



THE AMERICAN ORTHOPAEDIC ASSOCIATION®

Leading the profession since 1887

AOA Critical Issues in Education

Use of the Behavior Assessment Tool in 18 Pilot Residency Programs

April D. Armstrong, MD, FAOA, Julie Agel, MA, Matthew D. Beal, MD, FAOA, Michael S. Bednar, MD, FAOA, Michelle S. Caird, MD, FAOA, James E. Carpenter, MD, FAOA, Stuart T. Guthrie, MD, FAOA, Paul Juliano, MD, FAOA, Matthew Karam, MD, FAOA, Dawn LaPorte, MD, FAOA, J. Lawrence Marsh, MD, FAOA, Joshua C. Patt, MD, FAOA, Terrance D. Peabody, MD, FAOA, Karen Wu, MD, FAOA, David F. Martin, MD, FAOA, John J. Harrast, MS, and Ann E. Van Heest, MD, FAOA

Background: The purpose of this study was to determine the feasibility and evaluate the effectiveness of the American Board of Orthopaedic Surgery Behavior Tool (ABOSBT) for measuring professionalism.

Methods: Through collaboration between the American Board of Orthopaedic Surgery and American Orthopaedic Association's Council of Residency Directors, 18 residency programs piloted the use of the ABOSBT. Residents requested assessments from faculty at the end of their clinical rotations, and a 360° request was performed near the end of the academic year. Program Directors (PDs) rated individual resident professionalism (based on historical observation) at the outset of the study, for comparison to the ABOSBT results.

Results: Nine thousand eight hundred ninety-two evaluations were completed using the ABOSBT for 449 different residents by 1,016 evaluators. 97.6% of all evaluations were scored level 4 or 5 (high levels of professional behavior) across all of the 5 domains. In total, 2.4% of all evaluations scored level 3 or below reflecting poorer performance. Of 428 residents, the ABOSBT identified 26 of 32 residents who were low performers (2 or more < level 3 scores in a domain) and who also scored "below expectations" by the PD at the start of the pilot project (81% sensitivity and 51% specificity),

continued

ABOS/CORD Behavior Work Group: Elizabeth E. Ames, MD: University of Vermont Medical Center, South Burlington, Vermont; Raffi Avedian, MD: Stanford University, Redwood City, California; Thomas L. Bradbury, MD: Emory University School of Medicine, Atlanta, Georgia; Gerard J. Cush, MD: Geisinger Medical Center, Danville, Pennsylvania; John T. Gorczyca, MD: University of Rochester Medical Center, Rochester, New York; Mark E. Hake, MD: University of Michigan, Ann Arbor, Michigan; David F. Hubbard, MD: West Virginia School of Medicine, Morgantown, West Virginia; Joel C. Klana, MD: Geisinger Medical Center, Danville, Pennsylvania; Barry McDonough, MD: West Virginia School of Medicine, Morgantown, West Virginia; Peter Murray, MD: Mayo Clinic, Jacksonville, Florida; Scott E. Porter, MD: University of South Carolina School of Medicine, Greenville, South Carolina; Afshin E. Razi, MD: Maimonides Medical Center, Brooklyn, New York; James Roberson, MD: Emory University School of Medicine, Atlanta, Georgia; Brian Scannell, MD: University of North Carolina School of Medicine, Charlotte, North Carolina; Thomas M. Schaller, MD: University of South Carolina School of Medicine, Greenville, South Carolina; and Joseph K. Weistroffer, MD: Western Michigan University School of Medicine, Kalamazoo, Michigan.

Disclosure: The **Disclosure of Potential Conflicts of Interest** forms are provided with the online version of the article (<http://links.lww.com/JBJSOA/A212>).

Copyright © 2020 The Authors. Published by The Journal of Bone and Joint Surgery, Incorporated. All rights reserved. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Non Commercial-No Derivatives License 4.0](https://creativecommons.org/licenses/by-nc-nd/4.0/) (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

including 13 of these residents scoring poorly in all 5 domains. Evaluators found the ABOSBT was easy to use (96%) and that it was an effective tool to assess resident professional behavior (81%).

Conclusions: The ABOSBT was able to identify 2.4% low score evaluations (<level 3) for all residents. The tool was concordant with the PD for 81% of the residents considered low performers or “outliers” for professional behavior. The 5-domain construct makes it an effective actionable tool that can be used to help develop performance improvement plans for residents.

Level of Evidence: Level II

Society expects orthopaedic surgery residents completing training programs to act professionally. Measuring resident professionalism is a challenge, and it is our responsibility as orthopaedic educators to provide effective feedback to our residents regarding their level of professionalism.

The American Board of Orthopaedic Surgery (ABOS) and American Orthopaedic Association's Council of Residency Program Directors (PDs) (American Orthopaedic Association's Council of Residency Directors [AOA/CORD]) continue with their collaboration to develop the knowledge, skills, and behavior project¹. The behaviors portion of this project deals with the actions that as a whole reflect the resident's degree of “professionalism”². Presently, professional behaviors are reported as part of milestone assessments every 6 months for each resident as part of Accreditation Council for Graduate Medical Education (ACGME) Milestone 1.0 requirements³. The tools that programs use in assembling their behavior assessments vary widely and are not consistent across all residency programs.

The American Board of Orthopaedic Surgery Behavior Tool (ABOSBT) provides residency programs and clinical competency committees more directed and focused assessment of resident behaviors, using language standardized nationally. The tool is not a “pass” or “fail” assessment but rather a resource to provide effective feedback to the resident regarding their professionalism and can be used to develop performance improvement plans. The goal was to have all orthopaedic surgery residents in the United States understand and exhibit acceptable professional behavior to become board-certified orthopaedic surgeons.

Wilkinson et al.⁴ defined a blueprint of 5 assessable components for measuring professionalism that were used as the core assessment domains for the ABOSBT (Table I). Descriptors were then developed and added to give the evaluators some guidance or “things to consider” or “anchors” when assessing each of these domains⁵⁻⁸. It is important to recognize that the measured construct is repeatedly described throughout this study as “professionalism,” and these 5 domains are all categorized broadly as “professionalism.” However, under this broad construct of professionalism, it should be recognized that the ABOSBT also provides assessment of the ACGME communication and problem-based learning core competencies.

We believed that most residents would likely score high on the ABOSBT, regardless of their year in training. The goal of development of the ABOSBT is to identify the poor performers in professional behavior or “outliers,” compared with their peer group. The purpose of this study was to determine the feasibility and evaluate the effectiveness of the ABOSBT for measuring professional behavior. We hypothesized that the ABOSBT would be easy to use by evaluators and would effectively identify the “outlier” residents who score low for professionalism, when compared with PD's initial assessment.

Materials and Methods

Eighteen orthopaedic residency programs were selected by the CORD/AOA to represent a range of orthopaedic residency programs by size and geographic location. Institutional review board review was obtained and ruled as exempt (Exemption University of MN HRP-312). Faculty and resident informational material were provided to launch, educate, and execute the ABOSBT in each respective residency program.

Assessments were requested using the same platform as the ABOS Surgical Skills Assessment Tool⁹ that was open to receive assessments July 1, 2018, to June 30, 2019. All of the completed assessments within an individual training program were available for the PD to review online. The evaluator name and time of evaluation were redacted from the evaluation so that the PD was blinded to the specific evaluator identity to preserve confidentiality.

At the outset of the study, each PD was asked to provide an evaluation score for each of their residents individually regarding their level of professionalism. They were instructed to use past milestone, 360° evaluations, or other assessment tools to guide this assessment. The PD used a 4-point scale to score each participating resident's professionalism as (1) unacceptable, (2) below expectations, (3) meets expectations, or (4) exceeds expectations, termed the Baseline Professionalism PD Score.

Each resident was given a unique sign in and was instructed to electronically request a “Behavioral Assessment” during the last week of each rotation from every faculty whom they interacted with on that rotation. No immediate feedback was provided to the resident. To maintain confidentiality of the evaluations, the report back to the resident was provided at the

TABLE 1 Description of the American Board of Orthopaedic Surgery Behavioral Tool

Professional Domain	Descriptors
1. The resident adheres to the ethical principles	<p><i>Demonstrates honesty and integrity</i> (i.e., worthy of the trust bestowed on us by the patients' and the public's good faith, reports and analyzes medical errors, maintain confidentiality, understands their scope of practice with appropriate use of knowledge and skills, and trustworthy)</p> <p><i>Exhibits ethical behavior in professional code of conduct</i> (i.e., the student recognizes that being an orthopaedic surgeon is a "way of life" that serves the patient and community, advocates in the best interest of the patient, goes "above and beyond," they "do the right thing," respects diverse patient populations, including but not limited to diversity in sex, age, culture, race, religion, disabilities, and sexual orientation)</p>
2. The resident communicates effectively with patients and with people who are important to those patients	<p><i>Shows compassion/empathy</i> (i.e., Collaborates with patient, enhances the relationship)</p> <p><i>Demonstrates communication and listening skills</i> (i.e., attentive, shows patience, respects patient autonomy and empowers them to make informed decisions, and manages communication challenges with patients and families)</p> <p><i>Shows respect for patient needs</i> (i.e., respects patients' viewpoints and considers his/her opinions when determining healthcare decisions, regards the patient as a unique individual, treats the patient in the context of his/her family and social environment, and takes time to educate the patient and their family)</p>
3. The resident effectively interacts with other people working within the health system	<p><i>Shows ability to work with faculty, peers, and medical students</i> (i.e., shows respect, supports faculty mission to provide quality patient care, works collaboratively, can work with a team and cares for other members of the team, able to resolve conflicts effectively, adapts to change, and creates effective personal interactions)</p> <p><i>Students' level of composure</i> (i.e., ability to handle difficult situations with ease, has good coping strategies, and manages stress well)</p> <p><i>Students' identity formation</i> (i.e., ability to "fit in" with their role as a student learner, shows maturity in their specific role as a student physician learner, and socialized to the medical environment)</p>
4. The resident is reliable	<p><i>Work ethic</i> (i.e., shows interest and availability, protects patients interests, driven, willingness to conduct patient care without prompting, and committed to maintaining quality of care)</p> <p><i>Punctuality</i> (i.e., arrives to the clinic, OR, conferences, and call cases on time)</p> <p><i>Level of responsibility/accountability</i> (i.e., ability of the resident to answer for his/her conduct, timely completion of medical records or other required tasks, acknowledges their limitations, strives for excellence, shows pride in their actions and thoroughness, and level of confidence that a task will be carried out)</p>
5. The resident is committed to autonomous maintenance and continuous improvement of competence in self, others, and systems	<p><i>Students' ability to self-assess</i> (i.e., the resident recognizes their limits, ability to self-reflect and hold themselves accountable, commits to life-long learning, identifies strengths, deficiencies, and limits in one's knowledge and expertise, personal responsibility to maintain emotional, physical, and mental health)</p> <p><i>Students' receptiveness to critique</i> (i.e., the resident responds to feedback by accepting criticism, looks at self objectively, and changes their actions)</p>

OR = Operating Room, and PGY = Post-Graduate Year.

end of the academic year with a summative evaluation report. The PD could determine whether a performance improvement plan was required to target any of the 5 domains. If a resident received a low performing score of 1 or 2, then the PD was alerted electronically in real time so that the intervention could be implemented immediately if needed.

At the end of the academic year, the ABOSBT was also pushed out to a cohort of individuals for a 360-like evaluation. The PD identified a group of "other evaluators" to include all residents (peers), 10 midlevel orthopaedic pro-

viders (fellows, nurse practitioners, and physician assistants), 10 orthopaedic operating room (OR) nursing staff, 10 inpatient nursing staff, 10 orthopaedic outpatient clinic staff, and 10 emergency department (ED) faculty. In addition, each resident was asked to self-select 2 individuals from each of the cohort categories to provide an evaluation. At the end of the academic year, the ABOSBT results were compared with Baseline Professionalism PD Score for concordance. A survey was sent to the faculty to assess their experience using the ABOSBT.

TABLE II Number of Completed Behavior Assessments by Residency Program

Recorded Site Number	No. of Completed Evaluations	Percentage
1	1,361	13.8
2	1,273	12.9
3	776	7.8
4	673	6.8
5	638	6.4
6	630	6.4
7	626	6.3
8	601	6.1
9	558	5.6
10	477	4.8
11	420	4.2
12	405	4.1
13	387	3.9
14	280	2.8
15	269	2.7
16	260	2.6
17	232	2.3
18	26	0.3
Total	9,892	100.0

Statistics

Data analysis was performed using SPSS v 26@IBM, Microsoft Access 2016 and Excel 2016 for all descriptive statistics. For the calculation of specificity and sensitivity, each resident was categorized based on the PD scores and their domain scores. Any poor (<4) domain score gave them a positive test categorization and any PD score <3 gave them a positive disease present categorization. Medcalc v 12 was used to calculate the sensitivity and specificity of the evaluations.

Results

Analysis of Evaluations

Nine thousand eight hundred ninety-two evaluations were completed for 449 different residents (range 1 to 56 residents completed for each resident) in 18 residency programs. The numbers of evaluations by institution are shown in Table II. Evaluations completed by year in training are shown in Table III.

One thousand sixteen different evaluators participated in completing evaluations. Each evaluator completed between 1 and 50 evaluations. For the resident-requested evaluations, 468 orthopaedic faculty completed evaluations; the faculty completed 1,702 evaluations requested by the resident at the end of their rotation. For the 360 evaluations, 650 evaluations were completed by nonorthopaedic faculty identified by each resident (360 resident requested), and 7,540 evaluations were completed by individuals that the program

identified as part of the resident education environment (Table IV).

For each of the behavior domains, evaluators were asked to rate the residents by the scale, strongly disagree (1), disagree (2), neutral (3), agree (4), or strongly agree (5). Low scores on the ABOSBT were considered a score of 1, 2, or 3 for each of the 5 domains. For the 9,892 evaluations over 5 domains, 49,460 domain scores are available. Domain scores were low in 2.4% of evaluations. Low domain scores were compared for 360 push (selected by the program), 360 push (selected by the resident), and end of rotation orthopaedic faculty (selected by the resident). Chi-square demonstrates that there is a statistically significant difference ($p < 0.0001$) in distribution of the low scores across the 3 groups of evaluators, with the highest percentage of low scores given by the program selected evaluators during the 360 push (Table V).

Evaluations by Domain

The rating results of all evaluations on all residents (449) by domain are shown in Table VI. Across all of the domains, 97.6% of evaluations were reported as “strongly agree” or agree for behaviors. Across all residents, the domain with the greatest number of low scores was interaction; the domain with the least number of low score evaluations was ethical behavior.

Evaluation Results per Resident (7 or More Evaluations)

Four hundred thirty-one residents had 7 or more evaluations; 18 residents had less than 6 evaluations and were not included in further analysis. In this group of 431 residents, low-scoring residents were identified. Low scores on the ABOSBT were considered a score of 1, 2, or 3. Low-score residents had a minimum of 2 or more low scores (within one domain). Low-score residents are shown in Table VII, for each of the 5 domains. The domain with the greatest number of residents exhibiting low scores was interaction.

Concordance of Traditional PD Evaluation with the ABOSBT

The Baseline Professionalism PD Score was available for all G2 to G5 level residents. Baseline Professionalism Score as assigned by the PD was below expectations for 35 residents. Three of these residents only had 1 evaluation and were removed from this analysis, leaving 32 residents who each had at least 7 or

TABLE III Behavior Evaluations Completed by Resident Year in Training

Resident Training Year	No. of Evaluations	Percentage (%)
PGY-1	1,558	15.8
PGY-2	1,990	20.1
PGY-3	1,921	19.4
PGY-4	2,179	22.0
PGY-5	2,244	22.7
Total	9,892	100.0

TABLE IV 360° Types of Evaluators

	No. of Evaluators	No. of Evaluations	No. of Total Domains	Percentage
ED faculty	70	368	1,840	4.5
Inpatient nurse	81	462	2,310	5.6
Nurse practitioner	21	226	1,130	2.8
OR nurse	77	420	2,100	5.1
Orthopaedic fellow	16	36	180	0.4
Outpatient staff	88	572	2,860	7.0
Physician assistant	65	467	2,335	5.7
Faculty	250	3,513	17,565	42.9
Resident	124	2,126	10,630	26.0
Total	792	8,190	40,950	100.0

OR = Operating Room.

more evaluations using the ABOSBT but who also had a low score from the PD with a baseline score of 1 or 2. For those 32 residents, the distribution of their low domain scores are shown in Table VIII, and the number of low domain scores is shown in Table IX.

Sensitivity and Specificity of the Behavior Assessment Tool

The ABOSBT identified the same low performing residents as the PD's in 26 of 32 instances (Table X). Thus, the *sensitivity* of the ABOSBT assessment when compared with the PD negative (poor) baseline assessment is 81% (95% confidence interval [CI] 64% to

93%) (*True Positive*); i.e., the ABOSBT will identify those residents whom the PD has identified as unprofessional is concordant 81% of the time.

On the other hand, the ABOSBT identified 176 residents as scoring low on at least one assessment of the 408, residents' PDs scored as meeting or exceeding expectations. Thus, the *specificity* of the ABOS Behavior Assessment when compared with the PD positive (good) baseline assessment is 51% (95% CI 45% to 56%) (*True Negative*), i.e., the ABOSBT will identify those residents whom the PD has identified as professional is concordant 57% of the time.

TABLE V Low Domain Scores by Source of Evaluation Request

	Sample	Low Scores (1, 2, 3) Numbers	Low Score % of Evaluations
360 push program selected	37,700	1,059	2.8%
360 push resident selected	3,250	35	1.1%
End of rotation faculty (resident selected)	8,510	84	1%
Total	49,460	1,178/49,460	2.4%

TABLE VI All Behavior Assessment Evaluation Results by Domain

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Total
Ethical behavior	30	13	106	472	9,271	9,892
Communication	23	28	202	838	8,801	9,892
Interaction	28	54	191	937	8,682	9,892
Reliability	27	46	179	733	8,906	9,891
Self-assessment	17	30	204	807	8,834	9,892
Total	125	171	882	3,787	44,494	49,459
Percent	0.3	0.3	1.8	7.6	90.0	100.0

TABLE VII Number of Residents with 2 or More Low Scores within a Domain

Domain	No. of Residents with ≥2 Low Scores within the Same Domain	Percentage
Ethical behavior	32	8%
Communication	61	14%
Interaction	63	15%
Reliability	57	13%
Self-assessment	59	14%

TABLE VIII Distribution of Low Domain Scores for 32 Residents with Low PD Baseline Assessment*

	No. of Baseline Professionalism PD Assessment Low Score Residents (N = 32) Also with Low Scores on the Behavior Tool
Ethical behavior	17
Communication	20
Interaction	23
Reliability	22
Self-assessment	21

*PD = Program Director.

Faculty Survey

148/468 (32% response rate) faculty completed the survey to evaluate the ABOSBT and their experience using the tool (Table XI). Eighty-six percent believed that the length of the assessment was “just right.”

Discussion

Our findings support the hypothesis that evaluators would find the ABOSBT easy to use (96%) and as an effective tool to assess resident professional behavior (81%) (Table XI).

The ABOSBT was in accordance with the PD initial assessment because it identified 26 of 32 residents who scored below expectations by the PD at the start of the project. Therefore, the ABOSBT was concordant with the PD for 81% of the residents with low scores and ongoing concerns regarding professional behavior. The ABOSBT had a specificity of 51%, identifying 196 residents of the 408 residents rated by PDs as meeting or exceeding expectations as low scoring in one or more domains.

As expected, 97.6% of all evaluations were scored level 5 (strongly agree) or scored level 4 (agree) in all 5 domains. This left 2.4% of all evaluations that were scored level 3 or below reflecting poorer performance and an opportunity for improvement across all the residents of these 18 programs. Unlike surgical skills that develop and improve over time, most residents showed excellent behavioral performance across all domains, regardless of year in training. The value of the ABOSBT is to identify those residents who would be considered “outliers.” Because the tool is divided into 5 different domains, the PD may develop a focused performance improvement plan for the resident based on the domain(s) that he or she showed lower performance. This creates a highly effective and actionable tool that could then be used to monitor progress in low-performing domains. This also serves to create an accurate record of behavioral deficiencies in the rare case that an adverse action against the resident is warranted. It is also interesting to note that not one domain substantially outranked another domain for poor performance. All 5 domains showed comparable numbers of low performance scores, 4% to 6% of residents with greater than 2 low scores in a single domain, suggesting that all 5 domains are relevant and important to measure. For the future, there is opportunity to develop programs that could help with remediation in each of these domains. Remediation could also extend beyond the “outliers” because the tool showed a low specificity. This could be considered another strength of the tool in that it identified residents with possible behavioral deficiencies that were not otherwise recognized by the PD.

When developing the ABOSBT, other measurement tools reported in the literature were explored. The P-Mex tool¹⁰⁻¹² and the University of Michigan, Department of Surgery Professionalism Assessment Instruments¹³ were considered, but it

TABLE IX Number of Low Domain Scores for Low Baseline Professionalism PD Assessment Score Residents (n = 32)*

No. of Residents with Low Score on Baseline Professionalism PD Assessment and No Low Score Behavior Tool Domains	No. of Residents with Low Score on Baseline Professionalism PD Assessment and 1 Low Score Behavior Tool Domains	No. of Residents with Low Score on Baseline Professionalism PD Assessment and 2 Low Scores on Behavior Tool Domains	No. of Residents with Low Score on Baseline Professionalism PD Assessment and 3 Low Score Behavior Tool Domains	No. of Residents with Low Score on Baseline Professionalism PD Assessment and 4 Low Score Behavior Tool Domains	No. of Residents with Low Score on Baseline Professionalism PD Assessment and 5 Low Score Behavior Tool Domains
7	2	1	3	5	14

*PD = Program Director.

TABLE X Specificity and Sensitivity of the ABOS Behavior Tool Compared with PD Baseline Assessment for All Participating Residents (n = 428)*

	PD Baseline Assessment Low Score (1 or 2)	PD Baseline Assessment High Score (3 or 4)
ABOS Behavior Assessment Low Scores (1, 2, 3)	26	196
ABOS Behavior Assessment High Scores (4, 5)	6	200

*ABOS = American Board of Orthopaedic Surgery, and PD = Program Director.

TABLE XI Faculty Survey Results

Survey Question	Agreed or Strongly Agreed
User interface was intuitive	98%
Easy to complete assessment	96%
Able to complete the assessments	97%
Behavior tool was beneficial compared to other methods	82%
Behavior tool was effective to assess resident professionalism	81%
Five domains of tool were effective	86%
Descriptors for 5 domains were helpful	89%

was determined that they were too lengthy. The faculty survey showed that 86% and 89% agreed or strongly agreed that the 5 domains of assessment for professional behavior were effective and that the descriptors for the 5 domains of assessment were helpful prompts to evaluate resident professional behavior, respectively.

A critical guiding principle when measuring behaviors and professionalism is to stay away from “a single evaluator at a single time point” approach¹⁴⁻²² (see Appendix). We found that the 360° program push evaluations were able to identify more low-score evaluations (2.8%) than a resident-driven 360 evaluation (1.1%) or the faculty evaluations at the end of the rotations (1%). We included the resident-chosen 360 evaluations to explore whether this would introduce a “selection bias.” The resident-chosen 360 evaluations identified fewer low score evaluations that was comparable with the end of rotation faculty evaluations. We would propose that the program-driven 360 push for evaluations once per year is an important component for a behavioral assessment program.

The ABOSBT provided 81% sensitivity with identifying 26 of the 32 residents whom the PDs also identified as low performers. However, the ABOSBT had 51% specificity because it identified 196 additional residents with low-performance scores in at least one domain by one evaluator. The 360 push provides the advantage of providing viewpoints from multiple providers in the education environment. Some evaluators were noted to straight-line negative performances for multiple residents within a program. Algorithms set up in this analysis included that a resident needed to have multiple evaluations (7 or more evaluations) with low scores by at least 2 evaluators. When the ABOSBT is used on a large scale across the country, such algorithms will be needed to safe guard from a single negative evaluator and to ensure that true patterns of unprofessional behavior are detected.

Strengths of this study are that the ABOS has developed the ABOSBT that will be available to all orthopaedic surgery residency programs in the United States. The ABOSBT is limited to 5 questions, which respects the educator's time in completion of the survey. All 5 domains included in the tool are important aspects of behavior, as evidenced by a comparison to other studies.

Limitations of this study include lack of data to assess the performance of the different evaluators. It is possible that there is variability in the severity of the evaluators, and over time, as more data continue to be collected, we will be able to “level set” the evaluator performance for severity. There is future opportunity to develop educational programs for evaluators, which could increase the reliability, and the ABOS has experience with developing “severity score indices” for examiners that give the annual Part II Oral Board Examination. We would foresee a similar approach being developed as we gain more experience using the ABOSBT. Use of the 360 tool requires the residency programs to identify multiple healthcare individuals in multiple environments who can evaluate resident performance. For large residency programs with multiple rotation sites, identifying individuals with adequate exposure to complete the assessment may be a challenge. The number of evaluations per resident varied considerably (1 to 56), this could help explain why the specificity of the tool was low (51%), and this could improve if more time was given to collect evaluations.

Conclusions

Assessment tools allow educators to provide feedback and guide performance to reach expected standards. Although orthopaedic educators often have considerable experience in assessing competency in knowledge and patient care skills, assessment of professional behaviors can be more challenging. Providing a common framework and language for assessing appropriate professional behavior is an important component of the ABOS collaboration with AOA/CORD in providing assessment tools for knowledge, skills, and behavior during residency training. The ABOSBT is an electronic web-based tool that is easy to use and effective real time, for measuring professionalism for orthopaedic residents across 5 domains of behavior. The 5-domain construct makes it a valuable actionable tool that can be used to help develop performance improvement

plans early in the residency training program, with a goal of educating competent, ethical, board-certified orthopaedic surgeons.

Appendix

 Supporting material provided by the authors is posted with the online version of this article as a data supplement at [jbjs.org \(http://links.lww.com/JBJSOA/A371\)](http://links.lww.com/JBJSOA/A371). ■

April D. Armstrong, MD, FAOA¹
 Julie Agel, MA²
 Matthew D. Beal, MD, FAOA³
 Michael S. Bednar, MD, FAOA⁴
 Michelle S. Caird, MD, FAOA⁵
 James E. Carpenter, MD, FAOA⁶
 Stuart T. Guthrie, MD, FAOA⁷
 Paul Juliano, MD, FAOA¹
 Matthew Karam, MD, FAOA⁸
 Dawn LaPorte, MD, FAOA⁹
 J. Lawrence Marsh, MD, FAOA⁸
 Joshua C. Patt, MD, FAOA¹⁰
 Terrance D. Peabody, MD, FAOA³
 Karen Wu, MD, FAOA⁴
 David F. Martin, MD, FAOA¹¹
 John J. Harrast, MS¹²
 Ann E. Van Heest, MD, FAOA¹³

¹Department of Orthopaedics and Rehabilitation, Bone and Joint Institute, Penn State Milton S. Hershey Medical Center, Hershey, Pennsylvania

²Department Orthopaedics and Sports Medicine, Harborview Medical Center, Seattle, Washington

³Department of Orthopaedic Surgery, North Western University, Chicago, Illinois

⁴Department of Orthopaedic Surgery and Rehabilitation, Loyola University Medical Center, Chicago, Illinois

⁵Department of Orthopaedic Surgery, University of Michigan, Ann Arbor, Michigan

⁶Department of Orthopaedic Surgery, Sports Medicine, Med Sport, Ann Arbor, Michigan

⁷Henry Ford Orthopaedics, Henry Ford Hospital, Detroit, Michigan

⁸Department of Orthopaedics and Rehabilitation, University of Iowa, Iowa City, Iowa

⁹Department of Orthopaedic Surgery, Johns Hopkins Medicine, Baltimore, Maryland

¹⁰CMC Orthopaedic Surgery, University of North Carolina School of Medicine, Charlotte, North Carolina

¹¹Department of Orthopaedic Surgery, Wake Forest Baptist Health, Winston-Salem, North Carolina

¹²Data Harbor Solutions, Chicago, Illinois

¹³Department of Orthopaedic Surgery, University of Minnesota, Minneapolis, Minnesota

E-mail address for A.D. Armstrong: aarmstrong@pennstatehealth.psu.edu

ORCID iD for A.D. Armstrong: [0000-0003-3729-9811](https://orcid.org/0000-0003-3729-9811)

ORCID iD for J. Agel: [0000-0002-4296-7389](https://orcid.org/0000-0002-4296-7389)

ORCID iD for M.S. Caird: [0000-0001-5865-4793](https://orcid.org/0000-0001-5865-4793)

ORCID iD for J.E. Carpenter: [0000-0002-3070-1751](https://orcid.org/0000-0002-3070-1751)

ORCID iD for S.T. Guthrie: [0000-0002-0678-493X](https://orcid.org/0000-0002-0678-493X)

ORCID iD for P. Juliano: [0000-0002-7705-4117](https://orcid.org/0000-0002-7705-4117)

ORCID iD for D. LaPorte: [0000-0001-6564-2707](https://orcid.org/0000-0001-6564-2707)

ORCID iD for J. L. Marsh: [0000-0002-3494-6289](https://orcid.org/0000-0002-3494-6289)

ORCID iD for J.C. Patt: [0000-0001-7323-7728](https://orcid.org/0000-0001-7323-7728)

ORCID iD for T.D. Peabody: [0000-0002-6669-4549](https://orcid.org/0000-0002-6669-4549)

ORCID iD for D.F. Martin: [0000-0001-8975-7031](https://orcid.org/0000-0001-8975-7031)

ORCID iD for J.J. Harrast: [0000-0003-3037-2748](https://orcid.org/0000-0003-3037-2748)

ORCID iD for A.E. Van Heest: [0000-0002-8064-5508](https://orcid.org/0000-0002-8064-5508)

References

- Nousiainen M, Incoll I, Peabody T, Marsh JL. Can we agree on expectations and assessments of graduating residents?: 2016 AOA critical issues symposium. *J Bone Joint Surg Am.* 2017;99:e56.
- Lynch DC, Surdyk PM, Eiser AR. Assessing professionalism: a review of the literature. *Med Teach.* 2004;26:366-73.
- Nousiainen MT, McQueen SA, Hall J, Kraemer W, Ferguson P, Marsh JL, Reznick RR, Reed MR, Sonnadara R. Resident education in orthopaedic trauma: the future role of competency-based medical education. *Bone Joint J.* 2016;98-B:1320-5.
- Wilkinson TJ, Wade WB, Knock LD. A blueprint to assess professionalism: results of a systematic review. *Acad Med.* 2009;84:551-8.
- Frohna A, Stern D. The nature of qualitative comments in evaluating professionalism. *Med Educ.* 2005;39:763-8.
- Kearney RA. Defining professionalism in anaesthesiology. *Med Educ.* 2005;39:769-76.
- Rabinowitz D, Reis S, Van Raalte R, Alroy G, Ber R. Development of a physician attributes database as a resource for medical education, professionalism and student evaluation. *Med Teach.* 2004;26:160-5.
- Wagner P, Hendrich J, Moseley G, Hudson V. Defining medical professionalism: a qualitative study. *Med Educ.* 2007;41:288-94.
- Van Heest AE, Agel J, Ames SE, Asghar FA, Harrast JJ, Marsh JL, Patt JC, Sterling RS, Peabody TD. Resident surgical skills web-based evaluation: a comparison of 2 assessment tools. *J Bone Joint Surg Am.* 2019;101:e18.
- Cruess R, Mclroy JH, Cruess S, Ginsburg S, Steinert Y. The professionalism mini-evaluation exercise: a preliminary investigation. *Acad Med.* 2006;81:S74-8.
- Karukivi M, Kortekangas-Savolainen O, Saxen U, Haapasalo-Pesu KM. Professionalism mini-evaluation exercise in Finland: a preliminary investigation introducing the Finnish version of the P-MEX instrument. *J Adv Med Educ Prof.* 2015;3:154-8.
- Tsugawa Y, Ohbu S, Cruess R, Cruess S, Okubo T, Takahashi O, Tokuda Y, Heist BS, Bito S, Itoh T, Aoki A, Chiba T, Fukui T. Introducing the Professionalism Mini-Evaluation Exercise (P-MEX) in Japan: results from a multicenter, cross-sectional study. *Acad Med.* 2011;86:1026-31.
- Gauger PG, Gruppen LD, Minter RM, Colletti LM, Stern DT. Initial use of a novel instrument to measure professionalism in surgical residents. *Am J Surg.* 2005;189:479-87.
- Hilton SR, Slotnick HB. Proto-professionalism: how professionalisation occurs across the continuum of medical education. *Med Educ.* 2005;39:58-65.
- Jha V, Bekker HL, Duffy SR, Roberts TE. Perceptions of professionalism in medicine: a qualitative study. *Med Educ.* 2006;40:1027-36.
- Swick HM. Toward a normative definition of medical professionalism. *Acad Med.* 2000;75:612-6.
- Van De Camp K, Vernooij-Dassen MJ, Grol RP, Bottema BJ. How to conceptualize professionalism: a qualitative study. *Med Teach.* 2004;26:696-702.
- Wass V. Doctors in society: medical professionalism in a changing world. *Clin Med (Lond).* 2006;6:109-13.

19. Hughes G. Understanding doctors: harnessing professionalism. *Emerg Med J.* 2008;25:788.

20. Chard D, Elsharkawy A, Newbery N. Medical professionalism: the trainees' views. *Clin Med (Lond).* 2006;6:68-71.

21. Hauck FR, Zyzanski SJ, Alemagno SA, Medalie JH. Patient perceptions of humanism in physicians: effects on positive health behaviors. *Fam Med.* 1990;22:447-52.

22. Cohen JJ. Professionalism in medical education, an American perspective: from evidence to accountability. *Med Educ.* 2006;40:607-17.

Update

This article was updated on March 17, 2022, because of previous errors on pages 1, 4, 5, 6, and 7. In the Abstract, Results, and Discussion sections, the specificity of the ABOS Behavior Assessment Tool that had read “57%” and “57% (95% CI 52% to 62%)” now reads “51%” and “51% (95% CI 45% to 56%).” In the Abstract, “1,012 evaluators” now reads “1,016 evaluators” and “431 residents” now reads “428 residents.” In the Results section entitled “Evaluation Results per Resident,” the sentence that had read “The domain with the greatest number of residents exhibiting low scores was ethical behavior.” now reads “The domain with the greatest number of residents exhibiting low scores was interaction.” In the Discussion section on pages 6 and 7, the number of residents with low scores in at least one domain that had read “176” now reads “196.” In Table III, the title that had read “Behavior Evaluations Completed by Resident Year in Training Source” now reads “Behavior Evaluations Completed by Resident Year in Training.” In Table VII, the column head that had read “No. of Residents with >2 Low Scores within the Same Domain” now reads “No. of Residents with ≥ 2 Low Scores within the Same Domain.” Also in Table VII, in columns 2 and 3, the values that had read “26 (6%), 23 (5%), 19 (4%), 18 (4%), and 20 (5%)” now read “32 (8%), 61 (14%), 63 (15%), 57 (13%), and 59 (14%).” In Table VIII, the column head that had read “No. of Baseline Professionalism PD Assessment Low Score Residents (N = 32) Also with Low Scores on the Behavior Tool by at least 2 Evaluators” now reads “No. of Baseline Professionalism PD Assessment Low Score Residents (N = 32) Also with Low Scores on the Behavior Tool.” In Table IX, the title that had read “Number of Low Domain Scores by at least 2 Evaluators for Low Baseline Professionalism PD Assessment Score Residents (n = 32)*” now reads “Number of Low Domain Scores for Low Baseline Professionalism PD Assessment Score Residents (n = 32)*.” Also, the values in the table that had read “6, 3, 2, 2, 6, 13” now read “7, 2, 1, 3, 5, 14.” In Table X, the title that had read “Specificity and Sensitivity of the ABOS Behavior Tool Compared with PD Baseline Assessment for All Participating Residents (n = 440)*” now reads “Specificity and Sensitivity of the ABOS Behavior Tool Compared with PD Baseline Assessment for All Participating Residents (n = 428)*.” Also in Table X, in the right column entitled “PD Baseline Assessment High Score (3 or 4),” the values that had read “176” and “232” now read “196” and “200,” respectively. Finally, a supplementary data file has now been included with the article that contains results that portray resident performance with at least two low scores in one domain by at least two different evaluators.

An erratum has been published: JBJS Open Access. 2022 Mar 31;7(1):e20.00103ER.