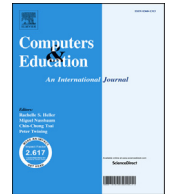


Contents lists available at [ScienceDirect](#)

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Automation of mathematics examinations

Christopher J. Sangwin ^{a,*}, Nadine Köcher ^b^a School of Mathematics, University of Edinburgh, United Kingdom^b Baden-Wuerttemberg Cooperative State University, Karlsruhe, Germany

ARTICLE INFO

Article history:

Received 1 April 2015

Received in revised form 23 November 2015

Accepted 25 November 2015

Available online 30 November 2015

Keywords:

Online assessment

Automatic marking

Mathematics

Examinations

ABSTRACT

Assessment is a key component of all educational systems, and automatic online assessment is becoming increasingly common for formative work in mathematics. This paper reports an investigation of the extent to which contemporary automatic assessment software can automatically mark answers to questions from existing high-stakes mathematics examinations. The questions are taken from a corpus of publicly available core mathematics questions designed for high-achieving students aged approximately eighteen at the school–university interface. We focus on the extent to which objective properties of each final answer may be automatically established and the extent to which automatic marking reasoning by equivalence supports assessment of students' methodology. Our results show that transcribing existing paper-based mathematics examinations into an electronic format is now feasible for a significant proportion of the questions as currently assessed. The most significant barrier to using contemporary automatic assessment is the requirement from examiners that students provide evidence that they have used an appropriate method.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Over the last twenty five years, but particularly over the last decade, there has been a concerted effort to develop software which automatically assesses students' answers to mathematics questions. A summary of early work in this field is given by [Sleeman et al., 1982](#) and a more recent survey is contained in [Sangwin \(2013\)](#). Some early systems relied only on multiple choice and numeric input questions, but for many years in mathematics students have been expected to type in an algebraic expression which constitutes their answer. More recently serious attempts have been made to automatically assess a student's ability to construct a chain of mathematical reasoning. Examples of such software will be provided in due course. In elementary mathematics, including high-school algebra and calculus, there are many situations where a student can provide an answer and the properties of this answer can be established objectively and automatically using a computer algebra system (CAS). The question we seek to answer in this paper is, to what extent can the assessment of existing mathematics examinations be automatically marked using contemporary automatic assessment software?

Our methodology is to take a corpus of published examination questions, together with the official mark scheme. We have examined the extent to which we can faithfully automatically mark these questions using selected representative contemporary software in a way which is faithful to the published mark scheme. The attempt to genuinely automate marking of

* Corresponding author.

E-mail addresses: C.J.Sangwin@ed.ac.uk (C.J. Sangwin), koecher@dhbw-karlsruhe.de (N. Köcher).

existing questions, using existing software, is a “litmus test” which is a long way beyond a purely theoretical or speculative approach.

Constructive alignment, [Biggs and Tang \(2011\)](#), starts with the outcomes we intend students to learn and seeks to align teaching and assessment to those outcomes. All assessments have to balance constructive alignment with other factors such as validity, reliability and practicality. The format of an assessment constrains what is practical and influences validity and reliability. Multiple choice is an extreme example, but paper based examinations are no exception.

One finding from the literature is that direct translation of paper-based assessments into online assessments is inappropriate; there is a need to revisit question formulation, reflecting on what it is intended to test. The process of creating CAA [Computer aided assessment] questions therefore raises fundamental issues about the nature of paper-based questions as well. ([Conole & Warburton, 2005](#), p. 21).

Therefore, to start with an existing examination format and merely translate questions into a new format without regard for the underlying educational construct they are seeking to test might seem incongruous. If our goal was to construct an online examination, working within the constraints and taking best advantage of the format, this concern would be appropriate. The purpose of the research reported in this paper is to understand the extent to which the published intentions of examiners can be faithfully automatically marked at this moment in time, with software actually in use. That is to say, we are not the examiners and we are not (for the purposes of this research at least) engaged in the process of writing valid examinations from scratch.

Indeed, it is out of a respect for experienced examiners, professionally engaged by a large examination board, that we have started with their questions. A failure to be able to faithfully automatically mark traditional examinations may point to serious deficiencies in contemporary software. The data we seek to obtain may therefore be very useful in setting priorities for developers of such software.

We note that the ability to faithfully repeat and examine all the steps required for passing a classical paper examination is not the gold standard of a computer based test. For many users of such systems the goal is formative practice. Other users have selected automatic marking because of the practical advantages of using computers and the internet, for example the ability to scale to large groups and to provide rapid feedback. However, the ability to automate the assessment process, while necessary, is not sufficient. Even if all aspects of a traditional examination could be sufficiently covered by a fully automated exam, it does not immediately follow that this is the most convenient way of performing such examinations. For example, the usability of the system could hinder the performance of all or some of the students so that the results are changed. Basic usability of the interface is important but usability could also relate to differences in computer skills and accessibility issues within the system. For example, [Galbraith and Haines \(1998\)](#) sought to disentangle attitudes related to mathematics from those associated with the technology for learning it. Lack of usability testing with students is a limitation of this study which is a question to be addressed by future research.

The previous experience of the authors strongly suggests that the task of devising automatic marking schemes sheds interesting light on assessment design and on what is currently assessed in practice. Indeed, mathematical proficiency consists of several different aspects, see [Kilpatrick, Swafford, and Findell \(2001\)](#), some of which can be automatically assessed with computer based examinations more readily than others. Contemporary software is developing rapidly. Examiners experienced in writing questions for traditional paper examinations may not be familiar with what is now possible online. Having established our results, a secondary purpose of our research is to inform examiners and teachers of the extent to which we may automatically mark questions which are currently examined. Whether these questions should continue to be used in examinations is a matter for debate, and is ultimately a personal value judgement.

Indeed, an underlying motivation for undertaking this research is a concern that existing examinations may be automatically marked without due regard to the educational constructs they are seeking to test.

The issue for e-assessment is not if it will happen, but rather, what, when and how it will happen. E-assessment is a stimulus for rethinking the whole curriculum, as well as all current assessment systems. ([Ridgeway, McCusker, & Pead, 2004](#), p. 4).

For the purposes of this research we have therefore made no serious attempt to evaluate whether the published questions truly align with stated course goals. Whether or not the corpus of published questions we have chosen do really align with course goals does not alter the fact that teachers will, and do, naturally look to specimen examinations for practical guidance on what and how to teach. Students, naturally, also look to specimen examinations for practice. Many authors, e.g. [Burkhardt and Swan \(2012\)](#), have stressed how important it is to align assessment with the curriculum, going as far as saying that in order to ensure teachers follow the intended curriculum the assessments must cover the goals in a balanced manner.

Similarly, this paper does *not* seek to address the important question of whether existing mathematics examinations actually constitute valid or reliable tests of mathematical expertise. There is a long-standing discussion on this issue. Deciding whether examinations in mathematics are valid is controversial because the decision reflects a set of subjective value judgements about such things as the extent to which students should be fluent in traditional procedures including calculation and algebraic manipulation. For the purposes of this research we are not seeking to define or discuss what constitutes mathematical expertise. Indeed, to do so would potentially confound our research as we have tried to suspend our value judgements and objectively evaluate the extent to which we can automate existing assessments. Instead, we confine ourselves to evaluating the extent to which a question can be automatically marked faithfully to its published mark scheme with contemporary software. By looking at existing mathematics examinations, the main contribution of this paper is data on the

Download English Version:

<https://daneshyari.com/en/article/6834948>

Download Persian Version:

<https://daneshyari.com/article/6834948>

[Daneshyari.com](https://daneshyari.com)