The American Journal of Surgery 221 (2021) 369–375

Contents lists available at ScienceDirect

The American Journal of Surgery

journal homepage: www.americanjournalofsurgery.com

Natural language processing and entrustable professional activity text feedback in surgery: A machine learning model of resident autonomy

Christopher C. Stahl [a], Sarah A. Jung [a], Alexandra A. Rosser [a], Aaron S. Kraut [b], Benjamin H. Schnapp [b], Mary Westergaard [b], Azita G. Hamedani [b], Rebecca M. Minter [a], Jacob A. Greenberg [a, *]

[a] *Department of Surgery, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA*
[b] *Department of Emergency Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI, USA*

ARTICLE INFO

Article history:
Received 18 May 2020
Received in revised form
19 November 2020
Accepted 21 November 2020

Keywords:
Surgery education
Natural language processing
Entrustable professional activities
Assessment
Feedback

ABSTRACT

Background: Entrustable Professional Activities (EPAs) contain narrative 'entrustment roadmaps' designed to describe specific behaviors associated with different entrustment levels. However, these roadmaps were created using expert committee consensus, with little data available for guidance. Analysis of actual EPA assessment narrative comments using natural language processing may enhance our understanding of resident entrustment in actual practice.

Methods: All text comments associated with EPA microassessments at a single institution were combined. EPA—entrustment level pairs (e.g. Gallbladder Disease—Level 1) were identified as documents. Latent Dirichlet Allocation (LDA), a common machine learning algorithm, was used to identify latent topics in the documents associated with a single EPA. These topics were then reviewed for interpretability by human raters.

Results: Over 18 months, 1015 faculty EPA microassessments were collected from 64 faculty for 80 residents. LDA analysis identified topics that mapped 1:1 to EPA entrustment levels (Gammas >0.99). These LDA topics appeared to trend coherently with entrustment levels (words demonstrating high entrustment were consistently found in high entrustment topics, word demonstrating low entrustment were found in low entrustment topics).

Conclusions: LDA is capable of identifying topics relevant to progressive surgical entrustment and autonomy in EPA comments. These topics provide insight into key behaviors that drive different level of resident autonomy and may allow for data-driven revision of EPA entrustment maps.

Introduction

Surgical education is constantly evolving to meet the needs of the field and its patients. There has been a push towards competency-based education in Surgery due to concerns about the ability of surgical graduates to operate independently upon completion of their training.[1] One of the fundamental challenges involved in competency-based education is the measurement of competency: how do we know when a resident is ready to safely practice independently?

Entrustable Professional Activities (EPAs) are a novel competency-based assessment framework developed to help define and standardize entrustment decisions in graduate medical education. EPAs represent the essential activities of a practicing physician in a given specialty. Five EPAs currently exist for general surgery, and each time a resident completes an EPA evaluation, faculty complete a microassessment. The microassessment consists of a numeric score (0–4, ranging from "observation only" to "supervising others") and an option for free text comments giving feedback on the professional activity.[2] Each EPA is published with a corresponding 'entrustment roadmap', a narrative text description of the behaviors demonstrated by the resident that correspond to the different 0–4 entrustment levels. These entrustment roadmaps are meant to guide score assignment and discussion of specific behaviors within the narrative comments.

* Corresponding author. Division of Minimally Invasive Surgery, University of Wisconsin School of Medicine and Public Health, 600 Highland Avenue, Clinical Science Center, Madison, WI, 53792, USA.
E-mail address: greenbergj@surgery.wisc.edu (J.A. Greenberg).

https://doi.org/10.1016/j.amjsurg.2020.11.044
0002-9610/© 2020 Elsevier Inc. All rights reserved.

These entrustment maps were carefully written by an expert committee with representation from the ABS, APDS, RRC, and RAS.[2] However, at that time there was very little data on what surgeon educators out in practice considered representative behaviors for each EPA entrustment level. An in-depth analysis of real-world EPA comments associated with each given entrustment level can offer important insight into what "boots on the ground" surgical faculty consider important for resident entrustment. Additionally, this data could be used to iteratively revise the EPA entrustment roadmaps to better reflect daily practice.

This study harnesses the power of natural language processing (NLP) to perform such an analysis. Latent Dirichlet allocation (LDA) topic modeling was used to identify latent topics within EPA narrative feedback. We assessed the ability of the LDA algorithm to identify topics within these comments that consistently mapped to different entrustment levels as defined within the Entrustable Professional Activities (EPA) framework. We then reviewed the identified topics to determine if the mapping of these computer-generated topics to entrustment levels was comprehensible to human raters.

Material and methods

EPA data

Our institution is participating in the national American Board of Surgery (ABS) pilot trial of Entrustable Professional Activities (EPAs) in surgical education. As previously described, our implementation strategy for EPA assessment involves a mobile phone application to which residents and faculty can submit EPA assessments containing both an entrustment score (0–4, ranging from "observation only" to "supervising others") and free text comments on the resident's performance.[3] Assessments were collected for the five general surgery EPAs currently in the pilot trial (General Surgical Consultation, Trauma, Gallbladder Disease, Right Lower Quadrant Pain, Inguinal Hernia) from July 2018 to January 2020. Only faculty assessments of residents were included in the analysis. A unique feature of our implementation strategy is multi-departmental collaboration, with faculty evaluations of residents performed by both Surgery and Emergency Medicine faculty. Assessments performed by both of these faculty groups were included in the analysis.

Topic modeling

Natural language processing (NLP) involves the utilization of computers to interact with natural language data. Topic modeling is a subset of NLP designed to uncover topics present within a corpus (body) of text data. On a more technical level, these are generative statistical models that provide a probabilistic framework for the frequency of term (word) occurrence in text.[4] Of note, despite significant advances in the fields of NLP and machine learning, most computational analyses of text still interact with data using a bag-of-words model. Future developments may allow computers to interact with text in a more humanly familiar way, such as a 'bag-of-concepts' or 'bag-of-narratives', but currently bag-of-words models predominate.[5] This means that the topics generated using these methods are lists of words ordered by likelihood of belonging to a given topic (a distribution over words) and still require human interpretation.[6]

Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is a common approach used for topic modeling. LDA is a generative, probabilistic, Bayesian, unsupervised machine learning algorithm used to identify hidden (latent) topics in text corpora using a Dirichlet distribution to allocate the words present in documents into topics.[4,7] A simplified view of the allocation process is as follows: the algorithm randomly assigns every word in the corpus of data to one of k topics (the number of topics [k] is pre-specified by the user). The algorithm then "throws out" the topic for a single word and re-calculates the topic to which that word is most likely to be assigned based on the distribution of all other words. It then reassigns that word to its most likely topic and repeats the process for the next word until it has reassigned all words in the corpus. This process is repeated thousands of times, until the words stop consistently getting reassigned to new topics (i.e., the estimates have reached a steady state).

Interpretation

After the LDA process is complete, it generates document-topic probabilities (gamma), and word-topic probabilities (beta).[8] LDA models each document as a mixture of topics—each document-topic pair has a gamma that represents proportion of the document made up of that topic. Gamma is reported like a probability, with a value of 1 meaning that 100% of that document is made up of the given topic, and a value of zero indicating 0%. Each topic consists of a distribution of words—each word has a word-topic probability (beta), with a higher beta representing a higher likelihood of that word being present in the corresponding topic. Each word may appear in multiple topics, albeit with varying probability. The topics identified using LDA are bags-of-words ordered by decreasing beta values. A summary of these terms can be found in Table 1.

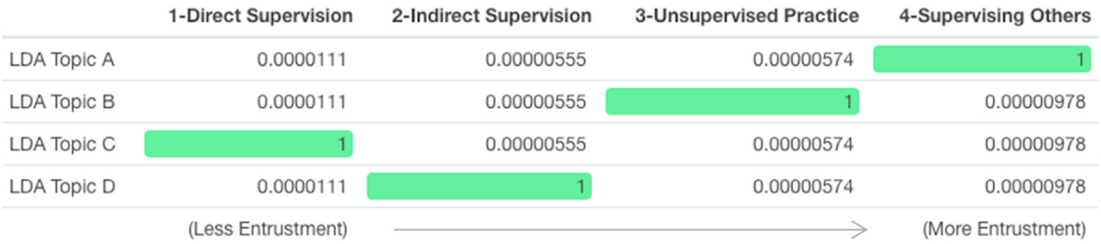
Analysis

EPA comments were cleaned for LDA analysis by removing contractions and special characters and then tokenized into individual words using the tm package in R.[9] Stop words (commonly used words in the English language that typically provide little meaning such as "the", "and", etc.) were removed, along with any

Table 1
Terms used to interpret LDA analysis.

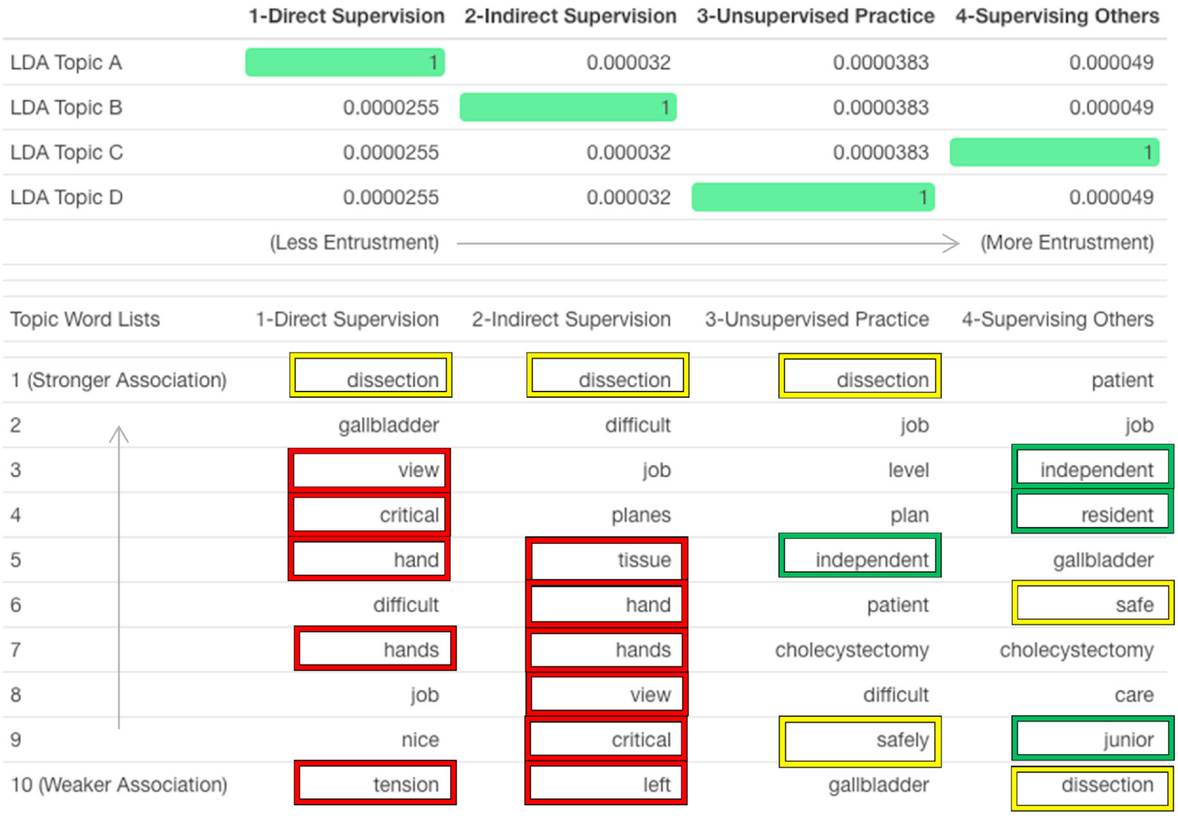
Term	Possible Values	Interpretation
EPA Entrustment Level	0[a],1,2,3,4	Attending assessment of resident entrustment—e.g Level 0: Resident can only observe task, not perform, Level 4: Resident can supervise others performing task
LDA Topic	A,B,C,D	Topics (patterns of word distributions) found in the data by the LDA algorithm. Letters randomly assigned by algorithm
Gamma	0–1	Relationship between a single EPA Entrustment Level and a single LDA Topic—e.g. a Gamma of 1 means that EPA level consists 100% of that Topic
Beta	0–1	Relationship between a single word and a single LDA topic—the higher the Beta, the more likely a word is to be found in a given topic

[a] Note: Level 0 was so infrequently assigned that it was omitted from this analysis.



	1-Direct Supervision	2-Indirect Supervision	3-Unsupervised Practice	4-Supervising Others
LDA Topic A	0.0000111	0.00000555	0.00000574	1
LDA Topic B	0.0000111	0.00000555	1	0.00000978
LDA Topic C	1	0.00000555	0.00000574	0.00000978
LDA Topic D	0.0000111	1	0.00000574	0.00000978
	(Less Entrustment)	→	→	(More Entrustment)

Fig. 1. EPA Entrustment Levels and Corresponding LDA Topics: All EPAs CombinedLegend
Legends:
Numeric values – gamma (document-topic probability). A gamma of 1 means the that topic constitutes 100% of that EPA level, a gamma of 0 means 0%.
Highlights – topic representing the largest proportion of the associated EPA entrustment level
LDA Topic – latent Dirichlet allocation topic.



	1-Direct Supervision	2-Indirect Supervision	3-Unsupervised Practice	4-Supervising Others
LDA Topic A	1	0.000032	0.0000383	0.000049
LDA Topic B	0.0000255	1	0.0000383	0.000049
LDA Topic C	0.0000255	0.000032	0.0000383	1
LDA Topic D	0.0000255	0.000032	1	0.000049
	(Less Entrustment)	→	→	(More Entrustment)

Topic Word Lists	1-Direct Supervision	2-Indirect Supervision	3-Unsupervised Practice	4-Supervising Others
1 (Stronger Association)	🟨 dissection	🟨 dissection	🟨 dissection	patient
2	gallbladder	difficult	job	job
3	🟥 view	job	level	🟩 independent
4	🟥 critical	planes	plan	🟩 resident
5	🟥 hand	🟥 tissue	🟩 independent	gallbladder
6	difficult	🟥 hand	patient	🟨 safe
7	🟥 hands	🟥 hands	cholecystectomy	cholecystectomy
8	job	🟥 view	difficult	care
9	nice	🟥 critical	🟨 safely	🟩 junior
10 (Weaker Association)	🟥 tension	🟥 left	gallbladder	🟨 dissection

Fig. 2. Gallbladder Disease: EPA Entrustment Levels and Corresponding LDA Topics
Legend:
The top of this figure displays the gamma values for each LDA-Topic:EPA Level pair. A gamma of 1 means the that topic constitutes 100% of that EPA level. Highlighted numbers represent the strongest LDA-Topic:EPA Level pair associations.
The bottom of this figure displays the top ten words associated with the above LDA-Topic:EPA Level pairs. Words are listed in order of decreasing frequency (the words with the strongest association to a topic are on top of the list). Colored boxes are used to represent manual researcher interpretations of words' entrustment levels: red (low entrustment), yellow (intermediate), green (high entrustment). The word lists are located directly underneath their associated entrustment level. Note the progression from red to green boxes as you move up entrustment levels. . (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

words less than three letters long. Duplicate words were not removed. Assessments with an entrustment level of 0 were removed due to low frequency ($n = 8$ out of 1015). The topicmodels package was used to perform the LDA analysis using the standard variational expectation-maximization (VEM) algorithm with k topics set at 4 to match the number of EPA entrustment levels ("1", "2", "3", "4") in the data.[4] Alpha, the parameter of the Dirichlet prior for the per-document topic distribution, was assigned a starting value per the topicmodels package default of $50/k$. This analysis was performed using 'unigram' (single word) items. The analysis was repeated using a combined unigram + bigram (two word item) analysis to help capture the impact of negative modifiers such as 'not ready' or 'not competent'. The addition of bigrams did not significantly impact the results and were omitted from this manuscript for simplicity. The mapping of LDA-generated topics to EPA entrustment levels was analyzed using document-topic probabilities (gammas). This analysis was repeated for each individual EPA (Trauma, General Surgical Consultation, etc.).

The top ten words associated with each topic (calculated using word-topic probabilities [betas]) were manually reviewed to determine if the topics coherently corresponded to the entrustment levels. A single surgery resident identified words that differed between entrustment levels and were associated with autonomy. These words were manually highlighted using boxes colored according to the level of entrustment they appeared to represent (low = red, intermediate = yellow, high = green) and then these assignments were reviewed by 4 other authors (surgeons and education scientists). It is important to note that this manual review was unstructured, and not systematic. There is no one way to interpret the topics created by topic modeling, and we recommend that any group attempting to do so select a method well suited to the needs of their specific project. This manual review was a way to quickly check that the computer-generated topics were understandable by humans—topics that perfectly sorted EPA assessment levels into groups but failed to provide any interpretable pattern or narrative would have limited real-world utility.

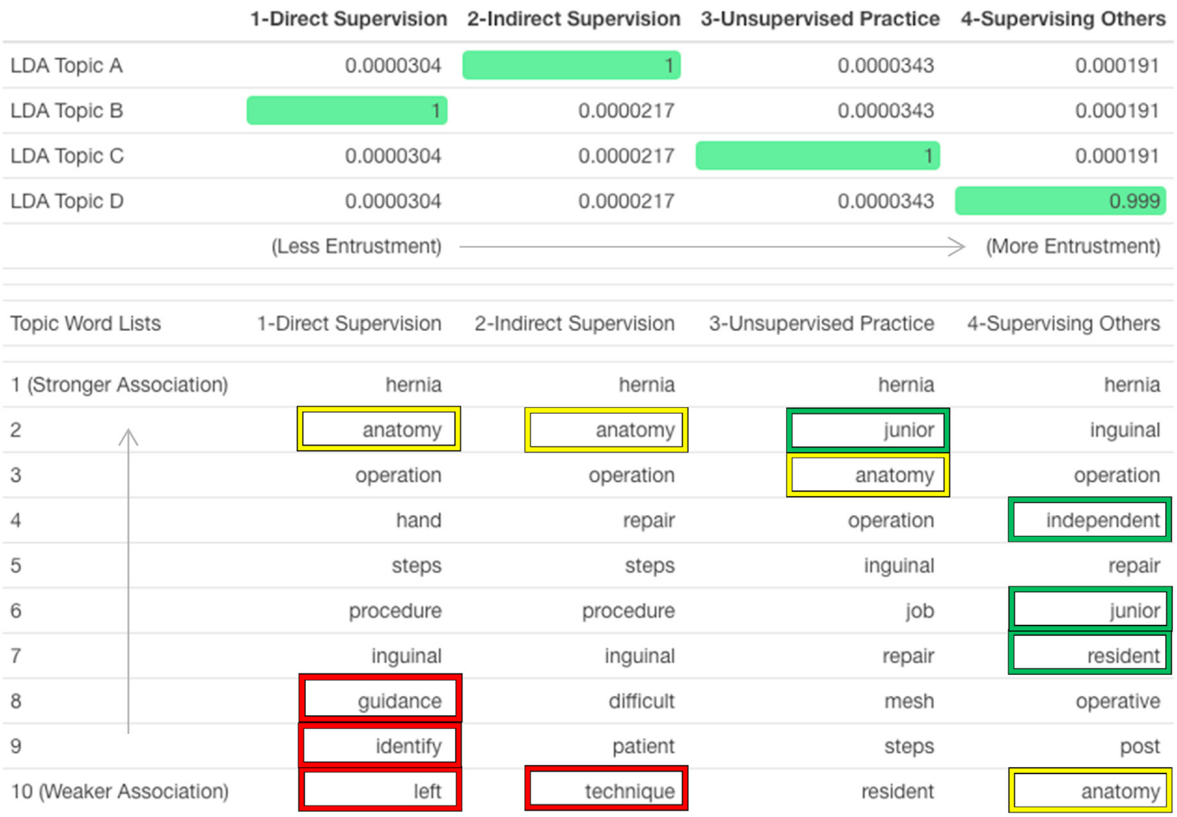
This project was reviewed by the institutional Health Sciences IRB and certified as exempt from formal review. Informed consent was waived. All data analysis and visualizations were performed using R 3.6.1 (R Foundation for Statistical Computing, Vienna, Austria).

Results

LDA-EPA correlation

1015 faculty EPA assessments were collected between July 2018 and January 2020; 97% of these assessments were associated with free text comments. These comments consisted of 32,215 words. After removal of stop words, words less than three letters in length, and level 0 assessments, 13,364 words remained. Words were well distributed across all EPAs, with each EPA containing at least 1831 words.

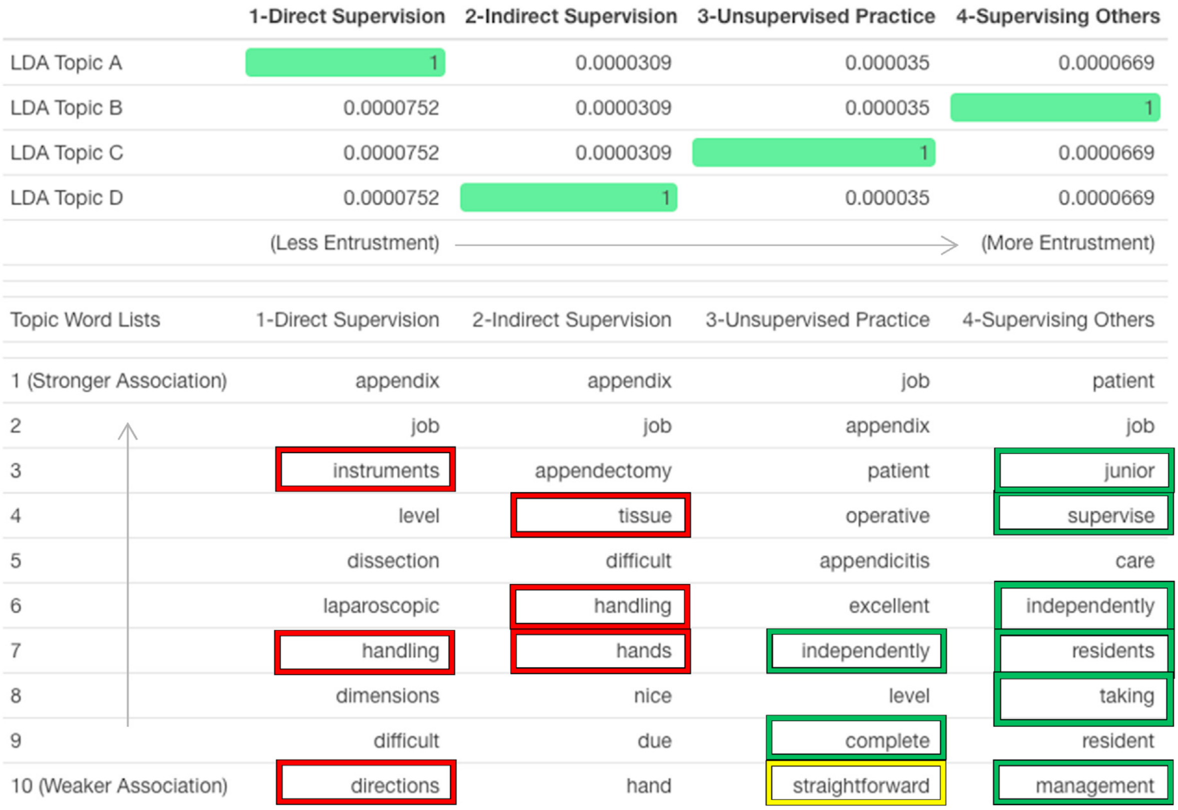
When all 5 EPAs were analyzed together as a single group, the LDA algorithm was able to identify topics that mapped 1:1 onto EPA entrustment levels 1–4 (all gammas >0.99) (Fig. 1). Unfortunately, these topics qualitatively reflected the different EPAs commonly assigned to each entrustment level, rather than differences in entrustment between levels. Therefore, the LDA analysis was repeated for each EPA individually to see if more meaningful topics would be generated. The highly effective discrimination between entrustment levels was replicated at the individual EPAs level (all gammas >0.99), and more comprehensible topics were generated (Figs. 2–6).



	1-Direct Supervision	2-Indirect Supervision	3-Unsupervised Practice	4-Supervising Others
LDA Topic A	0.0000304	🟩 1	0.0000343	0.000191
LDA Topic B	🟩 1	0.0000217	0.0000343	0.000191
LDA Topic C	0.0000304	0.0000217	🟩 1	0.000191
LDA Topic D	0.0000304	0.0000217	0.0000343	🟩 0.999
	(Less Entrustment)	→		(More Entrustment)

Topic Word Lists	1-Direct Supervision	2-Indirect Supervision	3-Unsupervised Practice	4-Supervising Others
1 (Stronger Association)	hernia	hernia	hernia	hernia
2 ↑	🟨 anatomy	🟨 anatomy	🟩 junior	inguinal
3	operation	operation	🟨 anatomy	operation
4	hand	repair	operation	🟩 independent
5	steps	steps	inguinal	repair
6	procedure	procedure	job	🟩 junior
7	inguinal	inguinal	repair	🟩 resident
8	🟥 guidance	difficult	mesh	operative
9	🟥 identify	patient	steps	post
10 (Weaker Association)	🟥 left	🟥 technique	resident	🟨 anatomy

Fig. 3. Inguinal Hernia: EPA Entrustment Levels and Corresponding LDA Topics
Legend:
The top of this figure displays the gamma values for each LDA-Topic:EPA Level pair. A gamma of 1 means the that topic constitutes 100% of that EPA level. Highlighted numbers represent the strongest LDA-Topic:EPA Level pair associations.
The bottom of this figure displays the top ten words associated with the above LDA-Topic:EPA Level pairs. Words are listed in order of decreasing frequency (the words with the strongest association to a topic are on top of the list). Colored boxes are used to represent manual researcher interpretations of words' entrustment levels: red (low entrustment), yellow (intermediate), green (high entrustment). The word lists are located directly underneath their associated entrustment level. Note the progression from red to green boxes as you move up entrustment levels. . (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



	1-Direct Supervision	2-Indirect Supervision	3-Unsupervised Practice	4-Supervising Others
LDA Topic A	1	0.0000309	0.000035	0.0000669
LDA Topic B	0.0000752	0.0000309	0.000035	1
LDA Topic C	0.0000752	0.0000309	1	0.0000669
LDA Topic D	0.0000752	1	0.000035	0.0000669
	(Less Entrustment)			(More Entrustment)

Topic Word Lists	1-Direct Supervision	2-Indirect Supervision	3-Unsupervised Practice	4-Supervising Others
1 (Stronger Association)	appendix	appendix	job	patient
2	job	job	appendix	job
3	instruments	appendectomy	patient	junior
4	level	tissue	operative	supervise
5	dissection	difficult	appendicitis	care
6	laparoscopic	handling	excellent	independently
7	handling	hands	independently	residents
8	dimensions	nice	level	taking
9	difficult	due	complete	resident
10 (Weaker Association)	directions	hand	straightforward	management

Fig. 4. Right Lower Quadrant Pain: EPA Entrustment Levels and Corresponding LDA Topics
Legend:
The top of this figure displays the gamma values for each LDA-Topic:EPA Level pair. A gamma of 1 means the that topic constitutes 100% of that EPA level. Highlighted numbers represent the strongest LDA-Topic:EPA Level pair associations.
The bottom of this figure displays the top ten words associated with the above LDA-Topic:EPA Level pairs. Words are listed in order of decreasing frequency (the words with the strongest association to a topic are on top of the list). Colored boxes are used to represent manual researcher interpretations of words' entrustment levels: red (low entrustment), yellow (intermediate), green (high entrustment). The word lists are located directly underneath their associated entrustment level. Note the progression from red to green boxes as you move up entrustment levels. . (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Topic interpretation

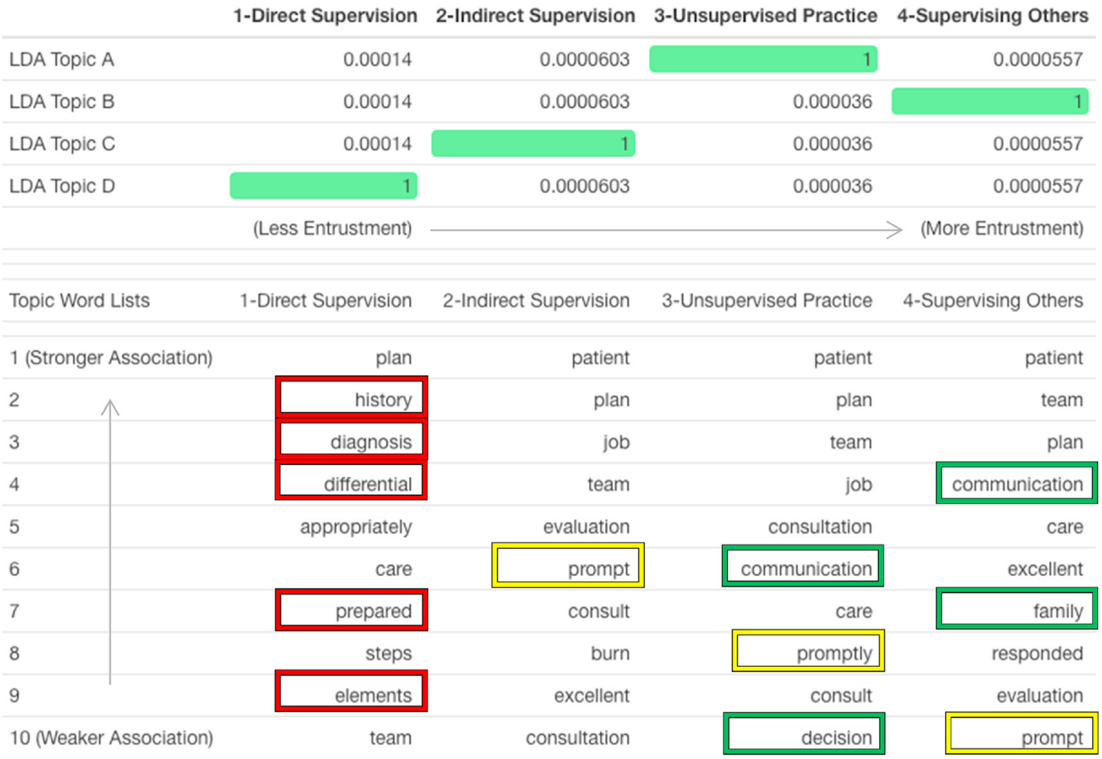
The topics uncovered for individual EPAs are shown underneath the corresponding entrustment levels (Figs. 2–6). Illustrative words were manually highlighted by a single reviewer using boxes colored according to the level of entrustment they appeared to represent (low = red, intermediate = yellow, high = green). There appears to be a logical trend from low entrustment words to higher ones for all of the individual EPAs examined (Figs. 2–6). For example, for the RLQ Pain EPA (Fig. 4) entrustment levels 1 and 2 seem to describe learning technical aspects of an appendectomy ("instruments", "hand/hands/handling", [following] "directions"), while levels 3 and 4 seem to describe independently completing cases, often while supervising junior residents ("independently", "complete", "supervise", "taking", "junior", "residents", "management"). Interestingly, level 3 appears to clarify the technical difficulty of the case ("straightforward"), which may highlight a critical difference between level 3 and 4 performance.

Similarly, the path to entrustment in the performance of Trauma EPAs may be traced from learning the fundamental evaluation ("initial", "assessment", "secondary", survey") to communicating effectively and ordering appropriate follow up studies ("communication", "imaging"), to finally performing all of the previous skills effectively and managing the trauma bay ("excellent", "calm") (Fig. 6). Similar coherent trends can be found in the remaining three EPAs evaluated (Figs. 2, 3 and 5).

Discussion

LDA is capable of identifying unique patterns of text feedback associated with different EPA entrustment levels. Topics generated using LDA map sensibly to EPA entrustment levels. This is an important proof-of-concept that also has practical applicability. First, the implementation of EPAs into surgical education is an evolving process. The current behaviors ascribed to each entrustment roadmap were created and refined by experts in surgical education based on their expert opinion.[2] While expert opinion is always a useful starting point, the EPA entrustment level—LDA topic pairs identified in this study provide data from a broad pool of educators working within an implemented EPA system on what actually differentiates the distinct entrustment levels. This provides important insight into what behaviors meaningfully correlate with trainee autonomy/entrustment and will allow for a data-driven revision of the existing entrustment descriptions and informed creation of future entrustment roadmaps.

On a broader scale, this manuscript demonstrates the potential utility of NLP strategies in surgical education. Not surprisingly, surgical educators use different vocabulary and structure in their comments on resident performance at different EPA entrustment levels. Critically, LDA was able to pick up on these important differences, and highlight key differentiating aspects of entrustment levels in ways that are interpretable to human raters. This is a promising first step in a broader integration of NLP into surgical education. As assessment frameworks that consolidate large



	1-Direct Supervision	2-Indirect Supervision	3-Unsupervised Practice	4-Supervising Others
LDA Topic A	0.00014	0.0000603	1	0.0000557
LDA Topic B	0.00014	0.0000603	0.000036	1
LDA Topic C	0.00014	1	0.000036	0.0000557
LDA Topic D	1	0.0000603	0.000036	0.0000557
	(Less Entrustment)	→	→	(More Entrustment)

Topic Word Lists	1-Direct Supervision	2-Indirect Supervision	3-Unsupervised Practice	4-Supervising Others
1 (Stronger Association)	plan	patient	patient	patient
2 ↑	🟥 history	plan	plan	team
3	🟥 diagnosis	job	team	plan
4	🟥 differential	team	job	🟩 communication
5	appropriately	evaluation	consultation	care
6	care	🟨 prompt	🟩 communication	excellent
7	🟥 prepared	consult	care	🟩 family
8	steps	burn	🟨 promptly	responded
9	🟥 elements	excellent	consult	evaluation
10 (Weaker Association)	team	consultation	🟩 decision	🟨 prompt

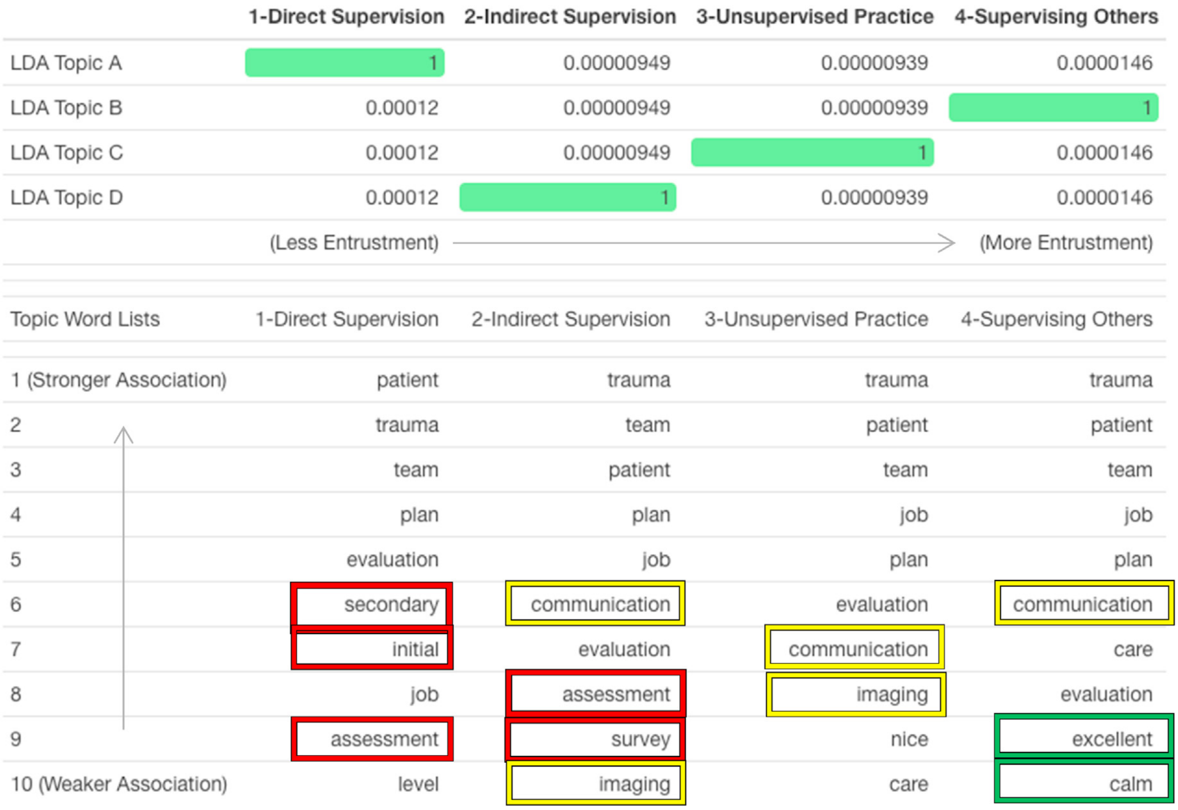
Fig. 5. Consultation: EPA Entrustment Levels and Corresponding LDA Topics
Legend:
The top of this figure displays the gamma values for each LDA-Topic:EPA Level pair. A gamma of 1 means the that topic constitutes 100% of that EPA level. Highlighted numbers represent the strongest LDA-Topic:EPA Level pair associations.
The bottom of this figure displays the top ten words associated with the above LDA-Topic:EPA Level pairs. Words are listed in order of decreasing frequency (the words with the strongest association to a topic are on top of the list). Colored boxes are used to represent manual researcher interpretations of words' entrustment levels: red (low entrustment), yellow (intermediate), green (high entrustment). The word lists are located directly underneath their associated entrustment level. Note the progression from red to green boxes as you move up entrustment levels. . (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

volumes of text-based data on resident performance are developed, strategies to interpret this data at scale are needed. NLP can help surgical education scale with its data. For example, extraction-based automated document summarization can highlight the most important comments within a corpus of narrative feedback provided to a resident—allowing for rapid review via CCC committee.[10,11] Predictive modeling techniques, such as supervised LDA (sLDA) can be trained using a body of text assessments with an associated response (e.g. EPA assessments with a paired entrustment score) with the goal of identifying latent topics *predictive of the response* (score) instead of the topics most adept at classifying the assessments into groups. Then an unlabeled assessment's score can be estimated using the model. Such a system would be able to provide a summative interpretation of a large body of text data for any individual resident, potentially assisting CCC decision making. Or one could imagine a system in which educators can focus only on providing formative free text feedback to their residents, while the job of summatively assessing this each EPA comment for a 'score' could be left to the computer.

Despite the potential promise for NLP in surgical education, this study has several limitations. Foremost, this was a single center study with a single implementation of EPA assessments, limiting generalizability. We hope to collaborate with other centers that incorporate free text data into their EPA assessments in the future to see how patterns in feedback might vary based on location or implementation strategy. This study also only assesses the first 5 EPAs created for general surgery, which cover only a fraction of the total knowledge of surgery which residents must accumulate by graduation. NLP techniques cannot make up for low quality text data. Evaluators will still need to provide high-quality text feedback to residents for the algorithms to analyze. Finally, current topic modeling techniques are still limited to the 'bag of words' model, requiring human interpretation of resulting topics. Different researchers or readers of this manuscript may disagree with our interpretations. We encourage this—our data and interpretations are meant to start discussions about entrustment and autonomy in EPAs, not function as a final authoritative arbiter. Eventually, computers may be able to distinguish semantics (bag of concepts) or pragmatics (bag of narratives) from natural language data to provide more useful input, but we see no reason to delay these important discussions until that is possible.

Conclusions

Topic modeling using latent Dirichlet allocation is capable of discriminating between EPA entrustment levels. Topics generated by LDA map coherently to entrustment levels, providing insight on how surgeon educators describe trainee autonomy and grant entrustment. This data can be used to inform the creation of future EPA entrustment roadmaps, and data-driven revision of existing ones. Further development of NLP methodologies in surgical education may allow surgeon educators to analyze large amounts of text assessment data in a scalable fashion.



	1-Direct Supervision	2-Indirect Supervision	3-Unsupervised Practice	4-Supervising Others
LDA Topic A	1	0.00000949	0.00000939	0.0000146
LDA Topic B	0.00012	0.00000949	0.00000939	1
LDA Topic C	0.00012	0.00000949	1	0.0000146
LDA Topic D	0.00012	1	0.00000939	0.0000146
	(Less Entrustment)	→	→	(More Entrustment)

Topic Word Lists	1-Direct Supervision	2-Indirect Supervision	3-Unsupervised Practice	4-Supervising Others
1 (Stronger Association)	patient	trauma	trauma	trauma
2	trauma	team	patient	patient
3	team	patient	team	team
4	plan	plan	job	job
5	evaluation	job	plan	plan
6	secondary	communication	evaluation	communication
7	initial	evaluation	communication	care
8	job	assessment	imaging	evaluation
9	assessment	survey	nice	excellent
10 (Weaker Association)	level	imaging	care	calm

Fig. 6. Trauma Evaluation: EPA Entrustment Levels and Corresponding LDA Topics
Legend:
The top of this figure displays the gamma values for each LDA-Topic:EPA Level pair. A gamma of 1 means the that topic constitutes 100% of that EPA level. Highlighted numbers represent the strongest LDA-Topic:EPA Level pair associations.
The bottom of this figure displays the top ten words associated with the above LDA-Topic:EPA Level pairs. Words are listed in order of decreasing frequency (the words with the strongest association to a topic are on top of the list). Colored boxes are used to represent manual researcher interpretations of words' entrustment levels: red (low entrustment), yellow (intermediate), green (high entrustment). The word lists are located directly underneath their associated entrustment level. Note the progression from red to green boxes as you move up entrustment levels. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Funding sources

Grant Support: Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number T32CA090217. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Declaration of competing interest

The authors report no conflicts of interest.

References

1. Greenberg J, Minter R. Entrustable professional activities: the future of competency-based education in surgery may already Be here. *Ann Surg*. 2019;269(3):407–408. https://doi.org/10.1097/SLA.0000000000003153.
2. Brasel KJ, Klingensmith ME, Englander R, et al. Entrustable professional activities in general surgery: development and implementation. *J Surg Educ*. 2019;76(5):1174–1186. https://doi.org/10.1016/j.jsurg.2019.04.003.
3. Stahl CC, Collins E, Jung SA, et al. Implementation of entrustable professional activities into a general surgery residency. *J Surg Educ*.. February 8, 2020 https://doi.org/10.1016/j.jsurg.2020.01.012. Published online.
4. Grün B, Hornik K. Topicmodels: an R package for fitting topic models. *J Stat Software*. 2011;40(1):1–30. https://doi.org/10.18637/jss.v040.i13.
5. Gross A, Murthy D. Modeling virtual organizations with Latent Dirichlet Allocation: a case for natural language processing. *Neural Network*. 2014;58:38–49. https://doi.org/10.1016/j.neunet.2014.05.008.
6. Cambria E, White B. Jumping NLP Curves: A Review of Natural Language Processing Research. :10.
7. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res*. 2003;3(Jan):993–1022.
8. Robinson JS and D. Text mining with R. https://www.tidytextmining.com/. Accessed April 7, 2020.
9. Feinerer I, Hornik K, Meyer D. Text mining infrastructure in R. *J Stat Software*. 2008;25(1):1–54. https://doi.org/10.18637/jss.v025.i05.
10. Erkan G, Radev DR. LexRank: graph-based lexical centrality as salience in text summarization. *J Artif Intell Res*. 2004;22:457–479. https://doi.org/10.1613/jair.1523.
11. Mihalcea R, Tarau P. *TextRank: bringing order into texts. Proc 2004 Conf Empir Methods Nat Lang Process*. July 2004:404–411. Published online.