Journal of Graduate Medical Education

Natural Language Processing to Estimate Clinical Competency Committee Ratings

--Manuscript Draft--

Manuscript Number:	
Full Title:	Natural Language Processing to Estimate Clinical Competency Committee Ratings
Short Title:	Natural Language Processing
Article Type:	Original Research
Keywords:	Natural language processing; artificial intelligence; assessment; clinical competency committee; milestones
Corresponding Author:	Kenneth L Abbott, MS University of Michigan Medical School Ann Arbor, MI UNITED STATES
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	University of Michigan Medical School
Corresponding Author's Secondary Institution:	
First Author:	Kenneth L Abbott, MS
First Author Secondary Information:	
Order of Authors:	Kenneth L Abbott, MS
	Brian C George, MD, MAEd
	Gurjit Sandhu, PhD
	Calista M Harbaugh, MD, MS
	Paul G Gauger, MD
	Erkin Ötleş, MEng
	Niki Matusko, BS
	Joceline V Vu, MD
Order of Authors Secondary Information:	
Abstract:	Background Residency program faculty participate in clinical competency committee (CCC) meetings, which are designed to evaluate residents' performance and aid in the development of individualized learning plans. In preparation for the CCC meetings, faculty members synthesize performance information from a variety of sources. Natural language processing (NLP), a form of artificial intelligence, might facilitate these holistic reviews. However, there is little research involving the application of this technology to resident performance assessments. Objective Examine whether NLP can be used to estimate CCC ratings. Methods We aggregated and analyzed text from end-of-rotation assessments for surgical residents who trained at one institution between 2014 and 2018. No residents were excluded. We created predictive models for 16 Accreditation Council for Graduate Medical Education (ACGME) Milestones. We compared the performance of models with and without NLP predictors. Results We analyzed 594 end-of-rotation assessments and 97 CCC assessments for 24

	general surgery residents. The mean (standard deviation) AUC was 0.84 (0.05) for models with non-NLP predictors, 0.83 (0.06) for models with NLP predictors, and 0.87 (0.05) for models with both NLP and non-NLP predictors. Conclusions NLP can identify language correlated with specific ACGME Milestone ratings. In preparation for CCC meetings, faculty could use information automatically extracted from text to focus attention on residents who might benefit from additional support and guide the development of educational interventions.

Natural Language Processing to Estimate Clinical Competency Committee Ratings

Authors

Kenneth L. Abbott, MS,[1] Medical Student; Brian C. George, MD, MAEd,[2] Assistant Professor of Surgery; Gurjit Sandhu, PhD,[2] Associate Professor of Surgery; Calista M. Harbaugh, MD, MS,[2] Chief Resident, General Surgery; Paul G. Gauger, MD,[2] Professor of Surgery and Program Director; Erkin Ötleş, MEng,[1] Medical Student; Niki Matusko, BS,[2] Statistician; Joceline V. Vu, MD,[2] Chief Resident, General Surgery

Affiliations

[1]University of Michigan Medical School, Ann Arbor, Michigan

[2]Department of Surgery, University of Michigan, Ann Arbor, Michigan

Contributions

KA served as lead investigator and analyst for this study. All authors contributed to study design and manuscript preparation.

Prior Oral Presentation

Abbott KL, Harbaugh CM, Matusko N, Sandhu G, Gauger PG, Vu JV. Use of natural language processing to interpret resident performance evaluations. Presented at the Association for Academic Surgery and Society of University Surgeons Academic Surgical Congress, February 2019, Houston, TX.

Funding

This project was funded by a University of Michigan Medical School Capstone for Impact grant.

Conflicts of Interest

The authors have no conflicts of interest to disclose.

Correspondence

Correspondence should be addressed to Joceline Vu, Department of Surgery, University of Michigan, Ann Arbor, MI; 48109; phone: (703) 200-8623; e-mail: vuj@med.umich.edu; Twitter: @jocelinevu.

Manuscript Type

Original Research

Word Counts

250 (abstract), 1439 (manuscript)

Abstract

Background

Residency program faculty participate in clinical competency committee (CCC) meetings, which are designed to evaluate residents' performance and aid in the development of individualized learning plans. In preparation for the CCC meetings, faculty members synthesize performance information from a variety of sources. Natural language processing (NLP), a form of artificial intelligence, might facilitate these holistic reviews. However, there is little research involving the application of this technology to resident performance assessments.

Objective

Examine whether NLP can be used to estimate CCC ratings.

Methods

We aggregated and analyzed text from end-of-rotation assessments for surgical residents who trained at one institution between 2014 and 2018. No residents were excluded. We created predictive models for 16 Accreditation Council for Graduate Medical Education (ACGME) Milestones. We compared the performance of models with and without NLP predictors.

Results

We analyzed 594 end-of-rotation assessments and 97 CCC assessments for 24 general surgery residents. The mean (standard deviation) for area under the receiver operating characteristic curve (AUC) was 0.84 (0.05) for models with non-NLP predictors, 0.83 (0.06) for models with NLP predictors, and 0.87 (0.05) for models with both NLP and non-NLP predictors.

Conclusions

NLP can identify language correlated with specific ACGME Milestone ratings. In preparation for CCC meetings, faculty could use information automatically extracted from text to focus attention on

residents who might benefit from additional support and guide the development of educational interventions.

Introduction

Residency programs use a system of assessments to track trainee progress and development. For example, a subset of faculty members participate in clinical competency committee (CCC) meetings, which occur every six months and are designed to evaluate performance and aid in the development of individualized learning plans and interventions.[1] In preparation for the CCC meetings, committee members synthesize performance information from a variety of sources—some formal (e.g., monthly end-of-rotation assessments) and some informal (e.g., conversations).

Artificial intelligence could support the CCC faculty performing these holistic reviews by guiding their attention to residents who may benefit from additional support. Natural language processing (NLP) is a form of artificial intelligence that interprets complex human language.[2] In general surgery, Milestones are used to structure CCC meeting discussion and resident assessment.[3,4] It is unknown whether NLP can identify language correlated with specific Accreditation Council for Graduate Medical Education (ACGME) Milestone ratings, but this could help faculty identify residents who may need additional support in a specific performance domain. For example, faculty could review predictions of Milestone ratings, gather additional information about residents who are predicted to have low Milestone ratings, and spend additional CCC meeting time discussing these residents.

With this study, we examine whether NLP can be used to estimate CCC Milestone ratings, using text from end-of-rotation assessments.

Methods

Data

We collected deidentified performance assessments for surgical residents who trained at one institution between 2014 and 2018. No residents were excluded. Assessments included monthly end-of-rotation assessments gathered via an online assessment system (MedHub, https://www.medhub.com/) and biannual CCC assessments. End-of-rotation assessments included nine numeric items with anchors that were generally related to the ACGME general surgery Milestones,[3,4] and asked faculty to rate trainees along multiple dimensions, using a 9-point Likert scale. End-of-rotation clinical assessments also included a tenth numeric item that asked faculty to rate a trainee's overall clinical competence, and one text field for general comments. The CCC assessments included a numeric rating for each of the 16 Milestones grouped within 6 competencies (*patient care*, *medical knowledge*, *systems-based practice, practice-based learning and improvement*, *professionalism,* and *interpersonal and communication skills*) and 8 domains (*care for diseases and conditions*, *coordination of care*, *performance of operations and procedures*, *self-directed learning*, *teaching*, *improvement of care*, *maintenance of physical and emotional health*, and *performance of administrative tasks*). CCC assessments also included a text field for comments for each Milestone.

Analysis

Figure 1 summarizes our analytic process. First, we identified and aggregated text from all the end-of-rotation assessments (not CCC assessments) delivered during each CCC assessment period. Since we aimed to detect low performance, we dichotomized CCC ratings into high (≥7) and low (<7) ratings.

Next, we used the *googleLanguageR* package[5] to connect to Google Cloud Natural Language[6] and complete sentiment analysis of text comments from end-of-rotation assessments. Sentiment analysis is a type of NLP whose demand has been driven by electronic commerce and other industries that wish to interpret large amounts of qualitative data, such as social media comments, product reviews, or restaurant reviews.[7] Sentiment analysis can extract information related to opinion and translate it into quantitative data, such as positive or negative numeric values for specific words; for example, in the phrase "excellent performance," the noun *performance* has positive sentiment, because *excellent* is positive, and the adjective *excellent* describes the noun *performance*. By contrast, in the phrase "terrible performance," the same noun *performance* has negative sentiment, because *terrible* is negative. Google's NLP software produces numeric scores between -1 and 1, in intervals of 0.1.

Then, we used the *tidytext* and *textstem* packages[8,9] to create a frequency matrix of words extracted from text comments. For example, a comment consisting only of "solid performance" would yield a 1 in the column for the word *solid*, a 1 in the column for the word *performance*, and 0 in all columns for other words. In creating this word frequency matrix, we discarded stop words, which are extremely common words of little value in NLP,[2] and used lemmatization, which is a means of identifying variants of the same word;[2] for example, singular *resident* and plural *residents* were both be mapped to *resident*.

Next, we used h2o.ai's Driverless AI[10] to estimate the probability of a low CCC assessment rating. This software automatically evaluates thousands of possible predictive models, which may involve a variety of machine learning algorithms, and then creates an ensemble of predictive models that yield the best performance. We created 48 models: 16 models with non-NLP predictors, 16 models with NLP predictors, and 16 models with all predictors. Outcome variables included each of the 16 numeric ratings on CCC assessments. NLP predictors included Google sentiment score for text comments from aggregated end-of-rotation assessments and the above-described word frequency matrix. We evaluated the performance of each of these models with 3-fold cross validation, using the resulting predictions to calculate area under the receiver operating characteristic curve (AUC).

We used *R* version 4.0.0[11] to aggregate and analyze all assessment data.

IRB Statement

This study was exempt from review by the University of Michigan Institutional Review Board (IRB).

Results

We analyzed 594 end-of-rotation assessments and 97 clinical competency assessments for 24 general surgery residents (Table 1). CCC assessment ratings varied by Milestone, with the prevalence of low ratings <7 ranging from 0.23 to 0.57 (Table 2); prevalence of low ratings was greatest for *performance of operations and procedures* under *patient care* and *performance of assignments and administrative tasks* under *professionalism*. Across all models, sensitivity for detection of low ratings ranged from 0.28 to 0.89; accordingly, AUCs ranged from 0.71 to 0.96 (Table 2). AUCs were comparable for models with NLP predictors, non-NLP predictors, and all predictors.

Table 1. Sample characteristics and clinical competency committee assessment ratings.

Variable	**Post-graduate year (PGY)**					***p***[a]
	PGY-1	PGY-2	PGY-3	PGY-4	PGY-5	
n	1	3	9	35	49	
Gender = female (%)	0 (0)	0 (0)	3 (33.3)	10 (28.6)	12 (24.5)	0.765
Ethnicity = non-white (%)	0 (0)	2 (66.7)	4 (44.4)	13 (37.1)	19 (38.8)	0.626
Patient care						
1. Care for diseases and conditions (mean (SD))	4 (NA)	5.33 (1.15)	6 (0)	7.6 (0.81)	7.92 (0.40)	<0.001
2. Care for diseases and conditions (mean (SD))	4 (NA)	5.33 (1.15)	6 (0)	7.54 (0.85)	7.8 (0.61)	<0.001
3. Performance of operations and procedures (mean (SD))	4 (NA)	4.67 (1.15)	5.78 (0.67)	6.51 (1.01)	7.27 (1.06)	<0.001
Medical knowledge						
1. Care for diseases and conditions (mean (SD))	6 (NA)	5.33 (1.15)	5.33 (1.00)	6.97 (1.12)	7.35 (1.11)	<0.001
2. Performance of operations and procedures (mean (SD))	4 (NA)	6 (0)	5.78 (0.67)	6.86 (1.00)	7.55 (0.84)	<0.001
Systems-based practice						
1. Coordination of care (mean (SD))	6 (NA)	5.33 (1.15)	6.22 (0.67)	7.66 (0.76)	7.71 (0.71)	<0.001
2. Improvement of care (mean (SD))	6 (NA)	6 (2.00)	5.78 (1.56)	7.03 (1.22)	7.43 (0.91)	0.001
Practice-based learning and improvement						
1. Teaching (mean (SD))	6 (NA)	6 (0)	5.89 (2.03)	7.6 (0.81)	7.63 (0.78)	<0.001
2. Self-directed learning (mean (SD))	6 (NA)	6 (0)	5.56 (1.33)	6.97 (1.40)	7.31 (1.19)	0.002
3. Improvement of care (mean (SD))	4 (NA)	6 (0)	5.11 (1.05)	7.2 (0.99)	7.8 (0.61)	<0.001
Professionalism						
1. Care for diseases and conditions (mean (SD))	6 (NA)	6 (0)	6.44 (1.33)	7.77 (0.65)	7.67 (0.75)	<0.001
2. Maintenance of physical and emotional health (mean (SD))	4 (NA)	5.33 (1.15)	5.56 (1.33)	7.49 (0.89)	7.71 (0.71)	<0.001
3. Performance of assignments and administrative tasks (mean (SD))	4 (NA)	3.33 (1.15)	5.11 (1.05)	6.17 (1.64)	7.31 (1.19)	<0.001
Interpersonal and communication skills						
1. Care for diseases and conditions (mean (SD))	4 (NA)	5.33 (1.15)	5.56 (0.88)	7.43 (1.04)	7.71 (0.71)	<0.001
2. Coordination of care (mean (SD))	6 (NA)	6.67 (1.15)	5.78 (0.67)	7.49 (0.89)	7.71 (0.71)	<0.001
3. Performance of operations and procedures (mean (SD))	4 (NA)	5.33 (1.15)	5.78 (0.67)	6.46 (0.98)	7.63 (0.78)	<0.001

PGY: post-graduate year; SD: standard deviation; NA: not applicable

[a]Analysis of rating change across PGY required exclusion of the lone PGY-1 observation, which had no standard deviation.

Table 2. Performance of models estimating clinical competency committee assessment ratings, with and without natural-language processing predictors.

Competency	**Prevalence**		**AUC**	
	Low ratings, mean (SD) = 0.36 (0.11)	Non-NLP predictors, mean (SD) = 0.84 (0.05)	NLP predictors, mean (SD) = 0.83 (0.06)	All predictors, mean (SD) = 0.87 (0.05)
Patient care				
1. Care for diseases and conditions	0.23	0.86	0.95	0.96
2. Care for diseases and conditions	0.27	0.93	0.88	0.92
3. Performance of operations and procedures	0.57	0.89	0.78	0.95
Medical knowledge				
1. Care for diseases and conditions	0.45	0.81	0.82	0.85
2. Performance of operations and procedures	0.45	0.83	0.82	0.81
Systems-based practice				
1. Coordination of care	0.26	0.79	0.81	0.83
2. Improvement of care	0.40	0.75	0.82	0.81
Practice-based learning and improvement				
1. Teaching	0.28	0.76	0.80	0.81
2. Self-directed learning	0.42	0.78	0.83	0.85
3. Improvement of care	0.33	0.83	0.92	0.92
Professionalism				
1. Care for diseases and conditions	0.23	0.88	0.87	0.94
2. Maintenance of physical and emotional health	0.29	0.86	0.82	0.83
3. Performance of assignments and administrative tasks	0.52	0.83	0.79	0.84
Interpersonal and communication skills				
1. Care for diseases and conditions	0.30	0.83	0.74	0.86
2. Coordination of care	0.29	0.89	0.89	0.90
3. Performance of operations and procedures	0.49	0.88	0.71	0.92

AUC: Area under receiver operating characteristic curve; NLP: natural language processing; SD: standard deviation

Discussion

We are aware of no previous research applying NLP to the ACGME Milestone rating process. In this study, we used NLP of end-of-rotation assessments to examine whether NLP could identify language correlated with specific Milestone ratings. We found that NLP could be used to estimate Milestone ratings on biannual CCC assessments. Information automatically extracted from text could help faculty focus attention on residents who might benefit from additional support.

Many prior studies have applied NLP to analysis of medical records,[12] but little research applies NLP to medical education. A recent review found only a handful of studies of NLP in medical education,[13] and only one of these involved performance assessments. That study classified text into six ACGME competencies,[14] but did not relate narrative data to ACGME Milestone ratings.[3,4] We found that NLP can be used to estimate Milestone ratings. This extends prior research into NLP in graduate medical education.

Faculty could use NLP to help prepare for CCC meetings. For example, automated analyses of numeric ratings and text comments could be used to predict the probability of a low Milestone rating or recommend a numeric Milestone rating. The scope of these analyses might include certain Milestones of interest, Milestones grouped according to competency or domain, or all Milestones. Before a CCC meeting, faculty could gather additional information about residents identified by these analyses, and during a CCC meeting, faculty could spend additional time discussing these residents. Faculty could also track estimates of CCC ratings over time. Since AUCs for models using NLP predictors are comparable to AUCs for models using all predictors, priority might be given to incorporating data sources that do not already include numeric information (e.g., messages existing outside of the MedHub performance assessment system). Priority might also be given to analysis of text that addresses gaps in numeric data (e.g., *improvement of care* under *systems-based practice*). Alternately, faculty rater training could be used to enhance the quality of text feedback for specific Milestones.

This study has limitations. First, the development of predictive models can entail tradeoffs between performance and interpretability (e.g., the ability to see how specific predictors account for variance in each Milestone rating). This increases the risk of an NLP model obscuring bias related to gender, ethnicity, or other variables that should have no bearing on performance ratings. Therefore, implementation of these methods should be preceded by attempts at detection and mitigation of biases that NLP might propagate from written assessments. Second, our study incorporated assessments from only 24 residents at a single institution and these findings might not generalize to other groups of residents. However, the pattern of high AUCs across models, despite such a small sample, is reassuring. Despite these limitations, our findings should provide medical educators with useful information on how NLP might support holistic review processes.

Conclusion

NLP can identify language correlated with specific ACGME Milestone ratings. In preparation for CCC meetings, faculty could use information automatically extracted from text to focus attention on residents who might benefit from additional support and guide the development of educational interventions.

References

1. Accreditation Council for Graduate Medical Education. Clinical competency committees. https://www.acgme.org/Portals/0/ACGMEClinicalCompetencyCommitteeGuidebook.pdf. Published 2020. Accessed May 6, 2020.
2. Manning CD, Schütze H. *Foundations of statistical natural language processing.* 2nd printing, with corrections. ed. Cambridge, Mass.: MIT Press; 2000.
3. Kuo LE, Hoffman RL, Morris JB, et al. A milestone-based evaluation system-the cure for grade inflation? *J Surg Educ.* 2015;72(6):e218-225.
4. Accreditation Council for Graduate Medical Education, American Board of Surgery. The general surgery milestone project. http://www.acgme.org/Portals/0/PDFs/Milestones/SurgeryMilestones.pdf?ver=2015-11-06-120519-653. Published 2015. Accessed July 22, 2018.
5. Edmondson M. googleLanguageR: Call Google's 'Natural Language' API, 'Cloud Translation' API and 'Cloud Speech' API. https://cran.r-project.org/web/packages/googleLanguageR/index.html. Published 2020. Accessed February 7, 2020.
6. Google LLC. Cloud Natural Language. https://cloud.google.com/natural-language/. Published 2020. Accessed February 7, 2020.
7. Pang B, Lee L. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval.* 2008;2(1–2):1-135.
8. Silge J, Robinson D. tidytext: Text mining and analysis using tidy data principles in R. http://dx.doi.org/10.21105/joss.00037. Published 2016. Accessed.
9. Rinker TW. textstem: Tools for stemming and lemmatizing text. http://github.com/trinker/textstem. Published 2018. Accessed February 7, 2020.
10. h2o.ai Inc. Driverless AI. https://www.h2o.ai/products/h2o-driverless-ai/. Published 2020. Accessed February 7, 2020.

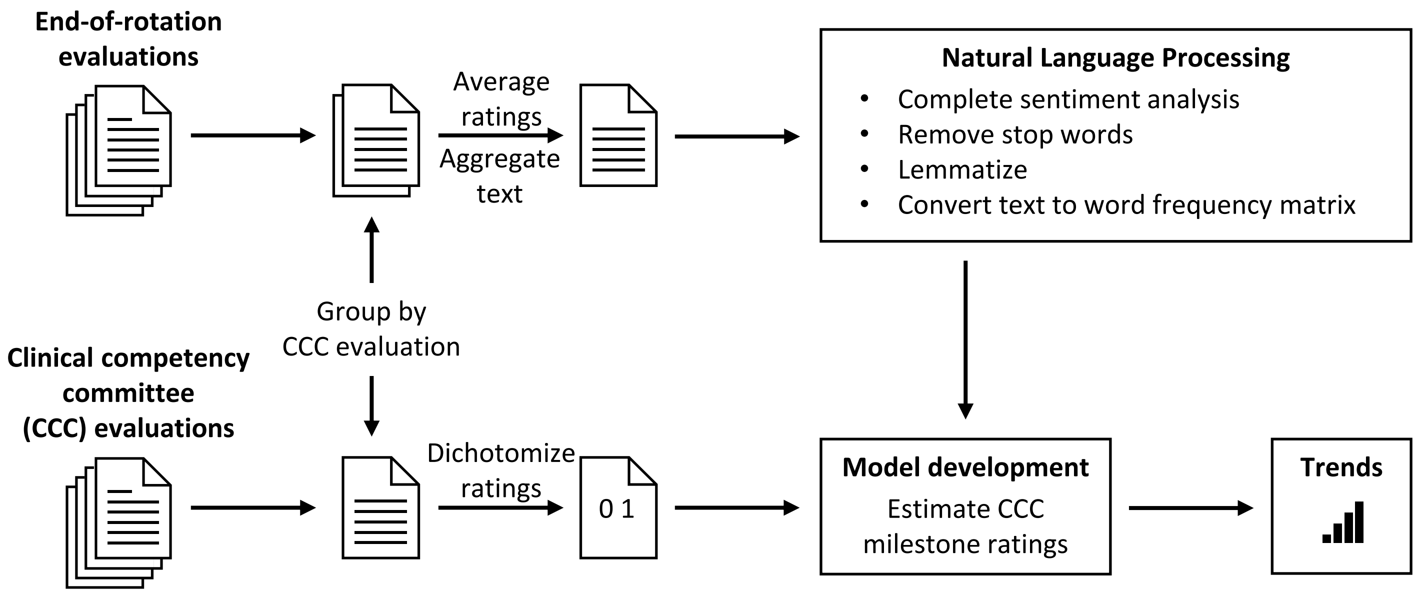
11. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. https://www.R-project.org/. Published 2020. Accessed May 11, 2020.

12. Dias RD, Gupta A, Yule SJ. Using machine learning to assess physician competence: A systematic review. *Acad Med.* 2018.

13. Chary M, Parikh S, Manini AF, Boyer EW, Radeos M. A review of natural language processing in medical education. *West J Emerg Med.* 2019;20(1):78-86.

14. Zhang R, Pakhomov S, Gladding S, Aylward M, Borman-Shoap E, Melton GB. Automated assessment of medical training evaluation text. *AMIA Annu Symp Proc.* 2012;2012:1459-1468.

Figure 1


End-of-rotation evaluations
Average ratings
Aggregate text
Natural Language Processing
Complete sentiment analysis
Remove stop words
Lemmatize
Convert text to word frequency matrix
Group by CCC evaluation
Clinical competency committee (CCC) evaluations
Dichotomize ratings
0 1
Model development
Estimate CCC milestone ratings
Trends

Sunday, September 27, 2020 Dr. Joceline Vu | Resident | Account | Log Off

MICHIGAN MEDICINE
UNIVERSITY OF MICHIGAN

Surgery Home Portfolio Schedules Procedures Evaluations Conferences Help

Home » Evaluations » Evaluations Form

Evaluations Form

Dr. Example - Peer to Peer Evaluation

Program: General Surgery HO Level: 1

Evaluator: Dr. Vu, Joceline

Service: Endocrin/MIS - Blue 1 & 2
Rotation: September (09/01-09/30/20)
Issue Date: 9/25/2020

Insufficient contact to evaluate (delete evaluation)

In evaluating the resident's performance, use as your standard the level of knowledge, skills and attitudes expected from the clearly satisfactory resident at this stage of training. For any component that is rated as 3 or less, please provide specific comments and recommendations in the comments field at the bottom of this form. Be as specific as possible, including reports of critical incidents and/or outstanding performance. Global adjectives or remarks, such as "good resident", do not provide meaningful feedback to the residents.

	Unsatisfactory			Satisfactory			Superior			Insufficient contact to judge
	1	2	3	4	5	6	7	8	9	N/A
1. Patient Care*	Incomplete, inaccurate medical interviews, physical examination, and review of other data: incompetent performance of essential procedures; fails to analyze clinical data and consider patient preferences when making medical decisions. Does not seek help when needed.								Superb, accurate, comprehensive, medical interviews, physical examinations, review of other data, and procedural skills; always makes diagnostic and therapeutic decisions based on available evidence, sound judgment, and patient preferences. Always seeks help when needed.	

	Unsatisfactory			Satisfactory			Superior			Insufficient contact to judge
	1	2	3	4	5	6	7	8	9	N/A
2. Medical Knowledge*	Limited knowledge of basic and clinical sciences; minimal interest in learning; does not understand complex relations, mechanisms of disease. No evidence of outside reading,								Exceptional knowledge of basic and clinical sciences, highly resourceful development of knowledge; comprehensive understanding of complex relationships, mechanisms of	

does not know pertinent literature.								disease. Clear evidence of outside reading, excellent knowledge of current literature.	

3. Practice-Based Learning and Improvement*

Unsatisfactory			Satisfactory			Superior			Insufficient contact to judge
1	2	3	4	5	6	7	8	9	N/A
Fails to perform self-evaluation; lacks insight; initiative, resists or ignores feedback; fails to use information technology to enhance patient care or pursue self-improvement. Loner, does not understand interdependencies and complex hospital and health care system.								Constantly evaluates own performance, incorporates feedback into improvement activities; effectively uses technology to manage information for patient care and self-improvement. Understands interdependencies and complexities of hospital and health care system.	

4. Interpersonal and Communication Skills*

Unsatisfactory			Satisfactory			Superior			Insufficient contact to judge
1	2	3	4	5	6	7	8	9	N/A
Does not establish even minimally effective therapeutic relationships with patients and families, does not demonstrate ability to build relationships through listening, narrative or nonverbal skills; does not provide education or counseling to patients, families, or colleagues. Condescending, demeaning, arrogant.								Establishes a highly effective therapeutic relationship with patients and families; demonstrates excellent relationship building through listening, narrative and nonverbal skills; excellent education and counseling of patients, families, and colleagues; always 'interpersonally' engaged. Modest and respectful.	

5. Professionalism*

Unsatisfactory			Satisfactory			Superior			Insufficient contact to judge
1	2	3	4	5	6	7	8	9	N/A
Lacks respect, compassion, integrity, honesty; disregards need for self-assessment; fails to acknowledge errors; does not consider needs of patients, families, colleagues; does								Always demonstrates respect, compassion, integrity, honesty, teaches/role models responsible behavior, total commitment to self-assessment;	

not display responsible behavior.								willingly acknowledges errors; always considers needs of patients, families, colleagues.	

6. System-Based Practice*

Unsatisfactory			Satisfactory			Superior			Insufficient contact to judge
1	2	3	4	5	6	7	8	9	N/A
Unable to access/mobilize outside resources, actively resists effort to improve systems of care; does not use systematic approaches to reduce error and improve patient care.								Effectively accesses/utilizes outside resources, effectively uses systematic approaches to reduce errors and improve patient care; enthusiastically assists in developing systems' improvement.	

7. Technical Abilities and Skills*

Unsatisfactory			Satisfactory			Superior			Insufficient contact to judge
1	2	3	4	5	6	7	8	9	N/A
Technical skills exceedingly poor. Unable to perform even the most rudimentary technical skills appropriate to their level of training. The resident makes little or no effort to improve on their skills.								Technical skills are outstanding. Able to perform skills and procedures well in advance of his/her level of training. The resident also makes a concerted effort to improve their technical skills and abilities.	

8. Leadership*

Unsatisfactory			Satisfactory			Superior			Insufficient contact to judge
1	2	3	4	5	6	7	8	9	N/A
No leadership skills. Cannot direct a group effectively. No conflict resolution skills. Cannot build consensus. Not respected by students, junior residents, peers, faculty or other ancillary health care personnel. Leadership often assumed by other team members.								Excellent leadership skills. Organized, effective. Excellent at conflict resolution. Respected by students, junior residents, peers, faculty, and other ancillary health care personnel. Delegates effectively.	

Unsatisfactory			Satisfactory			Superior			Insufficient contact to judge
1	2	3	4	5	6	7	8	9	N/A

9. Teaching Skills and Ability*

Resident has poor teaching skills, cannot transmit even simple concepts or skills. Minimal to no patience. Condescending, demeaning.								Resident has excellent teaching skills. Can transmit complex ideas and/or skills. Is very skillful at explaining concepts in multiple ways in order to facilitate understanding. Exceedingly patient. Never condescending or demeaning. Recognized at multiple levels as an excellent teacher.	

Unsatisfactory			Satisfactory			Superior			Insufficient contact to judge
1	2	3	4	5	6	7	8	9	N/A

10. Resident's overall clinical competence*

11. General Comments: (no responses)

* Required fields Option description (place mouse over field to view)

Reset Form Submit completed evaluation Submit