THE FUTURE OF NATURAL LANGUAGE PROCESSING (NLP) IN MEDICAL EDUCATION ASSESSMENT

NBME

INTRODUCTION

Medical education is ever evolving, and medical education assessment must evolve with it to prepare each generation of physicians not only with scientific knowledge, but also with the skills and behaviors to address a wide range of patient needs. NBME is committed to improving and developing assessments that evaluate skills and competencies, as well as researching innovative technologies that may improve the accuracy, efficiency and effectiveness of assessment practices in medical education.

These highlights from recent articles published by NBME researchers discuss their progress utilizing Natural Language Processing (NLP) and machine learning. These capabilities have the potential to transform assessment development and scoring, especially for more complex behaviors such as clinical reasoning.

What is NLP?

NLP is a subfield of Artificial Intelligence that explores the processing of natural language (e.g., English, French, Chinese) by computer systems. Despite being a relatively young field that traces its roots back to the 1960's, it has seen tremendous advances over the past decade largely thanks to the availability of vast quantities of digital text data and computational advances that allow its processing.

In medicine, clinical NLP has spearheaded the discovery of patterns within entire Electronic Health Record databases; however, data confidentiality and the cost of the required medical expertise of annotation to develop the system slows the pace of development compared to other areas.

Uses of NLP in assessment include automated scoring of free-text or spoken responses that are challenging to score versus multiple-choice questions. Other benefits include:

- capturing of new types of data, which provides unique evidence of specific skills and behaviors, and leads to new assessment methods such as computerized simulations of patient encounters

- facilitating student learning and growth by providing more timely feedback that can supplement educators' observations

> "
> Working in the **intersection between clinical NLP and assessment**, we are facing specific challenges such as lack of shared datasets, the need to develop highly specialized annotation and evaluation procedures, and the need for high accuracy of the NLP output, among many others. Meeting all these requirements at once necessitates **creative approaches and strategic identification** of areas where NLP research **can make the biggest difference.**"
>
> **Victoria Yaneva,**
> Manager of NLP Research at NBME

NBME

Identifying gaps to improve clinical reasoning

Researchers have explored the use of NLP in the assessment of clinical reasoning, a complex competency that poses unique measurement challenges. Using NLP captures novel aspects of learners' clinical reasoning abilities that cannot be obtained through traditional methods such as multiple-choice questions.

NBME and the University of Pennsylvania collaborated to develop an NLP system, INCITE, to detect the presence of clinical reasoning concepts in written text and enable meaningful feedback. This approach may allow educators to provide better targeted instruction that enhances student learning. For example, an NLP system could help identify learners who are missing critical aspects of taking a thorough patient history, a key competency of clinical reasoning. Missing these aspects may result in making a medical error. If a student is consistently under-reporting information from the patient's history, this may be evidence that additional instruction is needed to ensure sound clinical reasoning.

FOR FURTHER READING

Runyon, C. R., Harik, P., & Barone, M. A. (2022). "Cephalgia" or "migraine"? Solving the headache of assessing clinical reasoning using natural language processing. *Diagnosis*.

READ THE STUDY →

Pinpointing the "why" of scores

The INCITE system bases scores on what clinical concepts are detected, which allows the system to clearly pinpoint why an examinee got a specific score (traceability). In a formative setting, this "why" could provide educators with specific information about how to best upskill or provide remediation for lower-scoring performers.

The system also serves as a case study for the construction of new NLP-based scoring systems for patient notes, providing valuable insights for future development in this domain. The inclusion of specific findings from the patient encounter increases an examinee's score, and the score increase is concept-dependent; not all concepts are scored equal. The case study's emphasis on transparency in scoring provides valuable insights for educators and assessment professionals looking to implement NLP-based scoring systems in their own assessments.

FOR FURTHER READING

Harik, P., Mee, J., Runyon, C., & Clauser, B. E. (2023). Assessment of clinical skills: a case study in constructing an NLP-based scoring system for patient notes. In *Advancing Natural Language Processing in Educational Assessment* (pp. 58–73). Routledge.

READ THE CHAPTER →

NBME

RECENT FINDINGS

NLP in clinical skills assessment

A comprehensive dataset created from the USMLE Step 2 Clinical Skills patient note corpus provides an abundance of clinical patient notes, allowing researchers to explore the application of NLP models in clinical skills assessment. An important part of the dataset contains information on the myriad of ways that important clinical concepts can be expressed. Providing the content specific-data needed to build and train an NLP system opens avenues for future research and innovation in the field.

The dataset, together with expert annotation of important concepts, is available to request for research purposes via NBME's Data Sharing Portal.
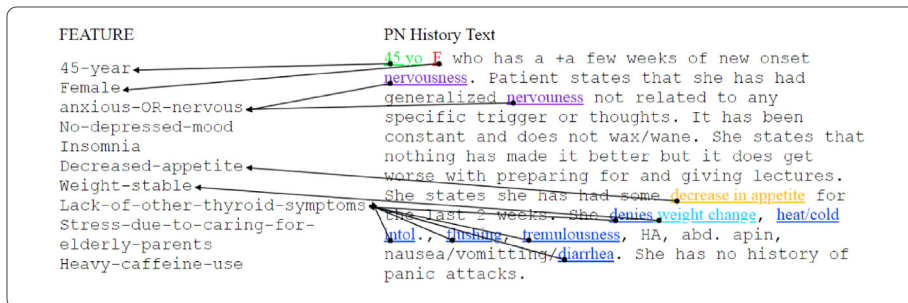
Figure 1: Key concepts from an exam rubric (left) and how these have been expressed in an examinee-written patient note (right). The more such concepts an examinee has covered in their patient note, the higher their score. NLP systems facilitate scoring by detecting these phrases automatically and mapping them to their corresponding concept from the rubric.

IN SUMMARY

> It's an exciting time to be a part of medical education assessment. For a long time, researchers have discussed the **importance of clinical reasoning and its assessment** but were bound by technological and methodological constraints. NLP mitigates some of these constraints and will **allow for novel and innovative assessments** to be developed."
>
> **Christopher Runyon,**
> Senior Measurement Scientist at NBME

The integration of NLP and machine learning technologies in medical education assessment holds tremendous promise. By embracing these innovative technologies, medical educators and assessment professionals can take advantage of new practices. These new approaches may lead to more accurate, efficient and meaningful evaluations of learners' knowledge, skills and behaviors, from providing more timely feedback to improving assessment of clinical reasoning.

NBME strives to advance the field of educational assessment by fostering innovation and promoting the adoption of novel approaches. Our research highlights the transformative potential of NLP and other emerging technologies to revolutionize assessment practices. Only by continuous evolution of our practices can we better assess and prepare students to meet the needs of their patients.