Natural Language Processing for Reflection and Feedback Quality in a Nationwide Otolaryngologic Workplace-Based Assessment Database

Presenting Author: Jeng-Wen Chen[1]

Co-Authors: Hai-Lun Tu[2], Ke-Hsin Chueh[2], Wei-Chung Hsu[3], Pa-Chun Wang[4], Chi-Chun Chou[5]

1. Deputy Superintendent at Cardinal Tien Hospital and Associate Professor at Fu Jen Catholic University, New Taipei City, Taiwan 2. Professor at Fu Jen Catholic University, New Taipei City, Taiwan
3. Professor at National Taiwan University Hospital, Taipei, Taiwan 4. Professor at Cathay General Hospital, Taipei, Taiwan 5. Superintendent at Cardinal Tien Hospital, New Taipei City, Taiwan

Abstract

Competency-based medical education (CBME) aims to train specialists in a way aligned with societal expectations. The Taiwan Society of Otolaryngology-Head and Neck Surgery (TSO-HNS), collaborating with the Joint Commission of Taiwan, launched the EMYWAY platform in 2021 and commenced nationwide Entrustable Professional Activities (EPAs) assessment in 2022. Over three years, the platform amassed over 37,000 EPA records. Specifically, from July 1, 2021 (pilot phase) to August 21, 2025, the platform accumulated 37,328 formative EPA assessments from 35 training programs, involving over 410 clinical teachers and 292 residents.

Introduction

Competency-Based Medical Education (CBME) aims to foster specialists who meet societal expectations. In Taiwan, the Taiwan Society of Otolaryngology-Head and Neck Surgery (TSO-HNS), in collaboration with the Joint Commission of Taiwan, initiated the EMYWAY platform in 2021 to integrate CBME frameworks, enabling digitally traceable workplace-based assessments (WBAs). Nationwide Entrustable Professional Activities (EPAs) assessment officially commenced in 2022. Over three years, the EMYWAY platform has accumulated a substantial dataset of over 37,000 formative EPA assessments from 35 training programs, involving over 410 clinical teachers and 292 residents. This rich dataset, with its structured digital format for EPA-based interactions, is crucial for advanced Natural Language Processing (NLP) analysis. This study aims to evaluate the effectiveness of deep learning NLP algorithms, specifically Bidirectional Encoder Representations from Transformers (BERT), in assessing the quality of resident reflections and faculty feedback within this nationwide WBA database, which is vital for enhancing feedback quality in CBME.

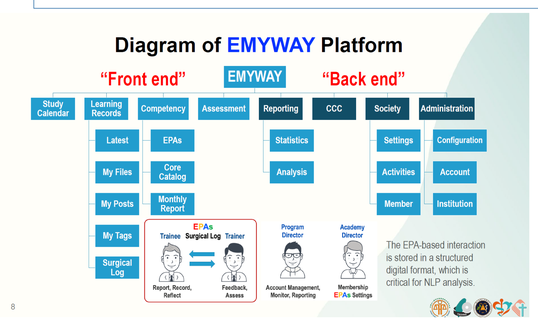

Figure 1. EMYWAY platform

Scan me!

Methods and Materials

This study analyzed 300 randomly selected narrative entries from the EMYWAY EPA database, focusing on workplace context, EPA titles, clinical diagnoses, and the narrative content of reflections and feedback. Two independent medical education experts performed expert coding to classify the quality of reflections and feedback based on their relevance, specificity, and the presence of reflective or improvement suggestions. Discrepancies were resolved through consensus or by a third reviewer. The coded data were then utilized for deep learning NLP. The quality of reflections and feedback was classified into four levels ('effective,' 'moderate,' 'ineffective,' or 'irrelevant') and subsequently into a binary classification of 'high-quality' and 'low-quality'. Three machine learning algorithms were compared: Logistic Regression (LR), Support Vector Machine (SVM), and BERT. A multilingual pre-trained BERT model was employed to effectively handle the mixed medical and everyday language prevalent in the narratives. The dataset was divided into a 240-entry training set and a 60-entry validation set.

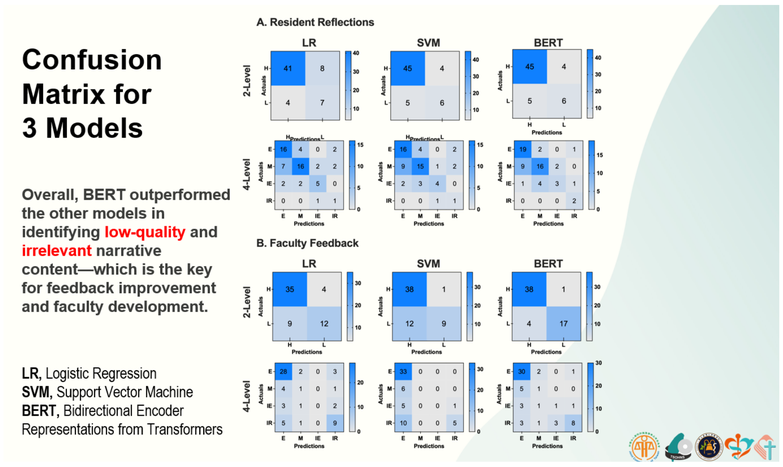

Figure 2. Confusion Matrix for 3 Models

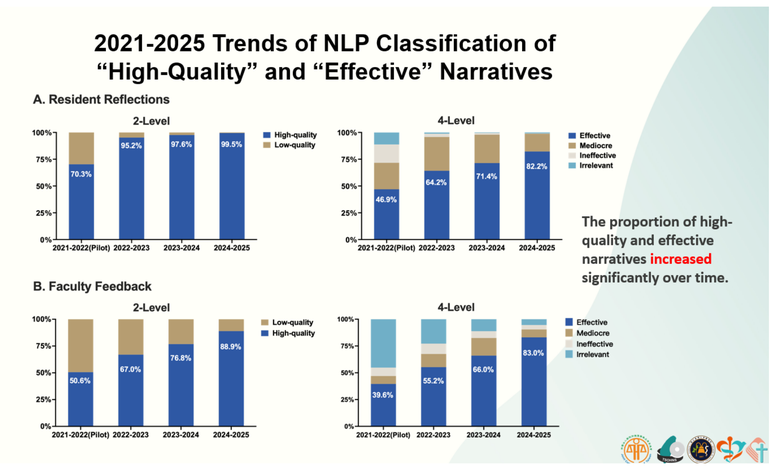

Figure 3. NLP Classification of "High-Quality" and "Effective" Narratives

Table 1. Key Performance Metrics (Validation Set)

Model	Classification	Accuracy	Precision	Recall	F1-Score
Resident Reflections	2-Level (High/Low Quality)	85%	85%	85%	85%
	4-Level	67%	67%	67%	65%
Faculty Feedback	2-Level (High/Low Quality)	92%	92%	92%	92%
	4-Level	67%	67%	67%	67%

Results

The BERT algorithm consistently demonstrated superior performance compared to LR and SVM across both resident reflections and faculty feedback narratives, particularly in binary classification.

- For resident reflection quality, the two-level BERT model achieved 85% accuracy. Previous study reported 90% accuracy with 67% precision, 50% sensitivity, and 96% specificity in a two-level model.
- For faculty feedback quality, the two-level BERT model achieved 92% accuracy. Previous study reported 80% accuracy with 100% precision and 60% sensitivity.
- In the four-level classification, BERT achieved 67% accuracy for both resident reflections and faculty feedback.
- BERT was particularly effective in identifying low-quality and irrelevant narrative content, which is crucial for targeted feedback improvement and faculty development.
- Longitudinal analysis from 2021 to 2025 indicated a significant increase in the proportion of high-quality and effective narratives over time.

Conclusions

This study demonstrates that deep learning NLP, especially the BERT algorithm, is highly effective in evaluating the quality of residents' reflections and faculty feedback within a nationwide WBA database. The results are comparable to traditional machine learning models even with a limited dataset. NLP provides a faster, scalable, and consistent approach to narrative evaluation compared to manual review, which, while capturing contextual nuances, is time-consuming and labor-intensive. The observed improvement in narrative quality over time further suggests the value of structured EPA frameworks and digital platforms in CBME implementation. While not a replacement for expert rating, NLP serves as a valuable adjunct to enhance CBME assessment systems and support faculty development. Future research should focus on cross-specialty and cross-cultural applications, as well as multimodal integration.

References

1. Chen, et al. *J Taiwan Otolaryngol Head Neck Surg* 2022;57:110-122.
2. Otles E, Kendrick DE, Solano QP, et al. *Acad Med* 2021;96(10):1457-1460. doi:10.1097/ACM.0000000000004153
3. Solano QP, Hayward L, Chopra Z, et al. J Surg Educ. 2021;78(6):e72-e77. doi:10.1016/j.jsurg.2021.05.012

Contact

Jeng-Wen Chen, MD, MSc
Cardinal Tien Hospital and Fu Jen Catholic University
086365@mail.fju.edu.tw