

# Analysis of the Applicability of General Scaling Laws on Course Size, Completion Rates, and Forum Activity in MOOCs

Thomas Staubitz  
thomas.staubitz@hpi.de  
Hasso Plattner Institute  
Potsdam, Brandenburg, Germany

Max Bothe  
max.bothe@hpi.de  
Hasso Plattner Institute  
Potsdam, Brandenburg, Germany

Mohamed Elhayany  
mohamed.elhayany@hpi.de  
Hasso Plattner Institute  
Potsdam, Brandenburg, Germany

Christiane Hagedorn  
christiane.hagedorn@hpi.de  
Hasso Plattner Institute  
Potsdam, Brandenburg, Germany

Sebastian Serth  
sebastian.serth@hpi.de  
Hasso Plattner Institute  
Potsdam, Brandenburg, Germany

Theresa Zobel  
theresa.zobel@hpi.de  
Hasso Plattner Institute  
Potsdam, Brandenburg, Germany

Christoph Meinel  
christoph.meinel@hpi.de  
Hasso Plattner Institute  
Potsdam, Brandenburg, Germany

## ABSTRACT

In 2017, Geoffrey West published his book *Scale* in which he examined universal laws of scale in different contexts. Inspired by his keynote in 2021's Learning@Scale conference, we investigated the applicability of these laws in the context of Massive Open Online Courses and learners' behavior. We tested these laws on different learning platforms from academic, enterprise and social, and research contexts. In this paper, we examine course characteristics, such as course size, the completion rate, and the forum activity. We observed that the number of issued certificates scales almost identically on all examined platforms, while forum participation scales slightly different on each of the platforms. In the future, we will perform a deeper analysis on the forum behavior that exceeds a mere quantitative analysis.

## CCS CONCEPTS

• **Applied computing** → **Education**; **E-learning**; **Interactive learning environments**; **Distance learning**; **Learning management systems**.

## KEYWORDS

MOOC, Scalability, Scaling Laws, Learning Analytics

### ACM Reference Format:

Thomas Staubitz, Max Bothe, Mohamed Elhayany, Christiane Hagedorn, Sebastian Serth, Theresa Zobel, and Christoph Meinel. 2022. Analysis of the Applicability of General Scaling Laws on Course Size, Completion Rates, and Forum Activity in MOOCs. In *Proceedings of the Ninth ACM Conference on Learning @ Scale (L@S '22)*, June 1–3, 2022, New York City, NY, USA. ACM, New York, NY, USA, 5 pages. <https://doi.org/10.1145/3491140.3528310>



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike International 4.0 License.

L@S '22, June 1–3, 2022, New York City, NY, USA.  
© 2022 Copyright held by the owner/author(s).  
ACM ISBN 978-1-4503-9158-0/22/06.  
<https://doi.org/10.1145/3491140.3528310>

## 1 INTRODUCTION

In 2017, Geoffrey West published his book *Scale* in which he examined *Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies*. He claims that there are universal laws of scale, which can be observed in various domains. In biology, he identified these universal laws regarding the size of animals in comparison to their metabolic rate or the length of their life. Also, in his book he describes similar patterns in economics or sociology [6]. For example, he observed that infrastructural facilities in many cities all over the world follow similar rules, e.g., larger cities need comparably less gas stations than smaller cities. So cities are economies of scale, the infrastructure scales sub-linearly. However, he also observed that socio-economic factors of cities scale differently: they scale in a super-linear way. In a following step, West tried to determine over- and under-performers within the examined cities and their deviation from the average city of their size [6]. West's theses are not completely undisputed. Fix [2] agrees with West in many aspects, but warns that the laws that are postulated might not be as universal as West claims.

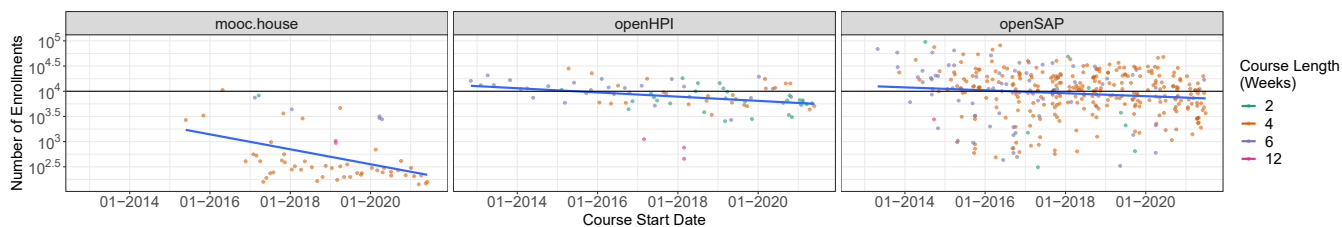
West's Keynote at L@S2021 [7] inspired us to determine whether similar laws can also be applied in the context of Massive Open Online Courses (MOOCs). Our basic hypothesis is that similar to West's observations<sup>1</sup> concerning the scaling of cities, we will be able to show that such scaling laws can also be found in our course data. For this paper, we focused on the following research questions:

RQ1: Is there a scaling law observable between the size of a MOOC and its completion rate?

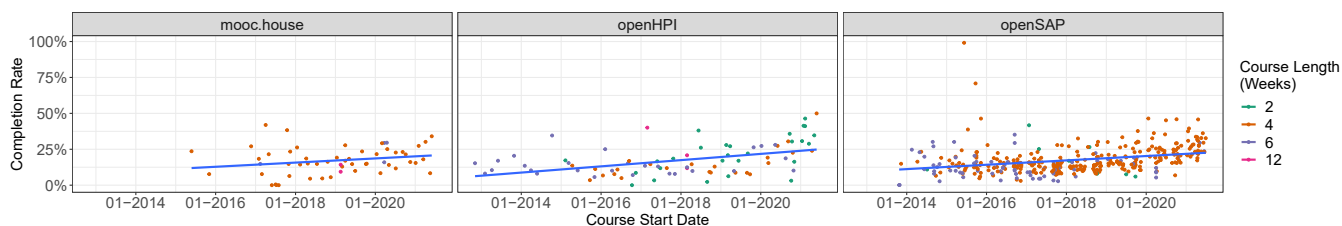
RQ2: Is there a scaling law observable between the size of a MOOC and its forum activity?

## 2 APPROACH

<sup>1</sup>A summary of West's theory can be found here: [https://ted.com/talks/geoffrey\\_west\\_the\\_surprising\\_math\\_of\\_cities\\_and\\_corporations](https://ted.com/talks/geoffrey_west_the_surprising_math_of_cities_and_corporations)



**Figure 1: Development of enrollments and course length (2012-2021).** Each dot represents a publicly accessible course on the examined platform. The black line has been added at the level of 10,000 enrollments as a reference point. (mooc.house: 59 courses, openHPI: 85 courses, openSAP: 372 courses)



**Figure 2: Percentage of course completion/number of graded certificates per course (2012-2021).** Each dot represents a publicly accessible course on the examined platform.

Our long-term evaluation of the laws of scale in the context of MOOCs is based on six learning platforms we operate for several partners in academic, enterprise, and social contexts. The data reflects the time frame from the start of each of the platforms until August 2021. While the learning features offered to course instructors and learners are almost identical on these platforms, the types of the courses and the target audiences differ. At the time of analysis, we operated six platforms for our partners. We excluded two of them as they were quite new and did not offer many courses yet. Therefore, only an insufficient amount of data was available. We excluded another platform which is mainly operated as a knowledge base rather than offering regular course formats.

Hence, for the analysis outlined in this paper, we used the following three platforms: (1) mooc.house is a white-label platform allowing small- to medium-sized partners to offer own courses for narrow target groups. (2) openHPI covers a wide range of topics about information technology (IT). As an academic platform, the courses range from introductory topics to current research topics and mainly target lifelong learners. (3) openSAP also offers courses in the field of IT and focuses on enterprise MOOCs including specific training on the services provided. Besides the different target groups, the platforms differ primarily in their size in terms of the number of courses, registered users, and course enrollments.

We extracted a set of variables for each publicly available course on all three platforms and examined it according to the approach that was used by West. We focused on three variables, which were (1) course size, (2) course completion as measured in the number of graded certificates, and (3) forum activity as measured in the number of forum posts per week. For the course size, we further differentiate between the total number of learners (referred to as *Number of Enrollments*) and the number of learners who enrolled

and, additionally, visited at least one item (referred to as *Shows*). Finally, we measured the distance from a course’s actual performance in comparison to the expected performance it should have according to its size.

### 3 EVALUATION AND DISCUSSION

To provide some context, we start with an overview on the general size of the examined courses and how they have evolved over time. Figure 1 shows the development of the course sizes over the years. Generally, the number of participants per course on all three platforms is slightly decreasing. The trend of participant numbers on the smaller platform mooc.house is not reflecting the actual development of this platform. The white-label platform mooc.house recently gained multiple partners offering smaller courses by design. On all platforms, we observe a tendency towards shorter courses, which can be explained by several factors, such as a better perception of shorter courses by learners and a more suitable schedule for teachers. Figure 2 shows that there is an increase in course completion over the years. This can be attributed either to the tendency towards shorter courses or to another tendency that we observed in our data: the number of new enrollments is decreasing stronger than the course sizes, which indicates that users are coming back and, therefore, are better accustomed to our course format and the platform.

Finally, we observed that the forum activity measured in average posts per participant per course week has significantly increased on openHPI over time, while it stayed constant on the other platforms.

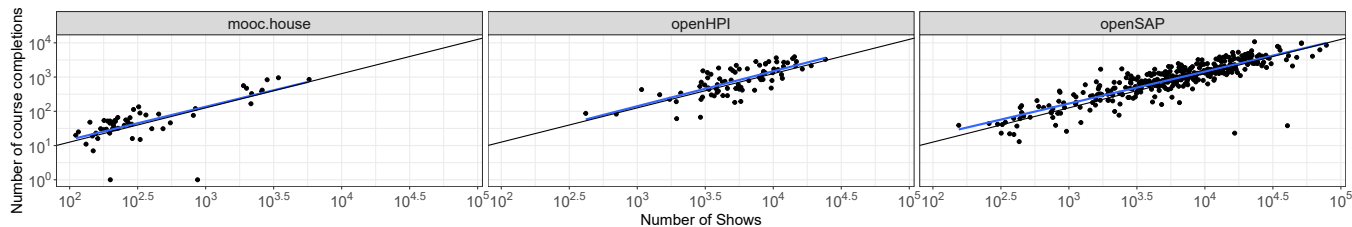


Figure 3: Course size correlated with graded certificates. Logarithmic scales on both axes. The black line represents a slope of 1 and has been added as a reference. Each of the dots represents a publicly available course.

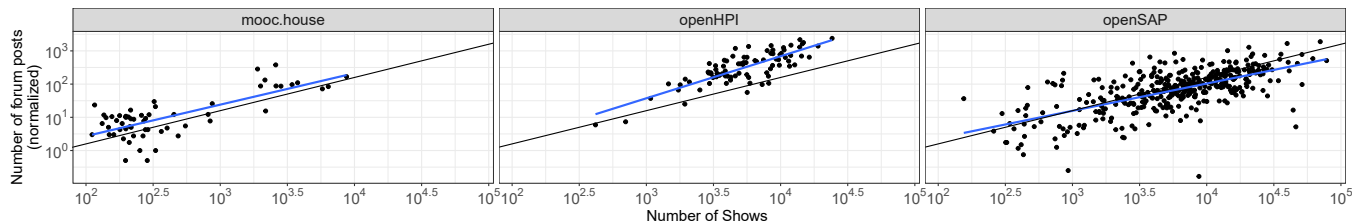


Figure 4: Course size correlated to number of forum posts (normalized by course length). Logarithmic scales on both axes. The black line represents a slope of 1 and has been added as a reference. Each of the dots represents a publicly available course.

### 3.1 Course Size vs. Completion Rate

To answer research question RQ1, we want to examine whether the course size itself has any effect on the course completion rate. Figure 3 represents the number of graded certificates that have been achieved by learners in the courses in relation to number of shows in the course. In accordance with West, a sub-linear slope here would indicate that the number of graded certificates earned in such a course would basically follow the curve of a sigmoid function, meaning that from a certain course size the number of earned certificates would stagnate [6]. Therefore, a sub-linear growth rate could be an explanation that we (with our comparably small courses) generally have higher completion rates than the much larger courses on the major US MOOC platforms, such as *Coursera* or *edX* [3]. A super-linear growth rate, however, would result in a super-exponential growth curve and, at a certain size, would have to result in the collapse of the system (West refers to Malthusian economics) [6].

Figure 3, however, shows that neither a sub- nor a super-linear scaling is the case. On all three platforms analyzed, the slope almost fits the reference line (slope = 1). So, at least in the context of the course formats as they are offered on our platforms, the success of the participants measured through the number of certificates achieved by learners is not dependent in any way of the course size. This implies that these courses can scale more or less indefinitely, without having an effect on the course completion. We double-checked these results with the excluded knowledge-based platform and observed the same behavior there, even though the nature of the courses and exams there differs completely from the three examined platforms.

### 3.2 Course Size vs. Forum Activity

For research question RQ2, we examined the scaling laws of the course size with regard to the forum activity. For this, Figure 4

displays the normalized number of forum posts within the courses correlated with the number of course shows. Our expectation here would be to find a super-linear scaling rate. When more people are active in a forum, posts are answered more quickly and discussions therefore gain traction more quickly, causing activity to rise. Figure 4 shows that this is actually the case for the courses on the openHPI. The slope of the course size in comparison to the forum activity is  $\approx 1.3$ . On the two other platforms, however, the slope is close to 1 with a tendency to a sub-linear scaling. So, according to West, we will basically just run out of human resources when a certain course size is reached. As we know from experience in teaching these courses, forum support can be scaled easily in the first step, e.g., with the help of student assistants, who just filter the noise and direct the more experienced teaching team members towards the tricky content questions and the discussions where the sentiment is taking a negative turn. Obviously, this approach is also limited as student assistants are a limited human resource.

According to West/Malthus [6], humanity is escaping the collapse with ever increasing innovation rates. Innovations enable us to continue a new super-exponential growth curve on top of the previous one right before it would have collapsed. For example, the implementation of a chatbot allowed us to handle the ever-increasing help desk requests on one of the platforms that we excluded from the work for the paper at hand [1]. Another factor that comes in useful in this context is that the nature of the teaching teams' forum tasks differ significantly in small and large courses. In small courses, the teaching teams have to act more frequently, while in larger courses the main task is reading and monitoring since many questions are already answered by other participants.

In our data we have also observed a correlation between the forum activity and the course completion. The slope of this correlation is approximately 0.6. As we have shown that both completion

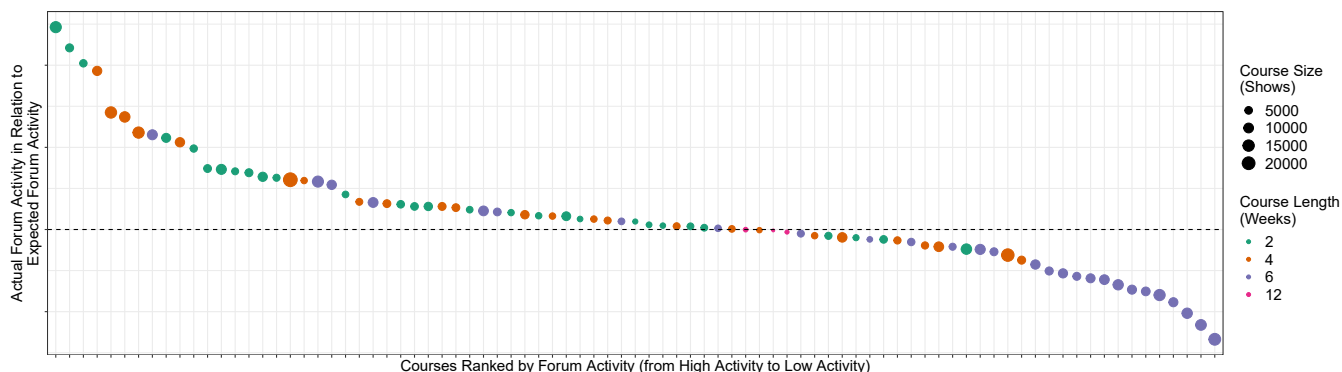


Figure 5: Forum activity performance: openHPI.

rate and forum activity generally are independent from the course size, we dare to say that a higher forum activity has a positive effect on the course completion. We have shown that there is such an effect in previous publications Staubitz et al. [5] where we also have suspected that there are differences in this effect due to the quality of the discussions, but we have not yet addressed this issue in depth. The sub-linear slope indicates that the dependency follows a sigmoid curve, which is plausible. Until the forum activity has gained a certain traction, it does have a low effect on the completion rate. At a certain point, there is an optimal spot where the amount of communication is just right and has a positive effect on the course completion. However, once that point is passed, the amount of communication just gets too much and rather distracts the participants than to contribute to their success.

## 4 OUTLOOK

Besides testing the scaling laws for the course completion and the forum activity, we additionally examined the over- and under-performers on openHPI. Figure 4 shows that there are courses with a higher forum activity than expected for their course size and other courses with a lower forum activity. To measure the over- and under-performance of these courses, we calculated the distance between the value that would be expected due to the number of participants in that course and the forum activity we measured<sup>2</sup>.

Figure 5 shows this performance for each of the examined courses on openHPI relative to the regression line. The courses are ordered by their relative performance (the difference of the actual and the expected activity). The size of the dots represents the number of course participants, while the color encodes the length of the course in weeks. On openHPI, the course length in most cases also has a semantic context: 2-week courses are either workshops or lectures that have been split into shorter modules and 4-week courses are hands-on programming courses. 6-week courses are our original course format, mostly derived from on-campus lectures from HPI professors. The 12-weeks courses are derivatives from the 4-week hands-on courses, which were stretched without adding new content to make them more applicable to be used in a school context. Figure 5 shows that the workshops and other 2-week-formats as

well as the hands-on courses are performing particularly well while the lecture-based MOOCs with few exceptions are located on the lower end.

The examined data set is by far too small to deduct universal laws from this. To do so, similar examinations would need to be done on many other MOOC platforms. A possible approach to do this would be to follow-up with the work that Ruipérez-Valiente et al. [4] have started in this direction.

Another interesting path worth following is a deeper, more qualitative examination of the forum data to get a better grasp why some of the courses are over-performers and others are under-performers in terms of the forum activity and the actual effect of the forum activity on the completion rate of the course. To do so we intend to implement a sentiment analysis of the forum discussions as well as a communication graph within the participants and to examine the effect of certain teaching team interventions as well as the teaching teams forum activity in general. Additionally to the writing activity in the forum, we will add the reading activity to the picture.

## 5 CONCLUSION

The starting point for this research was our interest in Geoffrey West's theories of Scale and if we can find similar laws in the context of MOOCs. For this paper, we focused on the variables *course completion* and *forum activity* in relation to the number of *course participants*. We have shown that both variables neither scale in a sub- nor a super-linear way. Except for the forum activity on openHPI, where we have a super-linear scaling factor of 1.3, the scale factor approximates the slope of 1. Therefore, we can conclude that the course size doesn't influence the completion rate and forum activity.

The results of our study already show the applicability of the scaling law described by West to MOOCs and outline the future aspects we are investigating.

## REFERENCES

- [1] Theresa Elfriede Isa Zobel, Jan Renz, and Christoph Meinel. 2020. Improving the Scalability of MOOC Platforms with Automated, Dialogue-based Systems. In *2020 IEEE Learning With MOOCs (LWMOOCs)*. IEEE, Antigua Guatemala, Guatemala, 42–46. <https://doi.org/10.1109/LWMOOCs50143.2020.9234359>

<sup>2</sup>A similar technique is used in West's work.

- [2] Blair Fix. 2021. A Review of Geoffrey West's 'Scale'. <https://economicsfromthetopdown.com/2021/10/04/a-review-of-geoffrey-west-s-scale/>. [Online; accessed 16-02-2022].
- [3] Katy Jordan. 2015. MOOC Completion Rates: The Data. <http://www.katyjordan.com/MOOCproject.html>. [Online; accessed 16-02-2022].
- [4] José A. Ruipérez-Valiente, Thomas Staubitz, Matt Jenner, Sherif Halawa, Jiayin Zhang, Ignacio Despujol, Jorge Maldonado-Mahauad, German Montoro, Melanie Peffer, Tobias Rohloff, Jenny Lane, Carlos Turro, Xitong Li, Mar Pérez-Sanagustín, and Justin Reich. 2022. Large scale analytics of global and regional MOOC providers: Differences in learners' demographics, preferences, and perceptions. *Computers & Education* 180 (2022), 104426. <https://doi.org/10.1016/j.compedu.2021.104426>
- [5] Thomas Staubitz, Christian Willems, Christiane Hagedorn, and Christoph Meinel. 2017. The gamification of a MOOC platform. In *2017 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, Athens, Greece, 883–892. <https://doi.org/10.1109/EDUCON.2017.7942952>
- [6] Geoffrey B. West. 2017. *Scale: The Universal Laws of Growth, Innovation, Sustainability, and the Pace of Life in Organisms, Cities, Economies, and Companies*. Penguin Press, New York.
- [7] Geoffrey B. West. 2021. Scale: A Window onto Universal Laws of Growth, Living and Dying in Organisms, Cities and Companies. <https://open.hpi.de/courses/hpi-learningatscale2021-emoocs2021/items/1qUnGyV7624M25htSL7Ans>. [Online; accessed 16-02-2022].