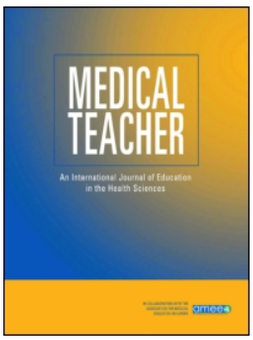

Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: http://www.tandfonline.com/loi/imte20

Not all unprofessional behaviors are equal: The creation of a checklist of bad behaviors

Michael J. Cullen, Mojca R. Konia, Emily C. Borman-Shoap, Jonathan P. Braman, Ezgi Tiryaki, Brittany Marcus-Blank & John S. Andrews

To cite this article: Michael J. Cullen, Mojca R. Konia, Emily C. Borman-Shoap, Jonathan P. Braman, Ezgi Tiryaki, Brittany Marcus-Blank & John S. Andrews (2016): Not all unprofessional behaviors are equal: The creation of a checklist of bad behaviors, Medical Teacher, DOI: 10.1080/0142159X.2016.1231917

To link to this article: http://dx.doi.org/10.1080/0142159X.2016.1231917

MEDICAL TEACHER, 2016
http://dx.doi.org/10.1080/0142159X.2016.1231917

MEDICAL TEACHER
Taylor & Francis
Taylor & Francis Group

Not all unprofessional behaviors are equal: The creation of a checklist of bad behaviors

Michael J. Cullen[a], Mojca R. Konia[b], Emily C. Borman-Shoap[c], Jonathan P. Braman[d], Ezgi Tiryaki[e,f], Brittany Marcus-Blank[a] and John S. Andrews[a]

[a]Graduate Medical Education, University of Minnesota, Minneapolis, MN, USA; [b]Department of Anesthesiology, University of Minnesota, Minneapolis, MN, USA; [c]Department of Pediatrics, University of Minnesota, Minneapolis, MN, USA; [d]Department of Orthopaedic Surgery, University of Minnesota, Minneapolis, MN, USA; [e]Minneapolis VA Health Care System, Minneapolis, MN, USA; [f]Department of Neurology, University of Minnesota, Minneapolis, MN, USA

ABSTRACT

Introduction: Professionalism is a key component of medical education and training. However, there are few tools to aid educators in diagnosing unprofessional behavior at an early stage. The purpose of this study was to employ policy capturing methodology to develop two empirically validated checklists for identifying professionalism issues in early-career physicians.

Method: In a series of workshops, a professionalism competency model containing 74 positive and 70 negative professionalism behaviors was developed and validated. Subsequently, 23 subject matter experts indicated their level of concern if each negative behavior occurred 1, 2, 3, 4, or 5 or more times during a six-month period. These ratings were used to create a "brief" and "extended" professionalism checklist for monitoring physician misconduct.

Results: This study confirmed the subjective impression that some unprofessional behaviors are more egregious than others. Fourteen negative behaviors (e.g. displaying obvious signs of substance abuse) were judged to be concerning if they occurred only once, whereas many others (e.g. arriving late for conferences) were judged to be concerning only when they occurred repeatedly.

Discussion: Medical educators can use the professionalism checklists developed in this study to aid in the early identification and subsequent remediation of unprofessional behavior in medical students and residents.

Introduction

In recent years, researchers have made important strides in defining the nature of professionalism in the medical sphere. Although differences exist across models, most researchers agree that professionalism includes a constellation of personal characteristics that include altruism, integrity, compassion, respect for others, accountability, self-awareness, a drive for excellence, and self-improvement (ABIM Foundation et al. 2002; AAMC & NBME 2003; Frank 2005; Royal College of Physicians 2005; Arnold & Stern 2006; Van Mook et al. 2009; Nasca et al. 2012). There is also widespread agreement regarding the importance of assessing professionalism in medical students, residents, and practicing physicians (Veloski et al. 2005; Wilkinson et al. 2009). Studies have shown that professional behaviors such as teamwork and respect are correlated with improved patient outcomes, higher patient satisfaction (Grumbach & Bodenheimer 2004), and greater adherence to treatment plans (Beach et al. 2005). Conversely, unprofessional behaviors are associated with negative faculty assessments of professionalism (Stern et al. 2005) and later disciplinary action by state medical boards (Papadakis et al. 2005). In addition to their effects on individuals, breaches of professionalism can have significant group-level costs by negatively affecting the reputation of the medical profession itself.

The serious consequences of physician misconduct on trainees, patients, and the profession argue in favor of an "early warning system" for detecting unprofessional behaviors. Papadakis et al. (2001) have created one such system in which preceptors and faculty rate students on a set of discrete, observable behaviors such as "fulfilling responsibilities in a reliable manner" and "maintaining honesty" during clerkships. Subsequently, these preceptors and faculty submit a "Physicianship Evaluation Form" for students who receive a less than satisfactory rating on these behaviors at

Practice points

- Professionalism training is an important component of medical education, but there are few tools to diagnose unprofessional behavior early in a physician's career.
- The egregiousness of an unprofessional behavior depends both on the behavior itself and the frequency with which it occurs.
- Of the seven professionalism dimensions investigated in this study, breaches of the integrity dimension were viewed to be most concerning.
- Further research should aim to discover whether the behaviors in these professionalism checklists predict important future outcomes, such as disciplinary action or professionalism milestone scores in residents.

CONTACT Dr. Michael J. Cullen cull0061@umn.edu Director of Evaluation for Graduate Medical Education, University of Minnesota Medical School, 420 Delaware Street S.E., Mayo Mail Code 293, Minneapolis, MN 55455, USA

Supplemental data for this article can be accessed here.

the end of any clerkship. If deficiencies in skills are identified in two or more clerkships, the dean's letter will document these areas of concern, and probation will follow. Faculty appreciate this system, and the authors believe it has helped to raise awareness of the core value of professionalism (Papadakis et al. 2001). In addition to formal systems for identifying professionalism issues, direct observation tools have been developed. These include tools such as the Mini-Clinical Evaluation Exercise (mini-CEX) (Norcini et al. 2003) and the Professionalism Mini-Evaluation Exercise (P-MEX) (Cruess et al. 2006). In the mini-CEX, a single faculty member observes a resident while that resident conducts a focused history and physical examination in a variety of settings. Subsequently, the faculty member rates the resident on a nine-point scale on various components of competence, including history-taking skill, physical examination skill, clinical judgment and synthesis, and humanistic/professionalism qualities. Behavioral anchors are provided for the ratings and include such behaviors as demonstrating respect, compassion, empathy and modesty, establishing trust, attending to a patient's needs of comfort, and maintaining confidentiality.

The P-MEX was designed for use in any situation where a medical student's behavior can be observed, including patient encounters, small group sessions, and sign-out rounds. It builds upon the mini-CEX by identifying a set of even more specific behaviors to be observed. Both of these observational tools have formats that are easy to use and provide useful feedback, and both have demonstrated good reliability, validity (Durning et al. 2002), and inter-examiner reliability (Norcini et al. 1997). Similar direct observation tools, such as the Standardized Direct Observation Tool, have been designed for use in specific settings, such as during Emergency Department patient encounters (Shayne et al. 2006).

One common feature of all direct observation tools for detecting professionalism issues (Arnold et al. 1998; Reed et al. 2008) is that they employ an observational "checklist" of positive or negative professionalism behaviors, which are either rated directly or serve as behavioral anchors in a scale. Such checklists have the value of being specific, to help guide remediation efforts. Several researchers have argued that when remediation is the goal, an assessment method that focuses on specific, observable unprofessional behaviors is helpful (Hawkins et al. 2009). An important limitation of the checklists designed so far is that they do not distinguish between the severity of the behaviors on the lists (Arnold 2002). In particular, existing professionalism checklists do not (a) distinguish between the *relative egregiousness* of behaviors that occur in isolation or (b) indicate whether the egregiousness of behaviors is affected by their *frequency* of occurrence. We believe these distinctions are important because it is possible that not all unprofessional behaviors are equally serious in isolation. For instance, some behaviors (e.g. displaying signs of substance abuse) may be so serious that a single occurrence of the behavior requires immediate remediation. In contrast, some behaviors (e.g. reacting defensively to criticism) may only be concerning when a pattern of misbehavior emerges over time. It would be desirable, therefore, to develop a professionalism checklist that takes these considerations into account.

In this study, we employed policy capturing methodology to develop two empirically validated professionalism checklists that take the relative egregiousness of behaviors and their frequency of occurrence into account. First, we created a "brief" checklist that identified the most egregious behaviors that required remediation even if they only occurred once. Second, we created an "extended" checklist that allowed for the identification of *patterns* of less-serious misbehavior that nonetheless required remediation. Policy capturing is a method employed by researchers to assess how decision makers use available information when making evaluative judgments. The purpose of this methodology is to capture individual judges' decision-making policies, that is, how they weight, combine, or integrate information. It involves asking decision makers to judge a series of scenarios describing various levels of one or more explanatory factors, and then using statistical methods to determine the emphasis decision makers give to each factor in their decision-making process (Zedeck 1977). Policy capturing has been used to assess judgments in a number of areas including compensation (Sherer et al. 1987), employee discipline (Klaas & Wheeler 1990), and employment interviews (Dougherty et al. 1986; Graves & Karren 1992). In this study, we asked education experts to judge how concerning a set of unprofessional behaviors would be if they occurred at different levels of frequency. Specifically, we captured program directors' level of concern if the unprofessional behaviors were repeated once, twice, three times, four times or five or more times in a six-month time period. We then used this *behavior* × *frequency* information to create the two professionalism checklists. We used a set of negatively (rather than positively) worded behaviors in this process because we believed it would be easiest to capture program directors' level of concern when the behaviors reflected the actual negative behaviors they witness.

Method

Competency model development and validation

To identify the unprofessional behaviors for the checklists, we first developed a professionalism competency model that aligned with the professionalism milestones of the Next Accreditation System (Nasca et al. 2012). The purpose of the competency model was to lay the foundation for developing professionalism assessments that support professional identity formation, applicant screening, resident development, and remediation. The first step was to identify the full set of professionalism behaviors representing important professionalism principles. We then sorted these behaviors into broader professionalism domains. We implemented a "bottom-up" approach in developing our model (Hawkins et al. 2009). To do this, the first author began by conducting a content analysis of the professionalism milestones within the ACGME's Next Accreditation System (Nasca et al. 2012). We used these milestones as a starting place because since 2013 the ACGME has required all programs to report biannually on their trainees' standing on these and other milestones to receive accreditation. Although each program's specialty created its own professionalism milestones, many important professionalism behaviors are repeated across specialties. Thus, by examining the milestones, we were able to provisionally identify a large set of important professionalism behaviors, and associated professionalism competencies, as programs

themselves viewed them. Subsequently, the first author supplemented this content analysis with a review of the medical education literature on professionalism models, tools, and attempts to synthesize existing models. We were greatly assisted in this task by systematic reviews of these models and tools by Wilkinson et al. (2009) and Van De Camp et al. (2004). We also examined large-scale efforts to define professionalism led by the Royal College of Physicians and the Charter of Medical Professionalism stemming from the Medical Professionalism Project. Finally, we examined comprehensive lists of professionalism behaviors generated by the National Board of Medical Examiners/Association of American Medical Colleges Workshop as well as existing professionalism checklists (Arnold et al. 1998; Papadakis et al. 2001; Fontaine & Wilkinson 2003; Norcini et al. 2003; Cruess et al. 2006; Reed et al. 2008).

Using the behaviors gleaned from this process, the first author rationally sorted them into seven "institution-level" professionalism dimensions and associated behaviors. In creating the model, the focus was on identifying the observable behaviors that could be reliably assessed by members of a trainee's interprofessional team. The dimensions and facets (in brackets) included: (1) Conscientiousness (dependability, planning/organizing, thoroughness), (2) Aspiring to Excellence (work commitment, motivation to learn), (3) Integrity (trustworthiness, discretion, personal conduct, organizational citizenship), (4) Accountability (personal responsibility, self-awareness), (5) Teamwork (cooperation, respectful interaction, team building), (6) Patient-Centeredness (compassion, respect for diversity, humanism), and (7) Stress Tolerance (situational stress tolerance, interpersonal stress tolerance). The first author used the behaviors contained within the dimensions to define each dimension and associated facet.

To refine the model, we held a series of three workshops with five Ph.D. educators and 17 faculty members (11 program directors) from 10 residency and fellowship programs, including representatives from our Internal Medicine (92 trainees), Diagnostic Radiology (41 trainees), Neurosurgery (14 trainees), Orthopaedic Surgery (38 trainees), Family Medicine (168 trainees in eight programs), Surgery (39 trainees), Pediatrics (69 trainees), Neonatal-Perinatal (five trainees), Psychiatry (25 trainees), and Neurology (24 trainees) programs. During the workshops, participants were presented with the professionalism competency model and definitions and were asked to review the behaviors within the model using two criteria: (a) does each behavior clearly measure the facet with which it was associated and (b) is each behavior distinct from the other behaviors. Based on consideration of these criteria, we dropped 36 behaviors from our model, reassigned a small number of behaviors to different facets within the model, and made minor changes to the definitions of dimensions and facets. We also added eight new behaviors that program directors had witnessed in their own programs. The final model contained 74 professionalism-related behaviors assigned to the seven dimensions and 19 facets.

At this stage, the behaviors were phrased positively. For instance, within the Integrity dimension, the model included such behaviors as "demonstrates honesty in interactions with patients, families, and other health care professionals" and "protects confidentiality of sensitive patient and co-worker information." To validate the model, 23 subject matter experts (i.e. program directors, faculty, residents, and medical educators) rated: (a) the extent to which each behavior measured at least one element of the intended facet (construct validation) and (b) the importance of each behavior for performance as an independent physician. Both ratings were made using an identical five-point Likert-type scale (1 = to a very small extent, 5 = to a very great extent). Prior to distributing the validation exercise, we had decided to delete from our model behaviors that achieved a mean rating below 3.5 on either scale, as this would indicate the behavior was either failing to measure the intended dimension or was judged not to be at least somewhat important for independent practice. However, none of the behaviors met this threshold. Accordingly, all behaviors were included in the final model. The final professionalism dimensions and facets are displayed in Table S1 (available online as Supplemental Material), and the final validation results are in Table S2 (available online as Supplemental Material).

Generating a list of unprofessional behaviors

In order to generate a list of unprofessional behaviors that aligned with the professionalism model, we asked workshop participants to translate each *positive* behavior in the model into an equivalent *negative* behavior. For instance, the positive behavior from the Conscientiousness dimension "demonstrates regular and punctual attendance" became "arrives late for conferences, rounds, or other work-related meetings." This translation exercise resulted in the generation of 70 non-overlapping negative professionalism-related behaviors.

Determining the relative egregiousness of unprofessional behaviors

To determine the relative egregiousness of these 70 unprofessional behaviors, we asked workshop participants to indicate how concerned they would be if a trainee engaged in each behavior 1, 2, 3, 4, or "5 or more" times during a six-month period. We chose six months as our time period since the ACGME Common Program Requirements require semi-annual reviews with trainees (Nasca et al. 2012). Thus, by setting a six-month time period to the scale, we hoped to make the checklists as useful as possible for these reviews. The concern scale ranged from 1 to 4 (1 = not concerned, 2 = a little concerned, 3 = somewhat concerned, and 4 = very concerned, a meeting with the program director is required). Thus, for each behavior, raters made five ratings, showing their level of concern for each of the five possible frequencies of occurrence. Using these ratings, we computed the mean concern rating for each behavior, for each possible frequency of occurrence. This procedure provided an overall egregiousness "score" for each behavior/frequency combination (Table S3, available online as Supplemental Material). To capture the "overall" level of egregiousness of each behavior across all frequencies of occurrence, we also created a "concern index" (CI) for each behavior by summing the scores for each behavior for each frequency of occurrence. Thus, the maximum possible score on the CI for a behavior was 20 (i.e. a maximum mean rating of 4 for each behavior for each of the five possible

frequencies of occurrence), while the minimum was 5 (i.e. a minimum mean rating of 1 for each behavior for each of the five possible frequencies of occurrence).

Creating the final professionalism checklists

We used the results from this rating exercise to choose the final behaviors for the two professionalism checklists. Those behaviors that achieved a mean concern rating of at least 3 (somewhat concerning) if they occurred at least once were placed on the "brief" professionalism checklist of serious unprofessional behaviors (see *dark* grey behaviors in Table S3). All other behaviors are part of the extended checklist. The extended checklist identifies the level of frequency at which less serious behaviors become at least somewhat concerning (i.e. reaches a level of 3 on the concern scale) when a pattern of misconduct emerges (see behavior × frequency combinations in *light grey* in Table S3).

Results

Table S2 displays the final professionalism dimensions, associated facets, and results of the validation exercise for the behaviors within each facet. Across all seven dimensions, the mean construct validation rating for the behaviors was 4.39 and the mean importance rating was 4.44. To estimate the reliability of ratings, we computed the Case II intraclass correlation formula for multiple raters (Shrout & Fleiss 1979). The results were 0.75 for the construct validation ratings and 0.74 for the importance ratings, demonstrating good reliability for both ratings exercises.

Table S3 contains the concern indices for all 70 unprofessional behaviors, for each of the five different possible frequencies of occurrence categories (1, 2, 3, 4, 5+) over a six-month period. Table S3 reveals that 14 behaviors were judged to be at least somewhat concerning if they occur even once. These behaviors included: (a) displaying obvious signs of substance abuse, (b) demonstrating abusive behavior toward co-workers, (c) using one's status as a doctor for personal gain, (d) failing to interact truthfully with patients, families, or other healthcare professionals, (e) discriminating against co-workers or patients, (f) failing to uphold ethical expectations of research and scholarly activity, (g) showing disrespect toward patients, (h) blaming co-workers for errors that were not their fault, (i) misrepresenting facts, or failing to present facts impartially, (j) knowingly disregarding site rules and procedures, (k) including erroneous information in the electronic medical record, (l) criticizing co-workers in public in a non-respectful manner, (m) putting one's individual needs above the needs of patients, and (n) stereotyping about groups of patients.

Table S3 reveals substantial variability across behaviors in the level of concern expressed when a behavior is witnessed once, but relative invariability in the level of concern when behaviors are witnessed multiple times. For instance, the range in concern ratings for the behavior rated the most and least concerning if it occurs once (i.e. displaying obvious signs of substance abuse, with a concern rating of 3.95, and arriving late for conferences, with a concern rating of 1.05) is 2.9. In contrast, the range in concern ratings for those behaviors rated the most and least concerning if they occur five or more times (the same behaviors, with ratings of 4.0 and 3.32) is 0.68. Thus, while only a subset of behaviors are considered to be concerning if they occur once, most behaviors are considered to be concerning if they occur frequently.

Table 1. Rank ordering of professionalism dimensions by overall concern rating.

Professionalism dimension	Average concern index for behaviors in dimension
1. Integrity	18.47
2. Accountability	17.32
3. Teamwork	17.04
4. Patient-centered care	17.00
5. Stress tolerance	16.83
6. Aspiring to excellence	16.11
7. Conscientiousness	14.77

The CI in Table S3 provides a rough proxy of the *overall* egregiousness of each unprofessional behavior. Table S3 indicates that, for the most part, the same behaviors that were judged to be concerning if they occur only once have the highest overall concern indices. Behaviors with the lowest concern indices included: (1) failing to follow up with patients to determine outcomes or satisfaction (CI =14.66), (2) not encouraging patients to ask questions (CI =14.60), (3) inappropriately dominating team interactions (CI =14.50), (4) approaching work tasks unsystematically (CI =12.23), and (5) arriving late for conferences, rounds, or other work-related meetings (CI =11.18).

As each behavior is linked to a professionalism dimension, it is possible to obtain an estimate of the perceived importance of breaches by dimension. Table 1 provides this CI for professionalism dimensions across all behaviors in each dimension. The professionalism dimension with the greatest overall average CI is Integrity (average CI =18.47) and the professionalism dimension with the lowest overall CI is Conscientiousness (average CI =14.77). Table 1 suggests that while all professionalism breaches are important, breaches of the Integrity domain are viewed to be most concerning.

Discussion

The purpose of this study was to employ a policy capturing approach to develop two empirically validated professionalism checklists for identifying professionalism issues at an early stage in a physician's career. We validated a professionalism competency model and created both a "brief" and "extended" professionalism checklist based on ratings by education experts. The "brief" checklist identifies those behaviors that are concerning, and require remediation, *if they occur at all*. The second "extended" checklist identifies *patterns* of misconduct in less serious behaviors that are concerning and require remediation.

This study makes several contributions to the literature. First, it demonstrates that unprofessional behaviors vary in their level of egregiousness. Although it seems obvious that some unprofessional behaviors are worse than others, our ranking provides support to this subjective impression. Second, our work creates a framework for comparing unprofessional actions that occur in isolation to those that represent a pattern of behavior. There is a relationship between the nature of an unprofessional behavior and how often it occurs that influences our level of concern. Mildly

unprofessional behaviors that occur regularly may lead to remedial action that is similar to that triggered by a grossly unprofessional behavior that occurs once. This is important, given that assessment of these behaviors is more subjective for some than others. For example, it may be necessary for a pattern to emerge before taking action on Conscientiousness behaviors that are mitigated by external factors such as workload and stress, whereas behaviors that are patently dangerous to patients require immediate intervention. Third, this study generated two empirically validated professionalism checklists that can be used for identifying and documenting early signs of trouble in physicians. Thus, this study responds to calls to create more rigorously-developed professionalism tools for early identification and remediation of problematic behaviors (Arnold 2002; Hodges et al. 2011; Cook et al. 2013).

There are limitations to this work. First, the development of the checklist is necessarily subjective. It would benefit from future efforts to broaden input across clinical disciplines, institutions, and most importantly, across cultures. Recent research suggests that there may be important regional differences regarding which behaviors are considered unprofessional (Chandratilake et al. 2012; Jha et al. 2015). Thus, the rankings of these behaviors may vary depending on the cultural context and unique properties of healthcare systems in different countries. Second, it would be worthwhile to further assess some of the psychometric properties of this tool (Veloski et al. 2005). In particular, once these checklists have been implemented, it would be useful to employ factor analytic techniques to determine which subset of behaviors is most predictive of future professionalism issues (Cruess et al. 2006). Finally, these tools are simply intended to improve detection of these behaviors. This is imperative since without detection, remediation cannot be initiated. However, to be useful it is imperative to assess outcomes following identification of the behaviors in the checklist. Important follow-up questions include: (a) which behaviors are remediable? (b) what type of remediation works best for different behaviors? (c) do the behaviors in these checklists manifest in the careers of physicians in whom they are identified during training?, and (d) are there important patient-directed behaviors or outcomes that are predicted by this checklist? In particular, future predictive studies should expand the set of criterion measures beyond the dichotomous criterion of disciplinary action investigated by Papadakis et al. (2005) to other clinically relevant outcomes such as patient satisfaction, patient outcomes, referrals, good record keeping, and appropriate prescribing (McLachlan 2010). McLachlan (2010) has begun this effort by demonstrating that conscientiousness-related behaviors are predictive of independent staff and student estimates of professionalism.

One important issue is how these professionalism checklists should be implemented in practice. We recommend that residency programs follow the lead of the General Medical Council in the United Kingdom, which requires medical schools to establish "fitness to practice" committees to address professionalism issues (Hilton & Slotnick 2005). Two key objectives of this committee will be to decide *when* checklist data should trigger remediation, and *how* to collect it efficiently. As mentioned above, it is critical for institutions to reflect on their own cultural identity and ensure that this tool – or any similar assessment – represents the actual behaviors that are objectionable in their organization.

Regarding the triggers for remediation, we have decided to intervene whenever any of the "brief checklist" behaviors occurs since even one occurrence of these behaviors is by definition at least "somewhat concerning". Similarly, remediation is indicated when a *pattern* of misconduct in any behavior in the extended checklist reaches the same level of concern. Functional implementation of these checklists is planned as follows: The brief form is intended for use at the end of each rotation. Regular use of the checklist will enable the evaluating faculty members to become familiar with the type of behaviors viewed as most concerning. The longer checklist should be used biannually to determine if there are patterns of less egregious behaviors that emerge over time. To ease the administrative burden on clinical faculty, different members of the trainee's interprofessional team could complete separate sections of the longer list. As an example, nurses, physician assistants or other members of a trainee's interprofessional team could rate the "teamwork", "integrity", and "stress tolerance" behaviors biannually, and clinical faculty could rate the others. Alternatively, all or parts of the extended checklist could be incorporated into peer evaluations. Lastly, it is possible to make completion of the extended checklist voluntary. In this approach, members of the trainee's interprofessional team would have the option to report observed behaviors to the committee by using the more comprehensive tool.

Regardless of the approach taken, best practices in assessment suggest that feedback for both checklists should come from multiple sources, and include as many members of a learner's interprofessional team (e.g. faculty, peers, supervising consultants, patients, allied health professionals, physicians, and clerical staff) as possible (Goldie 2013). Such multisource feedback systems are used frequently in industry, but less commonly in medical settings (Hawkins et al. 2009). By involving members of a learner's interprofessional team in assessment, a fuller and more reliable picture of that learner's actual behavior in different medical contexts emerges than is possible from information emerging from one rating source alone (Norcini 2003; Archer et al. 2005; Dannefer et al. 2005). In addition, including insights from multiple raters allows for a comparison of how different rater groups view the learner, which can be helpful when providing feedback to learners. The provision of feedback to learners about behaviors observed is a key part of any remediation plan, and expertise in providing that feedback should not be assumed. Feedback providers should be trained to provide specific, concrete examples that focus on the behaviors observed and their effect on individual or team performance (Kluger & DeNisi 1996). The trainee's perspective should be invited, and the trainee and feedback provider should both provide input into a time-limited action plan to improve any noted issues. Due to the reluctance of faculty to give negative evaluations of trainees (Albanese 2000; Boon & Turner 2004), it will be essential to train faculty about why documenting unprofessional behavior is important. Documentation of professionalism issues is crucial if a pattern of misbehavior requires probation, or in rare cases when termination is required. Finally, it will be important for institutional leadership to set the correct tone for the use of these checklists. For maximum

effect, they should be implemented in the spirit of continuous institutional improvement (Arnold 2002).

Although the primary goal of our professionalism checklists is early identification of unprofessional conduct, they could also be used for providing summative feedback in the areas of professional identity formation and the development of professionalism knowledge (Goldie 2013). As Hilton and Slotnick (2005) have observed, professionalism is not a static trait, but an acquired state developed over time. By introducing learners to these checklists during medical school or residency orientation, they may help learners to create a mental model of professionalism early in their career that can be reinforced over time with multisource feedback, role modeling, and active self-reflection. In addition, when appropriately combined with other professionalism tools, such as global ratings, peer ratings, patient surveys, and self-administered reflection scales, both checklists could play an important role in helping to triangulate understanding of a trainee's overall professionalism knowledge, skill, and attitudes at different points in time (Hawkins et al. 2009; Goldie 2013). Thus, these checklists could be one relevant source of input into professionalism milestones ratings in residency and fellowship programs (Nasca et al. 2012).

Professionalism is a multidimensional construct, which can be assessed for many purposes, using multiple methods (Hawkins et al. 2009; Goldie 2013). One of the most important reasons for assessing professionalism is early detection and remediation of unprofessional behavior. When combined with other appropriate professionalism assessments and remediation strategies, the checklists developed in this study can form an important part of an early warning system of physician misconduct. We encourage researchers to explore creative ways to combine these checklists with other professionalism measures to create new and effective approaches for detecting and managing professionalism issues in learners.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the article.

Glossary

Policy capturing: A method employed by researchers to assess how decision makers use available information when making evaluative judgments. The purpose of this methodology is to capture individual judges' decision-making policies, that is, how they weight, combine, or integrate information. It involves asking decision makers to judge a series of scenarios describing various levels of one or more explanatory factors, and then using statistical methods to determine the emphasis decision makers give to each factor in their decision-making process. The results indicate the relative importance of the various factors for the decision makers.

Notes on contributors

Dr. Michael J. Cullen, PhD is the Director of Evaluation for Graduate Medical Education at the University of Minnesota Medical School, Minneapolis, United States.

Dr. Mojca R. Konia, MD, PhD, is an Associate Professor and Vice Chair of Education and Residency Program Director in the Department of Anesthesiology at the University of Minnesota Medical School, Minneapolis, United States.

Dr. Emily C. Borman-Shoap, MD, is the Residency Program Director and Assistant Professor in the Department of Pediatrics at the University of Minnesota Medical School, Minneapolis, United States.

Dr. Jonathan P. Braman, MD, is the former Residency Program Director and is an Associate Professor in the Department of Orthopaedic Surgery at the University of Minnesota Medical School, Minneapolis, United States.

Dr. Ezgi Tiryaki, MD, is the former Residency Program Director and an Associate Professor of Neurology at the University of Minnesota Medical School and the Associate Chief of Staff for Education for the Minneapolis VA Health Care System, Minneapolis, United States.

Brittany Marcus-Blank, BA, is a doctoral student in the Industrial-Organizational Psychology Department at the University of Minnesota – Twin Cities, Minneapolis, United States.

Dr. John S. Andrews, MD, is the Associate Dean for Graduate Medical Education and the Designated Institutional Official (DIO) at the University of Minnesota Medical School, Minneapolis, United States.

References

AAMC, NBME. 2003. Embedding professionalism in medical education: assessment as a tool for implementation. Rep Int Conf Cosponsored Assoc Am Med Coll AAMC Natl Board Exam NBME. Washington (DC): National Board of Medical Examiners. p. 15–17.

ABIM Foundation, American Board of Internal Medicine, ACP-ASIM Foundation, American College of Physicians-American Society of Internal Medicine, European Federation of Internal Medicine. 2002. Medical Professionalism in the New Millennium: A Physician Charter. Ann Intern Med. 136:243–246.

Albanese MA. 2000. Challenges in using rater judgements in medical education. J Eval Clin Pract. 6:305–319.

Archer JC, Norcini JJ, Davies HA. 2005. Use of SPRAT for peer review of paediatricians in training. Br Med J. 330:1251–1253.

Arnold EL, Blank LL, Race KE, Cipparrone N. 1998. Can professionalism be measured? The development of a scale for use in the medical environment. Acad Med. 73:1119–1121.

Arnold L, Stern DT. 2006. What is medical professionalism? In: Stern DT, editor. Meas Med Prof [Internet]. Oxford: Oxford; [cited 2016 Apr 21]. p. 15–37. Available from: https://books.google.com/books?hl=en&lr=&id=GUA23tGPDaYC&oi=fnd&pg=PA15&dq=What+is+Medical+Professionalism%3F&ots=eZOQDT_cRN&sig=HHa2u7KXkEFET34FygV5SDW_g7c

Arnold L. 2002. Assessing professional behavior: yesterday, today, and tomorrow. Acad Med. 77:502–515.

Beach MC, Sugarman J, Johnson RL, Arbelaez JJ, Duggan PS, Cooper LA. 2005. Do patients treated with dignity report higher satisfaction, adherence, and receipt of preventive care? Ann Fam Med. 3:331–338.

Boon K, Turner J. 2004. Ethical and professional conduct of medical students: review of current assessment measures and controversies. J Med Ethics. 30:221–226.

Chandratilake M, McAleer S, Gibson J. 2012. Cultural similarities and differences in medical professionalism: a multi-region study. Med Educ. 46:257–266.

Cook AF, Sobotka SA, Ross LF. 2013. Teaching and assessment of ethics and professionalism: a survey of pediatric program directors. Acad Pediatr. 13:570–576.

Cruess R, McIlroy JH, Cruess S, Ginsburg S, Steinert Y. 2006. The professionalism mini-evaluation exercise: a preliminary investigation. Acad Med. 81:S74–S78.

Dannefer EF, Henson LC, Bierer SB, Grady-Weliky TA, Meldrum S, Nofziger AC, Barclay C, Epstein RM. 2005. Peer assessment of professional competence. Med Educ. 39:713–722.

Dougherty TW, Ebert RJ, Callender JC. 1986. Policy capturing in the employment interview. J Appl Psychol. 71:9.

Durning SJ, Cation LJ, Markert RJ, Pangaro LN. 2002. Assessing the reliability and validity of the mini-clinical evaluation exercise for internal medicine residency training. Acad Med. 77:900–904.

Edwards PJ, Roberts I, Clarke MJ, DiGuiseppi C, Wentz R, Kwan I, Cooper R, Felix LM, Pratap S. 2009. Methods to increase response to postal and electronic questionnaires. In: The Cochrane Collaboration, editor. Cochrane Database Syst Rev [Internet]. Chichester (UK): John Wiley & Sons, Ltd; [cited 2016 Apr 21]. Available from: http://doi.wiley.com/10.1002/14651858.MR000008.pub4

Fontaine S, Wilkinson TJ. 2003. Monitoring medical students' professional attributes: development of an instrument and process. Adv Health Sci Educ Theory Pract. 8:127–137.

Frank J. (2005). Re: the CanMEDS 2005 physician competency framework. Better standards. Better physicians. Better care. Ottawa: The Royal College of Physicians and Surgeons of Canada.

Goldie J. 2013. Assessment of professionalism: a consolidation of current thinking. Med Teach. 35:e952–e956.

Graves LM, Karren RJ. 1992. Interviewer decision processes and effectiveness: an experimental policy-capturing investigation. Pers Psychol. 45:313–340.

Grumbach K, Bodenheimer T. 2004. Can health care teams improve primary care practice? JAMA. 291:1246–1251.

Hawkins RE, Katsufrakis PJ, Holtman MC, Clauser BE. 2009. Assessment of medical professionalism: who, what, when, where, how, and why? Med Teach. 31:348–361.

Hilton SR, Slotnick HB. 2005. Proto-professionalism: how professionalisation occurs across the continuum of medical education. Med Educ. 39:58–65.

Hodges BD, Ginsburg S, Cruess R, Cruess S, Delport R, Hafferty F, Ho M-J, Holmboe E, Holtman M, Ohbu S, et al. 2011. Assessment of professionalism: recommendations from the Ottawa 2010 Conference. Med Teach. 33:354–363.

Jha V, Mclean M, Gibbs TJ, Sandars J. 2015. Medical professionalism across cultures: a challenge for medicine and medical education. Med Teach. 37:74–80.

Klaas BS, Wheeler HN. 1990. Managerial decision making about employee discipline: a policy-capturing approach. Pers Psychol. 43:117–134.

Kluger AN, DeNisi A. 1996. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychol Bull. 119:254–284.

McLachlan J. 2010. Measuring conscientiousness and professionalism in undergraduate medical students. Clin Teach. 7:37–40.

Nasca TJ, Philibert I, Brigham T, Flynn TC. 2012. The next GME accreditation system-rationale and benefits. N Engl J Med. 366:1051–1056.

Norcini JJ, Blank LL, Arnold GK, Kimball HR. 1997. Examiner differences in the mini-CEX. In: Adv Med Educ [Internet]. the Netherlands: Kluwer Academic Publishers; [cited 2016 Apr 21]. p. 170–172. Available from: http://link.springer.com/chapter/10.1007/978-94-011-4886-3_49

Norcini JJ, Blank LL, Duffy FD, Fortna GS. 2003. The mini-CEX: a method for assessing clinical skills. Ann Intern Med. 138:476–481.

Norcini JJ. 2003. Peer assessment of competence. Med Educ. 37:539–543.

Papadakis MA, Loeser H, Healy K. 2001. Early detection and evaluation of professionalism deficiencies in medical students: one school's approach. Acad Med. 76:1100–1106.

Papadakis MA, Teherani A, Banach MA, Knettler TR, Rattner SL, Stern DT, Veloski JJ, Hodgson CS. 2005. Disciplinary action by medical boards and prior behavior in medical school. N Engl J Med. 353:2673–2682.

Reed DA, West CP, Mueller PS, Ficalora RD, Engstler GJ, Beckman TJ. 2008. Behaviors of highly professional resident physicians. JAMA. 300:1326–1333.

Royal College of Physicians. 2005. Doctors in society: medical professionalism in a changing word. Report of a working party of the royal college of physicians of London. Suffolk, Great Britain: Lavenham Press Ltd.

Shayne P, Gallahue F, Rinnert S, Anderson CL, Hern G, Katz E, Group CSS. 2006. Reliability of a core competency checklist assessment in the emergency department: the Standardised Direct Observation Assessment Tool. Acad Emerg Med. 13:727–732.

Sherer PD, Schwab DP, Heneman HG. 1987. Managerial salary-raise decisions: a policy capturing approach. Pers Psychol. 40:27–38.

Shrout PE, Fleiss JL. 1979. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 86:420–428.

Stern DT, Frohna AZ, Gruppen LD. 2005. The prediction of professional behaviour. Med Educ. 39:75–82.

Van De Camp K, Vernooij-Dassen MJ, Grol RP, Bottema BJ. 2004. How to conceptualize professionalism: a qualitative study. Med Teach. 26:696–702.

Van Mook WNKA, van Luijk SJ, O'Sullivan H, Wass V, Zwaveling JH, Schuwirth L, van der Vleuten CPM. 2009. The concepts of professionalism and professional behaviour: conflicts in both definition and learning outcome. Eur J Intern Med. 20:85–89.

Veloski JJ, Fields SK, Boex JR, Blank LL. 2005. Measuring professionalism: a review of studies with instruments reported in the literature between 1982 and 2002. Acad Med. 80:366–370.

Wilkinson TJ, Wade WB, Knock LD. 2009. A blueprint to assess professionalism: results of a systematic review. Acad Med. 84:551–558.

Zedeck S. 1977. An information processing model and approach to the study of motivation. Organ Behav Hum Perform. 18:47–77.