Measuring Student's Proficiency in MOOCs: Multiple Attempts Extensions for the Rasch Model

Dmitry Abbakumov [1, 2], Piet Desmet [1], & Wim Van den Noortgate [1]

[1] KU Leuven (University of Leuven), Leuven, Belgium

[2] National Research University Higher School of Economics, Moscow, Russia

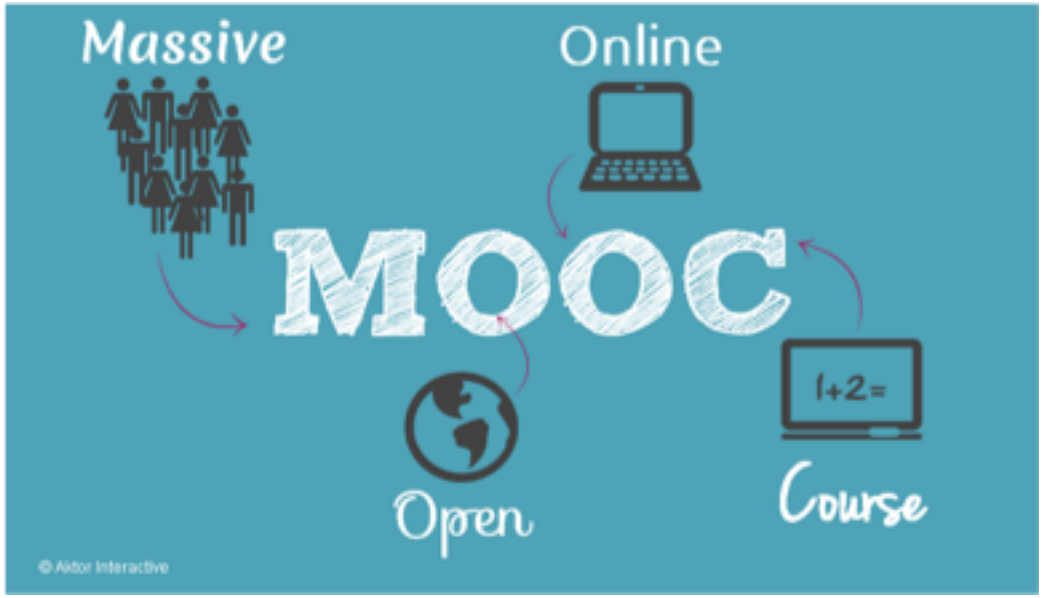
MOOCs

A **M**assive **O**pen **O**nline **C**ourse is an online course with (open) access and unlimited participation of students…

… consists of pre-recorded video lectures, reading assignments, assessments, and forums.

… is developed by universities and run on Coursera, edX, XuetangX, FutureLearn, Udacity, MiriadaX.

In 2017, 800 universities offered 9,400 MOOCs. The same year, Coursera achieved the milestones of 30 million students and 2,700 courses.



Context

Students, professors and universities have an interest in accurate student proficiency measuring.

Problem

Assessments are dynamic: items can be added, removed or replaced by a course author at any time

+

Students are allowed to make several attempts within one assessment

+

Assessments may include an insufficient number of items for accurate individual-level conclusions

=

Common psychometric models and techniques of Classical Test Theory (CTT) and **Item Response Theory (IRT)** do not serve perfectly to measure proficiency

Framework

The Rasch model

$$Logit(\pi_{ij}|\theta_j) = \ln(\pi_{ij}/1-\pi_{ij}) = \boldsymbol{\theta_j} - \boldsymbol{\delta_i} \; and \; Y_{ij} \sim Bernoulli(\pi_{ij})$$

The reformulation presented by Van den Noortgate, De Boeck, & Meulders (2003)

$$Logit(\pi_{ij}) = \boldsymbol{b_0} + \boldsymbol{u_{1j}} + \boldsymbol{u_{2i}} \; and \; Y_{ij} \sim Bernoulli(\pi_{ij})$$

where $u_{1j} \sim N(0, \sigma^2_{u1})$ and $u_{2i} \sim N(0, \sigma^2_{u2})$

→ **very flexible for making extensions**

Extension One

We consider two components of proficiency within weekly summative assessment: the **fixed component** and the **dynamic component**.

$$Logit(\pi_{ij}) = b_0 + (\boldsymbol{b_{10}} + \boldsymbol{b_{1j}}) * attempt_{ij} + \boldsymbol{u_{1j}} + u_{2i} \; and \; Y_{ij} \sim Bernoulli(\pi_{ij})$$

Extension Two

The effect of additional attempt may **vary from item to item** (*e.g. the difference between multiple choice and open-ended items*)

$$Logit(\pi_{ij}) = b_0 + (b_{10} + b_{1j} + \boldsymbol{b_{1i}}) * attempt_{ij} + u_{1j} + u_{2i} \; and \; Y_{ij} \sim Bernoulli(\pi_{ij})$$

(Explanatory) Extensions Three and Four

We assume the more attempts a student makes, the less his/her **effect of attempt** (*e.g. multiple guessing vs. learning from tips and hints*)

$$Logit(\pi_{ij}) = b_0 + (b_{10} + b_{1j} + b_{1i}) * attempt_{ij} + b_2 * class_j + \mathbf{b_3} * attempt_{ij} * class_j + u_{1j} + u_{2i}$$
$$and\ Y_{ij} \sim Bernoulli(\pi_{ij})$$

How does the student's **activity** in the course link to his/her proficiency?

$$Logit(\pi_{ij}) = b_0 + (b_{10} + b_{1j} + b_{1i}) * attempt_{ij} + b_2 * class_j + b_3 * attempt_{ij} * class_j + \mathbf{b_4} * formative.assessment.performance_j + \mathbf{b_5} * lecture.activity_j + u_{1j} + u_{2i}\ and\ Y_{ij} \sim Bernoulli(\pi_{ij})$$

Illustration - Data

“Economics for Non-Economists” MOOC on Coursera

N=1,609 from Russia (72%), Ukraine (8.4%), Kazakhstan (3.9%), Belarus (3.2%), USA (1.2%), other countries (11.3%)

The weekly summative assessment includes 10 items

51,550 responses

Attempts: *M*=2.04 *SD*=1.52.

Illustration - Results

			Basic Mod.	Extension 1	Extension 2	Extension 3	Extension 4
Fixed	Intercept		0.63 (0.22)	0.27 (0.25)	0.29 (0.27)	0.65 (0.27)	-0.75 (0.30)
	Attempt			0.92 (0.03)	0.90 (0.06)	1.34 (0.07)	1.34 (0.07)
	Class					-0.48 (0.03)	-0.42 (0.03)
	Attempt * Class					-0.33 (0.03)	-0.33 (0.03)
	Lect. Act.						0.32 (0.07)
	Pract. Perf.						1.16 (0.09)
			SD	SD	SD	SD	SD
Random	Intercept	Student	0.78	0.92	0.96	0.86	0.77
		Item	1.01	1.12	1.22	1.22	1.22
	Attempt	Student		0.60	0.59	0.52	0.52
		Item			0.21	0.21	0.21
AIC			58779	55401	54929	54487	54260

Cross-Validation

3 MOOCs, 9 assessments with 10-15 items

N=9,484, >500K responses

training/test proportion 75/25, 10 replications

	Overall		Course 1		Course 2		Course 3	
	M	SD	M	SD	M	SD	M	SD
Rasch (1st att.)	.710	.07	.695	.03	.638	.02	.797	.02
Basic Model	.739	.06	.719	.03	.688	.01	.811	.02
Extension 1	.768	.04	.751	.02	.734	.02	.819	.02
Extension 2	.770	.04	.753	.02	.737	.02	.820	.02
Extension 3	.770	.04	.753	.02	.737	.02	.821	.02
Extension 4	.771	.04	.754	.02	.738	.02	.821	.02

Key notes:

- more accurate measures and conclusions on the proficiency
- the transition from analysis of responses to analysis of behavior (in the course)

Limitations and further work:

- peer-assessments and programming assignments

Thank you!