University of St Andrews

UNIVERSITY OF Southampton

Clustering of learners' behaviour in the Understanding Language MOOC

Adriana Wilde

agw5@st-andrews.ac.uk

University of St Andrews

University of Southampton

Clustering of learners' behaviour in the Understanding Language MOOC

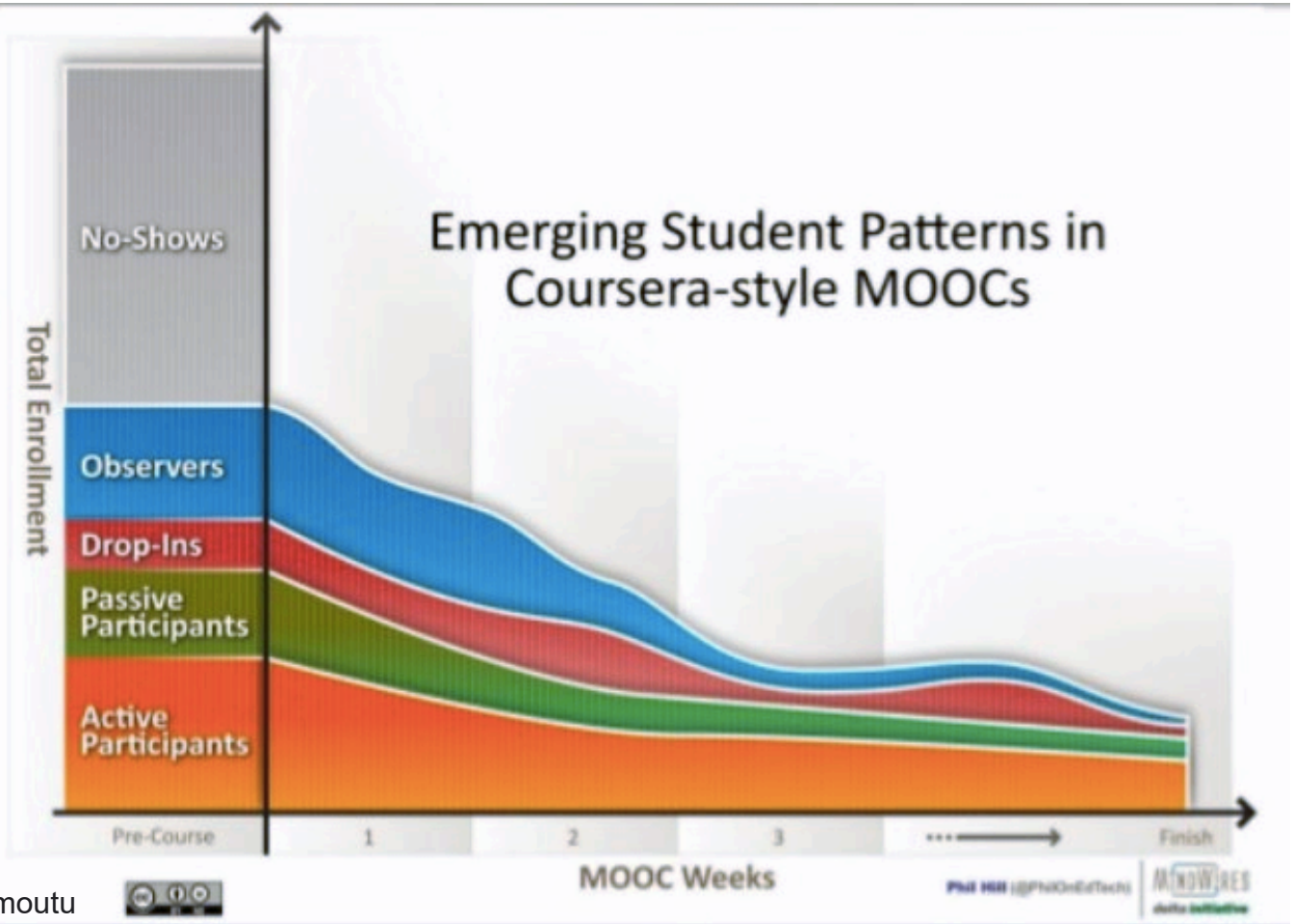
Adriana Wilde and Xin Zhang

agw5@st-andrews.ac.uk
xz5n15@ecs.soton.ac.uk

Remember when…

Average MOOC **COMPLETION RATE** is just....

6.8%

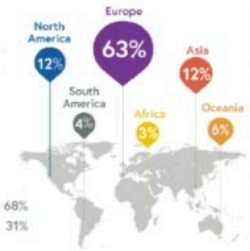


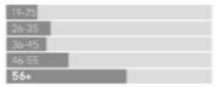
Characterising the diversity of learners

Vitalisers

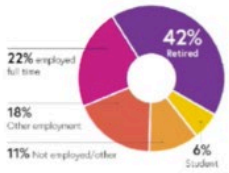
"I'm a lifelong learner and use FutureLearn to keep my brain active and gain knowledge in areas that interest me."

Vitalisers learn as a **hobby** and for the **love of learning**. They may be proud to call themselves **lifelong learners**, occupying themselves by learning anything of personal interest. They see learning as a **enjoyable** and **stimulating** activity, perhaps even an **indulgence**. They also feel it's a good, **constructive use of their time**; keeping them **mentally active**. They were one of the groups least motivated by communicating with other learners.





Female 68%
Male 31%



number of enrolments	3.96
activation rate	38.5%
full participation rate	8.64%
purchase rate	0.44%

Example needs

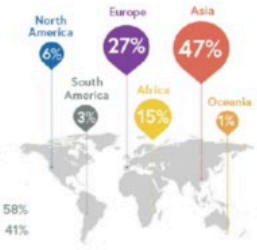
- Regular supply of diverse and interesting courses, allowing them to learn regularly
- To spend personal and luxury time learning
- Sound, accessible content and good reference materials
- To feel stimulated, mentally active and entertained
- To feel good about themselves, their identity as a learner and their good use of time

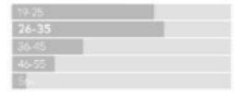
Preparers

"FutureLearn helps me to stay ahead of peers by offering courses with certification and ties with world's top universities."

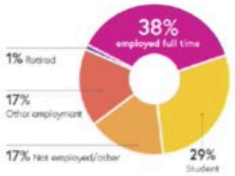
Preparers tend to be **starting out in jobs, careers or related study**, having chosen what they want to do. They have specific career or study goals and learn to improve their chances of success, **improve career prospects, stand out from the crowd and increase their confidence.**

Example goals are learning English, succeeding in interviews, preparing for a career or to study, achieving good grades and doing well in assignments.

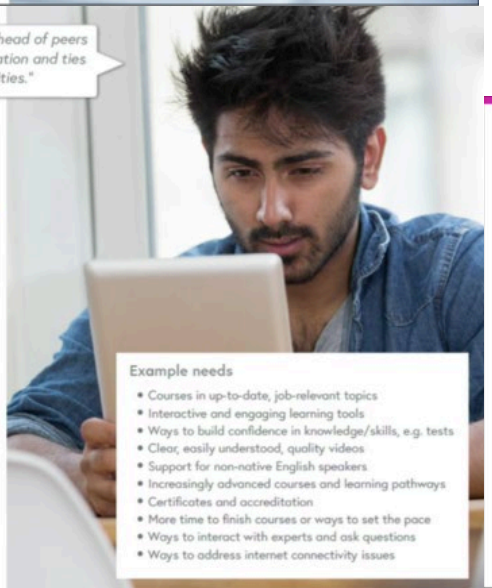




Female 58%
Male 41%



number of enrolments	1.8
activation rate	34.5%
full participation rate	5.52%
purchase rate	0.76%



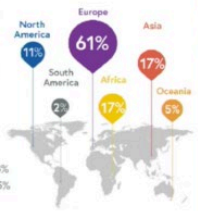
Example needs

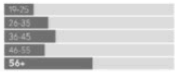
- Courses in up-to-date, job-relevant topics
- Interactive and engaging learning tools
- Ways to build confidence in knowledge/skills, e.g. tests
- Clear, easily understood, quality videos
- Support for non-native English speakers
- Increasingly advanced courses and learning pathways
- Certificates and accreditation
- More time to finish courses or ways to set the pace
- Ways to interact with experts and ask questions
- Ways to address internet connectivity issues

Hobbyists

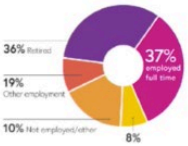
"Starting up a blog made me want to gain more knowledge about my areas of interest, to subsequently draw from when writing"

Hobbyists learn to **support their existing personal projects, leisure activities, pastimes & community activities**. Our courses complement, inform and enable their activities. Their need to learn may be ongoing or relevant at a certain point in time. Although they're one of the least likely groups to be motivated by communicating with other learners, their hobbies could be socially orientated. For example, they may want to share what they learn with others, including online and local communities.

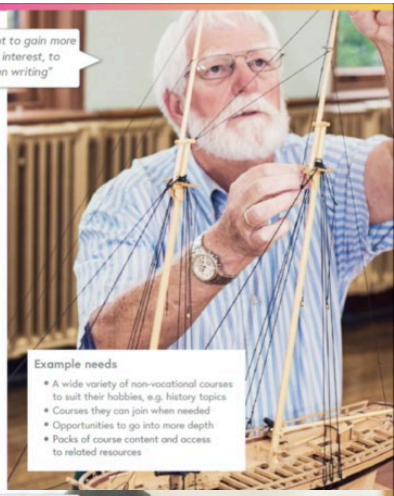




Female 54%
Male 45%



number of enrolments	2.26
activation rate	39.26%
full participation rate	8.89%
purchase rate	0.49%



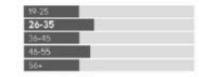
Example needs

- A wide variety of non-vocational courses to suit their hobbies, e.g. history topics
- Courses they can join when needed
- Opportunities to go into more depth
- Packs of course content and access to related resources

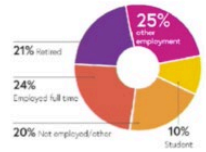
Fixers

"I'm currently on maternity leave so I wanted something to keep my brain active and also learn about the best way to care for my son."

Fixers learn in order to **understand or manage current aspects of their personal life**. This could include the physical or mental **health** of themselves or those close to them, **political or cultural issues**, situations requiring **practical life skills**, or **major life changes** such as bereavement, parenthood, retirement or redundancy. Their need to learn could be one-off or ongoing and might be related to a personal, serious or emotive situation.



Female 58%
Male 40%



number of enrolments	2.0
activation rate	38.65%
full participation rate	7.14%
purchase rate	0.13%

Example needs

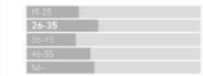
- A range of quality, up-to-date courses to support needs
- Pathways to learn more if desired
- Empathy and understanding
- Confidence and empowerment
- Accessible, credible content and expert advice
- Tools, techniques and information to apply and use
- Courses they can start when they need to
- Downloadable content for reference

Flourishers

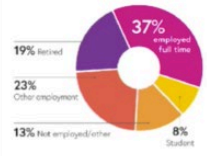
"If I am calmer and more focused I will work better and be more productive!"

Flourishers enjoy **self-help learning** in order to be **happy and healthy in their personal and professional lives**. They may learn to be calm, manage stress, be enriched, build self-esteem, gain motivation, help others, share what they learn, keep their brain active, improve relationships and improve their mental and physical well-being.

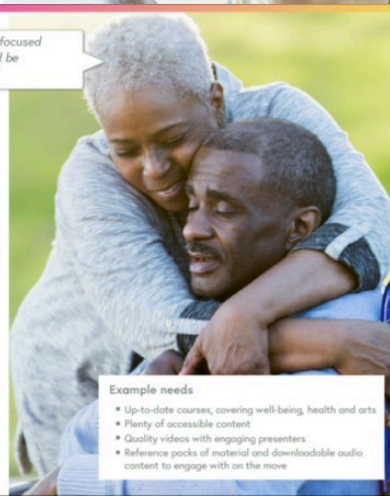
This group overlapped with the Vitalisers, suggesting learning for enjoyment can improve wellbeing generally.



Female 66%
Male 33%



number of enrolments	3
activation rate	36.4%
full participation rate	6.29%
purchase rate	0.51%

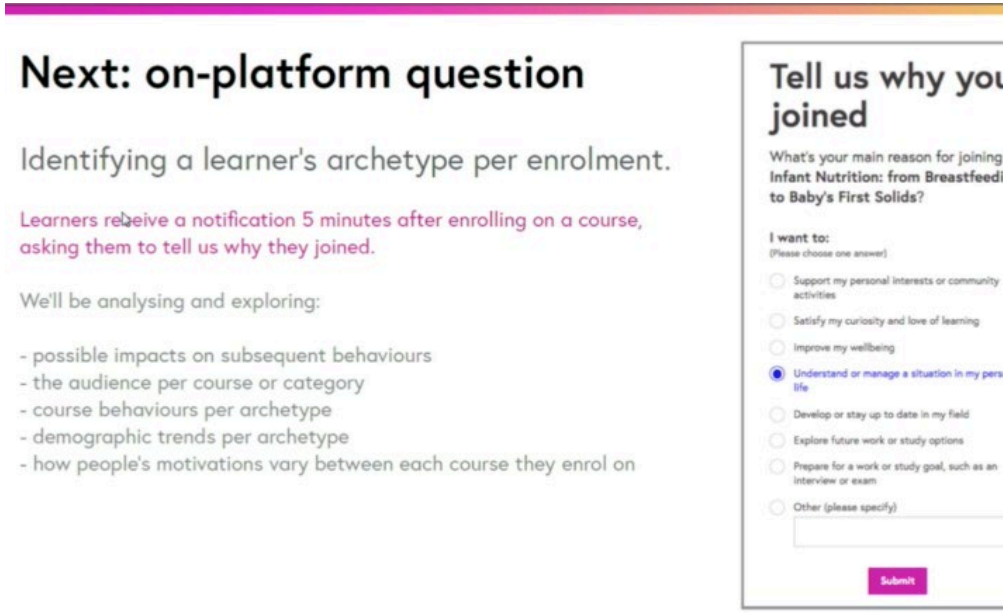


Example needs

- Up-to-date courses, covering well-being, health and arts
- Plenty of accessible content
- Quality videos with engaging presenters
- Reference packs of material and downloadable audio content to engage with on the move

Characterising the diversity of learners

FutureLearn “Archetypes” (Walker et al., 2017) aid understanding both the **motivations** and **needs** of participants falling in these categories.



In principle, this is only possible by collating and analysing self-reported data on learners’ **motivation**.

However, in practice, such self-reported data is **rare** in comparison with the wealth of data available on learners’ interaction with and within the platform.

Characterising the diversity of learners

"Sub-populations" (Kizilcec et al., 2013; Ferguson & Clow, 2015) can still be identified only by observing the behaviour in the platform, and arguably represent a very similar classification of learners to that in the archetype analysis.

Kizilcec et al. (2013)

Learners engagement per assessment was labelled:

- "**On track** = 3" (completed the assessments on time)
- "**Behind** = 2" (undertook the assessments late)
- "**Auditing** = 1" indicates they did not take any assessment but watching videos or doing quizzes.
- "**Out** = 0" indicates they did not do any assignment or watch videos in the course.

Kizilcec, R.F., Piech, C. and Schneider, E., 2013, April. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 170-179). ACM.

Kizilcec et al. (2013)

Learners were identified into four clusters:

- "**Completing**": the learners finished most of the assessments in the courses.
- "**Auditing**": the learners completed the assessments infrequently, but prefer to watch videos.
- "**Disengaging**": the learners engage the courses at the beginning, then decreasing their engagement.
- "**Sampling**": the learners only watched few videos

Kizilcec, R.F., Piech, C. and Schneider, E., 2013, April. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the third international conference on learning analytics and knowledge* (pp. 170-179). ACM.

Ferguson and Clow (2015)

Score	Description
1	Only visited content
2	Commented but visited no new content
3	Visited content and commented
4	Did the assessment late and did nothing else that week
5	Visited content and did the assessment late
6	Did the assessment late, commented, but visited no new content
7	Visited content, commented, late assessment
8	Assessment early or on time, but nothing else that week
9	Visited content and completed assessment early/ on time
10	Assessment early or on time, commented, but visited no new content
11	Visited, posted, completed assessment early/ on time

Ferguson, R. and Clow, D., 2015, March. Examining engagement: analysing learner subpopulations in massive open online courses (MOOCs). In Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (pp. 51-58). ACM.

The name of clusters	Description
Samplers	The learners who any explored few weeks (most at week1), and few of them post comments and submitted assessments.
Strong starter	All the learners completed the first assessment and then dropped out sharply in the cluster. And over third of them posted comments.
Returners	The learners who finished the assessment at week1 and week2. And most of them completed the first assessment late. Then dropped out.
Mid-way Dropouts	The learners who dropped out at the middle of the courses.
Nearly There	The learners can complete the assessments constantly, but cannot finish the course.
Late Completers	The learners in the cluster submitted the final assessment and majority of other assessments. However, they cannot complete on time.
Keen Completers	All the learners engage in the courses. They can finish all the assessments and most of them completed on time.

So what did we do?

MOOC in focus

- Understanding Language
  - University of Southampton and the British Council
  - 4/5 weeks, up to 20 learning objects per week (“steps”)
  - 6 runs

https://www.futurelearn.com/courses/understanding-language

MOOC in focus: Understanding Language

Understand key concepts in the effective teaching and learning of languages.

What is language? How do we learn meaning in a new language? What is easy and hard about learning another language? And what is the best way to teach other languages?

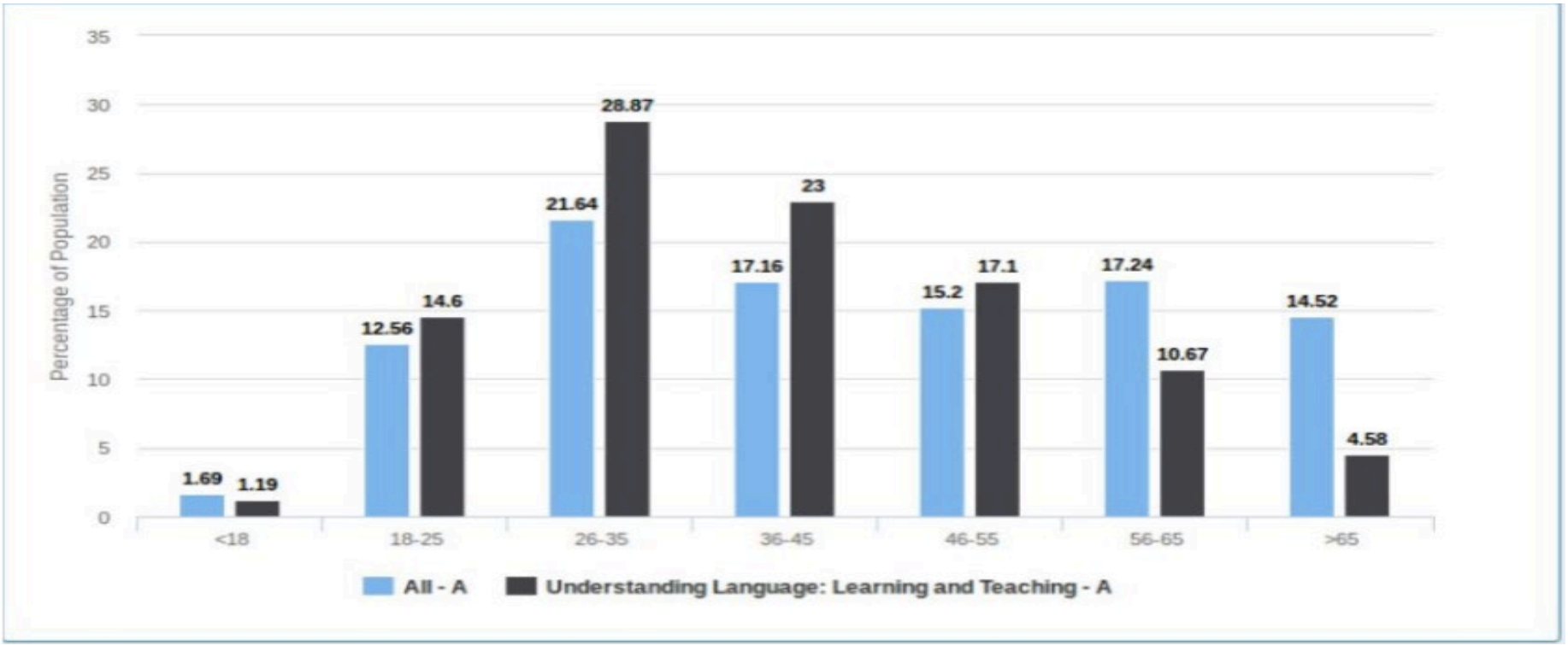
This free online course suggests some answers to these questions. It has been developed by the University of Southampton and the British Council, and draws on their exciting joint online course, MA in English Language Teaching.

Overview of the dataset

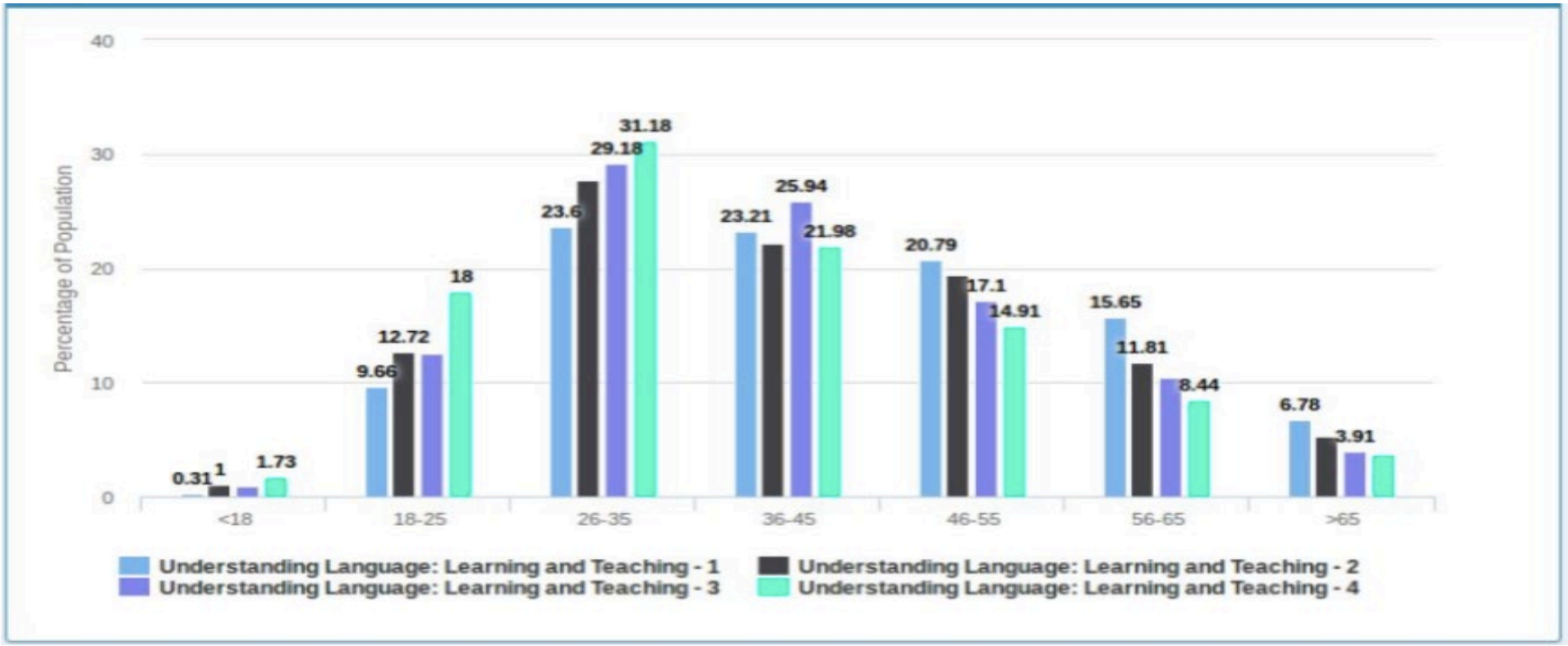
	Run1	Run2	Run3	Run4	Run5	Run6
The name of the course	Understanding Languages: Learning and Teaching					
Start	11/17/2014	4/20/2015	10/19/2015	4/4/2016	10/17/2016	4/24/2017
Weeks	4	4	4	4	5	5
Participants	58721	41874	44250	25569	19840	10260

Digging into the data
(first with exploratory statistics)

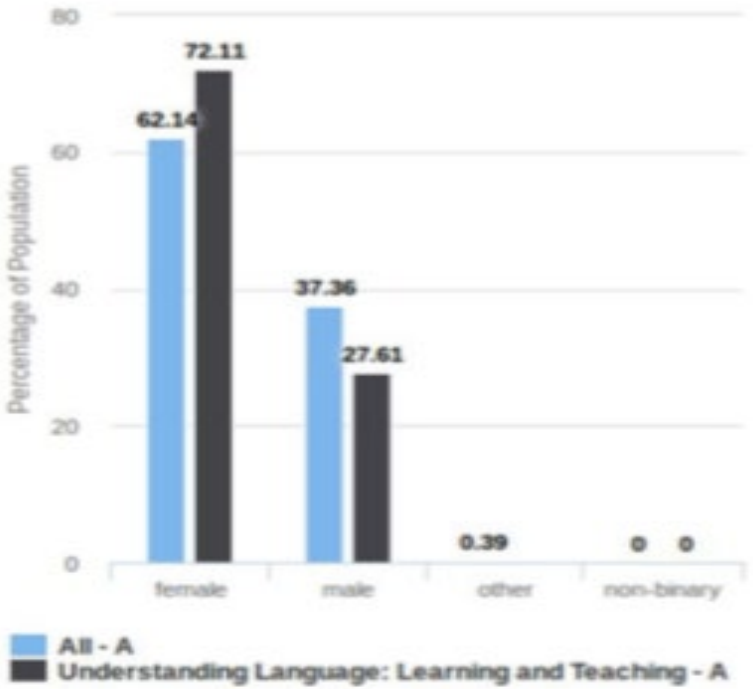
Age distribution

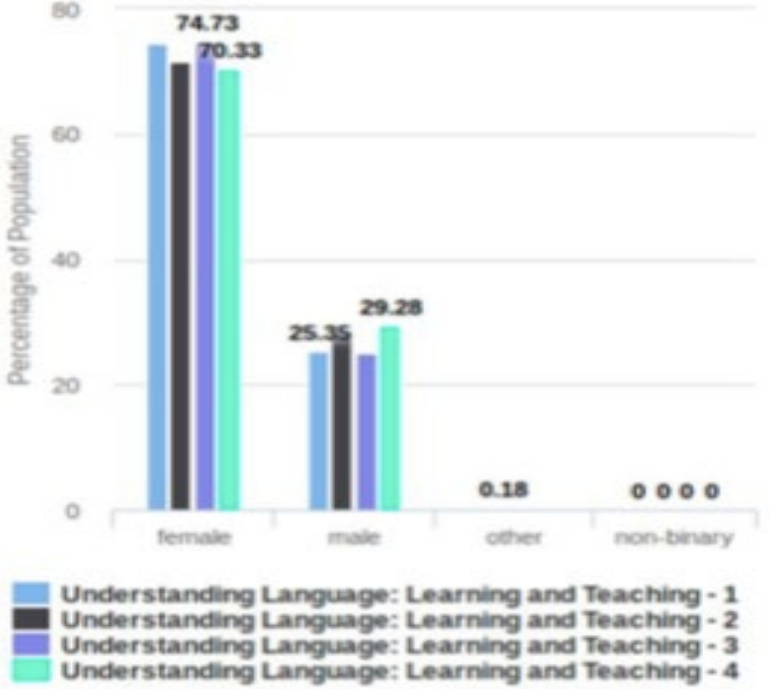


Comparing various UL runs (4 weeks long)

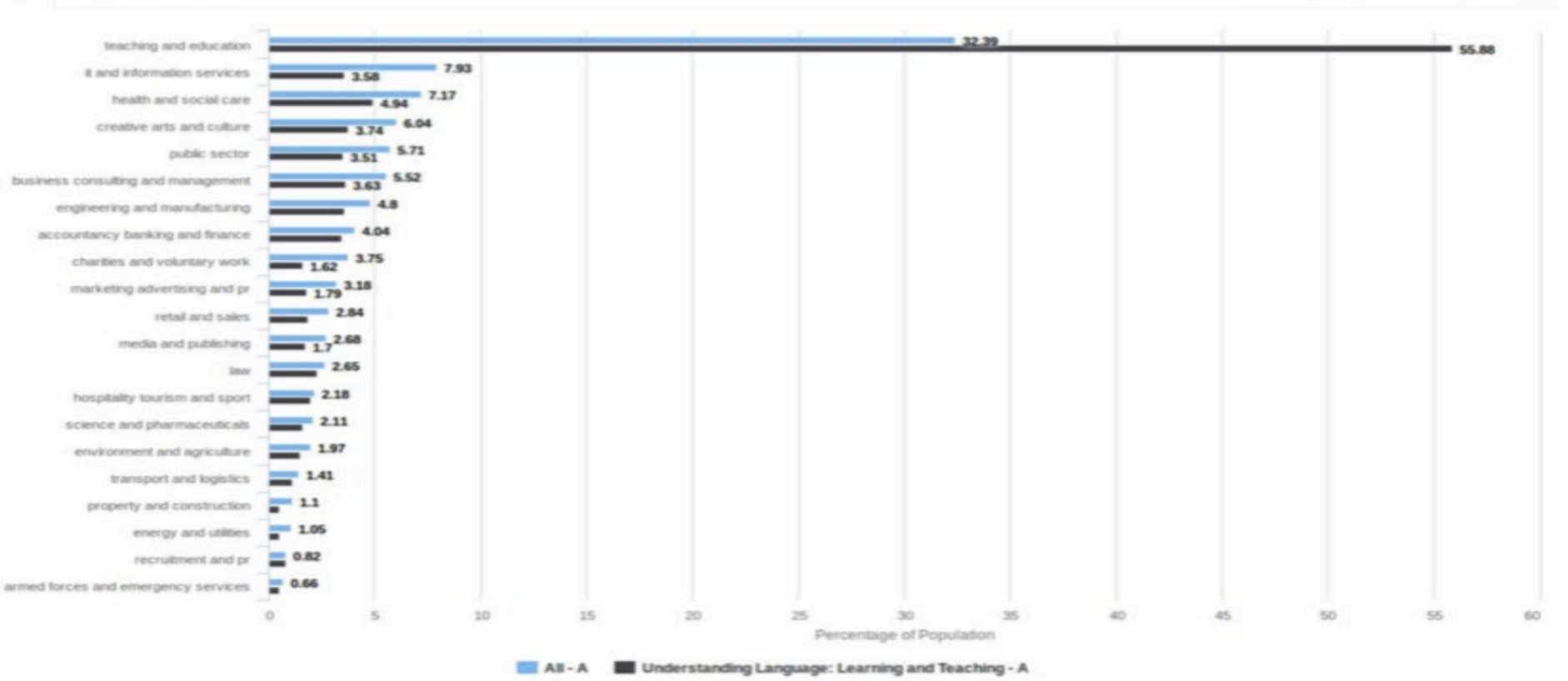


Gender distribution

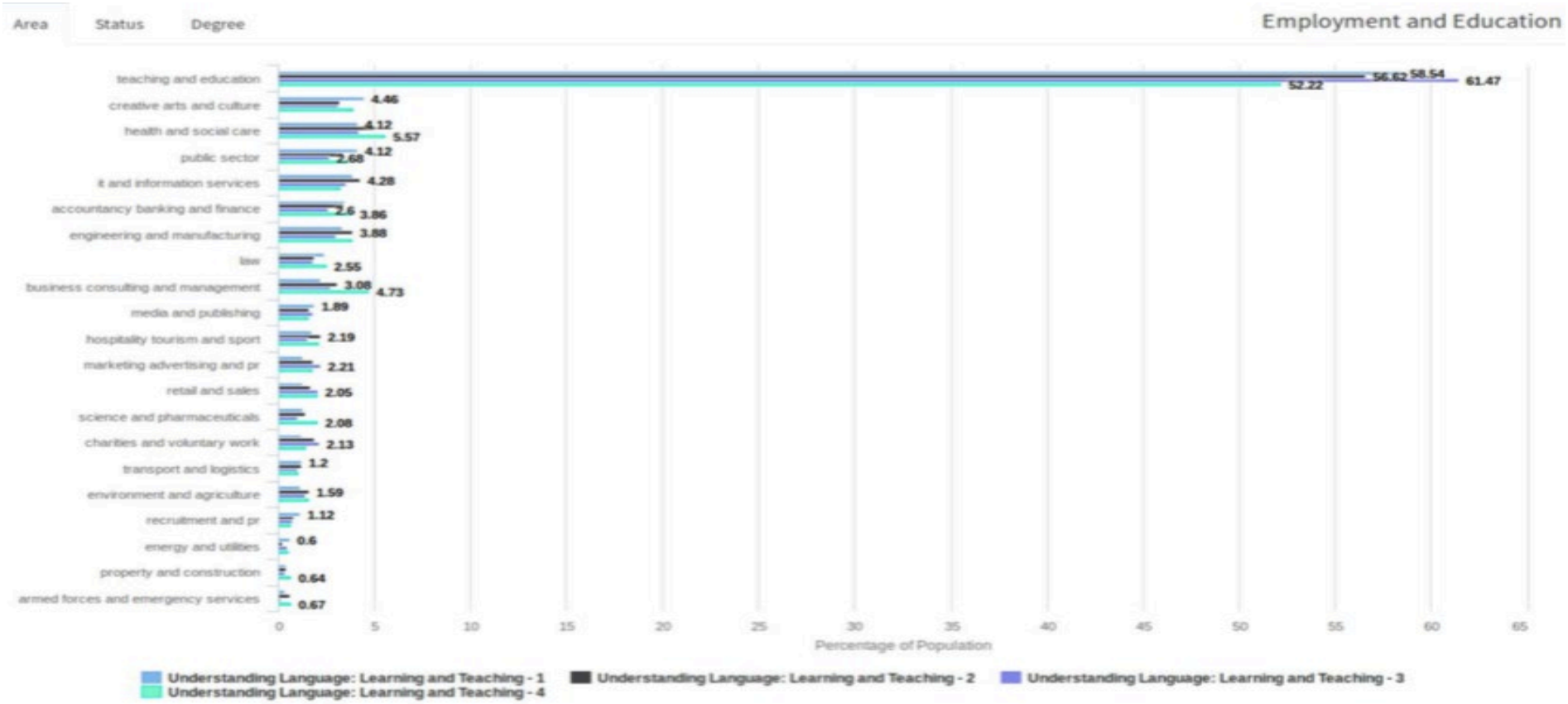




Teaching and Education!



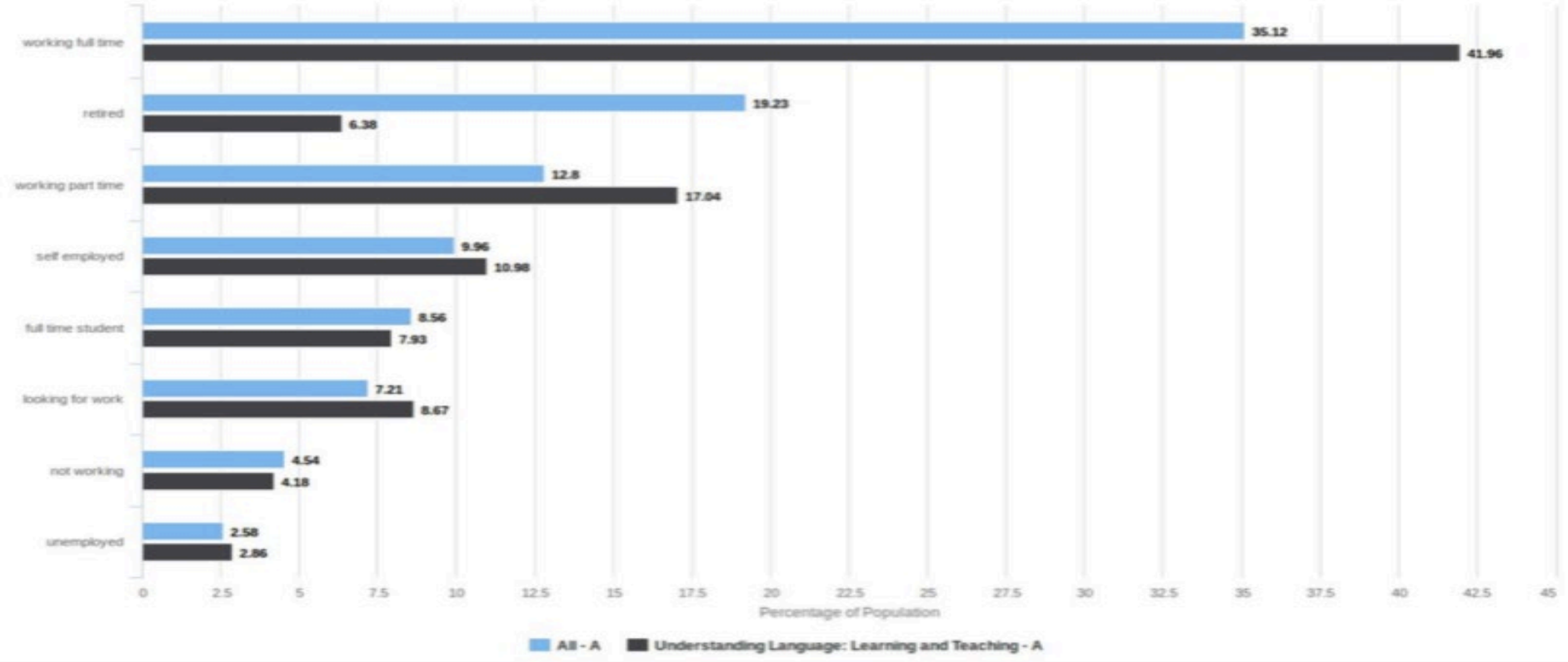
Teaching and Education!



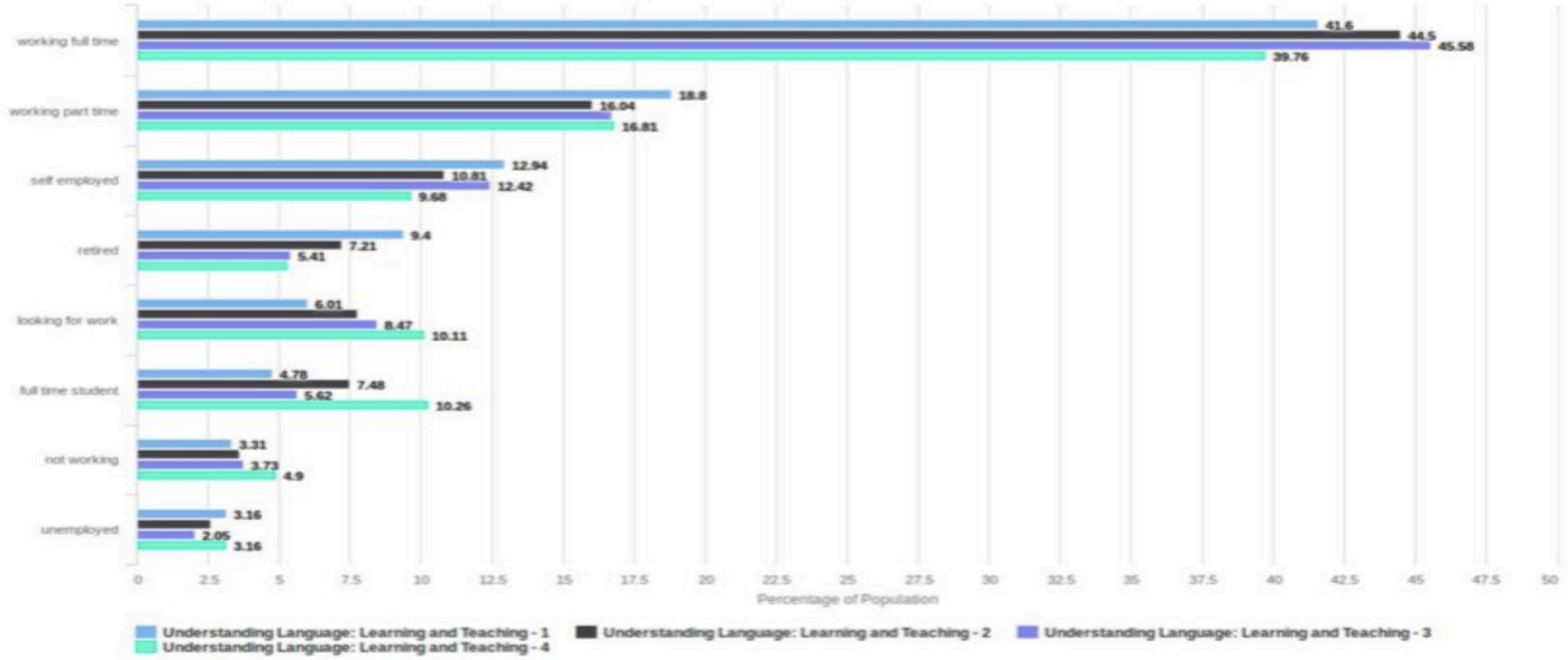
Teaching OR Education?

- Hard to tell with the data available!
- But we could see whether they were in employment and…
- …what is their highest level of education.

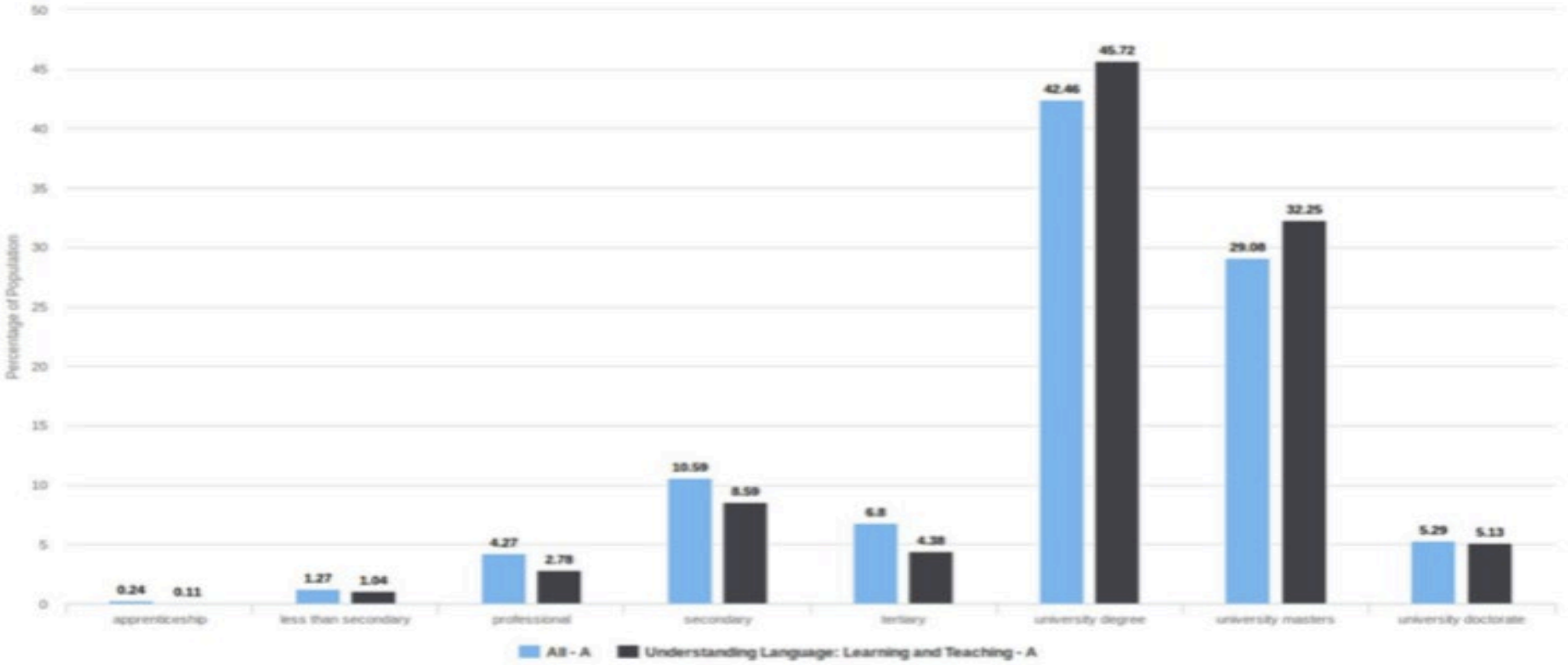
More likely working!
(full time or part time or self-employed)



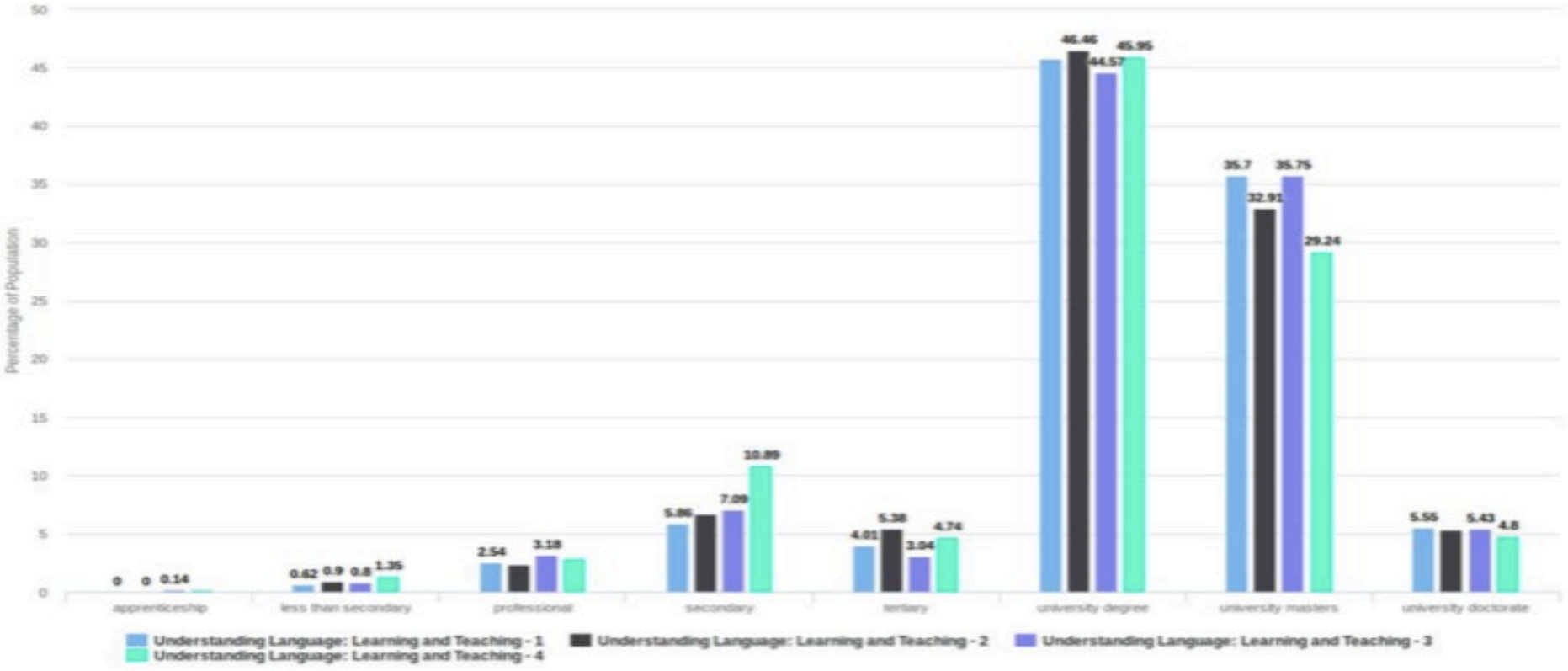
More likely working!
(full time or part time or self-employed)



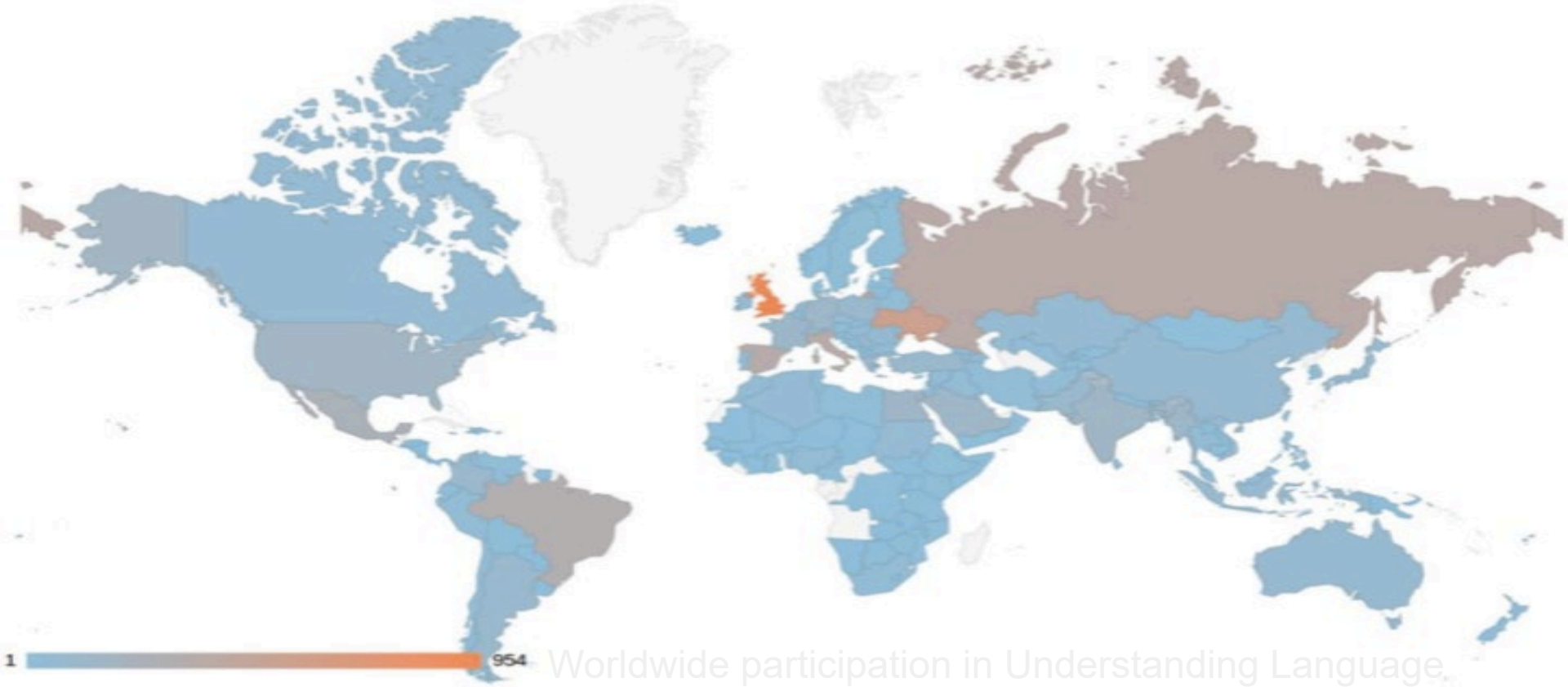
More likely with a degree!
(less likely in secondary school or less)



More likely with a degree!
(less likely in secondary school or less)



Where are the learners?



The typical learner in this MOOC is…

- … more likely to be female
- … more likely to have a degree
- … more likely to be working...
- ... as a teacher
- ... elsewhere (other than in the UK)

…compared to learners in any of our MOOCs

Learning activities per run (1 & 2)

	Run2				Run1			
Types	Week1	Week2	Week3	Week4	Week1	Week2	Week3	Week4
Videos	9	9	6	6	8	10	5	8
Articles	6	3	4	10	5	2	5	7
Audios	1	0	0	0	1	0	0	0
Discussion	1	4	2	2	3	3	2	2
assessment	1	0	0	2	0	0	0	3

Learning activities per run (3 & 4)

	Run4				Run3			
Types	Week1	Week2	Week3	Week4	Week1	Week2	Week3	Week4
Videos	7	8	6	9	8	9	7	6
Articles	7	4	3	6	7	4	3	6
Audios	1	0	0	0	1	0	0	0
Discussion	2	4	3	3	1	3	2	6
assessment	1	0	0	2	1	0	0	2

Learning activities per run (5 & 6)

Types	Week1	Week2	Week3	Week4	Week5
Videos	7	8	6	8	7
Articles	7	4	3	5	3
Audios	1	0	0	0	0
Discussion	2	4	3	3	2
assessment	1	0	0	2	0

Learning activities per run (5 & 6)

Types	Week1	Week2	Week3	Week4	Week5
Videos	7	8	6	8	7
Articles	7	4	3	5	3
Audios	1	0	0	0	0
Discussion	2	4	3	3	2
assessment	1	0	0	2	0

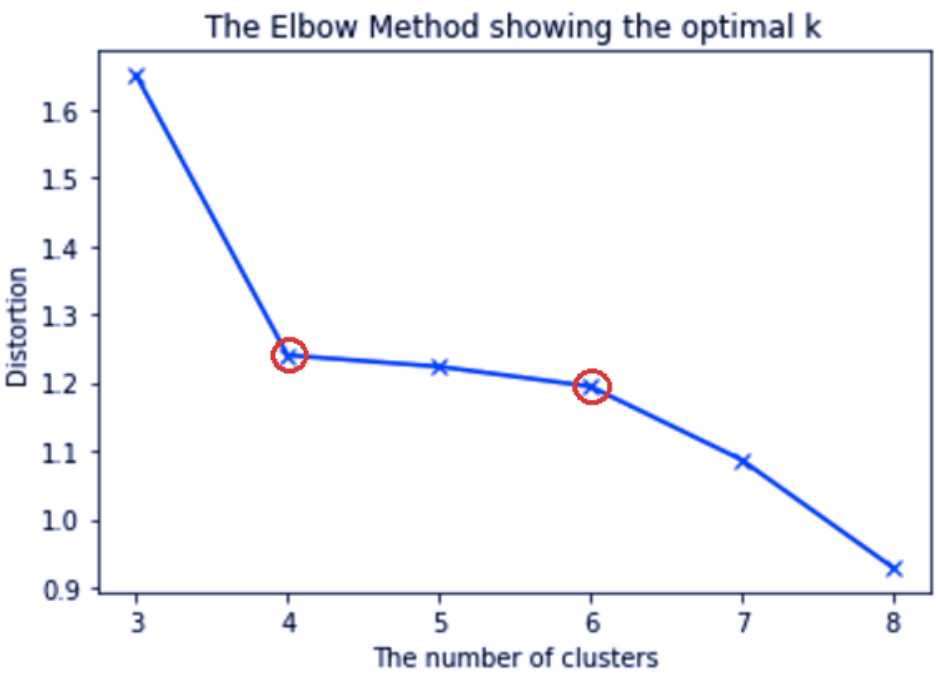
Digging **deeper** into the data
(in search of a characterisation of the diversity of learners)

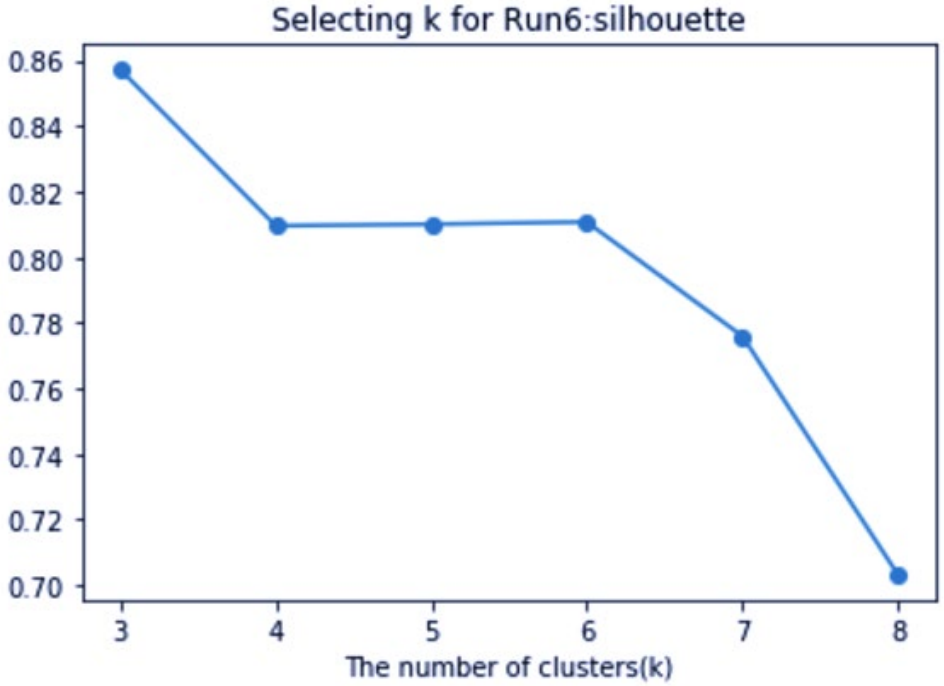
Selecting a clustering algorithm

Algorithms	Silhouette coefficient	Time cost
K-means (k = 6, 100 times initial centroids)	0.81	2.81s
Hierarchical clustering (Euclidean distance, ward linkage)	0.79	6.21s
Affinity propagation (preference = median, damping = 0.5)	0.68	921.83s

K-means was the best performing one!

Why 6 clusters?





Cluster 1: Samplers

Learners **seldom visited the contents of the course**. Normally, they just explored very few videos or articles at the beginning.

Samplers made up the most part of all the runs, accounting for **79.62% - 89.2%** of students.

The majority (50.37% - 70.75%) of learners in the cluster did nothing (70.75% on run 3), but few (9.77% - 24.9%) of them took one step of the course.

Furthermore, there is no one who finished all steps of first week in all the runs, but few of learners completed the assessment of first week. Very few samplers (4.56% - 7.34%) posted comments during the whole course

Cluster 2: Strong starters

The students in the cluster was engaged in the first week, but then explored articles or videos infrequently. **6.2% - 11.25%** learners are strong starters.
Over half (52.97%-86.25%) of them **finished all the activities in the first week.** However, there are only 18.59% - 23.59% learners completed steps in second week in run 4, 5 and 6, and no one finished steps in the second week in run 2 and 3. 51%-59% strong starters posted a comment (72% on run 1) at first week, then the number of comments are reduced week to week.

Cluster 3: Unsocial starters

Learners in the cluster **completed almost all steps** in the course, but typically did **not post any comments**.

Unsocial learners made up **2.3% - 5%** of students, and most of learners (54% - 75%) finished the whole steps in the course. However, there are only 6.7%-29% of learners who posted a comment at each week, those learners who did comment also did not get many likes from other learners.

Cluster 4: “Typical” learners

These do not have any outstanding points compared with popular and complete engagement groups. The group made up 1% - 2.01% learners in all runs except run 5 which does not have this cluster. The majority of learners (75% - 84%) in the cluster **completed all activities**, and 89% - 97% learners **posted comments each week**. Nearly half of learners who posted the number of comments over the average comments (10.18 – 13.69). Furthermore, most of learners who commented steps also got likes from other learners, and the average number of likes is between 7.8 and 16.5.

Cluster 5: Popular

The learners in the cluster finished the most steps of the courses and got **many likes from other learners**. Popular exists in four runs, it made up 0.11% in run 1, 0.08% in run 3, 1% in run 5 and 0.16% in run 6 separately. 82%-90% of them **completed all steps**, and almost all the learners in the group **commented every week**. The obvious feature of the group is the large number of likes they got compared with other clusters` except the completed engagement cluster. The average number of likes is between 30 and 37 in all runs, except 13.5 in run 5

Cluster 6: Completely engaged

The learners in the cluster not only finished all the steps in the course, but also engaged in commenting within steps and received many likes from others.

However, there is very few learners in the cluster, only one learner in the cluster in run 6, who finished all the videos, articles and assessments in the course, and posted 366 comments. Furthermore, this learner got 740 likes from others, but the reason might be the high number of comments.

Distribution of clusters in each run

Cluster	Run1	Run2	Run3	Run4	Run5	Run6
Samplers	79.62%	82.47%	88%	83.37%	89.2%	87.68%
Typical learners	2.01%	1.32%	1%	1.1%	-	1.19%
Strong starters	4.06%	11.25%	6.2%	10.56%	6.62%	7.06%
Unsocial learners	4%	4.91%	2.3%	5%	3.56%	3.9%
Popular	0.11%	-	0.08%	-	1%	0.16%
Complete engagement	-	-	-	-	-	0.01%
Run 1 extra cluster: easy drop-out	10.18%	-	-	-	-	-
Run 2 extra cluster: Suddenly drop-out	-	0.002%	-	-	-	-
Run 3 extra cluster: Mid-drop out	-	-	2.42%	-	-	-

Conclusions

We have studied learners' engagement in the 5-week FutureLearn course "Understanding Language" on its first six offerings (from 2014-2017), facilitated by the British Council in Collaboration with the University of Southampton.

Using three clustering algorithms on the related datasets with only step-activity, enrolments and comments (including number of likes), we have identified six clusters: Samplers, Strong Starters, Unsocial Learners, Popular, Fully Engaged and Atypical Learners. Samplers take the largest part of learners in all runs of the course,

Thank you!