



Article scientifique

Article

2021

Accepted version

Open Access

This is an author manuscript post-peer-reviewing (accepted version) of the original publication. The layout of the published version may differ .

A Meaningful and Actionable Professionalism Assessment: Validity
Evidence for the Professionalism Mini-Evaluation Exercise (P-MEX)
Across 8 Years

Bajwa, Nadia Masood; Nendaz, Mathieu; Posfay Barbe, Klara; Yudkowsky, Rachel; Park, Yoon Soo

How to cite

BAJWA, Nadia Masood et al. A Meaningful and Actionable Professionalism Assessment: Validity Evidence for the Professionalism Mini-Evaluation Exercise (P-MEX) Across 8 Years. In: Academic medicine, 2021, vol. 96, n° 11S, p. S151–S157. doi: 10.1097/ACM.0000000000004286

This publication URL: <https://archive-ouverte.unige.ch//unige:165069>

Publication DOI: [10.1097/ACM.0000000000004286](https://doi.org/10.1097/ACM.0000000000004286)

A Meaningful and Actionable Professionalism Assessment: Validity Evidence for the Professionalism Mini-Evaluation Exercise (P-MEX) Across Eight Years

Nadia M. Bajwa, MD, MHPE; Mathieu R. Nendaz, MD, MHPE; Klara M. Posfay-Barbe, MD, MS; Rachel Yudkowsky, MD, MHPE, and Yoon Soo Park, PhD

Author Bios:

Nadia M. Bajwa, MD, MHPE, is Residency Program Director in the Department of General Pediatrics at the Children's Hospital, Geneva University Hospitals in Geneva, Switzerland and Faculty Member at the Unit of Development and Research in Medical Education (UDREM), Faculty of Medicine, University of Geneva, Switzerland. ORCID: <http://orcid.org/0000-0002-1445-4594>

Mathieu R. Nendaz, MD, MHPE, is Professor and Director of the Unit of Development and Research in Medical Education (UDREM), Faculty of Medicine, University of Geneva, and Attending Physician in the Division of General Internal Medicine, University Hospitals of Geneva, Switzerland. ORCID: <http://orcid.org/0000-0003-3795-3254>

Klara M. Posfay-Barbe, MD, MS is Professor and Chairperson of the Department of General Pediatrics at the Children's Hospital, Geneva University Hospitals in Geneva, Switzerland. ORCID: <https://orcid.org/0000-0001-9464-5704>

Rachel Yudkowsky, MD, MHPE, is Professor in the Department of Medical Education, College of Medicine at the University of Illinois at Chicago in Chicago, Illinois, USA. ORCID: <https://orcid.org/0000-0002-2145-7582>

Yoon Soo Park, PhD is Associate Professor at Harvard Medical School and Director of Health Professions Education Research at Massachusetts General Hospital in Boston, Massachusetts, USA. ORCID: <http://orcid.org/0000-0001-8583-4335>.

Corresponding Author:

Nadia M. Bajwa
Département de l'enfant et de l'adolescent
Rue Willy-Donzé 6
1211 Genève 14
Switzerland
Phone: +41 22 372 30 82
E-mail : Nadia.Bajwa@hcuge.ch
Twitter handle: @nadiabajwa1, @Hopitaux_unige

Key Words: Medical Residency, Internship and Residency/organization and Administration, Personnel Selection, School Admissions Criteria, Educational Measurement, Patient Simulation, Professionalism, Reproducibility of Results, Clinical Competence/Standards

Abstract

Purpose: With the growing importance of professionalism in medical education, it is imperative to develop professionalism assessments that demonstrate robust validity evidence. The Professionalism Mini-Evaluation Exercise (P-MEX) is an assessment that has demonstrated validity evidence in the authentic clinical setting. Identifying the factorial structure of professionalism assessments determines professionalism constructs that can be used to provide diagnostic and actionable feedback. This study examines validity evidence for the P-MEX, a focused and standardized assessment of professionalism, in a simulated patient setting.

Method: The P-MEX was administered to 275 pediatric residency applicants as part of a 3-station standardized patient encounter, pooling data over an 8-year period (2012 to 2019 residency admission years). Reliability and construct validity for the P-MEX were evaluated using Cronbach's alpha, exploratory factor analysis (EFA), and confirmatory factor analysis (CFA).

Results: Cronbach's alpha for the P-MEX was 0.91. The EFA yielded 4 factors: doctor-patient relationship skills, interprofessional skills, professional demeanor, and reflective skills. The CFA demonstrated good model fit with a root mean square error of approximation (RMSEA) of .058 and a comparative fit index (CFI) of .92, confirming the reproducibility of the 4-factor structure of professionalism.

Conclusions: The P-MEX demonstrates construct validity as an assessment of professionalism, with four underlying subdomains in doctor-patient relationship skills, interprofessional skills, professional demeanor, and reflective skills. These results yield new confidence in providing

diagnostic and actionable subscores within the P-MEX assessment. Educators may wish to integrate the P-MEX assessment into their professionalism curricula.

The social construct of professionalism has taken a central place in the discussion of medical student and resident training during recent years. The assessment of professionalism has gained a growing evidence base that supports the measurement of professional behaviors in our trainees. It is complex to assess professionalism as it is a social construct created by the norms and values of the local culture. While almost all institutions would insist that professionalism is a core value, the definition of professionalism is widely debated. The definition has evolved from “I know it when I see it” to a more discriminating and finer observation of traits and behaviors. Epstein and Hundert define professionalism as the “habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served”.¹ This definition has been expounded upon by the Association of American Medical Colleges (AAMC) and the Accreditation Council for Graduate Medical Education (ACGME) to include characteristics of the competency such as compassion, integrity, respect, altruism, accountability, and sensitivity while linking these attributes to professionalization, professional conduct, humanism, and cultural competence in the milestones.² The Royal College of Physicians and Surgeons of Canada also includes professionalism an element of the CanMEDS framework where the domain of professionalism includes 29 different key concepts.^{3,4} Most accrediting bodies require that the domain of professionalism be systematically assessed for both formative and summative assessment.

In 2019, the Ottawa Consensus Group on the Assessment of Professionalism published recommendations for the assessment of professionalism and themes for future programs of research that would explore innovative professionalism assessments.⁵ Professionalism may be assessed at the individual level based on observable traits or behaviors, the inter-personal level during social interactions, and the societal-institutional level measured by the ability of students

to meet professional standards.^{5,6} The purpose of the assessment of professionalism is manifold. For trainees, professionalism assessment may provide useful feedback and foster self-reflection as a means of promoting self-regulation. For the curriculum, assessment is a method to monitor progress in the development of professional competencies, provides a guarantee of the competencies of graduating trainees, provides insight into the hidden curriculum of the program, and serves as an impetus for curricular change. The assessment may demonstrate whether expectations are clear and whether students have taken full advantage of the learning experience. For the institution, the assessment may be a way of maintaining autonomy by selecting trainees for further advancement, a means of expressing institutional values, and a way to promote shared educational values among educators. For the public, the certification of competence of graduates is an important safety measure to maintain public trust. Assessment sends a clear message to all stakeholders that professionalism is a core value and is one that is constantly promoted at all levels of learning and evaluation. Ultimately, upholding the values of professionalism can prevent conflicts that can corrupt the doctor patient relationship and result in distrust.

Assessment of professionalism knowledge may utilize methods such as multiple-choice questions, surveys, case-based discussions, and structured interviews. Multiple systematic reviews have described the various ways to measure aspects of professionalism.⁷⁻¹² However, validity evidence to justify the use of some of these assessments is often lacking.^{8,9} While there may not be a gold standard, in the creation of a robust assessment program it is necessary to understand the advantages and limitations of each assessment to ensure the quality of the application of the assessment to produce results that are useful and defensible. Higher end assessments may require the use of lower levels of knowledge and application in order to succeed. However, scores on a lower order assessment may not predict performance on a higher

order assessment.⁷ Simply conducting the assessment may not always lead to adequate consequences. Faculty report multiple barriers to assessing professionalism; a failure to fail phenomenon may occur due to uncertainty about the remediation process, ambiguity considering the incident, the consequences for the learner, and the time that reporting and remediation would take.¹³

A professionalism assessment that is capable of examining a learner at the “shows” and “does” level of Miller’s pyramid is The Professionalism Mini-Evaluation Exercise (P-MEX). The P-MEX is a 21-item direct-observation work-place based professionalism assessment that assesses doctor-patient relationship skills, reflective skills, time-management skills, and inter-professional skills; see supplemental digital content 1.¹⁴⁻¹⁷ In addition, the exercise includes an assessment of the learner’s overall performance and documents whether a critical incident occurred during the encounter. Faculty at McGill University established content validity for the 21-item P-MEX through a rigorous consensus process that identified observable professional behaviors.¹⁴ Through their work, they eliminated three redundant items from the original 24 items and three double-barreled items were reworded.¹⁴ The P-MEX effectively and reliably assesses both medical student and resident professionalism behaviors in the clinical setting.^{10,14,18-20} In a prior study, we were able to demonstrate validity evidence for the integration of the P-MEX in a residency admissions process using standardized patients.¹⁹ Furthermore, the P-MEX demonstrated predictive validity evidence for future professional behavior in residency training.²¹

Ideally, P-MEX scores and subscores should be used to provide meaningful and corrective feedback for learners. Moreover, scores should be followed over time to demonstrate progression and improvement in professional behaviors. And as such, identifying the robustness of fine-grained subdomains of professionalism can yield diagnostic and actionable information to

learners that is essential and goes beyond prior practices that have relied on overall professionalism scores. In this study, we set out to establish construct validity evidence, specifically internal structure validity evidence through factorial structure, for the use of the P-MEX in a simulated setting.

Method

Data were collected over an eight-year period from 2012 to 2019, as part of a standardized patient (SP) assessment using the P-MEX in the pediatrics residency admissions process.

The Institutional Review Board at the Geneva University Hospitals and the University of Illinois at Chicago granted an exemption for ethical approval for this study. We obtained written informed consent to analyze applicants' de-identified admission data from all participants.

Participants in the study were applicants to the Geneva University Hospitals Pediatric Residency Program. Applicants' noncognitive competencies were assessed using scores from two standardized letters of recommendation, one structured interview with two faculty members, and three standardized patient P-MEX scenarios rated by six raters (board-certified pediatricians, two raters per case). We described the adaptation of the P-MEX for simulated settings in a prior study and reported validity evidence for the implementation of the P-MEX following Messick's unified sources of validity evidence: content, response process, relations to other variables, internal structure, and consequences as operationalized in Downing and in the *Standards for Educational and Psychological Testing*.^{19,22,23} Details of the development of the blueprint, the admissions process, and the creation of a composite score have been published previously.¹⁹

Professionalism Mini-Evaluation Exercise (P-MEX)

In our admissions process, the P-MEX was employed in a simulated setting using three 13-minute SP cases representing pediatric professionalism challenges that involved conflicts of competing values based on Ginsburg's professionalism framework.²⁴ Over the eight-year study period, we created and used 11 different professionalism cases to assess the applicants. Typically, among the three cases used, one case would focus on a difficult parent encounter, the second an interprofessional conflict, and the third would involve a challenging adolescent consultation. A blueprint for the standardized patient cases can be found in Box 1. Frame of reference rater training to use the P-MEX instrument involved rating videos of volunteer residents portraying varying performance levels and discussing unacceptable behaviors.¹⁹ Items were scored on a 4-point scale: (1) unacceptable, (2) below expectations, (3) meets expectations, and (4) above expectations. Items not pertinent to the case were marked "not applicable". Items "solicited feedback" and "accepted feedback" were excluded from the analysis because there was no debriefing of the P-MEX stations and the item "was on time" was excluded due to the simulated setting. Each applicant was rated by two trained faculty raters for each SP encounter, while sitting behind a two-way mirror and using a paper form. Six P-MEX forms were generated per applicant. Responses were entered in duplicate using Data Scan©. Data were verified by the primary author (NB).

(Box 1 around here)

Validity evidence

Messick's unified validity framework, as operationalized in Downing²² and in the *Standards for*

*Educational and Psychological Testing*²³, was used to gather internal structure validity evidence for the use of P-MEX scores in the admissions process.²⁵

Internal structure

Descriptive statistics (means, standard deviations, and ANOVA) for the P-MEX subscores were calculated for each year in the study period as well as for the cumulative data. To determine the reliability of the P-MEX, Cronbach's alpha was conducted. To examine construct validity for the sample we carried out an exploratory factor (EFA) and then confirmatory factor analysis (CFA) through structural equation modeling on all P-MEX items, following recommended guidance from Fabrigar and colleagues.²⁶ The sample was randomly divided in half to conduct each step of the analysis, to empirically examine the factorial structure of the P-MEX, given its administration in a simulated learning environment as well as revised item structure, relative to the original scale.²⁶ We ran the factor analyses sequentially using the EFA and the CFA, respectively, to examine the stability of factor solutions across two random-halves of the data, informed by an exploratory approach driven by the data as well as confirming the hypothesized factorial structure. Retained factors for the EFA were based on the scree plot and had an eigenvalue >1. Factor labels were determined iteratively by the authors; these factor labels were presented, discussed, and confirmed by admissions committee members for verification and member checking. For the confirmatory factor analysis, the goodness of fit of the model was determined by the comparative fit index (CFI) and the root mean square error of approximation (RMSEA). CFI values >.90 indicate a good fit of the data.²⁷ A RMSEA <.05 indicates a close approximate fit, whereas a RMSEA <.10 indicates an acceptable fit.²⁸ Due to the nested structure of our data a second CFA analysis was conducted for sensitivity analysis and model fit consistency by aggregating scores for rater and case to check for possible overfitting of the model.

We performed the EFA with SPSS 26 for Macintosh (IBM Corp., Armonk, New York) and used StataSE 16 for Macintosh (Stata Corp, College Station, Texas) for the descriptive statistics, Cronbach's alpha, and the CFA.

Results

Among 275 interviewed applicants, 247 (90%) were women and 80 (29%) were foreign graduates from other countries within Europe. The P-MEX data consisted of 1650 forms generated by the 275 applicants (6 forms per applicant). P-MEX subscores for each of the domains are listed by year in Table 1. Scores varied minimally over the eight-year period. However, changes were significant for interprofessional skills and professional demeanor. Reliability of the P-MEX was high with a Cronbach's alpha of 0.91. Subscore reliabilities varied from 0.89 for doctor-patient relationship skills, to 0.74 for interprofessional skills, to 0.73 for professional demeanor, and 0.76 for reflective skills.

(Table 1 around here)

Internal structure

EFA revealed four factors with an eigenvalue >1 . We labeled the four factors as doctor-patient relationship skills, inter-professional skills, professional demeanor, and reflective skills. These dimensions captured 61% of the total variance with 29% of the variance attributed to doctor-patient relationship skills, 18% to inter-professional skills, 8% to professional demeanor, and 6% to reflective skills. The Kaiser-Meyer-Olkin measure of sampling adequacy was satisfactory at .81. Factor loadings are reported in Table 2.

(Table 2 around here)

Construct validity for the P-MEX was evaluated using the CFA. The CFA rendered a model that had a suitable model fit with an RMSEA of .058 and a CFI of .92; see Figure 1. Standardized coefficients ranged from .52 to .82. A second CFA was conducted using aggregated scores across rater and case to check for overfitting of the model. This model also demonstrated consistency in suitable model fit with an RMSEA of .075 and a CFI of .89.

(Figure 1 around here)

Discussion

Our results demonstrate construct validity evidence for the P-MEX in a simulated patient setting. Our initial exploratory factor analysis demonstrated four factors but a different factorial pattern than that initially described in the original factor analysis on the P-MEX (doctor-patient relationship skills, reflective skills, time-management skills, and inter-professional skills). Similar to the initial description of the P-MEX, we found that the same items came together for doctor-patient relationship skills and for interprofessional relationship skills. In addition, our results showed that “completed tasks in a reliable fashion” and “used health resources appropriately” loaded to doctor-patient relationship skills and that “admitted errors/omissions” and “was available to colleagues” loaded to interprofessional relationship skills. However, we interpreted that the grouping of items related to “maintaining appropriate boundaries”, “maintaining composure”, “maintaining appropriate appearance”, and “avoiding derogatory language” represented professional demeanor. Previously, these items were categorized as interprofessional relationship skills and reflective skills. In our analysis, we chose to include the global evaluation in our factor structure. The global evaluation loaded to “doctor-patient relationship skills” which

may be an indication that the examiners' overall perception of the performance during the simulation was influenced by the large number of items related to doctor-patient relationship skills in the scale. The CFA confirmed the model structure with a good model fit, and despite the nested structure of our dataset, the model using aggregated scores across rater and case also demonstrated suitable model fit and consistency in the factorial structure. Compared to two other studies that performed a CFA using the P-MEX, our model demonstrated more robust model fit, with lower RMSEA and comparable CFI.^{18,20}

This study provides evidence for the robustness of the subscore domains for the P-MEX. These subscores can be used reliably to provide personalized feedback for learners completing the P-MEX. Repeating the assessment at multiple occasions also provides for an opportunity to document the remediation and/or progression of the learner's professionalism competence. The information provided by the P-MEX may serve as a fundamental component of an individual learning plan or an educational portfolio.²⁹ The specific behavior-based P-MEX items allow supervisors to provide concrete suggestions for improving the learner's professional competence. To ensure that this feedback is meaningful and actionable, it is important that faculty observers are well prepared to deliver the feedback and that they have consensus on what constitutes professional and unprofessional behavior. Doing so will ensure that the P-MEX assessment falls into the category of "assessment for learning" instead of "assessment of learning".³⁰

This study contributes to the growing body of professionalism assessments that demonstrates validity evidence from which inferences about the learner can be made. Professionalism assessments during medical school have been shown to be predictive of future unprofessional events and disciplinary action by the state medical board as shown in the landmark work by Papadakis and colleagues.^{31,32} Integrating the assessment of professionalism early on in the

curriculum may help to identify deficiencies and lead to more effective remediation programs. Yet, we caution against using the P-MEX as the only assessment of professionalism in a curriculum. Professionalism is a multi-faceted construct that merits to be taught and assessed using multiple diverse methods to capture the professionalism skills of our learners. The P-MEX is an example of a behavioral-based instrument that makes it possible to give learners focused and specific feedback. The items included in the P-MEX seem to be representative of those behaviors that are the most easily observable during a simulation or real clinical encounter. It is possible that other relevant items may better reflect professionalism behaviors; future qualitative observational research would be needed to identify those behaviors.

Ideally, a professionalism curriculum should also include teaching and assessment of the values-based framework of professionalism and also focus on professional identity formation.³³

Integrating these constructs in addition to the P-MEX may give learners a perspective of professionalism that is well-rounded and nourish reflection on effective strategies for managing professionalism dilemmas. Further research on the P-MEX and other professionalism assessments should explore what strategies learners use to deal with uncertainty in a professionalism dilemma, the factors that may promote or hinder professional behavior, and what methods are effective at remediating professionalism difficulties. In our study, professionalism competencies were measured as a one-off occurrence during a simulation while in reality our learners build relationships with their patients and colleagues over time. Considering the competencies that favor the development of professional relationships may enrich the teaching and assessment of professionalism. Another possibility of future research would be to identify specific professionalism behaviors that are best suited to certain types of professional challenges

or conflicts. Future research will be needed to develop assessment tools that are able to measure these aspects of professionalism.

A strength of our Swiss-based study is that it provides additional cross-cultural validity evidence for use of the P-MEX which has already been employed in Canada, Japan, Iran, and Finland.^{14,18,20,34} Our results indicate that the professionalism items in the P-MEX are interpreted similarly in different contexts, thus contributing to response process evidence. There have been calls to provide more cross-cultural evidence for professionalism assessment and this study allows us to contribute to the growing body of evidence for the P-MEX.⁵ Our study is also strengthened by the large sample size which represents eight cohorts of applicants demonstrating varying levels of professionalism abilities. However, our study faces some limitations. Any assessment is limited by threats to validity such as construct under-representation and construct irrelevant variance. In our case, our original study showed that using the P-MEX with three cases and two raters per case would lead to a G-coefficient of 0.52 and would need six cases to arrive at a G-coefficient of 0.70.²¹ Because our study was conducted in a simulated setting we were able to control for outside factors that may influence scoring. However, scores were influenced by rater variations and case specificity as demonstrated in our prior study.¹⁹ Our study was also not longitudinal as each applicant was assessed at one moment in time. Dory et al. recently demonstrated that the P-MEX had low generalizability when used multiple times over a one-year period, possibly being due to students receiving high scores from the outset with little variation over the year.³⁵ Another potential limitation of our study is the complex data structure of our data set with observations being nested in applicants, cases, and raters. Our analysis did not take into account this multilevel structure and future analyses should explore whether multilevel confirmatory factor analyses would reveal new insights into the structure of the P-MEX.

Conclusions

The P-MEX demonstrates construct validity evidence for the assessment of professionalism.

Educators may wish to employ the P-MEX assessment as part of their curriculum on professionalism both in medical school and during residency training.

Disclosures

Acknowledgements: The authors wish to thank the applicants and residents that participated in this study for their time and collaboration.

Funding/Support: None

Other Disclosures: None

Ethical approval: Exemption from ethical review was provided by the Internal Review Board at the Geneva University Hospitals (September 5, 2018) and the University of Illinois at Chicago (October 25, 2018, Research Protocol # 2018-1167).

Disclaimers: None

Previous Presentations: none

References:

1. Epstein RM, Hundert EM. Defining and assessing professional competence. *JAMA*. 2002;287(2):226-235.
2. Ludwig S. Domain of competence: Professionalism. *Acad Pediatr*. 2014;14(2 Suppl):S66-69.
3. Cruess SR, Johnston S, Cruess RL. "Profession": A Working Definition for Medical Educators. *Teach Learn Med*. 2004;16(1):74-76.

4. Frank J, Snell L, Sherbino J, Editors. *CanMEDS 2015 Physician Competency Framework*. Ottawa: Royal College of Physicians and Surgeons of Canada;2015.
5. Hodges B, Paul R, Ginsburg S. Assessment of professionalism: From where have we come – to where are we going? An update from the Ottawa Consensus Group on the assessment of professionalism. *Med Teach*. 2019;1-7.
6. Roberts C, Wilkinson TJ, Norcini J, Patterson F, Hodges BD. The intersection of assessment, selection and professionalism in the service of patient care. *Med Teach*. 2019;1-6.
7. Hawkins RE, Katsufakis PJ, Holtman MC, Clauser BE. Assessment of medical professionalism: Who, what, when, where, how, and why? *Med Teach*. 2009;31(4):348-361.
8. Lynch DC, Surdyk PM, Eiser AR. Assessing professionalism: a review of the literature. *Med Teach*. 2004;26(4):366-373.
9. Veloski JJ, Fields SK, Boex JR, Blank LL. Measuring professionalism: a review of studies with instruments reported in the literature between 1982 and 2002. *Acad Med*. 2005;80(4):366-370.
10. Wilkinson TJ, Wade WB, Knock LD. A Blueprint to Assess Professionalism: Results of a Systematic Review. *Acad Med*. 2009;84(5):551-558.
11. Arnold L. Assessing professional behavior: yesterday, today, and tomorrow. *Acad Med*. 2002;77(6):502-515.
12. Goldie J. Assessment of professionalism: A consolidation of current thinking. *Med Teach*. 2012.

13. Ziring D, Frankel RM, Danoff D, Isaacson JH, Lochnan H. Silent Witnesses: Faculty Reluctance to Report Medical Students' Professionalism Lapses. *Acad Med.* 2018;93(11):1700-1706.
14. Cruess R, McIlroy JH, Cruess S, Ginsburg S, Steinert Y. The Professionalism Mini-Evaluation Exercise: A Preliminary Investigation. *Acad Med.* 2006;81(Suppl):S74-S78.
15. Rodriguez E, Siegelman J, Leone K, Kessler C. Assessing professionalism: summary of the working group on assessment of observable learner performance. *Acad Emerg Med.* 2012;19(12):1372-1378.
16. Green ML, Holmboe E. Perspective: the ACGME toolbox: half empty or half full? *Acad Med.* 2010;85(5):787-790.
17. Norcini JJ, Blank LL, Arnold GK, Kimball HR. The mini-CEX (clinical evaluation exercise): a preliminary investigation. *Ann Intern Med.* 1995;123(10):795-799.
18. Tsugawa Y, Ohbu S, Cruess R, et al. Introducing the Professionalism Mini-Evaluation Exercise (P-MEX) in Japan: Results From a Multicenter, Cross-Sectional Study. *Acad Med.* 2011;86(8):1026-1031.
19. Bajwa NM, Yudkowsky R, Belli D, Vu NV, Park YS. Improving the residency admissions process by integrating a professionalism assessment: a validity and feasibility study. *Adv Health Sci Educ Theory Pract.* 2016.
20. Amirhajlou L, Bidari A, Alipour F, et al. Validity, reliability and acceptability of Professionalism Mini-Evaluation Exercise (P-MEX) for emergency medicine residency training. *J Med Ethics Hist Med.* 2019;12:12-12.
21. Bajwa NM, Nendaz MR, Galetto-Lacour A, Posfay-Barbe K, Yudkowsky R, Park YS. Can Professionalism Mini-Evaluation Exercise Scores Predict Medical Residency Performance? Validity Evidence Across Five Longitudinal Cohorts. *Acad Med.* 2019;94(11S Association

- of American Medical Colleges Learn Serve Lead: Proceedings of the 58th Annual Research in Medical Education Sessions):S57-S63.
22. Downing SM. Validity: on meaningful interpretation of assessment data. *Med Educ.* 2003;37(9):830-837.
 23. AERA, APA. *Standards for Educational and Psychological Testing.* NCME. 2014.
 24. Ginsburg S, Regehr G, Hatala R, et al. Context, conflict, and resolution: a new conceptual framework for evaluating professionalism. *Acad Med.* 2000;75(10 Suppl):S6-S11.
 25. Messick S. Standards of validity and the validity of standards in performance assessment. *Educ Meas.* 1995;14(4):5-8.
 26. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. *Psychol Methods.* 1999;4(3):272.
 27. Bentler PM. Comparative fit indexes in structural models. *Psychol Bull.* 1990;107(2):238-246.
 28. Steiger JH. Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivar Behav Res.* 1990;25(2):173-180.
 29. O'Sullivan H, Van Mook W, Fewtrell R, Wass V. Integrating professionalism into the curriculum: AMEE Guide No. 61. *Med Teach.* 2012;34(2):e64-e77.
 30. Yudkowsky R, Park YS, Downing SM. *Assessment in health professions education.* Routledge; 2019.
 31. Papadakis MA, Hodgson CS, Teherani A, Kohatsu ND. Unprofessional behavior in medical school is associated with subsequent disciplinary action by a state medical board. *Acad Med.* 2004;79(3):244-249.
 32. Papadakis MA, Teherani A, Banach MA, et al. Disciplinary action by medical boards and prior behavior in medical school. *N Engl J Med.* 2005;353(25):2673-2682.

33. Irby DM, Hamstra SJ. Parting the Clouds: Three Professionalism Frameworks in Medical Education. *Acad Med*. 2016;91(12):1606-1611.
34. Karukivi M, Kortekangas-Savolainen O, Saxén U, Haapasalo-Pesu K-M. Professionalism Mini-Evaluation Exercise in Finland: A preliminary investigation introducing the Finnish version of the P-MEX instrument. *J Adv Med Educ Prof*. 2015;3(4):154-158.
35. Dory V, Gomez-Garibello C, Cruess R, Cruess S, Cummings B-A, Young M. The challenges of detecting progress in generic competencies in the clinical setting. *Med Educ*. 2018;52(12):1259-1270.

Box 1. Blueprint of standardized patient cases used with the P-MEX over the eight-year study period

Description of case	Encounter with a sexually active adolescent	Febrile infant that is in need of a work-up	Well-child checkup	Sick visit for a sore throat	Discharge of a hospitalized patient	Discharge of a hospitalized patient	Care of a hospitalized patient	Consultation in the emergency room	Consultation in the emergency room	Adolescent visit post ED visit for an alcohol intoxication	Collaboration with a colleague
Professionalism challenge	Explanation of a breach in confidentiality	Reticent mother who refuses care	Parent who refuses vaccination of their child	Parental pressure to prescribe antibiotics	Explanation of an error concerning the vaccination of the patient	Difficulty discharging the patient due to the mother's difficult social situation	Explanation of an error related to a medication overdose due to an interprofessional communication problem	The patient does not speak the language and an interpreter is not available	Conflict between two parents	The adolescent questions their sexuality	Dispute over the division of work
Doctor-patient relationship skills	√	√	√	√	√	√	√	√	√	√	
Inter-professional skills	√	√			√		√				√
Professional demeanor	√	√	√	√	√	√	√	√	√	√	√
Reflective skills	√	√	√	√	√	√	√	√	√	√	√

Table 1. Mean P-MEX subscores from 2012-2019 with a one-way ANOVA comparison between years.

	2012	2013	2014	2015	2016	2017	2018	2019	2012-2019 Cumulative	P- value
Doctor-Patient Relationship Skills	3.16 (.04)	3.15 (.03)	3.06 (.03)	3.12 (.03)	3.15 (.02)	3.13 (.03)	3.17 (.03)	3.13 (.03)	3.13 (.40)	.14
Inter- professional skills	3.14 (.04)	3.03 (.02)	3.13 (.03)	3.14 (.03)	3.11 (.03)	3.05 (.03)	3.17 (.04)	3.08 (.02)	3.10 (.37)	.003
Professional demeanor	3.14 (.03)	3.10 (.02)	3.08 (.02)	3.15 (.02)	3.13 (.02)	3.09 (.02)	3.16 (.03)	3.06 (.02)	3.11 (.30)	.003
Reflective skills	3.03 (.05)	3.01 (.03)	3.01 (.03)	3.01 (.03)	3.04 (.03)	2.97 (.03)	3.05 (.04)	3.00 (.03)	3.01 (.43)	.80

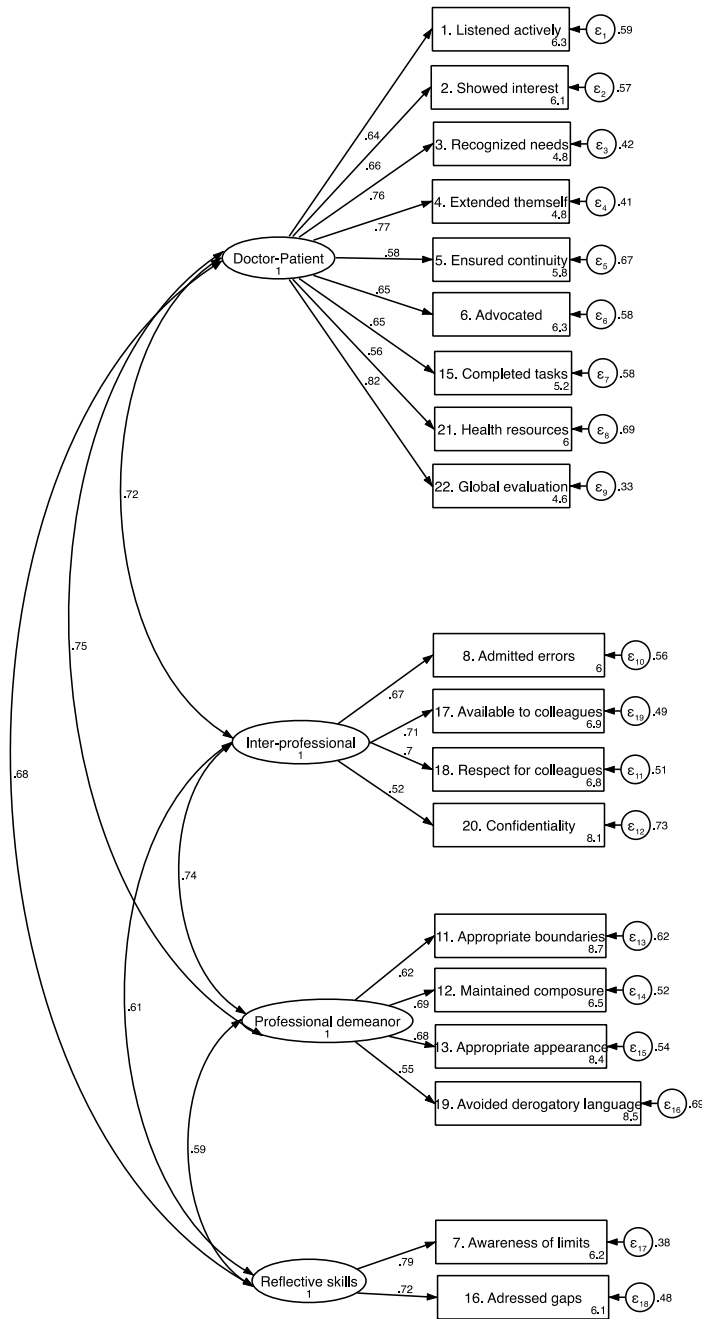
Table 2. Rotated Factor Matrix Solution for Exploratory Factor Analysis of the P-MEX

	Factor			
	1	2	3	4
Doctor-patient relationship skills				
1. Listened actively to patient	.489	-.004	.211	.062
2. Showed interest in patient as a person	.650	-.021	.100	-.042
3. Recognized and met patient needs	.877	.027	-.050	.060
4. Extended his/herself to meet patient needs	.680	-.179	-.041	.059
5. Ensured continuity of patient care	.671	.097	-.113	-.131
6. Advocated on behalf of a patient	.400	-.044	.117	-.077
15. Completed tasks in a reliable fashion	.541	.076	-.066	-.044
21. Used health resources appropriately	.223	-.165	.048	.017
22. Global evaluation	.798	.073	.021	.072
Inter-professional skills				
8. Admitted errors/omissions	.001	.850	.024	.090
17. Was available to colleagues	-.004	.999	.017	-.031
18. Demonstrated respect for colleagues	.066	.896	.004	.023
20. Maintained patient confidentiality	-.052	.116	-.063	-.020
Professional demeanor				
11. Maintained appropriate boundaries	.098	-.053	.558	.110
12. Maintained composure in a difficult situation	.063	-.029	.734	.034
13. Maintained appropriate appearance	-.179	.026	.885	-.045
19. Avoided derogatory language	.208	.154	.396	-.161
Reflective skills				
7. Demonstrated awareness of limitations	.016	.013	-.042	.894
16. Addressed own gaps in knowledge and skills	-.089	.106	.008	.743

^aExtraction method: Maximum Likelihood. Rotation method: Promax with Kaiser normalization.

^bRotation converged in 7 iterations.

Figure 1. Confirmatory Factor Analysis for the Professionalism Mini-Evaluation Exercise (P-MEX)



Observed variables are indicated in the rectangles and the latent variables are indicated in the ellipses. The model demonstrates good fit with a RMSEA of .058 and a CFI of .92.