

INTERNATIONAL
SYMPOSIUM
ON MEASUREMENT,
SELECTION AND
PLACEMENT

4-6 OCTOBER 2024 ANKARA

Venue: Bilkent Hotel and Conference Center

ULUSLARARASI
**ÖLÇME,
SEÇME VE
YERLEŞTİRME
SEMPOZYUMU**

4-6 EKİM 2024 ANKARA

Yer: Bilkent Otel ve Kongre Merkezi

SEMPOZYUM PROGRAM KİTABI

SYMPOSIUM PROGRAM BOOK

#ISMSP

#UOSYS



Klinik Akıl Yürütme Becerisinin Değerlendirmesinde Otomatik Soru Üretim Yöntemlerinin Karşılaştırılması

Betül Nalan Karahan¹, Emre Emekli²

Tıp fakültesi lisans eğitiminde klinik akıl yürütme (KAY) becerisinin kazandırılması gerekmektedir ve klinik öncesi dönemde KAY değerlendirmesi çoktan seçmeli sorular (ÇSS) ile yapılmaktadır (Daniel ve ark.,2019). ÇSS, yapay zeka (YZ) ile üretilen şablon tabanlı otomatik soru üretimi (ŞT-OSÜ) ve yazılımlarla üretilen şablon tabanlı olmayan OSÜ (ŞTO-OSÜ) olmak üzere iki yöntemle otomatik üretilebilir (Gierl ve ark.,2013). Bu çalışmanın amacı ŞT-OSÜ, ŞTO-OSÜ ile üretilen ve halihazırda kullanılan sorular hakkında uzman görüşü olarak, soruların KAY değerlendirilmesinde kullanılabilirliğini araştırmaktır. Abdominal aciller hakkında soru üretilmesi kararlaştırıldı. ŞT-OSÜ (içeriğin belirlenmesi, soru şablonu oluşturulması, yazılım ile soru üretilmesi) ile yaş, cinsiyet, semptom, öykü, laboratuvar, fizik muayene başlıkları belirlendi ve olası değişkenler yazıldı. Soru şablonu oluşturularak 7260 soru üretildi. ŞTO-OSÜ için ChatGPT-4o ile 17 soru üretildi (Kıyak&Emekli,2024). Rastgele altışar soru (18) seçilerek form ile değerlendirildi (a-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, e-Çeldiriciler mantıklıdır, f-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, g-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, h-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, i-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, j-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, k-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, l-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, m-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, n-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, o-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, p-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, q-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, r-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, s-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, t-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, u-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, v-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, w-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, x-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, y-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir, z-Soruda verilen bilgiler, doğru cevabı bulmak için yeterlidir). Formu 14 kişi (radyoloji hekimi) cevapladı. Katılımcılar soruların KAY ölçtüğünü ifade etmekteydi (ŞT-OSÜ 66/84 (%78,57), güncel sorular 69/84 (%82,14), ŞTO-OSÜ (%84,52)). Evet cevabı ortalamaları ŞT-OSÜ, güncel sorular, ŞTO-OSÜ için sırasıyla "a" 11,83-11,67-13; "b" 8,33-9,17-11,5; "c" 7,17-7,67-9; "d" 8-9,33-10,17; "e" 6,67-8,67-10,67; "f" 2,83-1-0,5; "g" 6,17-6,5-8,17 şeklindeydi. Üç soru üretme yöntemi ile de KAY ölçebilecek sorular üretilebilmektedir. Soru metninin anlaşılabilirliği, klinik açıdan uygunluğu ve doğru cevabı bulmak için verilen bilgileri ölçen başlıklar üç soru üretme tekniği için de yüksektir. ŞTO-OSÜ ile üretilen sorulara birçok parametrede daha çok evet cevabı verilmiştir. YZ ile ölçmede kullanılacak soru üretilebilir.

Anahtar Kelimeler: Otomatik Soru Üretimi, Klinik Akıl Yürütme, Yapay Zeka

Kaynakça

- Daniel, M., Rencic, J., & Durning, S. J., et al. (2019). Clinical reasoning assessment methods: A scoping review and practical guidance. *Academic Medicine*, 94(6), 902-912.
- Gierl, M. J., Lai, H., & Turner, S. R. (2012). Using automatic item generation to create multiple-choice test items. *Medical Education*, 46(8), 757-765.
- Kıyak, Y. S., & Emekli, E. (2024). ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: A literature review. *Postgraduate Medical Journal*, qgae065

¹ nalanksgl@gmail.com, ORCID: 0000-0003-4958-330X, Eskişehir Osmangazi Üniversitesi, Eskişehir, Türkiye

² emreemekli90@gmail.com, ORCID: 0000-0001-5989-1897, Eskişehir Osmangazi Üniversitesi, Eskişehir, Türkiye

Comparison of Automatic Question Generation Methods in Assessing Clinical Reasoning Skills

Betül Nalan Karahan¹, Emre Emekli²

Clinical reasoning (CR) skills are essential in undergraduate medical education, and in the preclinical period, CR assessment is conducted using multiple-choice questions (MCQs) (Daniel et al., 2019). MCQs can be automatically generated using two methods: template-based automatic item generation (TB-AIG) with artificial intelligence (AI) and non-template-based AIG (NTB-AIG) produced with software (Gierl et al., 2013). This study aims to investigate the usability of questions generated by TB-AIG, NTB-AIG, and currently used questions in assessing CR, based on expert opinions. For TB-AIG, headings such as age, gender, symptoms, history, laboratory, and physical examination were defined, and potential variables were written. A question template was created, and 7260 questions were generated. For NTB-AIG, 17 questions were generated using ChatGPT-4 (Kiyak & Emekli, 2024). A random selection of six questions (18 in total) was evaluated using a form, assessing the understandability, clinical appropriateness, single correct answer, sufficiency of information, logical distractors, difficulty for experts, and appropriate difficulty for students. Fourteen radiology physicians participated. They indicated that the questions assessed CR effectively: TB-AIG 66/84 (78.57%), current questions 69/84 (82.14%), NTB-AIG (84.52%). The average "Yes" responses for TB-AIG, current questions, and NTB-AIG were as follows: "a" 11.83-11.67-13; "b" 8.33-9.17-11.5; "c" 7.17-7.67-9; "d" 8-9.33-10.17; "e" 6.67-8.67-10.67; "f" 2.83-1-0.5; "g" 6.17-6.5-8.17. Questions capable of measuring CR can be produced using all three question generation methods. The headings measuring the understandability of the question text, clinical appropriateness, and the sufficiency of the information provided to find the correct answer are high for all three question generation techniques. Questions generated by NTB-AIG received more "Yes" responses in many parameters. Questions that can be used in assessments with AI can be produced.

Keywords: Automatic Question Generation, Clinical Reasoning, Artificial Intelligence

Reference

- Daniel, M., Rencic, J., & Durning, S. J., et al. (2019). Clinical reasoning assessment methods: A scoping review and practical guidance. *Academic Medicine*, 94(6), 902-912.
- Gierl, M. J., Lai, H., & Turner, S. R. (2012). Using automatic item generation to create multiple-choice test items. *Medical Education*, 46(8), 757-765.
- Kiyak, Y. S., & Emekli, E. (2024). ChatGPT prompts for generating multiple-choice questions in medical education and evidence on their validity: A literature review. *Postgraduate Medical Journal*, qgae065

¹ nalanksgl@gmail.com, ORCID: 0000-0003-4958-330X, Eskişehir Osmangazi Üniversitesi, Eskişehir, Türkiye

² emreemekli90@gmail.com, ORCID: 0000-0001-5989-1897, Eskişehir Osmangazi Üniversitesi, Eskişehir, Türkiye