Fairly Assessing U.S.- and Foreign-Trained Physician Performance Using 360-Degree Surveys

Julie J Lanz
University of Nebraska Kearney  https://orcid.org/0000-0003-2090-5475

Paul Gregory  (paul@pdplorida.com)
Physicians Development Program Inc / PULSE 360 Program  https://orcid.org/0000-0003-0476-2611

Larry Harmon
Physicians Development Program Inc / PULSE 360 Program

Research article

Keywords: 360-degree performance appraisal, 360-degree survey, multi-source rater feedback, interpersonal and communication skills, emotional intelligence

DOI: https://doi.org/10.21203/rs.3.rs-69954/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

Background

With a growing number of foreign-trained physicians joining the United States workforce, there is a need to fairly assess their job performance. The purpose of this study was to explore the fairness of a 360-degree competency assessment on U.S.- and foreign-trained physicians.

Methods

We conducted a non-experimental retrospective analysis on physicians working in the United States (n = 258) who participated in a physician assessment and education program between 2007 and 2017.

Results

There were no significant differences in performance outcomes of teamwork, motivating or discouraging behaviors, technical practice, and patient interactions based on demographic differences.

Conclusions

The PULSE 360 is a powerful tool that can be used to evaluate physician performance without bias in demographic differences including: gender, country of physician medical training, physician native language, or age.

Background

In 2017, foreign-trained physicians made up over 25% of practicing physicians in the United States.\(^1\) Given their substantial contributions to the U.S. healthcare system,\(^2,3\) fairly assessing their performance is critical. Policy suggestions have been developed recently to promote the benefits of employing international health care workers, such as treating them transparently and fairly.\(^4\) However, comparing the performance of U.S. trained physicians (USTP) to foreign trained physicians (FTPs) is not well-understood.\(^5\) Exploring soft skills like professionalism, interpersonal and communication skills, teamwork, and patient interactions is critical, and physician competency assessments should be fair regardless of demographic differences like age, sex, and nationality. The purpose of this paper is to explore the fairness of a 360-degree competency assessment program when evaluating U.S.- and foreign-trained physicians working in the United States.

Assessing Physician Competence
Fair assessment of physician competencies is a key step in improving job performance. Reasons for evaluating physician performance range from appraisal to recertification, identifying high-risk physicians, and remediating those with a previous history of poor performance. Maintaining quality patient care is important because 6-12% of physicians are referred to remediation for poor clinical skills. The Institute of Medicine estimates that physician dyscompetency is one contributor to preventable medical errors at an estimated cost of $17 billion. To determine which physicians should be referred for dyscompetency, one model for performance remediation starts with an assessment of the physician's competence. A common framework for maintaining physician competency is the American Board of Medical Specialties, which developed the Maintenance of Certification (ABMS MOC). This four-part framework includes: maintaining licensure, lifelong learning, cognitive expertise, and quality improvement.

**Multisource Feedback**

Multisource, or 360-degree, feedback is the use of physicians' team members (other physicians, nurses, and staff) to evaluate job performance. The scope and depth of multisource feedback is valuable given the argument that patient evaluations of physician performance are subjective at best. For example, patient evaluations have been found to be influenced by the race and gender of the physician such that only physicians who were white and male benefited from a customer satisfaction judgment, even after controlling for objective measures of performance. Beyond clinical skills, physician performance is based on a combination of individual differences including specialty area, gender, and age. Evidence suggests that biases against international medical graduates (IMGs) may lead to more complaints against physicians and disciplinary outcomes, but findings on biased physician performance evaluations are mixed. Given the inconclusive evidence, having two examiners appears to mitigate potential sex or ethnic biases against physicians who are being evaluated based on their clinical performance.

Some research has explored the use of multirater assessments on international medical graduates and found them reliable, but little research has examined bias in physician assessment as a function of training country (i.e., USTPs versus FTPs). Of the research on assessing physician performance, one experiment found that after holding education, experience, and personality consistent, foreign-born physicians were rated more poorly than those who had born in the prospective patients’ home country. However, physicians who had been trained in an industrialized and high-income country benefited on their evaluations. There are no significant differences in mortality rates for international versus national practitioners, but differences may exist in regard to the soft skills of communication, teamwork, and ethical issues. Part of this bias may be a function of the examiners themselves. In one study, IMGs have lower mortality rates than USMGs. Further, there is evidence that in Canada, international medical graduates are disciplined for misconduct more frequently than North American medical graduates. In Australia, IMGs receive more complaints and disciplinary adverse findings. Thus, there is a critical need for fair tools to evaluate USTPs and FTPs on their job performance.
Hypothesis 1: There will be no significant differences in PULSE 360 physician performance based on: a) gender; b) country where training occurred; c) first language spoken; or d) age.

Methods

Design

A non-experimental retrospective analysis of data was conducted for two hundred and fifty-eight physicians (n=258) who participated in a physician education program between 2007 and 2017.

Statistical Analyses

Independent samples t-tests and an analysis of variance (ANOVA) were conducted in order to evaluate potential biases in PULSE 360 scale scores due to demographic differences including: gender, country in which training occurred, cultural background (first language spoken), etc.; see Tables 1-5.

Participants

Seventy percent (80%) of the physicians were males (n=206), 62% were trained in the United States (n=157), 66% had English as their first language (n=166), the average physician age was 61 (range 41-84), and 78% were board certified (n=202). Age was separated into three ranges: 1. 55 or younger (n=76, 30%), 2. 56 to 65 (n=85, 34%), and 3. 66 or older (n=90, 36%). There were thirty (n=30) different specialties represented within the sample of physicians, including internal medicine, obstetrics and gynecology, anesthesiology etc.; see Table 1). Specialties were collapsed into 3 practice type areas: Primary Care (n=109, 42%), Specialists (non-surgical) (n=42, 16%), and Surgeons (n=107, 41%), to allow for greater power in statistical comparisons. All physicians in the sample were participating in an assessment conducted through a physician education program on the west coast of the United States. Physicians had been referred to this program for a variety of reasons (e.g., competency assessment, anger management).

PULSE 360 Survey

The PULSE 360 Survey is an assessment of leadership, teamwork, communication, professionalism, and other physician behaviors based on multisource feedback from other physicians, advanced practice providers, clinical staff, and administrative staff members who interact with a physician. Variations of the survey are based on n=96 behavioral items, 5 performance domains, including a total composite performance score known as the Teamwork Index (TI) Score with internal consistency reliability estimates ranging between $\alpha = .77$ to .85 across dimensions. TI scores typically range from 0 to 100 with a national mean score of 68.9 for physicians. Prior research has demonstrated both the internal and external validity of PULSE 360 scores in relation to important physician outcomes such as malpractice risk and patient satisfaction.\textsuperscript{11,28-32}
Results

In support of hypothesis 1a there were non-significant differences in the mean PULSE 360 scores for male vs. female physicians (see Tables 2-4 for independent samples t-test results). In support of hypothesis 1b there were non-significant differences in the mean PULSE 360 scores for US-trained vs. foreign-trained physicians. In support of hypothesis 1c there were non-significant differences in the mean PULSE 360 scores for native English speakers vs. non-native English speakers. In support of hypothesis 1d there were non-significant differences in the mean PULSE 360 scores between age ranges (see Table 5 for ANOVA results). Additionally, all post hoc comparisons amongst age ranges yielded non-significant differences in mean scores on all PULSE 360 Scale scores.

Discussion

Given the growing demand for FTPs in the United States, there is a need to fairly select, train, and support this diverse group of international physicians. Physicians in this study were evaluated on their performance using the PULSE 360 Survey and were compared across gender, country of training, native language, age, and board certification status. There were no significant differences in their reported performance on professionalism, teamwork, motivating behaviors, discouraging behaviors, technical practice style, or patient interactions. These findings suggest that there are valid and reliable tools established to fairly evaluate the performance of both U.S.-trained physicians and foreign-trained physicians. This is valuable given that previous research has found that some measures discriminate against some protected classes.

The physicians in our current sample were recruited to the physician assessment and education program for a variety of reasons that may not be representative of practicing physicians in the United States. However, the use of 360-degree data allows us to move beyond self-reported performance to a more comprehensive view of physicians’ performance on the job. The important takeaway is that within our sample, there were no significant variations in scoring patterns attributable to protected class membership.

Conclusions

The use of 360-degree feedback can provide a fair and unbiased assessment of others’ perceptions of physician behavior and performance within the healthcare team.

Abbreviations

United States trained physicians (USTP), foreign-trained physicians (FTP), USMG (United States medical graduates), international medical graduates (IMGs), analysis of variance (ANOVA)

Declarations
Ethics approval and consent to participate:

Because this study used de-identified archival data, it was deemed unnecessary to ask participants for informed consent. This study was approved by the UNK IRB #041320-2.

Consent for publication:

Not applicable.

Availability of data and materials:

The dataset used and analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests:

Two of the authors are employees of the Physicians Development Program Inc / PULSE 360 Program, Miami, FL, USA.

Funding:

Not applicable.

Authors’ contributions:

Study concept and design: PG and LH. Analysis and interpretation of data: PG, JL, and LH. Drafting and revising of the manuscript: JL, PG, and LH. All authors read and approved the final version.

Acknowledgements:

Not applicable.

References

1 American Immigration Council. Foreign-trained doctors are critical to serving many U.S. communities. https://www.americanimmigrationcouncil.org/research/foreign-trained-doctors-are-critical-serving-many-us-communities. January 17, 2018. Accessed May 4, 2020.

2 Jolly P, Boulet J, Garrison G, Signer MM. Participation in US graduate medical education by graduates of international medical schools. Acad Med 2011;86(5):559-564. doi:10.1097/ACM.0b013e318212de4d
3 Pinsky WW. The importance of international medical graduates in the United States. *Ann Intern Med* 2017;166(11):840-841. doi:10.7326/M17-0505

4 Chen PG, Auerbach DI, Muench U, Curry LA, Bradley EH. Policy solutions to address the foreign-educated and foreign-born health care workforce in the United States. *Health Affair* 2013;32(11):1906-1913. doi:10.1377/hlthaff.2013.0576

5 Rao NR, Yager J. Acculturation, education, training, and workforce issues of IMGs: Current status and future directions. *Acad Psychiatr* 2012;36(4):268-270. doi:10.1176/appi.ap.12050095

6 Humphrey C. Assessment and remediation for physicians with suspected performance problems: An international survey. *J Contin Educ Health Prof* 2010;30(1):26-36. doi:10.1002/chp.20053

7 Williams BW. The prevalence and special educational requirements of dyscompetent physicians. *J Contin Educ Health Prof* 2006;26:173–191. doi:10.1002/chp.6

8 Kohn LT, Corrigan J, Donaldson MS. Committee on Quality of Health Care in America, Institute of Medicine. *To err is human: Building a safer health system*. Washington, DC: National Academy Press, 1999

9 Hauer KE, Ciccone A, Henzel TR, Katsufrakis P, Miller SH, Norcross WA, Papadakis MA, Irby DM. Remediation of the deficiencies of physicians across the continuum from medical school to practice: a thematic review of the literature. *Acad Med* 2009;84(12):1822-1832. doi:10.1097/ACM.0b013e3181bf3170

10 Hawkins RE, Lipner RS, Ham HP, Wagner R, Holmboe ES. American Board of Medical Specialties Maintenance of Certification: Theory and evidence regarding the current framework. *J Contin Educ Health Prof* 2013;33(S1):S7-S19. doi:10.1002/chp.21201

11 Lanz JJ, Gregory PJ, Menendez ME, Harmon L. Dr. Congeniality: Understanding the importance of surgeons’ nontechnical skills through 360° feedback. *J Surg Educ* 2018;75(4):984-992. doi:10.1016/j.jsurg.2017.12.006

12 Castanelli D, Kitto S. Perceptions, attitudes, and beliefs of staff anaesthetists related to multi-source feedback used for their performance appraisal. *Brit J Anaesth* 2011;107(3):372-377. doi:10.1093/bja/aer152

13 Glickman SW, Schulman KA. The mis-measure of physician performance. *Am J Manag Care* 2013;19(10):782-785. PMID:24304157

14 Hekman DR, Aquino K, Owens BP, Mitchell TR, Schilpzand P, Leavitt K. An examination of whether and how racial and gender biases influence customer satisfaction. *Acad Manage J* 2010;53(2):238-264. doi:10.5465/amj.2010.49388763
15 Grace ES, Wenghofer EF, Korinek EJ. Predictors of physician performance on competence assessment: Findings from CPEP, the Center for Personalized Education for Physicians. *Acad Med* 2014;89(6):912-919. doi:10.1097/ACM.0000000000000248

16 Wenghofer EF, Williams AP, Klass DJ. Factors affecting physician performance: Implications for performance improvement and governance. *Healthc Policy* 2009;5(2):e141-e160. PMID:21037818

17 Elkin K, Spittal MJ, Studdert DM. Risks of complaints and adverse disciplinary findings against international medical graduates in Victoria and Western Australia. *Med J Australia*, 2012;197(8):448-452. doi:10.5694/mja12.11702

18 Denney ML, Freeman A, Wakeford R. MRCGP CSA: are the examiners biased, favouring their own by sex, ethnicity, and degree source?. *Br J Gen Pract* 2013;63(616):e718-e725. doi:10.3399/bjgp13X674396

19 McManus IC, Elder AT, Dacre J. Investigating possible ethnicity and sex bias in clinical examiners: an analysis of data from the MRCP (UK) PACES and nPACES examinations. *BMC Med Educ* 2013;13(103):1-11. doi:10.1186/1472-6920-13-103

20 Nair BR, Alexander HG, McGrath BP, Parvathy MS, Kilsby EC, Wenzel J, Frank IB, Pachev GS, Page GG. The mini clinical evaluation exercise (mini-CEX) for assessing clinical performance of international medical graduates. *Med J Australia* 2008;189(3):159-161. PMID:18673104

21 Louis WR, Lalonde RN, Esses VM. Bias against foreign-born or foreign-trained doctors: Experimental evidence. *Med Educ* 2010;44(12):1241-1247. doi:10.1111/j.1365-2923.2010.03769.x

22 Norcini JJ, Boulet JR, Dauphinee WD, Opalek A, Krantz ID, Anderson ST. Evaluating the quality of care provided by graduates of international medical schools. *Health Aff (Millwood).* 2010;29:1461-1468. doi:10.1377/hlthaff.2009.0222

23 Slowther A, Lewando Hundt GA, Purkis J, Taylor R. Experiences of non-UK-qualified doctors working within the UK regulatory framework: A qualitative study. *J R Soc Med* 2012;105:156-67. doi:10.1258/jrsm.2011.110256

24 Ko DT, Austin PC, Chan BTB, Tu JV. Quality of care of international and Canadian medical graduates in acute myocardial infarction. *Arch Intern Med* 2005;165:458-463. doi:10.1001/archinte.165.4.458

25 Yeates P, Woolf K, Benbow E, Davies B, Boohan M, Eva K. A randomised trial of the influence of racial stereotype bias on examiners’ scores, feedback and recollections in undergraduate clinical exams. *BMC Med* 2017;15(179):1-11. doi:10.1186/s12916-017-0943-0

26 Tsugawa Y, Jena AB, Orav EJ, Jha AK. Quality of care delivered by general internists in US hospitals who graduated from foreign versus US medical schools: Observational study. *BMJ* 2017;356:j273:1-8. doi:10.1136/bmj.j273
27 Alam A, Matelski JJ, Goldberg HR, Liu JJ, Klemensberg J, Bell CM. The characteristics of international medical graduates who have been disciplined by professional regulatory colleges in Canada: A retrospective cohort study. Acad Med 2017;92(2):244-249. doi:10.1097/ACM.0000000000001356

28 Gregory PJ, Robbins B, Schwitzberg SD, Harmon L. Leadership development in a professional medical society using 360-degree survey feedback to assess emotional intelligence. Surg Endosc 2017;31(9):3565-3573. doi:10.1007/s00464-016-5386-8

29 Gregory PJ, Ring D, Rubash H, Harmon L. Use of 360° feedback to develop physician leaders in orthopaedic surgery. J Surg Orthop Adv 2018;27(2):85-91. PMID:30084813

30 Hageman MG, Ring DC, Gregory PJ, Rubash HE, Harmon L. Do 360-degree feedback survey results relate to patient satisfaction measures? Clin Orthop Relat Res 2015;473(5):1590-1597. doi:10.1007/s11999-014-3981-3

31 Hammerly ME, Harmon L, Schwitzberg SD. Good to great: Using 360-degree feedback to improve physician emotional intelligence. J Healthc Manag 2014;59(5):354-366. PMID:25647955

32 Lagoo J, Berry WR, Miller K, Neal BJ, Sato L, Lillemoe KD, Doherty GM, Kasser JR, Chaikof EL, Gawande AA, Haynes AB. Multisource evaluation of surgeon behavior is associated with malpractice claims. Ann Surg 2019;270(1):84–90. doi:10.1097/SLA.0000000000002742

33 Peile E. Selecting an internationally diverse medical workforce. BMJ 2014;348(g2696):1-2. doi:10.1136/bmj.g2696

Tables

Table 1. Distribution of Physician Participants by Specialty/Sub-Specialty.
| Specialty/Sub-Specialty Description                  | Count | Percentage | Category                          |
|-----------------------------------------------------|-------|------------|-----------------------------------|
| 1. Anesthesiology                                   | 13    | 5.0%       | Specialists (non-surgical)        |
| 1. Cardiology                                       | 3     | 1.2%       | Specialists (non-surgical)        |
| 1. Cardiothoracic Surgery                           | 4     | 1.6%       | Surgeons                          |
| 1. Colon and Rectal Surgery                         | 1     | 0.4%       | Surgeons                          |
| 1. Critical Care Medicine                           | 1     | 0.4%       | Specialists (non-surgical)        |
| 1. Dermatology                                      | 1     | 0.4%       | Specialists (non-surgical)        |
| 1. Emergency Medicine                               | 4     | 1.6%       | Specialists (non-surgical)        |
| 1. Family Medicine / General Practice               | 44    | 17.1%      | Primary Care Physicians           |
| 1. Gastroenterology                                 | 1     | 0.4%       | Specialists (non-surgical)        |
| 1. General & Gastrointestinal Surgery                | 22    | 8.5%       | Surgeons                          |
| 1. Geriatrics                                       | 1     | 0.4%       | Specialists (non-surgical)        |
| 1. Internal Medicine                                | 44    | 17.1%      | Primary Care Physicians           |
| 1. Maternal Fetal Medicine                          | 1     | 0.4%       | Specialists (non-surgical)        |
| 1. Maxillofacial Surgery                            | 1     | 0.4%       | Surgeons                          |
| 1. Neurological Surgery                             | 2     | 0.8%       | Surgeons                          |
| 1. Neurology                                        | 2     | 0.8%       | Specialists (non-surgical)        |
| 1. Not Boarded (General Practice)                   | 13    | 5.0%       | Primary Care Physicians           |
| 1. Obstetrics & Gynecology                          | 44    | 17.1%      | Surgeons                          |
| 1. Ophthalmology                                   | 4     | 1.6%       | Surgeons                          |
| Specialty/Sub-Specialty Description          | Count | Percentage | Category          |
|---------------------------------------------|-------|------------|-------------------|
| 1. Orthopaedic Surgery                      | 14    | 5.4%       | Surgeons          |
| 1. Otolaryngology                           | 2     | 0.8%       | Surgeons          |
| 1. Pediatrics                               | 8     | 3.1%       | Primary Care Physicians |
| 1. Physical Medicine & Rehab                | 1     | 0.4%       | Specialists (non-surgical) |
| 1. Plastic Surgery                          | 5     | 1.9%       | Surgeons          |
| 1. Podiatry                                 | 2     | 0.8%       | Surgeons          |
| 1. Psychiatry                               | 11    | 4.3%       | Specialists (non-surgical) |
| 1. Radiology - Diagnostic                   | 3     | 1.2%       | Specialists (non-surgical) |
| 1. Radiology - Interventional               | 2     | 0.8%       | Surgeons          |
| 1. Urology                                  | 3     | 1.2%       | Surgeons          |
| 1. Vascular Surgery                         | 1     | 0.4%       | Surgeons          |
| **TOTAL**                                   | **258** | **100%**  |                   |

|                               |       |            |                   |
|                               | **109** | **42.2%** | Primary Care Physicians |
|                               | **42**  | **16.3%** | Specialists (non-surgical) |
|                               | **107** | **41.5%** | Surgeons          |

*Table 2. t-Test Comparison of Mean PULSE 360 Scale Scores by Gender.*
| Gender          | - | - |
|-----------------|---|---|
| *Male* (n=206)  | *Female* (n=52) | *Mean Comparison* (df=256) |
| PULSE 360 Scale Score | m | sd | m | sd | t  | p  |
| 1. Teamwork Index Score | 66.5 | 24.0 | 65.0 | 23.4 | .401 | .69ns |
| 1. Motivating Behavior Score | 81.4 | 10.2 | 81.2 | 10.2 | .153 | .88ns |
| 1. Discouraging Behavior Score | 28.2 | 9.9 | 29.2 | 9.4 | -.625 | .53ns |
| 1. Technical Practice Score | 86.2 | 10.3 | 87.9 | 8.3 | -1.14 | .25ns |
| 1. Patient Interaction Score | 87.1 | 9.7 | 87.5 | 9.1 | -.225 | .82ns |

Table 3. *t*-Test Comparison of Mean PULSE 360 Scale Scores by Country of Training Status.

| Physician trained in the U.S.? | - | - |
|----------------------------------|---|---|
| *Yes* (n=157)                   | *No* (n=97) | *Mean Comparison* (df=252) |
| PULSE 360 Scale Score | m | sd | m | sd | t  | p  |
| 1. Teamwork Index Score | 65.3 | 22.9 | 67.5 | 25.6 | -.712 | .48ns |
| 1. Motivating Behavior Score | 81.1 | 9.6 | 81.7 | 11.1 | -.433 | .67ns |
| 1. Discouraging Behavior Score | 28.9 | 9.6 | 27.7 | 10.2 | .946 | .35ns |
| 1. Technical Practice Score | 86.0 | 9.7 | 87.3 | 10.5 | -.952 | .34ns |
| 1. Patient Interaction Score | 87.0 | 9.1 | 87.4 | 10.3 | -.324 | .75ns |

*missing n=4*

Table 4. *t*-Test Comparison of Mean PULSE 360 Scale Scores by Native English Speaker Status.
| Native English Speaker? | Yes (n=166) | No (n=84) | Mean Comparison (df=248) |
|------------------------|-------------|-----------|-------------------------|
| **PULSE 360 Scale Score** | m | sd | m | sd | t | p |
| 1. Teamwork Index Score | 66.4 | 23.3 | 66.8 | 24.8 | -.122 | .90<sup>ns</sup> |
| 1. Motivating Behavior Score | 81.4 | 9.9 | 81.5 | 10.7 | -.032 | .98<sup>ns</sup> |
| 1. Discouraging Behavior Score | 28.3 | 9.6 | 28.1 | 9.8 | .207 | .84<sup>ns</sup> |
| 1. Technical Practice Score | 86.6 | 9.5 | 86.1 | 10.9 | .359 | .72<sup>ns</sup> |
| 1. Patient Interaction Score | 87.3 | 9.3 | 87.4 | 9.9 | -.031 | .98<sup>ns</sup> |

*missing n=8*

**Table 5. ANOVA of Mean PULSE 360 Scale Scores by Age Range.**

| Age Range of Physicians | 55 or younger (n=76) | 56 to 65 (n=85) | 66 or older (n=90) | ANOVA* (df=2, 248) |
|-------------------------|---------------------|------------------|--------------------|--------------------|
| **PULSE 360 Scale Score** | m | sd | m | sd | m | sd | F | P |
| 1. Teamwork Index Score | 66.9 | 23.1 | 68.1 | 19.9 | 63.7 | 27.1 | .796 | .452<sup>ns</sup> |
| 1. Motivating Behavior Score | 81.8 | 10.1 | 82.1 | 8.6 | 80.2 | 11.2 | .800 | .451<sup>ns</sup> |
| 1. Discouraging Behavior Score | 28.2 | 9.3 | 27.6 | 8.1 | 29.3 | 11.3 | .669 | .513<sup>ns</sup> |
| 1. Technical Practice Score | 86.3 | 10.5 | 87.1 | 9.7 | 86.3 | 9.4 | .199 | .820<sup>ns</sup> |
| 1. Patient Interaction Score | 87.4 | 9.0 | 87.9 | 8.4 | 86.3 | 10.7 | .629 | .534<sup>ns</sup> |

*missing n=7*

*All post hoc comparisons for between group differences were non-significant for all scale scores*