

English Proficiency and MOOC Performance: A Log Study of Videos and Forums

Tanguy Dubois, Damien De Meyere, Patrick Watrin & Magali Paquot

Background

MOOC accessibility

- **MOOCs (Massive Open Online Courses)** aim to democratize learning but have dropout rates as high as **90%** (Furukawa et al. 2020).
- Many MOOCs are **taught in English**, limiting access for **non-native English speakers**, especially from developing countries (Karp Gershon et al. 2021).

Challenges for non-native speakers

- **Lower Completion Rates:** More dropouts, less active in discussion forums, lower grades (Wen et al. 2014, Duru et al. 2019).

MOOC technology helpful for non-native speakers?

- **Different Interaction Patterns:** More pauses, slower video playback, skipping sections without visual aids (Uchidiuno et al. 2018).

Key gap in research

- Prior studies distinguish **native vs. non-native speakers** but **ignore language proficiency**.

RQs & Hypotheses

- ❖ **RQ1:** **How does participants' proficiency level correlate with their performance in MOOCs?**
 - H1:** Participants with lower English proficiency will have lower performance overall
- ❖ **RQ2:** **How does participants' proficiency level influence the relation between their performance and their interactions with video content and discussion forums?**
 - H1:** Participants with lower English proficiency should benefit more from video-viewing strategies (e.g., pausing, changing the playback rate, closed captions etc.) (Costley et al. 2021)

Our data:

- **Two** UCLouvain **MOOCs** on EdX about law: International Law & International Humanitarian Law
- Multiple **editions:** 2020, 2021, 2022, 2023
- **Two data sources:** pre- and post-course **surveys** and **log files** collected automatically by the EdX platform

Building the dataset

1. Retrieve the self-assessed **proficiency level** of MOOC participants from optional pre- and post-course surveys, ranging from **B1** to **C2** on the **CEFR** scale.

Other relevant survey variables:

- Active or passive participation
- No prior, limited or extensive MOOC experience
- Level of education
- Age
- Nationality

2. Calculate the **performance** of these participants, which reflects the **degree** to which **participants mastered** the educational **content** of the MOOC.

Performance is **operationalized** based on the CFA **score** participants obtained on optional **quiz questions** at the end of each MOOC section.

3. Generate aggregate **measures** of participants' **interactions** with **video** content and discussion **forums** using data from log events:

Measures include the following:

- Number of pauses
- Number of seek backs
- Number of seek forwards
- Playback rate
- Closed caption time
- Watch ratio
- Message length

Proficiency	Participants
B1	87 (9%)
B2	132 (14%)
C1	151 (16%)
C2	336 (37%)
Native	213 (23%)
Total	919

Analyzing the data

To analyze the data, we used **mixed-effects linear regression** using the *lme4* package (Bates et al. 2015) in R

- **Performance** as the dependent variable
- **Survey** and **log-based** variables as **predictors**
- **Interactions** between **predictors** and **proficiency** level reveal how proficiency modulates performance
- **Treatment** coding with **native** speakers as **reference** level, to be compared against other proficiency levels
- **Random effects** to control for **nationality** and **course + edition**

Results

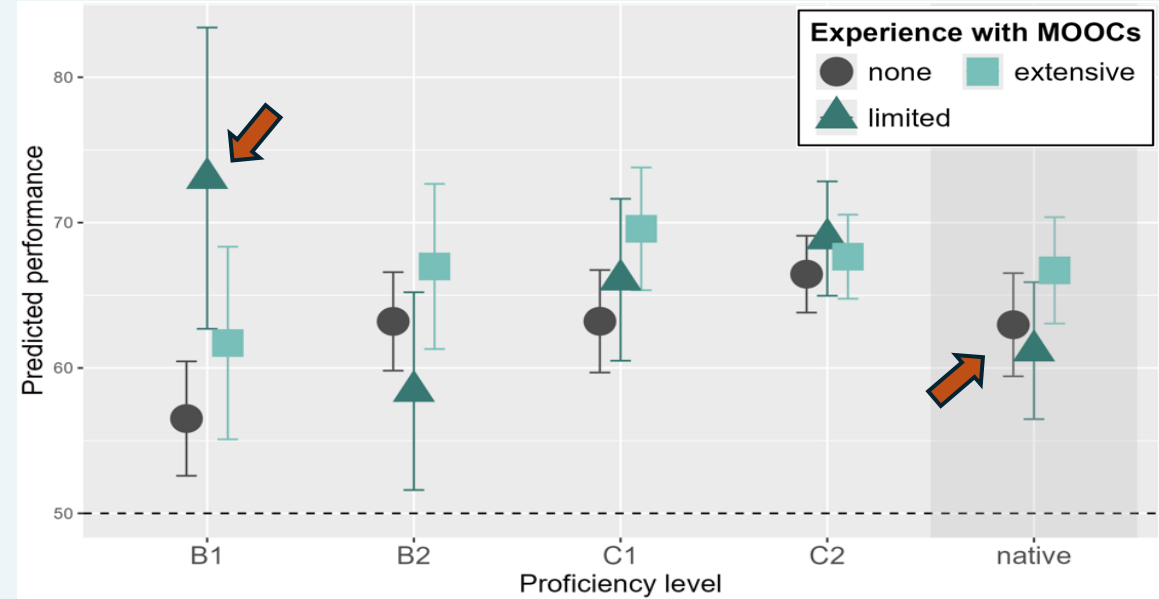
Variables that influence native and non-native speakers in the same way:

Better performance is associated with:

- **Education level:** Holding a degree from higher education
- **Age:** Being older
- **Number of pauses:** Pausing more often
- **Watch ratio:** Watching larger portions of videos
- **Message length:** Writing long messages in discussion forums

Variables that influence native and non-native speakers differently:

- **Proficiency level:** B1 learners perform less well than native speakers.
- **Prior MOOC experience:** B1 learners benefit more from having limited MOOC experience than native speakers:



Discussion

H1: Participants with lower English proficiency will have lower performance overall

TRUE: B1 participants perform less well than more proficient participants and native speakers.

H1: Participants with lower English proficiency should benefit more from video-viewing strategies

FALSE: All participants benefit from video-viewing strategies to the same extent. Only the effect of prior MOOC experience is modulated by proficiency level.

Possible explanation? Low-proficiency participants may find other ways to overcome language difficulties, unrelated to watching videos:

Reading is easier than listening (Buck 2001)

- **Low-proficiency participants** may prefer to **read transcripts** and **handouts** than watching **videos**
- **Handouts** are more **amenable** to the use of **auxiliary** (translation) **tools**

Future research

- Replicate study with **more data** from each proficiency level, especially B1
- Investigate **relation** between use of **transcripts** and video **content type**
- Use **Educational Process Mining** techniques to consider MOOC **engagement** over **time** (Bogarín et al. 2018)

References

