONS

Faculty evaluations of women IM residents were more likely to contain the term confidence (or its iterations), which persisted after adjustment for PGY level, faculty gender, service type, and mean numeric rating. Evaluations of women were more likely to contain both references to "too little confidence" and positive references to confidence.

Acknowledgments:

Contributors: None

Corresponding Author: Janae K. Heath, MD, MS; Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA (e-mail: janae.heath@penmedicine.upenn.edu).

Declarations:

Conflict of Interest: The authors declare that they do not have a conflict of interest.

REFERENCES

- Colbert-Getz JM, Fleishman C, Jung J, Shilkofski N. How do gender and anxiety affect students' self-assessment and actual performance on a high-stakes clinical skills examination? *Acad Med*. 2013;88(1):44-48.
- Sawdon M, Finn G. The 'unskilled and unaware' effect is linear in a real-world setting. *J Anat*. 2014;224(3):279-285.
- Hoops H, Heston A, Dewey E, Spight D, Brasel K, Kiraly L. Resident autonomy in the operating room: does gender matter? *Am J Surg*. 2019;217(2):301-305.
- Meyerson SL, Sternbach JM, Zwischenberger JB, Bender EM. The effect of gender on resident autonomy in the operating room. *J Surg Educ*. 2017;74(6):e111-e118.
- Fonseca AL, Reddy V, Longo WE, Udelsman R, Gusberg RJ. Operative confidence of graduating surgery residents: a training challenge in a changing environment. *Am J Surg*. 2014;207(5):797-805.
- Flyckt RL, White EE, Goodman LR, Mohr C, Dutta S, Zanotti KM. The use of laparoscopy simulation to explore gender differences in resident surgical confidence. *Obstet Gynecol Int*. 2017;2017.
- Aryana M. Relationship between self-esteem and academic achievement amongst pre-university students. *J Appl Sci*. 2010;10(20):2474-2477.
- Sterling AD, Thompson ME, Wang S, Kusimo A, Gilmartin S, Sheppard S. The confidence gap predicts the gender pay gap among STEM graduates. *Proc Natl Acad Sci*. 2020;117(48):30303-30308.
- Sterling A. Gender pay gap among STEM graduates. *PNAS*. 2020;117(48):29993-29994.
- Hutman E. The relationship between self-esteem and job performance (Order No. 1394458); 1999. Available from ProQuest Dissertations & Theses Global. (304578795). Retrieved from <https://proxy.library.upenn.edu/login?url=https://www.proquest.com/dissertations-theses/relationship-between-self-esteem-jobperformance/docview/304578795/se-2?accountid=14707>.
- Billick M, Rassos J, Ginsburg S. Dressing the Part: Gender Differences in Residents' Experiences of Feedback in Internal Medicine. *Acad Med*. 2022;97(3):406-413. <https://doi.org/10.1097/acm.0000000000004487>.
- Herbst THH. Gender differences in self-perception accuracy : the confidence gap and women leaders' underrepresentation in academia. *SA J Ind Psychol*. 2020;46(1):1-8. <https://doi.org/10.4102/sajip.v46i0.1704>
- Zander L, Höhne E, Harms S, Pfost M, Hornsey MJ. When Grades Are High but Self-Efficacy Is Low: Unpacking the Confidence Gap Between Girls and Boys in Mathematics. *Front Psychol*. 2020;11:552355. Accessed January 28, 2022. <https://www.frontiersin.org/article/10.3389/fpsyg.2020.552355>
- Wang Z, Wang Y, Redmiles D. Competence-Confidence Gap: A Threat to Female Developers' Contribution on GitHub. In: *2018 IEEE/ACM 40th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS)*. ; 2018:81-90.
- Sarsons H, Xu G. Confidence Men? Evidence on Confidence and Gender among Top Economists. *AEA Pap Proc*. 2021;111:65-68. <https://doi.org/10.1257/pandp.20211086>
- Exley CL, Kessler JB. The Gender Gap in Self-Promotion. *Q J Econ*. Published online January 21, 2022:qjac003. <https://doi.org/10.1093/qje/qjac003>
- Charness G, Rustichini A, van de Ven J. Self-confidence and strategic behavior. *Exp Econ*. 2018;21(1):72-98. <https://doi.org/10.1007/s10683-017-9526-3>
- Lerchenmueller MJ, Sorenson O, Jena AB. Gender differences in how scientists present the importance of their research: observational study. *BMJ*. 2019;367:l6573. <https://doi.org/10.1136/bmj.l6573>
- Hubble C, Zhao J. Gender differences in marathon pacing and performance prediction. *J Sports Anal*. 2016;2(1):19-36. <https://doi.org/10.3233/JSA-150008>
- Chang S, Guindani M, Morahan P, Magrane D, Newbill S, Helitzer D. Increasing promotion of women faculty in academic medicine: impact of national career development programs. *J Womens Health*. 2020;29(6):837-846.
- Palepu A, Carr PL, Friedman RH, Amos H, Ash AS, Moskowitz MA. Minority faculty and academic rank in medicine. *JAMA*. 1998;280(9):767-771.
- Marsh HW, Kuyper H, Seaton M, et al. Dimensional comparison theory: an extension of the internal/external frame of reference effect on academic self-concept formation. *Contemp Educ Psychol*. 2014;39(4):326-341. <https://doi.org/10.1016/j.cedpsych.2014.08.003>
- Pinxten M, De Fraine B, Van Damme J, D'Haenens E. Causal ordering of academic self-concept and achievement: Effects of type of achievement measure. *Br J Educ Psychol*. 2010;80(4):689-709.
- Michie F, Glachan M, Bray D. An Evaluation of Factors Influencing the Academic Self-concept, Self-esteem and Academic Stress for Direct and Re-entry Students in Higher Education. *Acad Psychol*. 2001;21(4):455-472. <https://doi.org/10.1080/01443410120090830>
- Holmboe ES, Huot SJ, Brienza RS, Hawkins RE. The association of faculty and residents' gender on faculty evaluations of internal medicine residents in 16 residencies. *Acad Med*. 2009;84(3):381-384.
- Rand VE, Hudes ES, Browner WS, Wachter RM, Avins AL. Effect of evaluator and resident gender on the American Board of Internal Medicine evaluation scores. *J Gen Intern Med*. 1998;13(10):670-674.
- Klein R, Ufere NN, Rao SR, et al. Association of Gender With Learner Assessment in Graduate Medical Education. *JAMA Netw Open*. 2020;3(7):e2010888-e2010888. <https://doi.org/10.1001/jamanetworkopen.2020.10888>
- Klein R, Julian KA, Snyder ED, et al. Gender bias in resident assessment in graduate medical education: review of the literature. *J Gen Intern Med*. 2019;34(5):712-719.
- Arkin N, Lai C, Kiwakyou LM, et al. What's in a Word? Qualitative and Quantitative Analysis of Leadership Language in Anesthesiology Resident Feedback. *J Grad Med Educ*. 2019;11(1):44-52. <https://doi.org/10.4300/jgme-d-18-00377.1>
- Kim C, McEwen LN, Gerzoff RB, et al. Is physician gender associated with the quality of diabetes care? *Diabetes Care*. 2005;28(7):1594-1598.
- Berthold H, Gouni-Berthold I, Bestehorn K, Böhm M, Krone W. Physician gender is associated with the quality of type 2 diabetes care. *J Intern Med*. 2008;264(4):340-350.
- Baumhäkel M, Müller U, Böhm M. Influence of gender of physicians and patients on guideline-recommended treatment of chronic heart failure in a cross-sectional study. *Eur J Heart Fail*. 2009;11(3):299-303.
- Andersen MR, Urban N. Physician gender and screening: do patient differences account for differences in mammography use? *Women Health*. 1997;26(1):29-39.
- Frank E, Dresner Y, Shani M, Vinker S. The association between physicians' and patients' preventive health practices. *CMAJ*. 2013;185(8):649-653.
- Franks P, Bertakis KD. Physician Gender, Patient Gender, and Primary Care. *J Womens Health*. 2003;12(1):73-80. <https://doi.org/10.1089/154099903321154167>
- Lurie N, Slater J, McGovern P, Ekstrum J, Guam L, Margolis K. Preventive Care for Women: Does the Sex of the Physician Matter? *Obstet Gynecol Surv*. 1994;49(2):102-103.
- Tsugawa Y, Jena AB, Figueroa JF, Orav EJ, Blumenthal DM, Jha AK. Comparison of hospital mortality and readmission rates for Medicare patients treated by male vs female physicians. *JAMA Intern Med*. 2017;177(2):206-213.
- Anderson C, Brion S. Overconfidence and the Attainment of Status in Groups. UC Berkeley: Institute for Research on Labor and Employment; 2010. Retrieved from <https://escholarship.org/uc/item/5zz0q2r0>.
- Kröner S, Biermann A. The relationship between confidence and self-concept—Towards a model of response confidence. *Intelligence*. 2007;35(6):580-590.
- Stankov L, Lee J, Luo W, Hogan DJ. Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learn Individ Differ*. 2012;22(6):747-758.
- Sander P, Sanders L. Understanding academic confidence. *Psychol Teach Rev*. 2006;12(1):29-42.
- Campbell JD. Self-esteem and clarity of the self-concept. *J Pers Soc Psychol*. 1990;59(3):538.

43. **Weller SC, Romney AK.** Systematic Data Collection. SAGE Publications, Inc; 1988:96.
44. **Ritchie J, Lewis J, Lewis PSPJ, Nicholls CMN, Ormston R,** Qualitative Research Practice: A Guide for Social Science Students and Researchers. SAGE Publications; 2013. <https://books.google.com/books?id=EQSIwAAQBAJ>.
45. **Heinze G, Wallisch C, Dunkler D.** Variable selection - A review and recommendations for the practicing statistician. *Biom J Biom Z.* 2018;60(3):431-449. <https://doi.org/10.1002/bimj.201700067>
46. **Stroud L, Freeman R, Kulasegaram K, Cil TD,** Ginsburg S. Gender Effects in Assessment of Clinical Teaching: Does Concordance Matter? *J Grad Med Educ.* 2020;12(6):710-716. <https://doi.org/10.4300/JGME-D-20-00145.1>
47. **Trix F, Psenka C.** Exploring the Color of Glass: Letters of Recommendation for Female and Male Medical Faculty. *Discourse Soc.* 2003;14(2):191-220. <https://doi.org/10.1177/0957926503014002277>
48. **Mueller AS, Jenkins TM, Osborne M, Dayal A, O'Connor DM, Arora VM.** Gender Differences in Attending Physicians' Feedback to Residents: A Qualitative Analysis. *J Grad Med Educ.* 2017;9(5):577-585. <https://doi.org/10.4300/JGME-D-17-00126.1>
49. **Heath JK, Weissman GE, Clancy CB, Shou H, Farrar JT, Dine CJ.** Assessment of Gender-Based Linguistic Differences in Physician Trainee Evaluations of Medical Faculty Using Automated Text Mining. *JAMA Netw Open.* 2019;2(5):e193520-e193520. <https://doi.org/10.1001/jamanetworkopen.2019.3520>

Publisher's Note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.