



# Measuring students' activity in MOOCs using extensions of the Rasch model

Dmitry Abbakumov <sup>1, 2</sup>, Piet Desmet <sup>1</sup>, & Wim Van den Noortgate <sup>1</sup>

<sup>1</sup> KU Leuven (University of Leuven), Leuven, Belgium

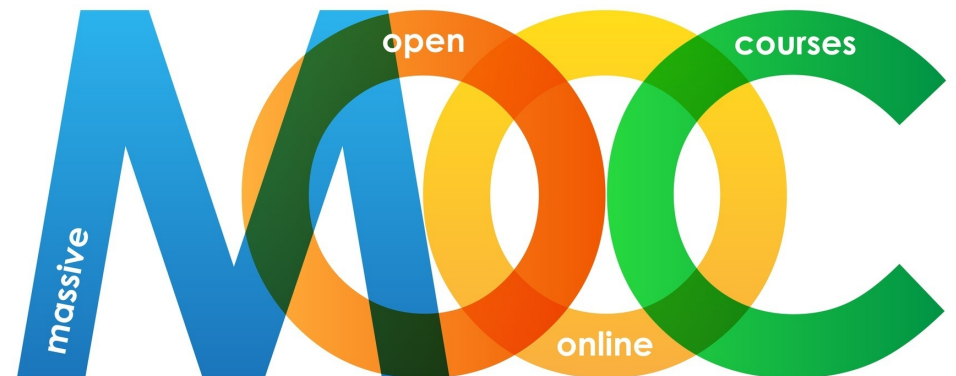
<sup>2</sup> National Research University Higher School of Economics, Moscow, Russia

A **M**assive **O**pen **O**nline **C**ourse is a large scale web-based course developed by a university, solely or in cooperation with industrial partners, in which anyone with internet access can participate.


A **M****O****O****C** consists of video lectures, reading assignments, assessments, and forums.

Now 800 universities offer more than 9,400 **M****O****O****C**s.

**Coursera**: 40 million students and 3,000 courses.



Students' activity in MOOCs is typically described through the proportions (videos viewed, assessments taken). These measures are simple and intuitive. However, by aggregating the information per student, *we lose information on how a student interacts with individual units of content.*



Addressing Large Hadron Collider Challenges by Machine Learning

View as learner

Course Checkpoints

Course checkpoints are predefined stages that a learner encounters as they progress through the course. The section below defines course checkpoints and lists the top 5 checkpoints causing the most drop-offs.

Course Checkpoint Details

The first and last checkpoint for every course are:

- Enrolled in Course
- Completed Course

To track progress within modules, we define 5 checkpoints for each module:

- Started Item: Learners starts any item in that module
- Completed Item: Learners completes any item in that module
- Started Assessment: Learner starts any assessment in that module
- Completed Assessment: Learner completes any assessment in that module
- Completed Module: Learner completes the module by passing all graded assessments in that module

So every course has 5 checkpoints per module, along with the first and last checkpoint.

Top 5 Drop Rates

Checkpoint Description	Module Number	Checkpoint Type	Total Eligible Learners Drop Rate	Paid Learners Drop Rate	Final Learners Drop Rate
1 Module 01- Started Item	1	Started Item	73.6%	49.7%	88.4
2 Module 01- Started Assessment	1	Started Assessment	40.4%	49.4%	41.8
3 Module 01- Completed Item	1	Completed Item	21.7%	23.0%	30.2
4 Module 02- Completed Module	2	Completed Module	17.1%	13.8%	20.0
5 Completed Course		Completed Course	10.8%	8.7%	12.5

# Framework

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

$Y_{ij}$  is the interaction of student  $j$  with a unit of content  $i$ , where 1 means that unit  $i$  was completed by student  $j$ , and 0 means that unit  $i$  was not completed by student  $j$  during the MOOC

$\theta_j$  is student's activeness

$\delta_i$  is the effect of a unit of content on the probability of interaction

The reformulation presented by Van den Noortgate, De Boeck, & Meulders (2003)  $Logit(\pi_{ij}) = b_0 + u_{1j} + u_{2i}$  and  $Y_{ij} \sim Bernoulli(\pi_{ij})$   
where  $u_{1j} \sim N(0, \sigma_{u1}^2)$  and  $u_{2i} \sim N(0, \sigma_{u2}^2)$

Very flexible for making extensions.

The dimensionality check did not reveal substantial evidence against unidimensionality ( $p = .37$ ).

# Extension

Extension 
$$\text{Logit}(\pi_{ij}) = b_0 + b_{10} * \text{type}_i + (b_{20} + b_{2j}) * \text{week}_{ij} + u_{1j} + u_{2i}$$

$\text{type}_i$  is a dummy variable, which distinguishes between two major types of content in a MOOC (where 0 stands for video lecture and 1 for reading assignment)

$\text{week}_{ij}$  is 0, 1, 2, ..., n-1 and means the first, the second, the third, ..., or n-th week of a MOOC at which student  $j$  interacts with unit of content  $i$  respectively

$b_{2j}$  and  $u_{1j}$  are assumed to follow the bivariate normal distribution  $N(\mathbf{0}, \Sigma_j)$

Thus, 
$$\theta_j = u_{1j} + (b_{20} + b_{2j}) * \text{week}_{ij}$$

$$\delta_i = (b_0 + b_{10} * \text{type}_i + u_{2i})$$

# Illustration: Data

MOOC: Introduction to Neuroeconomics: How the Brain Makes Decisions

Language: English

Length: 11 weeks

N = 11,826

Female (48%)

Africa (3.4%), Asia (21.5%), Europe (35.5%), N.America (28.2%), Oceania (2.5%), and S.America (8.9%)

Age: 18-24 (18.2%), 25-34 (49.5%), 35-44 (18.7%), 45-54 (7.7%), other age groups (5.9%)

We used the data from five weeks (modules) from the course

27 videos and 10 reading assignments

437,568 observations

# Illustration: Results

			Rasch Model	Extension	Extension with Type*Week Interaction		
Fixed	Intercept		-1.55 (0.16) ***	1.67 (0.15) ***		1.41 (0.19) ***	
	Type			-0.90 (0.17) ***		-0.30 (0.28)	
	Week			-4.32 (0.08) ***		-4.18 (0.10) ***	
	Type*Week					-0.48 (0.14) ***	
Random	Intercept	Student	3.43	3.11	Corr.	3.11	Corr.
	Week	Student		3.51	.27	3.52	.27
	Intercept	Content	2.29	0.54		0.52	
AIC			245,952	206,819		206,811	

# Illustration: Results

			Rasch Model	Extension	Extension with Type*Week Interaction		
Fixed	Intercept		-1.55 (0.16) ***	1.67 (0.15) ***		1.41 (0.19) ***	
	Type			-0.90 (0.17) ***		-0.30 (0.28)	
	Week			-4.32 (0.08) ***		-4.18 (0.10) ***	
	Type*Week					-0.48 (0.14) ***	
Random	Intercept	Student	3.43	3.11	Corr.	3.11	Corr.
	Week	Student		3.51	.27	3.52	.27
	Intercept	Content	2.29	0.54		0.52	
AIC			245,952	206,819		206,811	



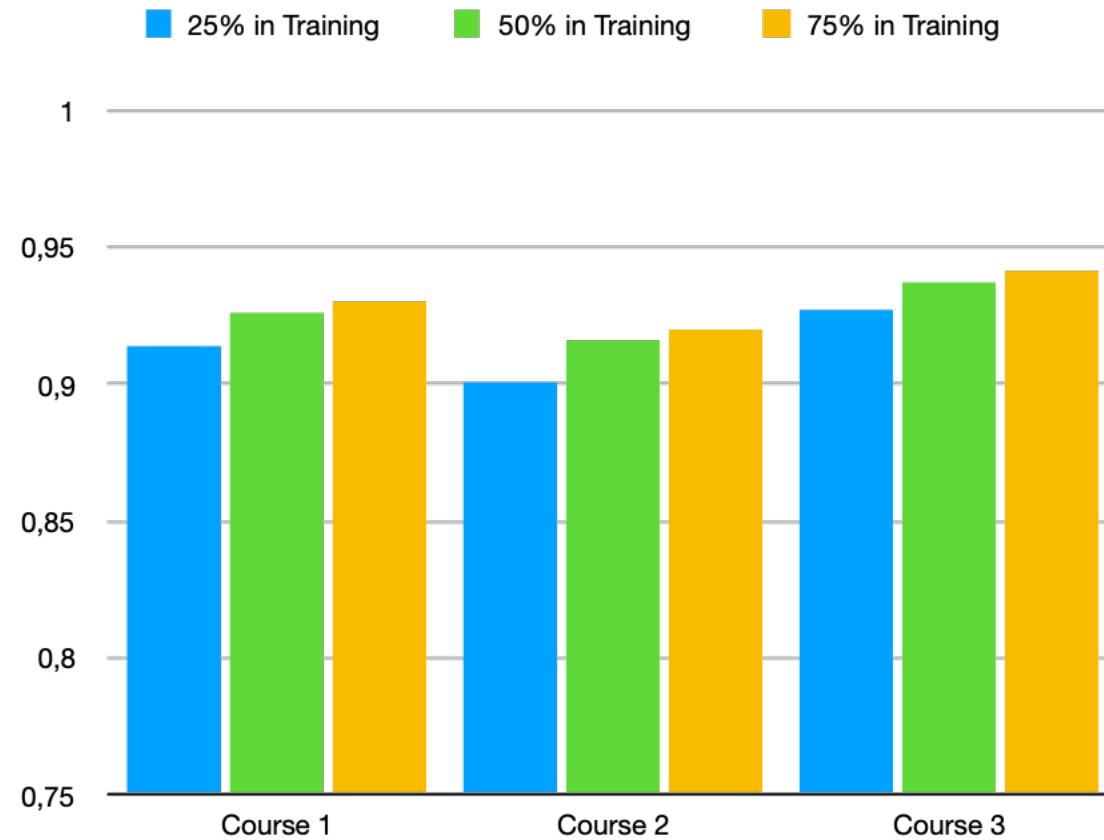
# Illustration: Results

			Rasch Model	Extension	Extension with Type*Week Interaction		
Fixed	Intercept		-1.55 (0.16) ***	1.67 (0.15) ***		1.41 (0.19) ***	
	Type			-0.90 (0.17) ***		-0.30 (0.28)	
	Week			-4.32 (0.08) ***		-4.18 (0.10) ***	
	Type*Week					-0.48 (0.14) ***	
Random	Intercept	Student	3.43	3.11	Corr.	3.11	Corr.
	Week	Student		3.51	.27	3.52	.27
	Intercept	Content	2.29	0.54		0.52	
AIC			245,952	206,819		206,811	

# Illustration: Results

			Rasch Model	Extension	Extension with Type*Week Interaction	
Fixed	Intercept		-1.55 (0.16) ***	1.67 (0.15) ***	1.41 (0.19) ***	
	Type			-0.90 (0.17) ***	0.22 (0.22)	
	Week			-4.32 (0.08) ***	-4.18 (0.10) ***	
	Type*Week				-0.48 (0.14) ***	
Random	Intercept	Student	3.43	3.11	3.11	Corr.
	Week	Student		3.51	3.52	.27
	Intercept	Content	2.29	0.54	0.52	
AIC			245,952	206,819	206,811	

# Cross-Validation



Key result:

- better insight on students' activity in a MOOC

Limitation:

- binary dependent variable

Practice:

- application in a recommender tool

Thank you!