Examining the Quality of Operative Performance Feedback at Scale (36)

Autonomy, Entrustment, Feedback: Addressing feedback to learners as well as ways to determine how and when learners can act independently in the clinical environment

Kayla Marcotte
University of Michigan Medical School

Background:
Feedback is a cornerstone of medical education. Despite its importance, most studies examine feedback at a relatively small scale, making it challenging to draw generalizable conclusions. Natural language processing (NLP) has created the opportunity to examine feedback at scale. For this investigation, we used NLP to examine differences in the quality of feedback across dozens of training programs and thousands of attending surgeons.

Actions, Methods or Intervention:
Using a validated NLP model, we assessed the quality of dictated feedback recorded in the Society of Improving Medical Professional Learning's (SIMPL) smartphone-based workplace-assessment app. Our primary outcome was the probability that the feedback was "high quality" by the NLP algorithm. We used linear mixed effects models to examine how feedback quality varied across programs, attending surgeons, trainees, procedures, autonomy granted, operative performance level, case complexity, and trainee's level of clinical training.

Results:
We used an NLP model to score 28,027 evaluations with dictated feedback. Approximately 42% of the variance in the quality feedback score was attributable to the attending surgeon, signalling that some faculty provide consistently high or consistently low quality feedback. After including fixed effects, feedback quality was lower at the the highest autonomy rating ("Supervision only," $B = -7.18$, $p < .001$) and highest performance rating ("Exceptional performance," $B = -16.88$, $p < .001$). We identified a small, significant trend related to faculty improving the quality of their dictated feedback as they complete more SIMPL evaluations ($B = .009$, $p < .001$) and more dictated feedback ($B = .016$, $p < .001$).

Lessons Learned:
Examining feedback at scale provides an opportunity to identify generalizable patterns. Using over 25,000 feedback instances, we found that feedback quality is strongly influenced by who provides the feedback and trainee performance. Feedback quality also tends to increase as faculty provide more evaluations.

Future Application and Next Steps:
NLP opens the door to identifying and developing actionable items for improving feedback.

Additional Authors:
Erkin Otles, Angela E Thelen, Rebecca Gates, Brian C George, Andrew Krumm