Calibration Bias and the Interpretation of Clinical Learning Environment Perceptions Surveys

David A. Deemer, MA
John M. Byrne, DO
Lawrence K. Loo, MD, MACP
David Puder, MD
Karina D. Torralba, MD, MACM
Sonny C. Lee, MD, FACP, FAAP
T. Michael Kashner, PhD, JD

ABSTRACT

Background The clinical learning environment (CLE) is frequently assessed using perceptions surveys, such as the AAMC Graduation Questionnaire and ACGME Resident/Fellow Survey. However, these survey responses often capture subjective factors not directly related to the trainee's CLE experiences.

Objective The authors aimed to assess these subjective factors as "calibration bias" and show how it varies by health professions education discipline, and co-varies by program, patient-mix, and trainee factors.

Methods We measured calibration bias using 2011–2017 US Department of Veterans Affairs (VA) Learners' Perceptions Survey data to compare medical students and physician residents and fellows (n = 32 830) with nursing (n = 29 758) and allied and associated health (n = 27 092) trainees.

Results Compared to their physician counterparts, nursing trainees (OR 1.31, 95% CI 1.22–1.40) and allied/associated health trainees (1.18, 1.12–1.24) tended to overrate their CLE experiences. Across disciplines, respondents tended to overrate CLEs when reporting 1 higher level (of 5) of psychological safety (3.62, 3.52–3.73), 1 SD more time in the CLE (1.05, 1.04–1.07), female gender (1.13, 1.10–1.16), 1 of 7 lower academic level (0.95, 1.04–1.07), and having seen the lowest tercile of patients for their respective discipline who lacked social support (1.16, 1.12–1.21) and had low income (1.05, 1.01–1.09), co-occurring addictions (1.06, 1.02–1.10), and mental illness (1.06, 1.02–1.10).

Conclusions Accounting for calibration bias when using perception survey scores is important to better understand physician trainees and the complex clinical learning environments in which they train.

Introduction

A critical component of medical education is the clinical learning environment (CLE) where trainees engage in supervised patient care to acquire competencies necessary to enter independent practice.[1] To evaluate CLEs for program accreditation, faculty evaluations, and program rankings, education leaders turn to perceptions surveys,[2] such as the AAMC Medical School Graduation Questionnaire[3] and the Accreditation Council for Graduate Medical Education (ACGME) Resident/Fellow Survey.[4] These surveys ask respondents to rate items on 5-point scales (satisfaction, agreement, excellence), with item responses grouped into domains reflecting CLE constructs such as supervision, interaction with faculty, clinical experience, scut work, research opportunities, working environment, personal experiences, and professionalism.[2,5,6] While perception surveys can reflect CLE qualities, critics charge that responses may also vary with how questions are framed,[7] surveys are designed,[8] and response options are quantified.[9] Importantly, respondents' subjective characteristics,[10] including personality traits, perceptions of personal support, peer morale, and autonomy,[11–13] and how respondents retrieve information, make judgments, and interpret survey questions,[14–16] have also been shown to impact perception survey responses.

In this study, we propose a theoretical framework that defines calibration bias as the difference between a trainee's self-reported rating from that of an actual rating had the trainee responded with the subjective characteristics of the average trainee respondent. If calibration bias were controlled, trainees would rate the same experience in exactly the same way. Well-validated data from the Department of Veterans Affairs (VA) national CLE surveys[10] are used to approximate calibration bias and assess if: (1) such biases exist with well-validated survey data, and if so, (2) does calibration bias vary by discipline and (3) by trainee and CLE factors.

Conceptual Model

Derived from the 9-criteria evaluation framework,[10] FIGURE 1 shows calibration bias as a mediator between the CLE as the object to be assessed, and domain scores used to assess the CLE. The bias is a result of subjective[17–19] factors that impact how a respondent's experiences are perceived, and threshold[9,20] factors

DOI: http://dx.doi.org/10.4300/JGME-D-20-00237.1

that impact how respondents value the 5-point response options they must select to rate those perceptions. These biases can lead respondents to over- or under-rate experiences compared to an "average" rater who, by definition, has no calibration bias. Based on our framework, possible remedies include changes in survey design, administration, scoring, and analyses.

Calibration bias is not directly observable. To test for its presence in validated data, we created a calibration index where respondents rate their satisfaction with selected CLE facility-level "calibrating items," such as parking, facility location, and electronic health record (EHR), where experiences are not likely to vary among trainees reporting on the same facility and academic year. The index equals the respondent's calibrating item score minus the average of all such scores from respondents to the given facility and academic year. If calibrating item experiences are invariant, then from FIGURE 1 any variation in index scores must be the result of mediating subjective and threshold factors. A second test does not depend on strict invariant item experiences. As shown in FIGURE 1, associations between the index score and trainee, patient, program, and other facility-level factors that are not expected to impact a trainee's calibrating item experiences, can only be observed if the respondent answered the survey in the presence of calibration biases and such biases are influenced by such factors.

What was known and gap
While subjective factors are believed to influence how residents and fellows rate their clinical learning environments, how to calibrate for these influences when using such ratings to rank programs by their performance is not well understood.

What is new
We measure calibration bias and show how biases vary by discipline, the trainee's program and facility factors, and the mix of patients that trainees see.

Limitations
Study data were limited to the Department of Veterans Affairs medical centers and to a limited set of predictor factors.

Bottom line
Educators must integrate calibration bias metrics into their perceptions surveys results in order to better understand their residents and fellows and the complex clinical learning environments in which they train.

Methods

Data Setting and Sample

Data came from the Department of Veterans Affairs' (VA) Learners' Perceptions Survey (LPS) for physician, nursing, and allied and associated health trainees,[21] collected from July 1, 2010, through August 30, 2017. Validated elsewhere,[10] LPS is an anonymous, voluntary, Office of Management and Budget approved, web-based perceptions survey administered annually to trainees who rotate through a VA medical center as part of a required curriculum for an accredited health professions education program. LPS respondents were solicited through advertising, capturing only 11% of all VA trainees. However, LPS findings have been well-published,[10] with physician resident respondents shown to be comparable by specialty, academic level, international status, and gender with US physician residents in ACGME-accredited non-pediatric and non-OB-GYN programs.[22]

Calibration Bias

Calibration bias is estimated by a proxy index based on how respondents rated their satisfaction with 3 calibrating items on a 5-point scale. Calibrating items are parking, location convenience, and EHRs. Item responses are scored as 1 for "very dissatisfied," 2 "somewhat dissatisfied," 3 "neutral," 4 "somewhat dissatisfied," and 5 "very satisfied." The index, C_{index}, is computed by taking the average of the 3 calibrating item scores and subtracting the mean of such averages computed for all survey respondents at the given facility.[22,23] C_{index} is also computed in standard deviates ($Cz = \frac{C_{index} - \mu}{\sigma}$) and binary ($C_{binary} = 1$ if $C\ z > 0$, and $= 0$ if $Cz \leq 0$). Positive C_{index} values indicate trainees whose subjective factors put them at risk of overrating their experiences compared to that of an average respondent, while negative values indicate trainees at risk of underrating their experiences.

The psychometric properties of C_{index} have been estimated for VA trainees before.[10] Calibration index values were found to have a mean (–0.06), range (–3.60 to 2.00), SD (0.84), facility-level clustering ICC (0.05), and test-retest reliability (ICC = 0.86). We also reported modest scalability ($H = 0.38$) and consistency (Cronbach's alpha = 0.59) among the 3 calibrating items. This is not surprising, as parking and location fall under working environment, and EHR falls under clinical environment. Calibration bias is expected to reflect the subjective properties of trainees, so combining different items together is tantamount to measuring illness severity by counting comorbidities, even though such diseases are clinically distinct and unrelated.[24]

Covariates

Trainee and CLE covariates were computed from LPS survey responses, previously shown to have high consistency (Cronbach's alpha) and test-retest reliability.[10] Trainee covariates included professional

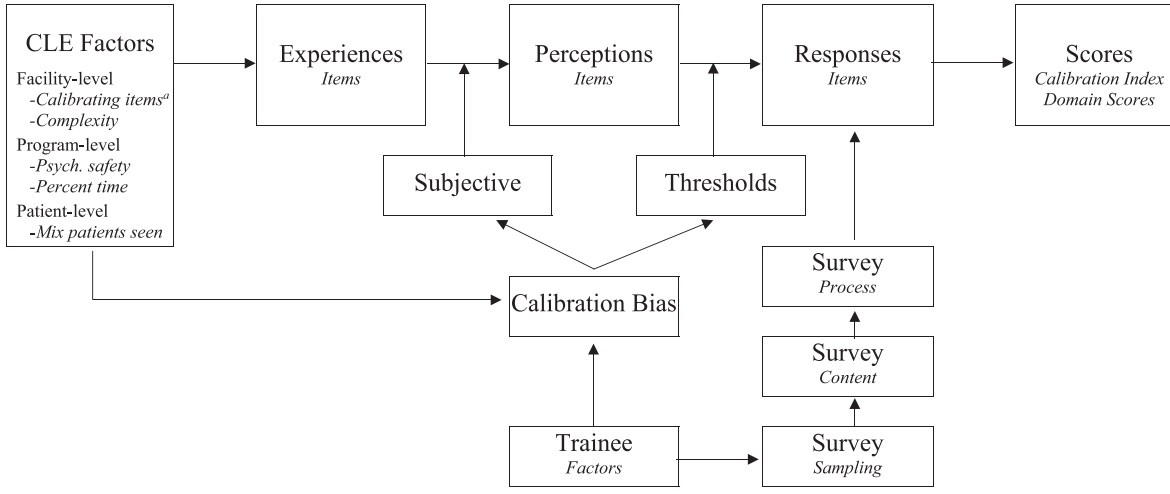

FIGURE 1
Role of Calibration Bias on Relationship Between Clinical Learning Environment and Trainee Perceptions Survey Scores

[a] Facility-level calibrating items for this example include parking, convenience of the facility location, and electronic health record. Respondent satisfaction with these calibrating items are scored as the calibration index.

discipline across 26 professions, academic level in years since high school, and gender. CLE covariates included the percent of time the trainee spent in VA, psychological safety[25] computed based on a 5-point agreement to: "Members of the clinical team of which I was a part are able to bring up problems and tough issues"; a 5-point VA facility service complexity score[26]; and mix of patients seen ranked into terciles by discipline, for patients "age 65 and over," with "chronic mental illness," "chronic medical illness," "multiple illnesses," "substance dependence," and "low income," and those who "lacked social support."

Analyses

Independent associations regressing calibration index on trainee and program factors were estimated using SPSS generalized linear models with an identity link function and Gaussian distribution for C_{index} and C*z*,

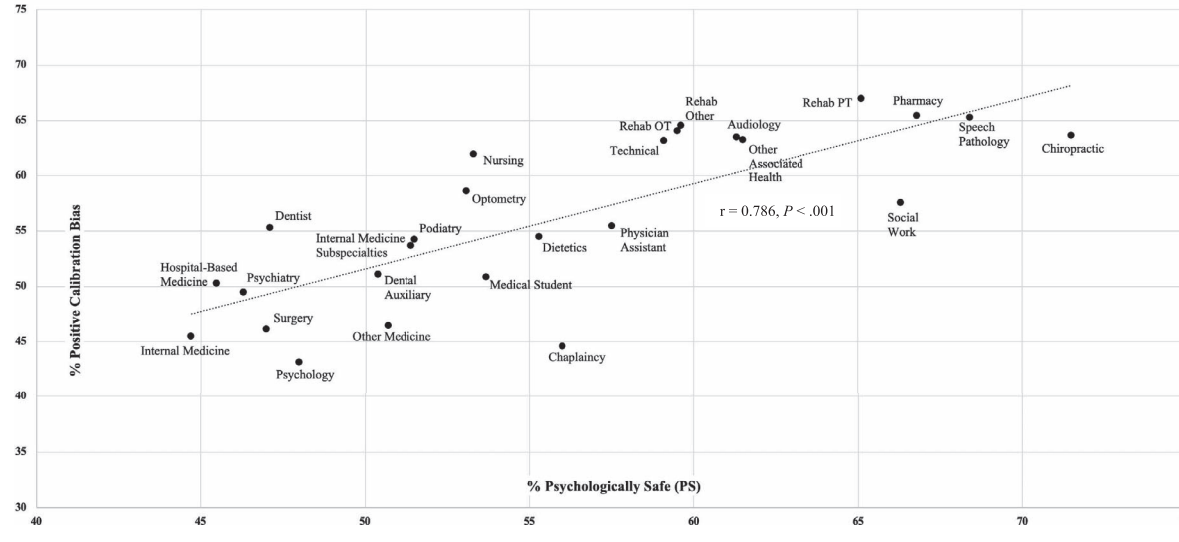

FIGURE 2
Percent Trainees Reporting Positive Calibration Index Values[a] and Psychologically Safe[b] Clinical Learning Environment by Professional Discipline

[a] Percent of respondents with C*z* > 0 (above sample mean).

[b] Percent of respondents who "strongly agreed" that their VA clinical learning environment was psychologically safe.

and a logit link function and binomial distribution for C_{binary}.

Results

Table 1 describes sample means, SDs, and frequencies of all study variables. C_{index} ranged from –3.563 to 1.628 with an SD of 0.7966 index values, consistent with theory that calibration biases exist and vary by trainee.

Table 2 shows how calibration, computed as the percent of respondents with positive index values ($Cz > 0$), varied by discipline ($P < .001$), ranging from 43.0% (psychology) to 66.9% (physical therapy). Table 3 shows allied and associated health trainees and nursing trainees were 17.8% and 30.5%, respectively, more likely to overreport favorable ratings ($C_{binary} = 1$) and had higher average index scores by 0.085 and 0.142 standard deviates, compared to their physician counterparts. These findings are consistent with our hypothesis that calibration biases vary by academic discipline.

Table 2 shows psychological safety, computed as the percent of trainees who strongly agreed that their CLE was psychological safe, was associated with calibration, on a discipline by discipline basis (except chiropractic). As with calibration, psychological safety also varied by discipline ($P < .001$). Figure 2 shows disciplines with a higher percentage of trainees who had a positive calibration index value also had a higher percentage of trainees who strongly agreed their CLE was psychologically safe ($r = 0.786$, $P < .001$). Table 2 also reveals that the associations between calibration and psychological safety varied by discipline ($P < .001$). Across disciplines, the size of the association between calibration and psychological safety was negatively correlated with the percent of a discipline's trainees who strongly agreed their CLE was psychologically safe ($r = –0.564$, $P = .003$).

Table 3 describes the independent association across all study variables between trainee and CLE factors and calibration based on how bias was scored. Psychological safety had overwhelmingly the largest association: one level increase in psychological safety 5-point scale was associated with an average increase in calibration (Cz) of 0.476 standard deviates. Calibration was also positively associated with lower academic level, female gender, percent of time the trainee was in VA, more complex facilities, and fewer patients the respondent sees than expected for the respondent's discipline with chronic mental illness, chronic medical illness, alcohol/substance dependence, low income, and lack social/family support.

TABLE 1
Sample Demographics

Item	No.[a]	Percent[a]
Sample	89 711	100.0%
Calendar year		
2011	11 558	12.9%
2012	12 718	14.2%
2013	13 121	14.6%
2014	13 465	15.0%
2015	13 327	14.9%
2016	11 863	13.2%
2017	13 659	15.2%
Trainee characteristics		
Trainee professional discipline		
Resident physicians	32 830	36.6%
Nursing	29 758	33.2%
Allied and associated health	27 092	30.2%
Academic level		
Pre-baccalaureate	17 775	19.8%
Baccalaureate	14 488	16.1%
Masters	9111	10.2%
Doctoral-1	8521	9.5%
Doctoral-2	12 943	14.4%
PostDoctoral-1	19 040	21.2%
PostDoctoral-2	7796	8.7%
Gender (female)	58 266	64.9%
CLE: Program		
Percent time in VA	mean 58.8%	SD 39.4
Psychologically safe[a]		
Strongly agree	48 280	53.8%
Agree	31 933	35.6%
Neither agree nor disagree	6839	7.6%
Disagree	1819	2.0%
Strongly disagree	840	0.9%
CLE: Facility		
Facility complexity		
Most complex	37 244	41.5%
Very complex	18 866	21.0%
Complex	17 270	19.3%
Medium complexity	10 173	11.3%
Least complex	6158	6.9%
CLE: Patients		
Mix of patients seen	mean	SD
Age 65 and over	67.3%	28.0%
Chronic mental illness	45.5%	30.8%
Chronic medical illness	73.0%	25.2%
Multiple conditions	74.4%	24.1%
Substance abuse/dependency	44.2%	27.1%
Low socioeconomic	48.6%	25.5%
Lack social support	39.7%	25.1%
Calibration Index (C_{Index})	−0.0161	0.7955

[a] Continuous data subsets report mean and SD. For example, our sample spent a mean 58.8% of their time at VA training sites; the percentage of patients aged 65 and over seen by the respondent was, on average, 67.3%.

TABLE 2
Percent Trainees with a Positive Calibration Index Score and a High Psychological Safety Rating, and Their Associations, by Discipline

Sample	Total	Cz > 0[a]	High PS Rating[b]	Cz - PS Association[c]			
		n (%)	n (%)	b	95% CI	Wald χ^2	*P*
Total	89 680	50 478 (56.3)	48 266 (53.8)	0.49	0.48, 0.50	10 810.0	< .001
Psychology	2854	1231 (43.1)	1370 (48.0)	0.27	0.22, 0.32	141.1	< .001
Chaplaincy	375	167 (44.5)	210 (56.0)	0.33	0.22, 0.43	34.4	< .001
Internal medicine	8435	3827 (45.4)	3771 (44.7)	0.59	0.55, 0.62	1282.0	< .001
Surgery	3845	1771 (46.1)	1807 (47.0)	0.64	0.59, 0.68	813.7	< .001
Other medicine	3098	1438 (46.4)	1571 (50.7)	0.50	0.44, 0.55	350.1	< .001
Psychiatry	2242	1107 (49.4)	1038 (46.3)	0.48	0.43, 0.53	333.3	< .001
Hospital-based medicine	2899	1454 (50.2)	1319 (45.5)	0.59	0.54, 0.64	536.9	< .001
Medical student	9436	4794 (50.8)	5069 (53.7)	0.52	0.50, 0.55	1327.6	< .001
Dental auxiliary	500	255 (51.0)	252 (50.4)	0.50	0.39, 0.61	82.9	< .001
Internal medicine subspecialty	2875	1540 (53.6)	1478 (51.4)	0.65	0.60, 0.69	709.5	< .001
Podiatry	790	428 (54.2)	407 (51.5)	0.50	0.42, 0.58	150.0	< .001
Dietetics	932	507 (54.4)	515 (55.3)	0.40	0.31, 0.49	73.4	< .001
Dentist	717	396 (55.2)	338 (47.1)	0.35	0.27, 0.43	78.7	< .001
Physician assistant	1003	556 (55.4)	577 (57.5)	0.48	0.39, 0.57	103.5	< .001
Social work	2928	1684 (57.5)	1941 (66.3)	0.37	0.31, 0.42	180.4	< .001
Optometry	1766	1034 (58.6)	938 (53.1)	0.42	0.36, 0.48	203.1	< .001
Nursing	29 758	18 415 (61.9)	15 855 (53.3)	0.46	0.44, 0.47	3432.7	< .001
Technical	1429	901 (63.1)	845 (59.1)	0.41	0.34, 0.47	134.4	< .001
Other associated health	2151	1360 (63.2)	1323 (61.5)	0.43	0.36, 0.49	188.8	< .001
Audiology	617	391 (63.4)	378 (61.3)	0.37	0.27, 0.46	53.9	< .001
Chiropractic	151	96 (63.6)	108 (71.5)	0.32	-0.02, 0.67	3.4	.066
Rehab occupational therapy	856	548 (64.0)	509 (59.5)	0.33	0.22, 0.43	36.7	< .001
Rehab other	406	262 (64.5)	242 (59.6)	0.42	0.29, 0.60	37.6	< .001
Speech pathology	535	349 (65.2)	366 (68.4)	0.29	0.19, 0.39	30.2	< .001
Pharmacy	7487	4900 (65.4)	5001 (66.8)	0.46	0.42, 0.50	534.5	< .001
Rehab physical therapy	1595	1067 (66.9)	1038 (65.1)	0.36	0.29, 0.43	93.5	< .001
Between sample		$\chi^2(25) = 1505.4$ $P < .001$	$\chi^2(25) = 1983.6$ $P < .001$	$\chi^2(25) = 359.1$ $P < .001$			

[a] Calibration Index scores [C index] are computed so higher scores indicate respondents who are more likely to give favorable clinical learning environment (CLE) domain ratings. Reported is the percent of respondents within a given discipline who over-reported favorable ratings, or Cz > 0.
[b] Psychological safety on the 5-point ordinal scale, where higher values indicate more psychological safety. Percentage reported is the percent of respondents who "strongly agreed" that their VA CLE was psychological safe.
[c] Cz and PS associations 'b' were computed as the average change in calibration bias standard deviates Cz per one level increase in the 5-point psychological safety scale.

Discussion

Our findings highlight the importance of accounting for calibration bias when interpreting CLE perceptions surveys scores. Calibration bias is viewed as subjective and threshold factors mediating between a trainee's CLE experience and their satisfaction rating of that CLE. We measured bias severity using an index scored that averages how trainees rated the 3 calibrating items, mean-centered by facility and academic year. We observed these index values varied by trainee and discipline, suggesting calibration biases exist, but only to the extent trainee experiences were invariant by facility and academic year. On exploratory analyses, we found patient, program, and trainee factors that were not expected to impact trainee experiences with parking, location, and EHR were consistently associated with index values, suggesting the presence of calibration biases.

Our findings are consistent with studies that have shown adjusting for C_{index}, under different names, has led to significant changes in reported results when assessing primary care continuity clinics,[27]

TABLE 3
Independent Associations Between Trainee and CLE Factors and Calibration Index, by How Calibration Is Scored[a]

Item	*z*-score Cz[b]				C_{binary}[c]			
	SD	95% CI	Wald χ^2	*P*	OR	95% CI	Wald χ^2	*P*
Trainee factors								
Trainee discipline								
Allied and associated health	0.085	0.065, 0.106	64.8	< .001	1.178	1.122, 1.236	43.3	< .001
Nursing	0.142	0.111, 0.173	82.5	< .001	1.305	1.215, 1.404	52.5	< .001
Physicians (referent)	0.000				1.000			
Academic level[d]	-0.025	-0.031, -0.019	65.0	< .001	0.948	0.934, 0.962	54.0	< .001
Gender (female)	0.051	0.038, 0.065	52.6	< .001	1.126	1.091, 1.157	52.2	< .001
CLE factors								
Program								
Percent of time in VA[e]	0.029	0.023, 0.035	80.8	< .001	1.049	1.035, 1.065	44.9	< .001
Psychological safety[f]	0.476	0.467, 0.485	10 271.4	< .001	3.623	3.521, 3.731	7910.8	< .001
Facility								
Facility complexity								
Least complex	-0.065	-0.087, -0.042	31.9	< .001	0.999	0.942, 1.059	0.0	.96
Medium complexity	-0.059	-0.079, -0.040	34.8	< .001	0.921	0.878, 0.965	11.8	.001
Complex	-0.003	-0.019, 0.014	0.1	.76	1.002	0.937, 1.042	0.0	.90
Very complex	0.000	-0.017, 0.016	0.0	.99	1.005	0.967, 1.044	0.1	.81
Most complex (referent)	0.000				1.000			
Patients								
Mix of patients seen[g]								
65 or older								
Highest tercile	-0.014	-0.029, 0.000	3.6	.06	0.973	0.939, 1.007	2.5	.12
Lowest tercile	-0.015	-0.033, 0.002	3.0	.08	0.954	0.916, 0.995	4.9	.027
Chronic mental illness								
Highest tercile	0.008	-0.009, 0.024	0.9	.35	1.004	0.966, 1.043	0.0	.83
Lowest tercile	0.024	0.008, 0.041	8.6	.003	1.056	1.016, 1.096	7.7	.006
Chronic medical illness								
Highest tercile	0.025	0.009, 0.042	9.2	.002	1.078	1.037, 1.121	14.3	< .001
Lowest tercile	0.032	0.012, 0.051	10.3	.001	1.063	1.015, 1.112	6.7	.010
Multiple medical illness								
Highest tercile	-0.054	-0.069, -0.038	45.9	< .001	0.892	0.860, 0.925	37.1	< .001
Lowest tercile	-0.003	-0.021, 0.016	0.1	.75	0.976	0.934, 1.019	1.2	.28
Alcohol/substance dependency								
Highest tercile	-0.012	-0.029, 0.005	2.0	.16	0.999	0.951, 1.029	0.3	.59
Lowest tercile	0.026	0.009, 0.043	9.3	.002	1.057	1.015, 1.100	7.3	.007
Low income/socioeconomic								
Highest tercile	-0.005	-0.021, 0.012	0.3	.59	0.999	0.962, 1.038	0.0	.96
Lowest tercile	0.020	0.004, 0.037	5.8	.016	1.046	1.005, 1.088	5.0	.026
Lack social/family support								
Highest tercile	-0.033	-0.050, -0.016	14.6	< .001	0.940	0.903, 0.977	9.8	.002
Lowest tercile	0.070	0.054, 0.086	75.7	< .001	1.161	1.119, 1.206	59.1	< .001

[a] Each association is adjusted for the linear effects of all other predictor variables. Associations are measured either as a change in C-zscore in standard deviates per unit change in the independent variable, or an odds ratio as the odds of reporting above the mean calibration index value (C_{binary}) for a given level in the independent variable divided by the odds for a lower level in the independent variable, and averaged over all levels.

[b] Computed as z-scores (standard deviates), with mean = -0.016 and SD = 0.796, with higher index values indicating higher favorable ratings.

[c] Scored as a binary variable equal to 1 if C-zscore > 0, and 0 if C-zscore ≤ 0.

[d] Measured on 7-point scale correlated with years of education following high school.

[e] Measured in standard variates.

[f] Computed as the change in Calibration Index scored in standard variates per 1 level increase in psychological safety. When Calibration Index is scored as binary score, psychological safety is measured as a binary design variable that equals 1 if psychological safety = 5 (strongly agree) and equals 0 otherwise. Association measured as an odds ratio that respondent is above the mean Calibration Index score when responding "strongly agree" to psychological safety versus when responding less than strongly agree to psychological safety.

[g] The proportion of patients the respondent reported seeing during their VA experience is classified into terciles by respondent's discipline. Measured associations are based on the middle tercile as the referent group.

psychological safety,[23] trainee preferences for education program elements,[22] and interprofessional team care.[28] Of note, calibration bias was accounted for by including the index as a covariate to explain a CLE domain score of interest.

The finding of a strong association between psychological safety and calibration bias is also consistent with the role psychological safety has been seen to play in the workplace[29–31] and in health professions education,[32,33] together with its connection to CLE satisfaction,[23] work-related communication,[34] team tenure,[35,36] perceived care,[37] self-awareness, burnout, civility,[38] and mental health.[39] Our findings suggest trainees who believe their CLE is psychologically unsafe will tend to underrate their CLE experiences below what a rater unaffected by psychological safety would otherwise have rated those same experiences.

Overall, resident physicians reported more negative calibration bias and lower levels psychological safety compared to their nursing and allied and associated health trainee counterparts. These findings are consistent with the pressures physician trainees face when engaged in the care of complex patients in situations with high levels of ambiguity and uncertainty,[40–42] where trainees assume the role of an apprentice with expectations of becoming independent practitioners.[43] Resident physicians advance their professional development by engaging in patient care in supervised environment,[44,45] where frequent risk-taking behavior is often taken in complex hierarchical situations fraught with uncertainty.[38,46–48]

There are study limitations. Our methods do not allow separate estimates for subjective and threshold biases. Computation rests on the assumption that calibrating item experiences are either invariant, or at minimum, not affected by other trainee and CLE factors. We also assumed respondents both comprehended the meaning of and recalled information relevant to answering the 3 index questions. Similarly, trainee responses may be subject to additional biases when assessing sensitive topics.[46–48] However, we believe the impact of such pressures may be minimal because the survey was administered nationally, with only aggregate scores reported to program directors. We also assumed calibration bias is a property of trainees. An alternative approach is to construct separate indices derived from experience-invariant items that are related to the CLE construct of interest.

Readers are also cautioned about extrapolating results to different clinical settings, as VA medical centers can differ from non-VA clinical settings. In addition, with an 11% sampling rate, it is unlikely this convenience sample represents all VA trainees.[49,50] However, our purpose here is to compare trainees by discipline. LPS resident physician sample has been shown to be comparable to US residents in non-pediatric and non-OB-GYN programs by international status, PGY level, and specialty.[22]

Finally, our list of C_{index} predictors was limited. Future studies should consider the prevalence of depressive disorder, burnout, and chronic anxiety among trainees and teaching faculty that have been shown to be associated with high pessimism, negative perceptions,[51] negative-selective memory,[52] lower satisfaction intensity,[53,54] increased frequency of medical errors,[55] and higher rates of medical negligence and malpractice litigation.[56,57]

Conclusions

This study offers evidence that a trainee's subjective and threshold factors introduce calibration biases that impact how responses to CLE perceptions surveys should be scored, analyzed, and interpreted. The integration of calibration bias metrics into CLE perceptions surveys should be an integral element in the quest to better understand medical trainees and the complex clinical learning environments in which they train.

References

1. Loo LK, Byrne JM. Towards robust validity evidence for learning environment assessment tools. *Acad Med*. 2015;90(6):698–699. doi:10.1097/ACM.0000000000000723.
2. Colbert-Getz JM, Kim S, Goode VH, Shochet RB, Wright SM. Assessing medical students' and residents' perceptions of the learning environment: exploring validity evidence for the interpretation of scores from existing tools. *Acad Med*. 2014;89(12):1687–1693. doi:10.1097/ACM.0000000000000433.
3. Carney PA, Rdesinski R, Blank AE, Graham M, Wimmers P, Chen HC, et al. Utility of the AAMC graduation questionnaire to study behavioral and social sciences domains in undergraduate medical education. *Acad Med*. 2010;85(1):169–176. doi:10.1097/ACM.0b013e3181c464c0.
4. Holt KD, Miller RS, Philibert I, Heard JK, Nasca TJ. Residents' perspectives on the learning environment: data from the Accreditation Council for Graduate Medical Education resident survey. *Acad Med*. 2010;85(3):512–518. doi:10.1097/ACM.0b013e3181ccc1db.
5. Jaffe RC, Bergin CR, Loo LK, Singh S, Uthlaut B, Glod SA, et al. Nesting domains: a global conceptual model for optimizing the clinical learning environment. *Am J Med*. 2019;132(7):886–991. doi:10.1016/j.amjmed.2019.03.019.

6. Jaffe RC, Bergin CR, Loo LK, Singh S, Uthlaut B, Glod SA, et al. Reactive, holistic, proactive: practical applications of the AAIM learning and working environment conceptual model. *Am J Med*. 2019;132(8):995–1000. doi:10.1016/j.amjmed.2019.04.004.
7. Schwartz N. Self-reports: how questions shape the answers. *Am Psychol*. 1999;54(2):93–105. doi:10.1037/0003-066X.54.2.93.
8. Phillips AW, Artino AR Jr. Lies, damned lies, and surveys. *J Grad Med Educ*. 2017;9(6):677–679. doi:10.4300/JGME-D-17-00698.1.
9. Yock Y, Lim I, Lim YH, Lim WS, Chew N, Archuleta S. Sometimes means some of the time: residents' overlapping responses to vague quantifiers on the ACGME-I resident survey. *J Grad Med Educ*. 2017;9(6):735–740. doi:10.4300/JGME-D-17-00187.1.
10. Kashner TM, Clarke CT, Aron DC, Byrne JM, Cannon GW, Deemer DA, et al. The 9-criteria evaluation framework for perceptions survey: the case of VA's Learners' Perceptions Survey. *Biostat Epidemiol*. 2020;4(1):140–171. doi:10.1080/24709360.2018.1553362.
11. Vasileva-Stojanovska T, Malinovski T, Marina V, Dobri J, Trajkovik V. Impact of satisfaction, personality, and learning style on educational outcomes in a blended learning environment. *Learn Indiv Diff*. 2015;38:127–135. doi:10.1016/j.lindif.2015.01.018.
12. Lam H-TC, O'Toole TG, Arola PE, Kashner TM, Chang BK. Factors associated with the satisfaction of millennial generation dental residents. *J Dental Educ*. 2012;76(11):1416–1426.
13. Perone JA, Fankhauser GT, Adhikari D, Mehta HB, Woods MB, Tyler DS, et al. It depends on your perspective: resident satisfaction with operative experience. *Am J Surg*. 2017;213(2):253–259. doi:10.1016/j.amjsurg.2016.09.042.
14. Artino AR Jr, La Rochelle JS, Dezee KJ, Gehlbach H. Developing questionnaires for educational research: AMEE guide no. 87. *Med Teach*. 2014;36(6):463–474. doi:10.3109/0142159X.2014.889814.
15. Krosnick JA. Survey research. *Annu Rev Psychol*. 1999;50:537–567. doi:10.1146/annurev.psych.50.1.537.
16. Karabenick SA, Woollery ME, Friedel JM, Ammon BV, Blazevski J. Cognitive processing of self-reports items in education research. *Educ Psychol*. 2007;42(3):139–151. doi:10.1080/00461520701416231.
17. Vlaev I, Chater N, Neil S, Brown GDA. Does the brain calculate value? *Trends Cogn Sci*. 2011;15(11):546–554. doi:10.1016/j.tics.2011.09.008.
18. Parducci A. *Happiness, Pleasure, and Judgement: The Contextual Theory and Its Applications*. Mahwah, NJ: Erlbaum Associates; 1995.
19. Smith RH, Diener E, Wedell DH. Intrapersonal and social comparison determinants of happiness: a range-frequency analysis. *J Pers Soc Psychol*. 1989;56(3):317–325. doi:10.1037//0022-3514.56.3.317.
20. Johnson TR. On the use of heterogeneous thresholds ordinal regression model to account for individual differences in response style. *Psychometrika*. 2003;68(4):563–583. doi:10.1007/BF02295612.
21. Keitz SA, Holland GJ, Melander EH, Bosworth HB, Pincus SH. The Veterans Affairs Learners' Perceptions survey: the foundation for education quality improvement. *Acad Med*. 2003;78(9):910–917. doi:10.1097/00001888-200309000-00016.
22. Kashner TM, Hettler DL, Zeiss RA, Aron DC, Bernett DS, Brannen JL, et al. Has interprofessional education changed learning preferences? A national perspective. *Health Serv Res*. 2017;52(1):268–290. doi:10.1111/1475-6773.12485.
23. Torralba KD, Loo LK, Byrne JM, Baz S, Cannon GW, Keitz SA, et al. Does psychological safety impact the clinical learning environment for resident physicians? Results from the VA's Learners' Perceptions Survey. *J Grad Med Educ*. 2016;8(5):699–707. doi:10.4300/JGME-D-15-00719.1.
24. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373–383. doi:10.1016/0021-9681(87)90171-8.
25. Edmondson, A. Psychological safety and learning behavior in work teams. *Admin Sci Quart*. 1999;44(2):350–383. doi:10.2307/2666999.
26. Peterson D. FY17 VHA Facility Complexity Model Overview. Veterans Health Administration Office of Productivity, Efficiency, and Staffing. November 6, 2017. https://www.va.gov/vhapublications/ViewPublication.asp?pub_ID=8802. Accessed December 3, 2020.
27. Byrne JM, Chang BK, Gilman SC, Keitz SA, Kaminetzky CP, Aron DC, et al. The learners' perceptions survey-primary care: assessing resident perceptions of internal medicine continuity clinics and patient-centered care. *J Grad Med Educ*. 2013;5(4):587–593. doi:10.4300/JGME-D-12-00233.1.
28. Byrne JM, Kashner TM, Gilman SC, Wicker AB, Bernett DS, Aron DC, et al. Do patient aligned medical team models of care impact VA's clinical learning environments. Health Services Research and Development/Quality Enhancement Research Initiative National Conference. Philadelphia, PA, July 8–10, 2015.
29. Frazier ML, Fainshmidt S, Klinger RL, Pezeshkan A, Vracheva V. Psychological safety: a meta-analytic review and extension. *Personnel Psychol*. 2017;70(1):113–165. doi:10.1111/peps.12183.

30. Newman A, Donohue R, Eva N. Psychological safety: a systematic review of the literature. *Human Res Manag Rev*. 2017;27(3):521–535. doi:10.1016/j.hrmr.2017.01.001.
31. Edmondson AC. *The Fearless Organization: Creating Psychological Safety in the Workplace for Learning, Innovation, and Growth*. New York, NY: Wiley; 2019.
32. Torralba KD, Puder D. Psychological safety among learners: when connection is more than just communication. *J Grad Med Educ*. 2017;9(4):538–539. doi:10.4300/JGME-D-17-00195.1.
33. Bynum WE, Haque TM. Risky business: psychological safety and the risks of learning medicine. *J Grad Med Educ*. 2016;8(5):780–782. doi:10.4300/JGME-D-16-00549.1.
34. Yanchus NJ, Derickson R, Moore SC, Bologna D, Osatuke K. Communication and psychological safety in veterans health administration work environments. *J Health Organ Manag*. 2014;28(6):754–776. doi:10.1108/jhom-12-2012-0241.
35. Koopman J, Lanaj K, Wang M, Zhou L, Shi J. Nonlinear effects of team tenure on team psychological safety climate and climate strength: implications for average team member performance. *J App Psychol*. 2016;101(7):940–957. doi:10.1037/apl0000097.
36. Rosenbaum L. Cursed by knowledge - building a culture of psychological safety. *N Engl J Med*. 2019;380(8):786–790. doi:10.1056/NEJMms1813429.
37. Binyamin G, Friedman A, Carmeli A. Reciprocal care in hierarchical exchange: implications for psychological safety and innovative behaviors at work. *Psychol Aesth Creat Art*. 2017;12(1):79–88. doi:10.1037/aca0000129.
38. Hernandez W, Luthanen A, Ramsel D, Osatuke K. The mediating relationship of self-awareness on supervisor burnout and workgroup civility & psychological safety: a multilevel path analysis. *Burnout Res*. 2015;2(1):36–49. doi:10.1016/j.burn.2015.02.002.
39. Woody RH. Psychological safety for mental health practitioners: suggestions from a defense lawyer. *Psychol Inj Law*. 2016;9(2):198–202. doi:10.1007/s12207-015-9239-x.
40. Domen RE. The ethics of ambiguity: rethinking the role and importance of uncertainty in medical education and practice. *Acad Pathol*. 2016;3:2374289516654712. doi:10.1177/2374289516654712.
41. Torralba KD, Baz S, Byrne JM, Kashner TM. Understanding Psychological Safety and its Role in Quality Improvement in Graduate Medical Education. In: *Textbook for Medical Education Programs*. 12th ed. Alexandria, VA: Alliance for Academic Internal Medicine; 2017.
42. Hillen MA, Gutheil CM, Strout TD, Smets EM, Han PK. Tolerance of uncertainty: conceptual analysis, integrative model, and implications for healthcare. *Soc Sci Med*. 2017;180:62–75. doi:10.1016/j.socscimed.2017.03.024.
43. Cruess RL, Cruess SR, Steinert Y. Medicine as a community of practice: implications for medical education. *Acad Med*. 2018;93(2):185–191. doi:10.1097/ACM.0000000000001826.
44. Goodwin D, Pope C, Mort M, Smith A. Access, boundaries and their effects: legitimate participation in anaesthesia. *Sociol Health Illn*. 2005;27(6):855–871. doi:10.1111/j.1467-9566.2005.00477.x.
45. Amalberti R, Vincent C, Auray Y, de Saint Maurice G. Violations and migrations in health care: a framework for understanding and management. *Qual Saf Health Care*. 2006;15(Suppl 1):i66–71. doi:10.1136/qshc.2005.015982.
46. Loo LK, Puri N, Kim DI, Kawayeh A, Baz S, Hegstad D. "Page me if you need me": the hidden curriculum of attending-resident communication. *J Grad Med Educ*. 2012;4(3):340–345. doi:10.4300/JGME-D-11-00175.1.
47. Farnan JM, Johnson JK, Meltzer DO, Humphrey HJ, Arora VM. Resident uncertainty in clinical decision making and impact on patient care: a qualitative study. *Qual Saf Health Care*. 2008;17(2):122–126. doi:10.1136/qshc.2007.023184.
48. Bush RW. Supervision in medical education: logical fallacies and clear choices. *J Grad Med Educ*. 2010;2(1):141–143. doi:10.4300/JGME-D-09-00095.1.
49. Fauth T, Hattrup K, Mueller K, Roberts B. Nonresponse in employee attitude surveys: a group-level analysis. *J Business Psychol*. 2013;28(1):1–16. doi:10.1007/s10869-012-9260-y.
50. Prins JT, Hoekstra-Weebers JEHM, Gazendam-Donofrio SM, Dillingh GS, Bakker AB, Huisman M, et al. Burnout and engagement among resident doctors in the Netherlands: a national study. *Med Educ*. 2010;44(3):236–247. doi:10.1111/j.1365-2923.2009.03590.x.
51. Beck AT, Rush JA, Shaw BF, Emery G. *Cognitive Therapy of Depression*. New York, NY: Guilford Press; 1979.
52. Williams JMG, Barnhofer T, Crane C, Hermans D, Raes F, Watkins E, et al. Autobiographical memory specificity and emotional disorder. *Psychol Bull*. 2007;133(1):122–148. doi:10.1037/0033-2909.133.1.122.
53. Kassam A, Horton J, Shoimer I, Patten S. Predictors of well-being in resident physicians: a descriptive and psychometric study. *J Grad Med Educ*. 2015;7(1):70–74. doi:10.4300/JGME-D-14-00022.1.
54. Dunn J, Ng SK, Breitbart W, Aitken J, Youl P, Baade PD, et al. Health-related quality of life and life satisfaction in colorectal cancer survivors: trajectories of adjustment. *Health Qual Life Outcomes* 2013;11(1):46. doi:10.1186/1477-7525-11-46.

55. Thomas NK. Resident burnout. *JAMA*. 2004;292(23):2880–2889. doi:10.1001/jama.292.23.2880.
56. Anagnostopoulos F, Liolios E, Persefonis G, Slater J, Kafetsios K, Niakas D. Physician burnout and patient satisfaction with consultation in primary health care settings: evidence of relationships from a one-with-many design. *J Clin Psychol Med Settings*. 2012;19(4):401–410. doi:10.1007/s10880-011-9278-8.
57. Zis PA, Anagnostopoulos F, Sykioti P. Burnout in medical residents: a study based on the job demands-resources model. *Sci World J*. 2014;2014:673279. doi:10.1155/2014/673279.

David A. Deemer, MA, is a Student, Loma Linda University School of Medicine; **John M. Byrne, DO,** is Associate Chief of Staff for Education and Informatics, VA Loma Linda Healthcare System, and Professor of Medicine, Loma Linda University School of Medicine; **Lawrence K. Loo, MD, MACP,** is Staff Physician, VA Loma Linda Healthcare System, and Vice Chair for Education and Faculty Development, Department of Medicine, and Professor of Medicine, Department of Medicine and Department of Medical Education, Loma Linda University School of Medicine; **David Puder, MD,** is Assistant Professor, Departments of Psychiatry and Internal Medicine, Loma Linda University School of Medicine; **Karina D. Torralba, MD, MACM,** is Staff Physician, VA Loma Linda Healthcare System, and Professor, Internal Medicine, and Chief and Fellowship Program Director, Division of Rheumatology, Department of Medicine, Loma Linda University School of Medicine; **Sonny C. Lee, MD, FACP, FAAP,** is Program Director, Loma Linda University Health Education Consortium Internal Medicine Residency, and Associate Professor, Departments of Internal Medicine and Pediatrics, Loma Linda University School of Medicine; and **T. Michael Kashner, PhD, JD,** is Health Science Specialist, Office of Academic Affiliations, Department of Veterans Affairs, and Research Professor of Medicine, Loma Linda University Medical School.

Funding: This study was funded by the Department of Veterans Affairs, Veterans Health Administration, in the Office of Research and Development, Health Services Research and Development Service IIR#14-071 and IIR#15-084, and by the MacPherson Society, Loma Linda University School of Medicine.

Conflict of interest: The authors declare they have no competing interests.

The authors would like to thank the following individuals for their help on the project: Christopher T. Clarke, PhD, David Bernett, BA, and Laura Stefanowycz, Office of Academic Affiliations Data Management and Support Center, St. Louis, MO; Grant W. Cannon, Utah VA Medical Center, University of Utah School of Medicine; Catherine P. Kaminetzky, Puget Sound VA Healthcare System, University of Washington School of Medicine; Sheri A. Keitz, University of Massachusetts School of Medicine; and Shanalee Tamares, MLIS, and Daniel Reichert, MD, Loma Linda University.

Corresponding author: T. Michael Kashner, PhD, JD, Jerry L. Pettis Memorial VA Medical Center, Loma Linda VA Healthcare System, 11201 Benton Street, Loma Linda, CA 92357, 909.825.7084 ext. 2853, fax 909.777.3861, michael.kashner@va.gov

Received March 27, 2020; revisions received July 28, 2020, and September 16, 2020; accepted September 21, 2020.