# Who is the Learner: Profiling the Engineering MOOC Student 

T. A. Hennis<br>Researcher<br>hennisNL<br>Amsterdam, The Netherlands<br>E-mail: info@hennis.nl<br>\section*{S. Topolovec}<br>Online Learning Analyst<br>Online Learning | Extension School, Delft University of Technology<br>Delft, The Netherlands<br>E-Mail: S.Topolovec@tudelft.nl<br>O. Poquet<br>PhD Candidate<br>University of South Australia<br>Adelaide, Australia<br>E-mail: skroy001@mymail.unisa.edu.au<br>P. de Vries<br>Research Scientist<br>Delft University of Technology<br>Delft, The Netherlands<br>E-mail: pieter.devries@tudelft.nl

Conference Key Areas: Continuing Engineering Education and Lifelong Learning, Open and Online Engineering Education, Gender and Diversity

Keywords: MOOCs, learner profile, engineering education, learning analytics

## INTRODUCTION

This paper addresses the analysis of demographic characteristics from the participants of DelftX Massive Open Online Courses (MOOCs) in 2014 and 2015. It aims to explore and better understand the learner profile of students in the TU Delft engineering MOOCs, especially in relation to their performance (grades). This study is a continuation and expansion of earlier exploratory research which covered a total of five first year MOOCs at the Delft University of Technology (DUT). This paper follows the same structure and themes addressed in the original paper and can be considered a replication study using a more recent and larger dataset of thirteen MOOCs with an enrolment that ranged from 4,187 to 38,000 students per course. The data and outcomes are analysed and compared with the earlier findings to provide additional insights into the profile of the MOOC learners. The analysis focuses on factors such as age, gender, the level of education, the experience with the domain and with online learning and culture. The main goal is to develop a consistent view of the user population and discuss the outcome in relation to the improvement of the educational design for the MOOC. It should be reiterated, though, that the analysis reported here are to be used and interpreted with caution, since the massive variety of the student
population and the diversity of MOOCs is very different from what higher education has experienced so far [1], [2], [3].

## 1 SUPPORTING LITERATURE

Online education has experienced a revival with the inclusion of MOOCs. This has caused numerous discussions, like on quality of online education compared to campus education, access and openness, and the learner profile. The difference between MOOCs and conventional online courses is that they are free, accessible by students not enrolled at the university, they do not provide students with university credit upon completion, and have unlimited enrolment. These features make MOOCs attractive for a diverse mix of participants from around the world, with a different background, learning experiences and motivations [1].
As such MOOCs are different from other university-sponsored online courses in that they provide education for a fraction of the cost. Therefore, MOOCs have been positioned as an alternative path for non-traditional students as a tool for social mobility [4]. Results indicate that MOOCs seem to be dominated by the same population as the universities in developed countries, which are the college-educated males [5]. Although this group embodies the 'typical' MOOC student, the profile represents less than onethird of the registrants [6]. A better understanding of the user groups is needed to make sense of the experiences with MOOCs to optimize the learning environment. Such understanding might inform sustainable, cost-effective business models that fulfil their potential for more equitable global access.
The DUT MOOCs are so called xMOOCs, which are more like large, lecture-based courses, which lack interaction with and reinforcement from teachers on which some students rely to maintain motivation, especially when they encounter difficulties. Initial evidence, primarily from interviews with MOOC learners (see [7], [8]), suggests that students who are most likely to benefit from MOOCs exhibit self-regulated learning, have flexible work-life schedules, possess digital literacies, and are proficient in English.

### 1.1 Contextual considerations

In this paper the attention goes primarily to the learner profile in an attempt to unravel the invisible learners conveyed by statistical reports. It is very important that institutional strategies are aligned with findings that can help MOOCs to contribute to expanding access, ensuring quality, and reducing cost. The ultimate goal of the analysis of the xMOOCs at DUT has been to support all stakeholders at the institution to make better decisions. These comprise the management team working on a strategic level, involved in organizational issues, business and financial decisions; the teachers and developers designing their courses, choosing media and instruments; the Media Center working on content design and production and the research team dealing with monitoring, datacollection and analysis. It is this context that elevates the usability and value of the findings for the improvement of the MOOC experience and online learning at large.

## 2 METHOD

This paper analysed data from 13 MOOCs by DUT that ran on the edX platform in 2014 and 2015. The course enrolment ranged from 4,187 to 38,029 students, with certification rate between $1.01 \%$ and $5,17 \%$. The grade limit for certification was usually $50-60 \%$, one course standing out with a limit set at $35 \%$. The full overview of courses and available data per course are presented in Table 1. In total, 223,393 cases were included in the analysis; 44,259 students were enrolled in more than one analysed course. However, since the majority of analyses is performed in relation to the grade,
and the overall results do not differ greatly if we exclude repeated students, all students, repeated and unique, were included.

Table 1. Information about analysed courses and available data

|  | Course information |  |  | Data ( $N$, \% of enrolled) |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | enrolled | certified | cert. limit | edX profile | edX grades | pre-survey |
| Credit Risk Management 2015 | 14,995 | 3.07\% | 50\% | $\begin{array}{r} 10,511 \\ 70.10 \% \end{array}$ | $\begin{array}{r} 10,305 \\ 68.72 \% \end{array}$ | $\begin{array}{r} 615 \\ 4.10 \% \end{array}$ |
| Delft Design Approach 2014 | 13,503 | 1.01\% | 60\% | $\begin{gathered} 10,212 \\ 75.63 \% \end{gathered}$ | $\begin{array}{r} 10,005 \\ 74.09 \% \end{array}$ | $\begin{array}{r} 742 \\ 5.50 \% \end{array}$ |
| Drinking Water Treatment 2014 | 10,543 | 2.67\% | 60\% | $\begin{array}{r} 7,867 \\ 74.62 \% \end{array}$ | $\begin{array}{r} 7,725 \\ 73.27 \% \end{array}$ | $\begin{array}{r} 988 \\ 9.37 \% \end{array}$ |
| Solar Energy 2014 | 28,564 | 4.52\% | 58\% | $\begin{gathered} 27,073 \\ 94.78 \% \end{gathered}$ | $\begin{array}{r} 26,727 \\ 93.57 \% \end{array}$ | $\begin{array}{r} 1,576 \\ 5.52 \% \end{array}$ |
| Data Analysis 2015-1 | 28,447 | 6.53\% | 60\% | $\begin{gathered} 26,452 \\ 92.99 \% \end{gathered}$ | $\begin{array}{r} 25,959 \\ 91.25 \% \end{array}$ | $\begin{array}{r} 3,397 \\ 11.94 \% \end{array}$ |
| Functional Programming 2014 | 38,029 | 5.17\% | 60\% | $\begin{array}{r} 27,067 \\ 71.17 \% \end{array}$ | $\begin{array}{r} 26,432 \\ 69.50 \% \end{array}$ | $\begin{array}{r} 3,979 \\ 10.46 \% \end{array}$ |
| Framing 2015 | 34,017 | 2.70\% | 50\% | $\begin{gathered} 27,196 \\ 79.95 \% \end{gathered}$ | $\begin{array}{r} 26,637 \\ 78.30 \% \end{array}$ | $\begin{array}{r} 1,550 \\ 4.56 \% \end{array}$ |
| Responsible Innovation 2014 | 10,824 | 3.66\% | 59\% | $\begin{array}{r} 7,603 \\ 70.24 \% \end{array}$ | $\begin{array}{r} 7,446 \\ 68.79 \% \end{array}$ | $\begin{array}{r} 434 \\ 4.01 \% \end{array}$ |
| Treatment of Urban Sewage 2015 | 10,725 | 4.38\% | 60\% | $\begin{array}{r} 8,336 \\ 77.72 \% \end{array}$ | $\begin{array}{r} 8,183 \\ 76.30 \% \end{array}$ | $\begin{array}{r} 964 \\ 8.99 \% \end{array}$ |
| Technology for Biobased Products 2014 | 9,606 | 3.62\% | 55\% | $\begin{array}{r} 7,238 \\ 75.35 \% \end{array}$ | $\begin{array}{r} 7,117 \\ 74.09 \% \end{array}$ | $\begin{array}{r} 544 \\ 5.66 \% \end{array}$ |
| Topology of Condensed Matter 2015 | 4,187 | 1.17\% | 35\% | $\begin{array}{r} 2,438 \\ 58.23 \% \end{array}$ | $\begin{array}{r} 2,379 \\ 56.82 \% \end{array}$ | $\begin{array}{r} 458 \\ 10.94 \% \end{array}$ |
| Solving Complex Problems 2014 | 32,424 | 4.31\% | 60\% | $\begin{array}{r} 24,804 \\ 76.50 \% \end{array}$ | $\begin{array}{r} 24,378 \\ 75.19 \% \end{array}$ | $\begin{array}{r} 2,183 \\ 6.73 \% \end{array}$ |
| Water \& Climate 2014 | 6,705 | 3.56\% | 60\% | $\begin{array}{r} 6,181 \\ 92.18 \% \end{array}$ | $\begin{array}{r} 6,070 \\ 90.53 \% \\ \hline \end{array}$ | $\begin{array}{r} 1,208 \\ 18.02 \% \end{array}$ |

Notes. Cert. limit - minimum grade needed to receive a certificate. Sample sizes of data sources include all students with at least one datum in a given source group.
This research focuses on a fraction of available data, namely (by data source):

1. edX profile data (self-reported): age (year of birth), gender, educational level
2. edX performance data (automatic): grades
3. Pre-survey data (self-reported; employed in the beginning of the course):

- Certificate importance: How important were the following factors for your choice for this course? - The possibility of earning a Statement of Accomplishment / Verified Certificate
- Domain experience: Do you have professional experience in this field?
- Online experience: How many online classes have you ever taken before? How many online classes have you ever completed?
- Nationality: What is your nationality (country)?
- Preference to work alone / with others: Do you prefer to do this course alone or with others?


## 3 RESULTS \& DISCUSSION

This paper explores demographic characteristics of students in MOOCs, especially in relation to their performance (grades), and aims to provide general insights into who MOOC learners are.

### 3.1 Age

Age was calculated based on stated year of birth in students' profiles (age $=2015-$ year of birth). For age-related analysis, students with stated age below 12 and above 80 years old were excluded for two reasons: (i) they are outliers; (ii) they are more likely falsely reported information. Students had an average age of 32.43 ( $S D=10.68$; median =30), more detailed information is presented in Table 2.
Table 2. Descriptive statistics of age ( $N=184,029$ )

|  | All | Within courses |  |  |  |
| :--- | :--- | :---: | :---: | :---: | :---: |
|  |  | $\min$ | $\max$ | $M$ | $S D$ |
| Min | 12.00 | 12.00 | 15.00 | 12.46 | 0.88 |
| Max | 80.00 | 79.00 | 80.00 | 79.54 | 0.52 |
| Median | 30.00 | 27.00 | 32.00 | 29.15 | 1.46 |
| Mean | 32.43 | 30.30 | 34.61 | 31.90 | 1.36 |
| $S D$ | 10.68 | 8.81 | 11.76 | 10.36 | 0.93 |

There was a significant, but very weak correlation (Spearman) between age and grade, both when we look at all students ( $r_{s}=0.08, p=0.000, N=179,177$ ), and participating students (grade above 0.01) ( $r_{s}=0.07, p=0.000, N=24,690$ ). To compare the results of older and younger students with results from the previous paper, we divided learners into two groups, younger (up to 25 years old) and older (above 25 years old). The density plot (Figure 1) suggests a similar, yet not completely the same picture as in the original paper, since the density of grades around 0.60 (i.e. around certification limit) for "older" learners in the original paper was lower. Nevertheless, results still show that grade density of "older" learners is high close to the grade 1.00 (100\%), while it is lower for "younger" learners. While density plots themselves can differ greatly between the courses, which can be connected to, for example, different certification limits, course design choices, or course difficulty, "older" learners outperform "younger" rather consistently, with only a few exceptions.
There are several possible interpretations, which can as well be intertwined. Older students may be more versed or better able to self-regulate, since studies suggest that self-regulatory abilities develop through childhood and adolescence (overview in e.g. [9]). Older students may also receive higher grades because of different motivation; for example, more of them might want to use learned knowledge in the workplace.


Fig. 1. Density plot of age for students up to 25 ("younger") and above 25 years old ("older") with grade above 0.01 ( $N=24,690$ )
If we investigate differences between different age groups further, we can see that with age, the density peak close to 1.00 increases, and the density peak around the certificate limit decreases for both age groups. We speculate that younger students may be more oriented towards receiving a certificate, so they have something tangible that can be used for the beginning of their careers, while older students may want to simply learn (lifelong learners) and/or use the knowledge (work-motivated), and are subsequently participating in the course more (and therefore receive a higher grade). Indeed, we observed a negative small, but significant correlation between the age and importance of receiving a certificate, i.e. older students rate the possibility of receiving a certificate as a less important factor for their enrolment in the course ( $r_{s}=-0.18$, $p=0.000, N=19,156)$. In the previous paper we also tried to relate the difference in performance in courses to difference in "performance" vs. "mastery orientation". While this is consistent with some studies which show that more mastery-oriented students outperform more performance-oriented peers (e.g. [10]), other research reveals the opposite picture (e.g. [11]). This may be interpreted in light of a study by Grant and Dweck [12], who also found that learning (or mastery) goals predicted higher achievement (in the face of challenge), as well as active coping and sustained motivation. They showed that it is important how performance goals are operationalized: if we define them as outcome goals (wanting to do well), this can be connected both to the learning (mastery) framework and to the performance framework. For example, doing well can either be a means of assessing the mastery of skills, or a means of demonstrating ability [12]. Further research is necessary to relate different (age-related) motivations with performance in MOOCs.
Complementary to that, the low density around the certification limit can be explained by the possibility that most students tend to work towards a 'passing grade', and they may stop making graded assignments (or quit the course in its entirety), when they realize they will not pass, which is supported by other research on assessment in MOOCs [13].

### 3.2 Gender

For gender-related analysis, we included only students that had specified their gender as either female or male. $25.04 \%$ of all students in analysed courses were female. The overview of results is presented in Table 3. The share of enrolled female participants was smaller than the share of male participants in all courses, which is not unexpected
since DUT is a technical university, offering mostly courses that traditionally attract a higher share of male students even on campus. It is also in line with the finding from our previous paper [1], as well as other studies that found the predominance of male learners in MOOCs in general [14]. In a report about two years of MOOCs by HarvardX and MITx, Ho et al. [14] showed a slight increase in female participation, which also seems to hold true for DUT courses; in our previous paper, which looked at five MOOCs in 2014 [1], the overall participation of female students was $19 \%$ ( $15 \%$ certified), while it has risen to $25 \%$ ( $18 \%$ certified) in this report. However, additional data in coming years will be needed to conclude this as a growing trend.
Table 3. Share of female students among enrolled ( $N=189,363$ ) and certified ( $N=8,794$ ) students, and the difference between the two

|  | All | Within courses |  |  |  |
| :--- | :--- | ---: | :--- | ---: | ---: |
|  |  | $\min$ | $\max$ | $M$ | $S D$ |
| \% F enrolled |  | $9.72 \%$ | $36.70 \%$ | $26.89 \%$ | $8.92 \%$ |
| \% F certified |  | $4.04 \%$ | $44.53 \%$ | $22.49 \%$ | $10.13 \%$ |
| Difference $^{\text {s }}$ | $-6.62 \%$ | $-11.32 \%$ | $14.44 \%$ | $-4.40 \%$ | $6.46 \%$ |

Note. ${ }^{\text {a }}=\%$ F certified $-\%$ F enrolled
More concerning is the result that an even lower percentage of female students receives a certificate, which was true in all but one course (18.42\% of certificatereceivers are female). The difference between percentages of enrolled and certified students is presented in Figure 2. The difference between percentages is not significant (Wilcoxon rank-sum test, $W=114, p=0.139$ ), even when we exclude the outlier with a positive change ( $W=103, p=0.078$ ). We must keep in mind, however, that the test was performed on data for only 13 courses.


Fig. 2. Percentage of female students enrolled and certified per course.
In the one course with a positive change, the share of female students that received a certificate was 14.44 percentage points higher than the share of enrolled female students. This may be connected to the subject of the course, which is about design, and may be considered more "female". While this result might seem positive in the light of other courses, it can be argued that this result may at the same time mean that male students were for some reason discouraged, which should also not be the goal of efforts to increase female participation. Also, since only 136 students received a certificate in
the end, the percentage might be a coincidence caused by small numbers. The evaluation of the second run of the course (which was slightly adapted from the first run), which is not included in this analysis, indeed showed that the share of female participants decreased by 2 percentage points from enrolled to certified students [1].

We were also interested in the performance of female vs. male students, and found a significant, but small difference between the two gender groups (Wilcoxon rank-sum test, $W=3,068,600,000, p=0.000, M_{f}=0.04, M_{m}=0.06, N_{f}=46,014, N_{m}=138,400$ ). The difference is small even when only compared with students that engaged in the course, i.e. had a grade above 0.01 ( $W=50,750,000, p=0.000, M_{f}=0.36, M_{m}=0.39$, $N_{f}=5,209, N_{m}=20,237$ ). A look into the density distribution of grades reveals a slightly different picture than was observed in the previous paper: performance of male students did still peak around both the certificate limit and close to 1.0, however, peaks were much less pronounced. Furthermore, the density distribution of female students shows a more even picture. Since we can observe great differences in density distributions between the courses, our results are probably influenced by the courses with the larger number of data. Therefore, we can only conclude that students of different genders perform differently in regard to the overall grade in general, while the comparison of grade distributions is more inconclusive.
A speculation based on the previous paper is that female students might perform less well in courses because they are in general younger; this is not supported by our data. Figure 3 shows that there is an interaction between age and gender, i.e. younger students of both genders perform similarly, however, the older the student, the greater the difference between male and female students is. At this point, we can only speculate why it may be so. It may be that since fewer older female students are employed relative to older male students [15], they may be taking the course more for fun, i.e. not to apply their knowledge or to obtain a certificate to prove they have learned. However, additional research would be needed to understand actual causes behind this result.


Fig. 3. Regression lines with confidence intervals for female and male students with a grade above 0.01 ( $N=25.446$ ).

### 3.3 Education

The majority of students report to have a Bachelor's degree, Master's degree, or high school degree as can be observed in Table 4. These three levels were used for analysing educational background in relation with performance to compare the results to
our previous paper. Unsurprisingly, results mirror the previous finding: the higher the education level, the higher the grade, although the effect is small ( $r_{s}=0.07, p=0.000$, $N=162,835$ ). This is true for both female and male students. A slight interaction with gender can be observed, similar to interaction observed with age, which is not unexpected since there is a significant medium to large correlation between the three educational levels and age ( $r_{s}=0.49, p=0.000, N=22,729$ ).

Table 4. Students' educational levels ( $N=190,154$ )

|  | All |  | Within courses |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: |
|  | $N$ | $\%$ |  | $\min$ | $\max$ | $M$ |
| None | 610 | $0.32 \%$ | $0.20 \%$ | $0.57 \%$ | $0.32 \%$ | $0.11 \%$ |
| Elementary school | 505 | $0.27 \%$ | $0.08 \%$ | $0.37 \%$ | $0.26 \%$ | $0.10 \%$ |
| Junior high school | 2,730 | $1.44 \%$ | $0.47 \%$ | $2.14 \%$ | $1.41 \%$ | $0.51 \%$ |
| High school | 34,350 | $18.06 \%$ | $9.31 \%$ | $23.18 \%$ | $18.23 \%$ | $3.63 \%$ |
| Associate degree | 6,500 | $3.42 \%$ | $1.67 \%$ | $4.86 \%$ | $3.29 \%$ | $0.76 \%$ |
| Bachelor's degree | 71,688 | $37.70 \%$ | $26.87 \%$ | $41.01 \%$ | $36.97 \%$ | $3.84 \%$ |
| Master's degree | 61,168 | $32.17 \%$ | $24.89 \%$ | $40.37 \%$ | $32.01 \%$ | $3.99 \%$ |
| Doctorate | 8,881 | $4.67 \%$ | $2.01 \%$ | $16.60 \%$ | $5.57 \%$ | $3.53 \%$ |
| Other | 3,722 | $1.96 \%$ | $1.25 \%$ | $2.50 \%$ | $1.94 \%$ | $0.36 \%$ |

We have also looked at density distributions for the three educational levels (Figure 4), and unsurprisingly, a density distribution was observed that can be deduced from the relation of grade distribution to age groups, i.e. all three educational levels have a somewhat heightened density around the certification limit, while the peak near grade 1.00 is the highest for master's degree (on average the oldest students), and lowest for high school degree (on average the youngest students).


Fig.4. Density distribution of grades for students with different degrees and a grade above 0.01 ( $N=162,835$ )

### 3.4 Experience

We have looked at two types of experience, namely experience with the subject itself (domain experience), and previous experience with online courses (online experience). Statistics of domain experience are presented in Table 5.

Table 5. Domain experience of students ( $N=18,058$ )

|  | All |  | Within courses |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $N$ | \% | min | max | M | SD |
| no experience | 8,672 | 48.02\% | 21.68\% | 75.43\% | 52.13\% | 13.02\% |
| some experience | 5,444 | 30.15\% | 18.46\% | 45.10\% | 27.54\% | 6.08\% |
| working in a related domain | 2,322 | 12.86\% | 4.46\% | 20.17\% | 12.29\% | 4.33\% |
| working in this domain | 1,620 | 8.97\% | 1.65\% | 14.92\% | 8.04\% | 4.34\% |

A significant, but small correlation was observed between the level of experience and the grade for all students $\left(r_{s}=0.10, p=0.000, N=14,633\right)$, but not for students with a grade above $0.01\left(r_{s}=0.003, p=0.74, N=8,382\right)$, which replicates the results from the previous paper. Again, an interaction with gender is observed, indicating that the effect of the experience level is more pronounced for male students. Male students in all experience levels also receive higher grades than female students.
The question remains, why there is no correlation between the levels of experience and the grade for only at least somewhat active students. We could debate whether levels of experience are truly ordinal, especially if we look at the third level, working in a related domain, which may not necessarily be more than having some experience. However, even if we exclude this level of experience, results only slightly change. In our internal evaluation processes (e.g. DUT, 2015) we regularly observe that more students with less experience drop out early (therefore receive a lower grade), which could explain the difference in grades when looking at the whole population. This would in turn mean that less experienced students that do persist in a course, are not at a disadvantage in regard to their performance. It may be that students with less experience are more easily discouraged by more difficult parts of the course, or they drop out early, because they realize the course was not what they were looking for. Additional research would be needed to understand the reasons why students with less experience dropout more likely and whether this is cause for concern.
Apart from domain experience, we have also looked at previous online experience of students; results are presented in the Table 6.
Table 6. Previous online experience of students ( $N=18,301$ )

|  | All |  | Within courses |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | $N$ | \% | min | max | M | SD |
| enrolled in none | 3,120 | 17.05\% | 11.28\% | 35.33\% | 19.83\% | 7.55\% |
| completed none | 2,319 | 12.67\% | 6.99\% | 14.39\% | 12.06\% | 2.06\% |
| completed 1+ | 12,862 | 70.28\% | 54.67\% | 76.69\% | 68.11\% | 7.40\% |

It seems as if the experience of finishing an online course could be more important than the mere experience of an online course; students that had previously enrolled in no course ( $M=0.23$ ) perform somewhat better than students that enrolled in at least one course, but completed none ( $M=0.17$ ), while the group with experience of completing performs the best ( $M=0.28$ ).


Fig. 5. Density plot for students with different levels of online experience and a grade above 0.01 ( $N=8,482$ )
A similar picture can be concluded from Figure 5. We can observe that students with the most experience have the highest density above the grade 0.50 , and peaks of density around certification limit and close to 1.00 can be observed. Students that enrolled in, but completed none of the courses before may be those that have more difficulties with online learning, for example, with self-regulation. It may also largely consist of students that enrol for different purposes than completing a course. There is a small correlation between the size of the gap between numbers of enrolled and completed courses, and grade, i.e. the larger the gap, the lower the grade ( $r_{s}=-0.13, p=0.000, N=12,142$ ). It must be noted that this gap is an approximation, since the last option for both questions in the survey was "enrolled / completed 4+ courses". Students with previous experience with completing courses may be students who have found their way around learning, and use it according to their needs (for example, to get a certificate, or to master the material). There is a small correlation between the number of completed courses and grade ( $r_{s}=0.12, p=0.000, N=12,212$ ).

### 3.5 Culture

Table 7. Culture of students based on Mensah \& Chen [15] ( $N=17,926$ )

|  | All |  |  | Within courses |  |  |  |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: | ---: |
|  | $N$ | $\%$ |  | $\min$ | $\max$ | $M$ | $S D$ |
| African | 762 | $4.25 \%$ |  | $1.14 \%$ | $7.93 \%$ | $4.60 \%$ | $2.20 \%$ |
| Anglo | 5,294 | $29.53 \%$ |  | $14.74 \%$ | $35.65 \%$ | $25.95 \%$ | $6.79 \%$ |
| Confucian Asian | 566 | $3.16 \%$ |  | $1.68 \%$ | $8.66 \%$ | $3.41 \%$ | $1.80 \%$ |
| Eastern European | 1,894 | $10.57 \%$ |  | $5.17 \%$ | $18.09 \%$ | $9.67 \%$ | $3.82 \%$ |
| German | 1,415 | $7.89 \%$ |  | $4.66 \%$ | $12.07 \%$ | $7.84 \%$ | $2.55 \%$ |
| Latin American | 2,613 | $14.58 \%$ |  | $8.21 \%$ | $25.52 \%$ | $16.53 \%$ | $5.27 \%$ |
| Latin European | 1,866 | $10.41 \%$ |  | $8.28 \%$ | $15.92 \%$ | $10.91 \%$ | $2.44 \%$ |
| Middle Eastern | 558 | $3.11 \%$ |  | $1.49 \%$ | $5.30 \%$ | $3.48 \%$ | $1.13 \%$ |
| Nordic | 467 | $2.61 \%$ |  | $0.62 \%$ | $5.63 \%$ | $2.08 \%$ | $1.47 \%$ |
| South-East Asian | 2,491 | $13.90 \%$ |  | $8.71 \%$ | $20.17 \%$ | $15.52 \%$ | $3.53 \%$ |

As in the previous paper, culture was defined based on self-stated nationality and on the country clustering presented in the GLOBE Extension Study [16], which takes into account several aspects, including racial/ethnic distribution, religious distribution, geographic proximity of the countries, major language distribution, and colonial heritage. We have observed that German and Latin European cultures show the best performance among all groups, with a large peak in grades close to 1.00 . Density plots of cultures consisting of developing countries show a more bell-shaped distribution above the grade 0.50 , rather than specific peaks. However, since some cultural groups consist of a relatively small number of students, and therefore results might be influenced by a specific course, these and other results, connected to culture, must be considered with caution

To further investigate characteristics of different cultures, we looked at preferences for working either alone or with others, results are presented in Figure 6. Preference to work alone is the strongest in various European and Anglo cultures, while it is the least strong in African cultures, followed by Middle Eastern, and Latin American. However, only in African cultures more than one half stated that they prefer to work with others. These results might be connected to differences between individualistic and collectivistic cultures (e.g. [17], [18]).


Fig. 6. Share of students that prefer to work alone compares to students that prefer to work with others per cultural group ( $N=13.119$ ).

## 4 CONCLUSIONS

This paper presented an overview of some characteristics of online students in relation to course performance, which can help course designers and researchers understand better who the learners are in online courses and how they perform, as well as provide ideas for additional research that might be needed to further shed light on online students. Results largely mirror findings presented in the original paper, but we also presented additional insights that broaden our view of this subject. They also highlight the importance of considering different student characteristics when designing online courses, such as age, gender, education, previous experience and culture.
Moreover, our insights showed the importance of not only understanding the overall, average picture of online course participants (such as average age, predominant gender, etc.), but rather, of understanding how different subsets of students with different characteristics can differ, for example, in regards to their performance. This is
important to keep in mind when we build and develop courses, so as to not deprive one group of students in favour of the other. Rather, we should try to further investigate some of the raised concerns, to be able to understand why one group underperforms and if, and how they could be helped through the course, so they can learn more and better. While we tried to provide some possible explanations, further research is needed to answer why these differences occur, and to be able to understand them in a broader context, for example related to other student characteristics, preferences, and attitudes.
However, our research is not without limitations. While we focused on a general, overall picture, we must keep in mind that we observed some important differences between the courses, not just in regard to the representation of characteristics (e.g. share of students of a specific courses), but also in relation to performance. Since some of our analyses are based on a relatively small amount of data, some results might be affected by courses with a predominant number of cases. Therefore, some of the findings may not necessarily hold true inside each of the courses, which might be connected to course design choices (e.g. placement and rules for assessment).

## REFERENCES

[1] Hennis, T., Skrypnyk, O., De Vries, P. Who is the Learner in the DelftX Engineering MOOCs? In proceedings of the International Joint Conference on the Learner in Engineering Education (IJCLEE 2015). San Sebastian, Spain, 2015.
[2\} J. DeBoer, A. D. Ho, G. S. Stump, and L. Breslow, "Changing 'Course': Reconceptualizing Educational Variables for Massive Open Online Courses," Educ. Res., vol. March, no. 43, pp. 74-84, 2014.
[3] J. Ross and S. Bayne, "The pedagogy of the Massive Open Online Course (MOOC): the UK view," High. Educ. Acad., pp. 1-76, 2014.
[4] C. R. Glass, M. S. Shiokawa-Baklan, and A. J. Saltarelli, "Who Takes MOOCs?," New Dir. Institutional Res., vol. 2015, no. 167, pp. 41-55, 2016.
[5] J. Reich, "Rebooting MOOC Research," Science (80-. )., vol. 347, no. 6217, pp. 34-35, 2015.
[6] A. D. Ho, J. Reich, S. O. Nesterko, D. T. Seaton, T. Mullaney, J. Waldo, and I. Chuang, "HarvardX and MITx : The First Year of Open Online Courses," 1, 2014.
[7] G. Veletsianos, "The Invisible Learners Taking MOOCs," Inside Higher Ed, 2015. [Online]. Available: https://www.insidehighered.com/blogs/higher-ed-beta/invisible-learners-taking-moocs. [Accessed: 28-May-2016].
[8] G. Veletsianos, A. Collier, and E. Schneider, "Digging deeper into learners' experiences in MOOCs: Participation in social networks outside of MOOCs, notetaking and contexts surrounding content consumption," Br. J. Educ. Technol., vol. 46, no. 3, pp. 570-587, 2015.
[9] G. Effeney, A. Carroll, and N. Bahr, "Self-regulated learning and executive function: Exploring the relationships in a sample of adolescent males.," Educ. Psychol., vol. 33, pp. 773-796, 2013.
[10] R. A. Mattern, "College Students ' Goal Orientations and Achievement," Int. J. Teach. Learn. High. Educ., vol. 17, no. 1, pp. 27-32, 2005.
[11] J. M. Harackiewicz, K. E. Barron, J. M. Tauer, and A. J. Elliot, "Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation.," J. Educ. Psychol., vol. 94, no. 3, pp. 562-575, 2002.
[12] H. Grant and C. S. Dweck, "Clarifying Achievement Goals and Their Impact.," J. Pers. Soc. Psychol., vol. 95, no. 3, pp. 541-553, 2003.
[13] O. Skrypnyk, T. A. Hennis, and P. De Vries, "Reconsidering Retention in MOOCs : the Relevance of Formal Assessment and Pedagogy," in eMOOCs 2015 -

Proceedings of the Third European MOOCs Stakeholder Summit, 2015, pp. 166172.
[14] A. D. Ho, I. Chuang, J. Reich, C. A. Coleman, J. Whitehill, C. G. Northcutt, J. J. Williams, J. D. Hansen, G. Lopez, and R. Petersen, "HarvardX and MITx: Two Years of Open Online Courses," 10, 2015.
[15] CEDEFOP, "Working and ageing. The benefits of investing in an ageing workforce," Luxembourg, 2012.
[16] Y. M. Mensah and H.-Y. Chen, "Global Clustering of Countries by Culture - An Extension of the GLOBE Study," SSRN Electron. J., 2013.
[17] F. E. Jandt, "Dimensions of Culture," in An introduction to intercultural communication: Identities in a global community, 5th ed., Thousand Oakes, CA: Sage Publications Ltd, 2007, pp. 159-181.
[18] G. Hofstede, "Dimensionalizing Cultures : The Hofstede Model in Context," Online Readings Psychol. Cult., vol. 2, no. 1, pp. 1-26, 2011.

