

Evaluating Nonclinical Performance of the Academic Pathologist: A Comprehensive, Scalable, and Flexible System for Leadership Use

Academic Pathology Volume 5: 1-7 © The Author(s) 2018 sagepub.com/journalsPermissions.nav DOI: 10.1177/2374289518754889 journals.sagepub.com/home/apc

(\$)SAGE

Austin Blackburn Wiles, MD¹, Michael O. Idowu, MD, MPH¹, Charles V. Clevenger, MD, PhD¹, and Celeste N. Powers, MD, PhD¹

Abstract

Academic pathologists perform clinical duties, as well as valuable nonclinical activities. Nonclinical activities may consist of research, teaching, and administrative management among many other important tasks. While clinical duties have many clear metrics to measure productivity, like the relative value units of Medicare reimbursement, nonclinical performance is often difficult to measure. Despite the difficulty of evaluating nonclinical activities, nonclinical productivity is used to determine promotion, funding, and inform professional evaluations of performance. In order to better evaluate the important nonclinical performance of academic pathologists, we present an evaluation system for leadership use. This system uses a Microsoft Excel workbook to provide academic pathologist respondents and reviewing leadership a transparent, easy-to-complete system that is both flexible and scalable. This system provides real-time feedback to academic pathologist respondents and a clear executive summary that allows for focused guidance of the respondent. This system may be adapted to fit practices of varying size, measure performance differently based on years of experience, and can work with many different institutional values.

Keywords

academic pathologist, guidance, nonclinical activity, practice leadership, promotion

Received August 21, 2017. Received revised October 26, 2017. Accepted for publication November 08, 2017.

Introduction

Academic pathologists not only perform clinical work, most also have substantial activities beyond direct patient care. For this report, nonclinical work is any activity that does not directly impact individual patients and may consist of scholarly activities, like research, teaching, administrative, or professional academic activities. While numerous measures of clinical performance exist, such as the relative value units (RVUs) of Medicare reimbursement, there are few systematic measures of nonclinical performance.¹ Although nonclinical activities usually do not contribute directly to patient care or clinical income, they are integral to the academic pathologist's position and advancement and, as such, they have real value. Unfortunately, measurement of this value is difficult.² Only a few

studies exist which address methods of evaluation of nonclinical work by the academic pathologist. Each system uses different assumptions and methods to capture efforts in a quantifiable form. These evaluation systems serve 2 purposes: (1) to function as yardsticks of performance for members of an institution and (2) to make explicit the goals of an organization

¹ Department of Pathology, VCU School of Medicine, Richmond, VA, USA

Corresponding Author:

Austin Blackburn Wiles, Department of Pathology, VCU School of Medicine, PO Box 980662, 1200 E Marshall Street, Gateway Building, 6-205, Richmond, VA 23298, USA. Email: austin.wiles@vcuhealth.org



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage). by translating abstract ideals into behaviors expected of individual members.^{3,4}

Evaluating clinical performance alone is insufficient for adequately appraising the performance of an academic pathologist.^{5,6} Decisions related to advancement in leadership roles, promotion, and tenure place significant emphasis on nonclinical activity. Despite its tremendous importance, there are few systems that adequately measure nonclinical performance. Often systems used by institutions have selected generic metrics for a "one size fits all" approach. An evaluation system that inadequately accounts for nonclinical activities may actually discourage scholarly activity by implicitly discouraging time spent on nonclinical activity.7 Positive outcomes, including increased physician satisfaction and increased utility, have been achieved with various methods related to continuing medical education.⁸ In addition, evaluation systems that have the capacity to recognize and reward the excellence of many diverse institutional members are stronger positive motivators than systems that simply reinforce the "elite performance" of a favored few.9 Therefore, a well-designed and balanced evaluation system can foster productivity of individuals as well as promote the aims of the institution.¹⁰⁻¹⁶

The method of evaluation itself should be understood by both respondent and supervisor. Transparency and clarity in evaluation systems will increase physician performance as subsequent evaluation cycles progress.¹⁷⁻¹⁹ Contrastingly, simply linking pay to academic performance does not have as clear a link to producing increased performance.^{20,21}

The format of measurement is important to the success of an evaluation system. Some systems measure units of time spent on various activities, reminiscent of RVUs, to monitor performance.²² These systems fail to establish substantive links between the reported information and real attending physician nonclinical productivity. Spent time does not equal valuable time. Time spent on research does not necessarily equate to valuable published literature. Such evaluation methods require meticulous, time-consuming, record keeping on the part of the academic pathologist. Detailed time-keeping systems may also be inaccurate due to conscious or unconscious misreporting.²³ In order to generate useful conclusions, a good evaluation system should be easy to complete from quantitative data readily at hand.

In 2009, our institution, Virginia Commonwealth University Health System, implemented a nonclinical performance review system, Faculty Activity Reporting and Evaluation System (FARES). This system requires multiple hours of preparation by the respondent as well as the supervising reviewer. It is used by all institutional physicians, regardless of specialization and therefore has very generic inputs. The FARES uses time in the form of hours per week averaged over a year to track activities. Due to the irregularity of academic pathologist schedules, alternating between clinical service, administration, educational activities, and research, the reported values were in fractions of hours per week, which was nonintuitive and required either guestimates or extraordinarily meticulous record keeping. Additionally, it offered no immediate feedback to the respondent who remained completely unaware of what level of performance review they might receive. While the form collected a large amount of information and had quantitative outputs, it was difficult to tell which pathologists were "good performers" according to the institutional and departmental missions and which pathologists were poor performers who needed growth in 1 or more specific areas.

Although our department used various additional metrics to evaluate individual achievement in academics, there was no formal, metric system. We developed a new format that addressed individual as well as group performance equitably with quantifiable metrics. This report describes this new method of evaluating nonclinical performance of academic pathologists that is easy to use and easily adaptable.

Materials and Methods

Presented herein is a simplified version of our institution's evaluation. Categories and activities have been adjusted for clarity of presentation.

Microsoft Excel (from Microsoft Office 2013 for Windows; for Mac Version 15.35) was used to create an academic pathologist reporting form and a grading rubric on separate worksheets within the same workbook file. The grading rubric, containing the grading benchmarks, and activities in the form were generated by senior administrative faculty in conjunction with a review of the literature and took into account contractual job descriptions for the target faculty, the institutional mission, and criteria for promotion and advancement. The form was presented during a faculty meeting and was met with immediate buy-in due to its intuitive use by respondents, ease of result interpretation, and respect for individual career paths of the faculty being evaluated. This single file, containing both worksheets, was e-mailed to respondents by their supervisor. The rubric worksheet was protected from editing while the respondent worksheet allowed editing. Due to the inclusion of the grading rubric, the respondent was immediately aware of their performance as the form updated instantaneously. Upon completion, individual respondents forwarded this self-evaluation to their supervisor and each individual respondent worksheet was collated within a master Excel file and referred to a single common grading rubric worksheet allowing for simple indexing of respondents over time. The respondent form was paginated for clear printability and subsequent physical storage.

Each page of the form had separate activity categories for evaluation that were specifically designed for our institution. These were "Quality," "Education," "Scholarship," "Administration," and "Compensation Plan Thresholds." Other institutions can adapt or replace these categories without compromise of the tool presented herein. Within each activity category, associated subcategories were created. For example, under the category "Education," subcategories "Resident Education" and "Lectures" were established. These subcategories were assigned constitutive, discrete observable actions with associated point values. The respondents selected and entered the appropriate point value in the adjacent response

	Quality		Education		Scholarship		Administrat	tion	Compensation Plan
Evaluation	QualLESS	QualMORE	EducationLESS	EducationMORE	ScholarLESS	ScholarMORE	AdminLESS	AdminMORE	Comp
Needs Improvement	0	0	0	0	0	0	0	0	0
Needs Improvement	3	3	3	3	0	1	0	1	2
Satisfactory	4	4	4	4	1	3	1	2	3
Very Good	5	6	5	6	2	5	2	4	4
Excellent	7	7	7	7	3	6	3	5	5
	LESS Fewer than 5 years practice MORE Equal to or More than 5 years in practice]				

Figure 1. The grading rubric. This is an independent spreadsheet within the workbook containing the evaluation form. The physician's responses will be compared against these values to determine performance. Four of the areas, "Quality," "Education," "Scholarship," and "Administration," are tiered for physicians practicing for greater or fewer than 5 years via the subcolumns within each category (eg, QualLESS for physicians practicing less than 5 years and QualMORE for physicians practicing more than 5 years). The "Compensation Plan" has the same thresholds of performance regardless of years in practice. More gradations for years in practice, or other practice styles, and other thresholds can be added easily. This spreadsheet within the workbook can also be locked to prevent editing by the respondent physician. The evaluation category of "needs improvement" is duplicated to define both the upper and lower boundaries of scores for this category for the technical necessity for the Excel INDEX and MATCH functions. This grading rubric is used in scoring the fictional respondent "Jane James" in Figures 2 and 3.

field commensurate with their level of activity in that subcategory. The response form summed these points for each activity category and displayed their level of performance in a summary field at the end of the form.

The levels of performance for each category were "needs improvement," "satisfactory," "very good," and "excellent." The point thresholds for these performance levels were set based on common clusters of behaviors, known to be attainable from previous evaluations systems, and deemed important by clinical and academic administration guidelines for our health system. The specific points assigned should be scaled for specific institutional concerns which is easily accomplishable with this tool. Two different grading schemes were used: 1 for pathologists who had practiced 5 years or fewer and 1 for pathologists in practice more than 5 years.

The primary Excel functions used in this form were INDEX, MATCH, and ERF. The nested INDEX and MATCH functions allowed for the completed respondent form to reference the grading rubric for the appropriate activity category accounting for the respondents' years of practice and display the level of performance automatically. The ERF function is the Excel Gaussian error function which served 2 purposes in the visualization of the respondent's overall performance. First, it mathematically normalized the scores between the different activity categories allowing for their symmetric and comparative display in a radar graph. Second, the ERF function accounted for the sigmoidal shape of the association between activity category raw scores and performance evaluation categories.

Results

This evaluation tool records and scores responding attending pathologist activities rather than time increments spent on activities. Completing this review takes approximately 10 minutes and the format and output are intuitive for both the reviewer and the respondent. A well-rounded respondent will have a complete symmetric pentagon, whereas an asymmetric or incomplete pentagon indicates areas for improvement or more focused attention.

Importantly, the scoring of activities occurs on 2 levels. First, points are awarded for activities. The grouping of activities, the activities themselves, and the points awarded to various levels of participation in the activities may be adjusted to address departmental and/or institutional values. Second, the overall evaluation of a category (ie, research) receives a nonquantitative score (ie, very good) for an accumulation of points. Both of these activities may be adjusted by the institution or department for an individual pathologist (eg, a part-time faculty member) or an entire group of physicians (eg, division of anatomic Pathology or division of clinical pathology) to produce a sensible output.

The accrual of points in a category is linear, but the evaluation of accrued points within a category is sigmoidal. A sigmoidal function has a roughly linear central component buttressed by 2 planes. An attending pathologist can achieve excellence in many ways. Two pathologists might achieve excellence but using different methods. The topology of the ERF function and the manner of point accrual accepts the axiom that excellence is a multidimensional plateau that pathologists can explore with their careers. The ERF function was deployed to reflect this and normalize the scale between different categories of performance balancing the necessity of actionable conclusions for an evaluation and encourage diversity in academic pursuits.

Images from the (Figures 1–4) schematic grading rubric, form, and summary are included. The grading rubric shows the evaluation for the accrual of points for attending physicians in 2 tiers: less than 5 years of practice and more than 5 years of practice (Figure 1). Points were accrued via filling out the form which clearly depicts the number of points associated with each activity (Figure 2). Our institution's actual form is not depicted here in order to clearly demonstrate the utility and function of



Figure 2. The complete, 3-page physician response form with sample data from a fictitious attending pathologist, "Jane James" who has been practicing for 3 years. The respondent selects the score commensurate with their activity level for each subcategory according to the rubric in Figure 1. Her executive summary is depicted in Figure 3, below. These activity point values can be adjusted independently of the grading rubric to allow for fine-tuning. The score for the categories is automatically tabulated. The form is paginated for easy printing for retention in physical files.



Figure 3. The summary of the individual physician's performance for the fictitious attending pathologist, "Jane James," continuing from Figure 2, who has been practicing for 3 years and is graded according to the rubric presented in Figure 1. It identifies the physician and their years in practice and summarizes their numerical scores, their evaluation, and demonstrates each of these dimensions graphically with a radar plot. The radar plot allows for rapid and multidimensional understanding of a physician's performance.

the tool. The activities in the form are sometimes mutually exclusive and not all activities must be performed in order to do well. For example, the "curriculum and development course" (Figure 2) at our institution is very time-consuming and not available to all faculty at all times. Therefore, "intent" to participate is valuable for both scheduling, and participation is valuable in terms of point accrual. These activities and their categories should be altered depending on the institution deploying the evaluation and the target respondent physician population. The dashboard executive summary summarizes the evaluation of the completed form under the grading scheme of the rubric (Figure 3). The form itself and executive summary were paginated for printing and physical storage. Finally, 3 years of noncumulative successive evaluation is overlaid for both a junior and senior faculty member to see progress and change with respect to their activities (Figure 4). These faculty are fictional and presented for schematic demonstration of the utility of this tool.

The terms deployed to summarize performance were skewed toward the positive. Three terms were overtly positive ("very good," "excellent," and "outstanding"), whereas one term was neutral ("satisfactory") and one was negative ("needs improvement"). This overall positive skewing of the category evaluations reflected an institutional desire for encouragement. This can be modified for deployment at other institutions. Indeed, information can be extracted easily to provide other visualizations, like histograms or scatter plots. Faculty provided feedback after participating in the initial evaluation which resulted in several changes. These changes were easy to incorporate into the working system, taking only a few minutes. The evaluation system was then resubmitted to the physician respondents in a second deployment. The second deployment was uniformly approved by faculty in the time and effort it took to complete, as well as the value of activities it collected to evaluate.

Discussion

The desire for a new, more accurate evaluation tool arose from the multiplicity of critical comments from pathologists regarding the current system. The most common complaints were the length of time it took to complete and the relative inaccuracy of the calculations. Although the evaluation template was identical for all pathologists, the interpretation and calculations individuals used to self-report varied. As such, comparisons among pathologists or between specific groups were problematic. The new evaluation tool was specifically designed to be intuitive to complete and interpret. This tool is intended to result in a "dashboard" that allows the administrator to quickly sort through the numerous other evaluations, metrics, and products of academic physician activity. Due to this intuitive design and ease of completion, more than 1 response cycle can occur during a designated evaluation period, even before a final appraisal of the results. In our department, after initial construction, the form was provided to attending pathologists who completed their self-evaluation and forwarded to their supervisor. Comments and criticisms were then elicited from both respondents and supervisors. This feedback was incorporated in the context of the preliminary results and comments and the form adjusted. The faculty uniformly embraced the new system and their criticisms addressed adjustment of scoring for some areas and the inclusion of some activities. The new form was then redistributed to respondent pathologists and new results obtained. The responders reported no more than 20 minutes time spent for both response cycles, compared to the hours necessary for other evaluation tool previously used at our institution. The ease of response and refined clarity following incorporation of pertinent suggestions reassured responders that they had assisted in the development of a meaningful evaluation tool. They were more comfortable with the procedural aspects of the form and believed the subsequent evaluations were more accurate and comparable. The form also prompted productive conversations about resources, time, and goals on both the side of the institution and the individual pathology attendings.

Although the example presented herein is excerpted from our specifically designed form, this simplified format may be edited and repurposed for other institutions. The general categories as well as subcategories may be changed to suit the needs of the surveyor. In our version, "well-rounded" performance appeared as a symmetrical pentagon. The activities and categories will be altered to fit the institutional mission.



Figure 4. Two fictional physicians, a junior attending passing 5 years of practice (A) and an established senior attending (B) during 3 years of monitoring. The junior physician (A) has maintained their Compensation Plan and Quality performance and has acquired a new administrative role and educational duties. While this junior attending has increased their Scholarship during the third and fourth years, they have not maintained Scholarship performance as the thresholds evolved as they passed the milestone of 5 years of practice. In contrast, the senior physician (B) demonstrates relative stability during their 15th through 17th years of practice. This senior physician has acquired new education roles at the expense of their scholarship which has decreased since year 15. Specific guidance, with respect to focused performance, easily deduced from these depictions, despite their vastly differing stage in professional development.

However, this can be accomplished easily. The grading rubric is also easy to adjust for institutional thresholds (Figures 1 and 2). Though we selected grading nomenclature that was skewed toward positive feedback (very good and excellent) with only 1 neutral grade (satisfactory) and 1 grade with a negative connotation (needs improvement), this too can be changed to reflect different nomenclature or levels of achievement.

An additional positive feature of this evaluation tool is that both the evaluation form and linked grading rubric are provided to pathologist–responders. This means that responders can see in real time the effect their reporting of each item in the subcategories has on their general category score and overall evaluation (Figure 3). In addition, the subcategories contain very specific metrics of achievement such that calculations are not required. Thus, inaccurate reporting is virtually nullified since these metrics are generally verifiable.

In general, our faculty performed "excellent" or "very good" in most categories. A good evaluation system should not simply reinforce what is already expected. A good evaluation system should uncover unexpected strengths to be nurtured and unearth weakness to be addressed. This tool made these outlier behaviors visible and allowed for quick indexing. The administration felt empowered to make specific recommendations to the evaluated physicians while appreciating the balance between the physician's individual multidimensional interests and the institutional mission for the target physician respondent group.

Because of the flexibility of this system, different category scoring schemes are deployable for differing levels of experience (Figure 4). In our version, we wished to have a different grading scheme for junior (less than 5 years in practice) and senior pathologists since there are different levels of expectation for nonclinical activity between these 2 groups. Thus, both junior and senior pathologists may receive a similar overall performance score of excellent but through very different metrics. The normalization steps provided by the ERF function and the establishment of tiered grading for differing years of practice allows pathologist progress to be tracked over time. The overall progression of an individual pathologist can be tracked throughout the years despite changes in expectations within the department or institution which may be implemented for attending pathologists based on their years of practice. Sideby-side or overlaid comparison of the summary pentagon is easily possible via Excel or simple comparison of the physical documents. In addition, the aforementioned INDEX and MATCH functions in Excel allow for as many tiers of practice experience as the administrator would wish to differentiate. Indeed, different grading schemes could easily be built in depending on different faculty work goals depending on individual contracts.

In summary, we present a new method of evaluating nonclinical performance of academic pathologists which captures observable activities in an auditable dashboard format, is intuitive and quick to complete, and has an easily understood immediate evaluation. This tool is flexible and scalable capturing specific, granular activities while providing a global perspective of a pathologist respondent.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Price PB, Taylor CW, Richards JMJ, Jacobsen TL. Measurement of physician performance. J Med Educ. 1964;39:203-211.
- Carey RM, Wheby MS, Reynolds RE. Evaluating faculty clinical excellence in the academic health sciences center. *Acad Med*. 1993;68:813-817.
- Robinson M, MacNeily A, Afshar K, et al. Leadership in Canadian urology: what is the right stuff? J Surg Educ. 2013;70: 606-612.
- Kumar RDC. Leadership in healthcare. Anaesth Inten Care Med. 2013;14:39-41.
- Andreae MC, Freed GL. Using a productivity-based physician compensation program at an academic health center: a case study. *Acad Med.* 2002;77:894-899.
- Gunderman RB. The perils of paying academic physicians according to the clinical revenue they generate. *Med Sci Monit*. 2004;10:RA15-RA20.
- Akl EA, Meerpohl JJ, Raad D, et al. Effects of assessing the productivity of faculty in academic medical centres: a systematic review. *Can Med Assoc J.* 2012;184:E602-E612.
- Davis DA, Thomson M, Oxman AD, Haynes R. Changing physician performance: a systematic review of the effect of continuing medical education strategies. *JAMA*. 1995;274:700-705.
- Schindler N, Corcoran JC, Miller M, et al. Implementing an excellence in teaching recognition system: needs analysis and recommendations. *J Surg Educ*. 2013;70:731-738.
- Nutter DO, Bond JS, Coller BS, et al. Measuring faculty effort and contributions in medical education. *Acad Med.* 2000;75:199-207.
- Schindler NM, Winchester DPM, Sherman HM. Recognizing clinical faculty's contributions in education. *Acad Med.* 2002; 77:940-941.
- Sonnino RE. Professional development and leadership training opportunities for healthcare professionals. *Am J Surg.* 2013;206: 727-731.
- 13. Raj A, Carr PL, Kaplan SE, Terrin N, Breeze JL, Freund KM. Longitudinal analysis of gender differences in academic

productivity among medical faculty across 24 medical schools in the United States. *Acad Med.* 2016;91:1074-1079.

- Schrijver I, Brady KJ, Trockel M. An exploration of key issues and potential solutions that impact physician wellbeing and professional fulfillment at an academic center. *PeerJ*. 2016;4:e1783.
- Holcombe RF, Hollinger KJ. Mission-focused, productivitybased model for sustainable support of academic hematology/ oncology faculty and divisions. J Oncol Pract. 2010;6:74-79.
- Freischlag JA. It is not a slide but it is a ladder. *Am J Surg.* 2013; 206:667-668.
- Scoggins CR, Crockett T, Wafford L, Cannon RM, McMasters KM. Improving clinical productivity in an academic surgical practice through transparency. J Am Coll Surg. 2013;217:46-51.
- Ragsdale JR, Vaughn LM, Klein M. Characterizing the adequacy, effectiveness, and barriers related to research mentorship among junior pediatric hospitalists and general pediatricians at a large academic institution. *Hosp Pediatr.* 2014;4:93-98.
- Pitt MB, Furnival RA, Zhang L, Weber-Main AM, Raymond NC, Jacob AK. Positive Peer-Pressured Productivity (P-QUAD): novel use of increased transparency and a weighted lottery to increase a division's academic output. *Acad Pediatr.* 2017;17: 218-221.
- Filler G, Burkoski V, Tithecott G. Measuring physicians' productivity: a three-year study to evaluate a new remuneration system. *Acad Med.* 2014;89:144-152.
- Ma OJ, Hedges JR, Newgard CD. The academic RVU: ten years developing a metric for and financially incenting academic productivity at Oregon Health & Science University. *Acad Med.* 2017;92:1138-1144.
- Hilton C, Fisher WJ, Lopez A, Sanders C. A relative-value-based system for calculating faculty productivity in teaching, research, administration, and patient care. *Acad Med.* 1997;72:787-793.
- Nora LM, Pomeroy C, Curry TEJ, Hill NS, Tibbs PA, Wilson EA. Revising appointment, promotion, and tenure procedures to incorporate an expanded definition of scholarship: the University of Kentucky College of Medicine Experience. *Acad Med.* 2000;75: 913-924.