Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220583

Software Architecture for Automated Assessment of Prescription Writing

Alireza KHATAMI[a], Anne HOLBROOK[b], Anthony J. LEVINSON[c], Karim KESHAVJEE[a, 1]

[a] *Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada*

[b] *Division of Clinical Pharmacology & Toxicology, McMaster University, Hamilton, Canada*

[c] *Division of e-learning Innovation, McMaster University, Hamilton, Canada*

Abstract. Prescribing skills are a crucial competency in medical practice considering the increasing numbers of medications available and the increasingly complex patients with multiple diseases faced in clinical practice. Medical students need to become proficient in these skills during training, as required by medical licensing colleges. Not only is teaching the fundamentals of safe and cost-effective prescribing to medical students challenging but evaluating their prescribing skills by faculty members is difficult and time consuming. The COVID-19 pandemic has accelerated the interest in clinically relevant online exams, including automated assessment of short answer style questions. The goal of this project was to design a software to automate the assessment of learners' prescriptions written during low stakes formative assessments. After establishing the components of a legal prescription with multiple medications, and identifying the sources of errors in prescribing and prescribing assessment, we designed and validated an architecture and developed a prototype for automated parsing of learner prescriptions.

Keywords. Online assessment, prescription marking, software architecture, prototype, prescribing skills

1. Introduction

Teaching medical students the art and science of safe and cost-effective prescribing is challenging yet essential for any high quality, sustainable health care system [1]. Aging patients with multimorbidities means that polypharmacy is common and necessary. An increasingly large armamentarium of effective medications means that medical students need to learn how to prescribe many more medications, frequently in combination with other medications. The stakes can be quite high. Increasingly fragile patients end up being hospitalized from iatrogenic causes. More than 700 million prescriptions are written annually in Canada for drugs from approximately 1100 therapeutic groups [2]. Medication errors are common, usually go unrecognized and can pose a serious patient safety hazard [3-7]. Medical students and residents are particularly at risk, with a 7-10% error rate in their prescriptions [4].

[1] Corresponding Author, Dr. Karim Keshavjee, Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada; E-mail: karim@infoclin.ca

In 2020, the Medical Council of Canada added Prescribing Skills to its specific list of mandatory training objectives for medical schools. However, the number of clinical pharmacology and toxicology specialists who can teach prescribing skills effectively is very small in Canada, making it necessary to maximize use of online resources and make better tools available for teaching prescribing skills.

Multiple choice questions (MCQs) with single best answer are the dominant testing method for most university health professions programs. However, MCQs have well-known limitations, including cueing effect, testing prompted memory rather than ability to generate an answer, and difficulty representing many important areas of medicine [8]. Recently, computer-readable very short answer (VSA) question responses have been validated as a novel pragmatic innovation with excellent results. VSAs better represent actual processes in clinical practice and test actual case-based knowledge and skills with better discrimination than MCQs with single best answer [8,9].

Marking prescriptions and providing high quality feedback to students is time consuming, labor intensive and requires expertise. There is a need for an automated solution that can assist educators to provide high quality feedback to trainees at different stages of training in a timely manner [10]. An automated prescription marker could also create opportunities for more frequent formative assessments and feedback cycles, helping improve prescribing skills faster. This approach has been attempted with some success in the UK and in Australia, but both projects are proprietary, to our knowledge.

Health professions training has relied on in-person, patient-based, preceptor-supervised learning in hospitals and clinics for centuries. However, over the past couple of years, the COVID-19 pandemic has led to the sudden loss of this in-person learning and accelerated the necessity for online curricula, low stakes assessments and high stakes exams. Recognized as one of the most advanced in the world, Canada's medical training system will depend increasingly on high quality eLearning and online evaluation opportunities with appropriate technology and the best faculty trained in delivering education in virtual environments over the next several decades.

2. Methods

Requirements for a Prescription Parser-Marker software were obtained through review of the literature and expert panel consensus. This included parsing of legal prescriptions into their component parts and creating a standardized prescription. A draft software architecture was developed (Figure 1). The Parser-Marker software was designed to handle a variety of question sets with a Marking File that had standard features but was customized for each new question set. This would enable the Parser-Marker to be used for multiple question sets without reprogramming.

A prototype was built using the RapidMiner® (RapidMiner GmbH, Dortmund, Germany) software platform version 9.1 using the built-in Regex functions for text matching. The prototype was tested on responses to 5 clinical scenarios (12 total prescriptions) developed with pre-set marking for alternatives, by an experienced prescribing competence expert (AH). Test scenarios were completed by 10 medical learners using Examplify examination software from ExamSoft (ExamSoft, Dallas, TX, USA) (https://examsoft.com). The prescriptions were typed in manually by students with no prompts or direction about how the prescription should be written. A Marking File was developed using the pre-set marking criteria. Student generated prescriptions were scored by the Parser-Marker software. The prototype-generated scores (N=840 = 10 learners * 12 prescriptions * 7 prescription elements) were examined by two co-authors (KK and AK) to identify issues and reasons why the scores were incorrect.

A root cause analysis was used to identify the reasons for incorrect marking of the prescriptions (Figure 2). Using this information, we redesigned the software architecture to address the issues identified and redeveloped the software to address the issues that lead to incorrect assessments using an automated approach (Figure 3).

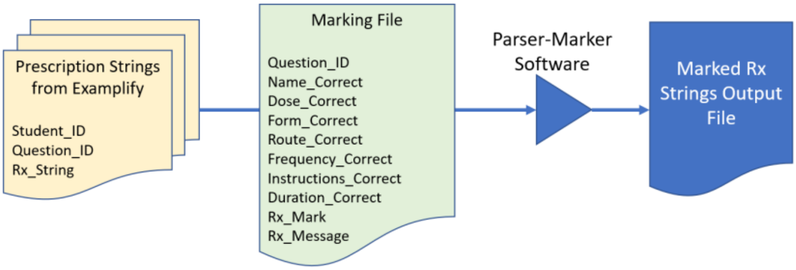

Figure 1. Draft software architecture for prescription marker

3. Root cause analysis of failure of automated prescribing assessment

Several factors affected the performance of the Prescription Parser-Marker. They are grouped into the following factors: student-related, exam software-related, question-related, response-related, marking file-related and software (parsing and matching)-related factors (Figure 2).

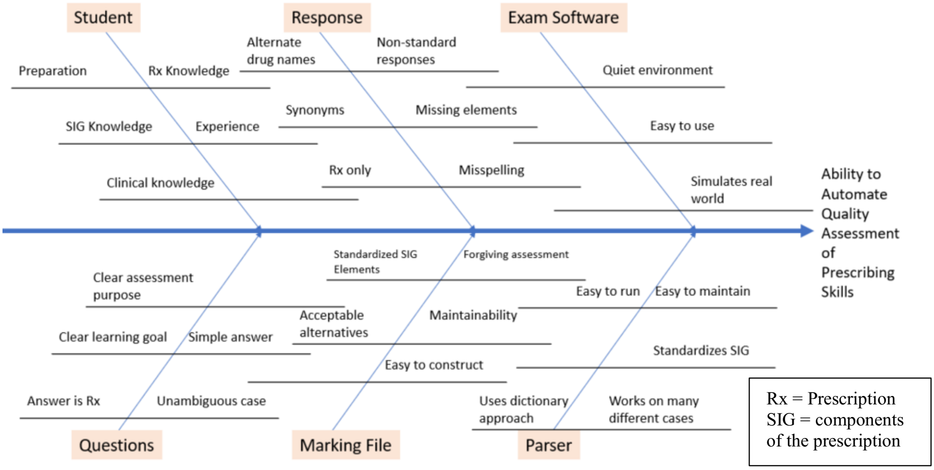

Figure 2. Fishbone diagram of root cause analysis for errors in prescribing assessment

4. Software requirements derived from root cause analysis

The root cause analysis made it clear that variability in the prescription strings (e.g., synonyms for drug names, synonyms for frequency of dosing) was the key issue creating problems with correct marking and accurate scores. Misspellings and non-responses (e.g., to questions which required multiple prescriptions) were other sources of problems.

Variability in students' completion of the prescription elements was a key factor in explaining poor scores on the exam; e.g., it is common for prescribers to leave out certain elements because they are assumed, such as form (e.g., tablet) or route of administration (usually oral route). These are either 'teachable moments' for feedback, or the exam instructions need to be more specific regarding expectations of what constitutes a valid prescription. Variability also arises when there are different possible treatments for a case or different allowed doses or when the correct response is a non-specific instruction (e.g., double the dose of X). Ambiguous questions led to higher variability in answers, threatening successful automation. Straightforward questions help students give unambiguous responses. Ideally, questions need to have a single prescription string as the correct answer. Representing all the potential variability in the Marking File was difficult and it also threatened the ease of creating new Marking Files for new question sets. In addition to being accurate for a single question set, a key requirement for the Parser-Marker software is for it to be easy to extend to new question sets without major reprogramming.

5. Resolution of Architecture for Prototype Prescription Parser-Marker

Variability arising from synonyms can be solved by standardizing the elements of the prescription strings in student responses before being marked. This was accomplished using a 'replace dictionary' (Standardization File) which pre-processes the student response file before passing it onto the Marking File (Figure 3). The Standardization File replaces synonyms with a single, standard word or phrase (e.g., po, by mouth and orally are all replaced with ORAL), making automated marking easier. Since updating the Standardization File is simpler than building variability into the Marking File, it facilitates Marking File generation for new question sets. The Standardization File can

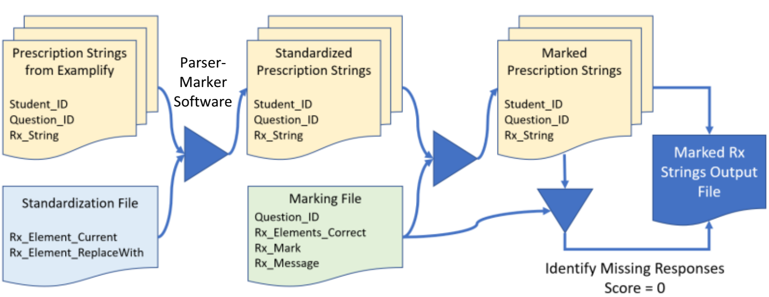

Figure 3. Final architecture for the Parser-Marker Software

be used for all new question sets. Errors in marking due to misspelling of drug names was solved by using the Levenshtein distance of drug names [11]. If the drug name the trainee provided was ≥ 75% distant from the correct response, the system would mark it

0. For example, Atorvastatin and Rosuvastatin have a Levenshtein distance of 75%. Lack of responses to some of the multi-prescription questions was handled by comparing the marked student responses against the original Marking File to identify questions that had not been answered.

Limitations of the current software include the limited pilot testing to date. Future experimental designs are needed for robust evaluation of the developed prototype.

6. Conclusion

Despite increasing use of prescription auto-completer systems, prescription writing competencies are essential parts of medical training and a free-text automated marking system can reduce the current burden on faculty and help students receive timely accurate feedback. We developed a prototype Prescription Parser-Marker which demonstrates a proof of concept for automated prescription writing assessment. The software architecture solves some key issues that threaten automated prescription marking, namely variability in prescription elements, misspellings, and non-responses. The Prescription Parser-Marker is scalable to new question sets by creating a customized Marking File for new question sets. The software can be used outside Canada with minor modification to comply with local requirements.

The next iteration will address: 1) scalability to additional question sets, 2) development of a version control system to better connect question sets to Marking Files, 3) a better user interface to enable users to run the software in their own environments, and 4) improved instructions for additional developers.

References

[1] Holbrook A, Liu JT, Rieder M, Gibson M, Levine M, Foster G, et al. Prescribing competency assessment for Canadian medical students: a pilot evaluation. Can Med Educ J. 2019 Mar 13;10(1):e103-e110. PMID: 30949264; PMCID: PMC6445319.

[2] Canada Health Infoway. Current prescribing and dispensing landscape in Canada: CHI; 2017 [Available from: https://infocentral.infoway-inforoute.ca/en/resources/docs/med-mgmt/1778-report-2017-current-prescribing-and-dispensing-landscape-in-canada. Accessed on 23 Apr 2019].

[3] Baker GR, Norton PG, Flintoft V, Blais R, Brown A, Cox J, et al. The Canadian Adverse Events Study: the incidence of adverse events among hospital patients in Canada. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2004;170(11):1678-86.

[4] Dornan T, Ashcroft D, Heathfield H, Lewis P, Miles J, Taylor D, et al. An in-depth investigation into causes of prescribing errors by foundation trainees in relation to their medical education: EQUIP study. London: General Medical Council. 2009:1-215.

[5] Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. Jama. 1998;279(15):1200-5.

[6] Ryan C, Davey P, Francis J, Johnston M, Ker J, Lee AJ, et al. The Prevalence of Prescribing Errors amongst Junior Doctors in Scotland. Basic and Clinical Pharmacology and Toxicology. 2011;109:35.

[7] Young H. Lack of pharmacological training causes overuse and misuse of drugs. CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne. 2008;178(3):276.

[8] Sam AH, Hameed S, Harris J, Meeran K. Validity of very short answer versus single best answer questions for undergraduate assessment. BMC Med Educ. 2016;16(1):266.

[9] Sam AH, Field SM, Collares CF, van der Vleuten CPM, Wass VJ, Melville C, et al. Very-short-answer questions: reliability, discrimination and acceptability. Med Educ. 2018;52(4):447-55.

[10] Mucklow J, Bollington L, Maxwell S. Assessing prescribing competence. Br J Clin Pharmacol. 2012;74(4):632-9.

[11] Wikipedia contributors. "Levenshtein distance." *Wikipedia, The Free Encyclopedia*. Wikipedia, The Free Encyclopedia, 26 Jan. 2022. Web. 26 Jan. 2022.