**Computer adaptive practice of Maths ability using a new item response model for on the fly ability and difficulty estimation**

**S. Klinkenberg\*, M. Straatemeier, H.L.J. van der Maas**

**What is the contribution made by this paper?**

<Identify what’s *really* important and omit mention of anything superfluous; a short paragraph should be sufficient to communicate this.>

The authors show that the Elo rating system is a valid and reliable model to select appropriate test items and measure the skill levels of users in the context of computer adaptive practice. The authors compared the measurement precision and bias of computer adaptive practice to standard CAT. For easy items, the use of a scoring model which combines speed and accuracy and the Elo rating system resulted in less loss in measurement precision and less bias than what is found in standard CAT estimation whereas the model without speed performed worse.

**Give a high-level description of the work?**

<A description of the important aspects described in the paper>

The authors use an item response model based on the Elo rating system to assign Maths problems to students and to measure their ability. They created an experiment where users and problems were given Elo ratings. Users and test items begin with a baseline rating and their ratings (both the students’ and the items’) self-adjust based on whether the student answers the questions correctly. The authors also incorporated a scoring rule for speed and accuracy where the faster a user answers a question, the larger the effect on their rating.

**What is the motivation for the work carried out?**

<Why is the work useful/necessary?>

Three problems of standard CAT based on IRT: in CAT, item difficulties need to be determined prior to testing, item difficulty needs to be similar to candidate ability, tradeoff between speed and accuracy is not taken into account. Their new CAT aims to solve those problems in the context of adaptive practice.

**What is your evaluation of the author’s claim?**

<Critically evaluate what the authors have said? Do you agree with them?>

They claim that Elo-based testing is reliable/valid. They claim that estimates on student’s Maths ability through Elo-based CAT correlate highly with students’ actual ability (indexed by norm-referenced CITO test scores (from standardised test)). This is somewhat valid, since correlation between Elo-rating and CITO norm score is around 0.8 (plusminus 0.02). However, Elo-rating varies significantly at lower ends of CITO scores, which undermines the usefulness of Elo-based CAT in the context of summative tests.

**What are the limitations of the work?**

<Limitations may be given by the authors, but you may also need to read between the lines and identify further limitations of their work.>

The main applications of this paper lie within the realm of computer adaptive practice which is different from computer adaptive testing. Another possible limitation is that study was conducted using elementary maths as the test items which is quite limited in terms of breadth and depth. As such, these results may not necessarily apply to other domains or for higher level mathematics.

**What additional interesting points are made in the paper?**

<Anything else that strikes you as important.>

They found that the established system can estimate the skill of new students within only 10 answers. They also found that the users (children) of both high and low ability were motivated to play the Maths games.

**What is the relevance of the research to your own work?**

<How might this work be used in conjunction with your own?>

We have more variety in the types of algorithms that are feasible to be used for CAT. Standard CAT is based on a 3 parameters logistic model (with item information curve and item characteristic curve), which is quite a heavy model to implement in practice. However, the Elo-model is designed to be simple to use and also easy to customise to calibrate better. Since the research showed feasibility of its use, we may extend our interest further in this direction.

**How might the work be developed further?**

<Next steps, perhaps mentioned by the authors, perhaps identified by you from the reading the paper.>

We may replicate the use of the Elo-model in CAT (perhaps by pre-calibrating the items and then administering the calibrated items in a CAT format?)

**Which references would be worth following up and why?**

<Could be interesting related work identified by the authors, pre-req work that is useful to read in order to better understand this paper, authors’ earlier that looks relevant and interesting.>

Optimal testing with easy or difficult items in computerized adaptive testing. <https://psycnet.apa.org/record/2006-10279-002>

Gives comparison between 1PL and 2PL model. This is more for CAT based on IRT, rather than Elo model, but still worth looking into.

The future of educational change: system thinkers in action (Journal of Educational Change, 7 (2006), pp. 113-122)

CAP maximises information about candidate ability, and ‘acting on data is critical’ in education. We can probe into Elo-model CAT/CAP as a tool for maximising learning