Predicting Adherence to an Online Peer Review Course: A Logistic Regression Analysis

Abstract:

An online peer review course was offered to African researchers and professionals in public health aimed at addressing the shortage of trained and qualified reviewers. The course aimed to equip participants with skills to critically assess manuscripts, provide constructive and ethical feedback. Of the 1,086 who enrolled, 651 accessed the course and 203 (31%) gained a certificate. Descriptive analysis revealed that the majority had no prior experience of formal training in peer review, but no differences in this or other measured variables including career stage and publication experience between those who did and did not access the course. Logistic regression analysis among those who accessed the course revealed only one weakly statistically significant predictor of certificate attainment, with males less likely to complete the course (odds ratio 0.68, $p = 0.04$). Other factors, including age, career stage and prior review or publication experience, showed no relationship. Given the high level of interest in this course, a reasonable rate of course completion and a failure to predict access or outcome, we conclude that there is a need for similar courses to be offered to a broad range of academics to help build capacity in peer review skills and practice.

Keywords: Peer Review Training, Online Peer Review Course, Python, Logistic Regression, Scikit-learn, Statsmodel, Diamond Open Access, African Researchers

Introduction

The peer review process plays a crucial role in maintaining the quality and credibility of scholarly publishing (Zimba & Gasparyan, 2021; Tennant & Ross-Hellauer, 2020). It assesses whether scientific findings are adequately supported by evidence to justify publication (Kelly & Adeli, 2014)). By ensuring that research undergoes rigorous evaluation before dissemination, peer review upholds research integrity and strengthens the reliability of published work (Tennant & Ross-Hellauer, 2020). However, many researchers and graduate students, particularly in Africa, receive little to no formal training in peer review, limiting their ability to contribute effectively as reviewers (Willis et al., 2023a; Buser et al., 2023; Publons, 2018; Mulder et al., 2014). This shortfall in peer review capacity is especially concerning for Diamond Open Access (OA) journals, which rely on a dedicated pool of skilled reviewers to maintain high standards while removing financial barriers for both authors and readers (Kuchma & Ševkušić, 2024).

Despite the growing number of OA journals in Africa, many struggle with delays in manuscript processing due to a shortage of qualified reviewers (Kuchma & Ševkušić, 2024). A high proportion of researchers decline review invitations not only because the manuscript is outside their expertise but also due to a lack of confidence or formal training in peer review (Willis et al., 2023b; Publons, 2018). Without sufficient reviewer engagement, manuscript evaluation can become slow and inconsistent, which may hinder the timely dissemination of important research and affect the credibility of OA journals. Addressing this challenge requires structured, accessible, and scalable peer review training programs tailored to African researchers.

Our previous work has indicated both a need to assist journals and researchers to make their work more accessible through open access publications (Agyei et al, 2023) and an expressed need for online courses in open publishing including review skills (Ruredzo et al, 2024).

To address this challenge, the Coalition for Open Access Publishing of Public Health in Africa (COPPHA) launched its first major initiative: an online peer review training course (WACREN, 2024). Developed and delivered by the West and Central African Research and Education Network (WACREN) and Peoples-Praxis, the COPPHA Peer Review Course was designed to equip African researchers with essential peer review skills (WACREN, 2024; Peoples-Praxis, 2024). By strengthening peer review capacity, the initiative seeks to support the growth and credibility of Diamond OA journals in Africa.

Although there is increasing demand for peer review training and a range of self-paced training opportunities available (Willis et al., 2023a), there is limited evidence on completion rates and the factors that influence successful completion of such training. This study aimed to explore course engagement, while identifying demographic and experiential factors associated with course completion. This study addressed the following research questions:

- Participant Engagement: What are the levels of participant engagement in the course, as measured by activities such as access to and completion of the course?
- Prediction of participation and completion: Can demographic and experiential factors measured on enrolment predict access to the course and gaining a certificate of completion?

Methods

Participant Recruitment and Enrollment

Participants were recruited via academic networks, mailing lists, and social media across Africa. Interested individuals completed an application form capturing demographic and professional details, including peer review experience and prior training. Social media handles were optional. A total of 1,088 individuals enrolled and formed a single cohort. Post-enrollment, participants updated their course profiles with demographic data (country, city, and gender).

Course Structure and Delivery

The five-week online course was hosted on the Peoples-Praxis platform (Moodle-based) and designed based on prior LMIC-focused capacity-building models (Heller et al., 2022). Weekly modules, facilitated by experienced instructors, covered:

1. Fundamentals and challenges of peer review
2. AI in peer review
3. Practical peer review skills
4. A hands-on review assignment using a structured proforma

Engagement included participation in three discussion forums, downloading five core resources, and submitting the final peer review assignment. Completion of all components was required to earn a certificate. Optional weekly Zoom sessions enhanced interaction.

Data Collection and Ethical Considerations

Data were collected from application forms completed before the course and learning analytics captured throughout the program. These data included demographics, academic background, and engagement metrics such as logins, downloads, forum activity, assignment submission, and certification status. Participants were informed that anonymized data would be analyzed to assess the course, with the option to withdraw. Two individuals requested data exclusion, resulting in a final dataset of 1,086 participants. No personally identifying information was included in the final analysis.

Data Analysis and Preprocessing

All analyses were conducted in Python (version 3.11) using standard scientific libraries. The data were first cleaned and processed using pandas to handle missing values, filter records, and encode variables. Ordinal and nominal categorical variables were imputed using the mode and encoded appropriately, ensuring the structure and meaning of the data were retained. A unified transformation pipeline using ColumnTransformer ensured consistent preprocessing and reproducibility across all modeling steps.

A logistic regression model was implemented using scikit-learn to examine the association between participant characteristics and certificate attainment, with class weights applied to account for imbalance in the outcome variable. For statistical inference, a parallel model was run using the Logit function from statsmodels, which allowed for the calculation of coefficient estimates, p-values, and confidence intervals. Categorical predictors were interpreted using odds ratios, with values above or below one indicating increased or decreased likelihood of course completion, respectively.

Exploratory Analysis and Handling of Missing Data

Initial analysis included frequency distributions and cross-tabulations to assess the structure and relationships among variables. Categorical associations were examined using Phi and Cramér's V coefficients derived from chi-square tests (Akoglu, 2018). Missing data were managed systematically by imputing the most frequent category for all categorical variables. All records were retained except for one participant identifying as "Other" in the gender field, which was excluded to maintain model integrity, resulting in a final analytical sample of 650 participants who accessed the course.

Modeling Approach

A logistic regression model was used to examine the relationship between participant characteristics and certificate attainment. The model was implemented using scikit-learn's LogisticRegression, with class weights applied to address imbalance in the binary outcome (Zhang et al., 2021: Pedregosa et al., 2011).

To obtain statistical inference, a parallel logistic regression model was fitted using the Logit function from the statsmodels library, leveraging the same preprocessed dataset (Seabold & Perktold, 2010). Unlike scikit-learn, which excels at predictive modeling but lacks inferential statistics capabilities, statsmodels provided coefficient estimates, p-values, and 95% confidence intervals (Abraham et al., 2014; Seabold & Perktold, 2010). The model was estimated through maximum likelihood estimation (MLE), with statistical significance evaluated at an alpha level of 0.05.

Predictive performance was assessed using the scikit-learn model, based on standard classification metrics including precision, recall, and F1-score.

Reproducibility Assurance

The entire analytical process from initial data cleaning to final model validation was implemented with strict reproducibility controls, including fixed random seed initialization (random_state = 42) for all stochastic operations (Baykal et al., 2024; Beam et al., 2020).

Results

Course Access and Participant Characteristics

Of the 1,086 registrants, 651 accessed the course (60%). Table 1 compares the distribution of information from the registration form between those who did and did not access the course site. There were few differences, and no further analysis was conducted on this comparison. The majority of participants were early-career researchers with no prior training in peer reviewing. Around half of the participants had not performed a peer review before but had submitted between one to five papers (1-5) for publication. A majority of the participants disclosed having access to social media, which was the largest source of information about the course.

Table 1: Comparison of Registration Information Between Participants Who Accessed and Did Not Access the Course Site.

		Accessed the course (n = 651)		Did not access the course (n = 435)	
		N	%	N	%
Career Stage	Early Career Researcher	481	75%	309	72%
	Mid-Career Researcher	141	22%	96	22%
	Senior Researcher	18	3%	23	5%
	Total	640		428	
Have social media	Yes	513	79%	320	74%
	No	138	21%	115	26%
	Total	651		435	
Previous training in peer review	Yes	120	19%	62	14%
	No	520	81%	366	85%
	Total	640		428	
Experience as a Peer Reviewer	None	295	46%	232	54%
	Limited (1-3 reviews)	216	34%	110	26%
	Moderate (4-10 reviews)	101	16%	62	14%
	Extensive (more than 10 reviews)	28	4%	24	6
	Total	640		428	
Number of Papers Submitted to Journals	0	167	26%	123	29%
	1-5	292	46%	175	41%
	6-10	87	14%	49	11%
	More than 10	94	15%	81	19%
	Total	640		428	
How did you hear about this course?	Social Media	267	42%	146	34%
	Academic Network	210	33%	161	38%
	Colleagues	79	12%	67	16%
	Professional Organizations	64	10%	44	10%
	Newsletter	20	3%	10	2%
	Total	640		428	

Further analysis focused on the 651 participants who accessed the course site. Table 2 presents the distribution of data collected at enrollment. Around half of the participants were from West Africa, with Nigeria having the highest representation (further analysis compared those from West Africa with those from elsewhere). Two-thirds of those who accessed the course were male and between 30 and 59 years old. Half held a master’s degree, and one-fifth were medical practitioners.

Table 2: Distribution of Demographic Characteristics at Enrollment.

		Number	% of total in each category
West Africa country	Yes	371	57%
	No	279	43%
Country*	Nigeria	240	37%
	Ghana	112	17%
	Kenya	78	12%
	Ethiopia	44	7%
	Uganda	32	5%
	Cameroon	26	4%
	Zambia	14	2%
	Zimbabwe	10	2%
	Total	650	
West Africa	Yes	371	57%
	No	279	43%
Gender	Male	349	63%
	Female	206	37%
	Total	555	
Age	Less than 30	170	33%
	30-59	337	66%
	60+	6	1%
	Total	513	
Highest education	High school	8	1%
	Undergraduate degree	142	27%
	Masters	254	47%
	Doctorate	131	24%
	Total	535	
Current occupation	Student	119	22%
	Medical practitioner	98	18%
	Other health professional	201	38%
	Others	117	22%
	Total	535	

*Data for those countries with 10 or more participants

Participant Engagement and Course Activity Completion

There were three discussion forums to which posting was a requirement for a certificate of completion, and 5 files for which downloading was also a requirement for a certificate. Of the 1086 participants, 234 posted to all three forums and an extra 69 posted to at least one forum, while 260 participants downloaded all of the files and an extra 88 downloaded at least one file. Table 3 shows the number of participants who completed the various course activities. 45% of the participants posted to the first forum, which was the first of the required course activities. There was drop-off during the course, with 32% submitting an assignment and 31% gaining a certificate of completion.

Table 3: Course Activities and the Number and Percentage of Participants Who Completed Each Activity

Course activities	Title	N completed	% of 651 who accessed the course
Discussion forums	1. Reflection on the peer review process and its challenges	295	45%
	2. Reflection: AI and the review process	264	41%
	3. Forum on how to review a journal article	239	37%
	Posted to all three forums required for a certificate	234	36%
Downloads	PREreview reviewer guide	324	50%
	PREreview bias reflection guide	291	45%
	PREreview review assessment rubric	287	44%
	Published paper for us to practice peer review	288	44%
	Review proforma (from COPPHA) to submit a peer review	281	43%
	Downloaded all 5 files required for a certificate	260	40%
	Model answer*	193	30%
Assignment submitted		210	32%
Certificate awarded		203	31%

* Access to the model answer was restricted to participants who submitted an assignment and was not required to obtain a certificate of completion.

Predicting Course Completion

To investigate the potential for predicting course completion, demographic features collected during registration and enrollment were analyzed. Figure 1 presents a correlation matrix heatmap showing relationships between demographic characteristics, course engagement metrics, and outcomes. There was a strong correlation between certificate attainment and course activities, given that earning a certificate was contingent on completing these activities. Moderate correlations were found between age and highest education level ($r = 0.43$), and between the number of papers submitted and peer reviews performed ($r = 0.40$). Most other demographic features showed weak correlations.

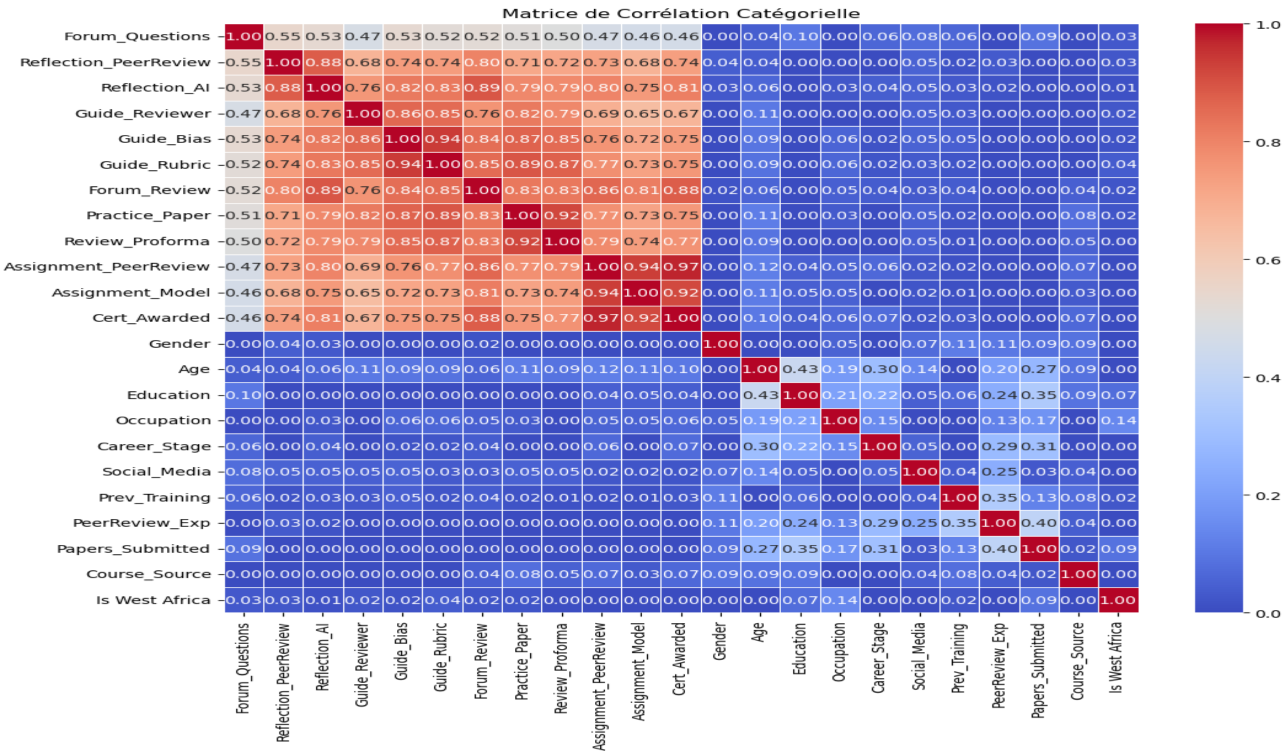

Figure 1. Correlation Matrix Heatmap of Demographic and Engagement Variables

Model Performance in Predicting Certificate Attainment (Scikit-learn)

The scikit-learn logistic regression model's performance in predicting certificate attainment was evaluated on a sample of 650 individuals, with 203 having gained a certificate (true positives) and 447 not gaining a certificate (true negatives). Table 4 shows that the model correctly identified 119 of the 203 certificate earners (a sensitivity of 59%) and 268 of the 447 non-earners (a specificity of 60%). However, it misclassified 179 non-earners as certificate earners (false positives) and 84 certificate earners as non-earners (false negatives).

The positive predictive value (PPV) was 40%, with 119 of the 298 participants predicted to gain a certificate actually doing so. The negative predictive value (NPV) was 76%, with 268 of the 352 participants predicted not to gain a certificate correctly identified as non-earners. The model achieved an overall accuracy of 60%.

Table 4. Scikit-learn Logistic Regression Classification Results for Certificate Attainment

		Truth (gained certificate or not)		
		Positive	Negative	
Test (predicted by logistic regression model)	Positive	119	179	298
	Negative	84	268	352
		203	447	650

Table 5 presents the Statsmodels results. Among all predictors included in the model, gender was the only statistically significant variable associated with certificate attainment. Male participants were significantly less likely to earn a certificate compared to female participants (coefficient = –0.383, p = 0.044; 95% CI: –0.755 to –0.011), corresponding to a 31.8% reduction in the odds of earning a certificate (OR = 0.682).

Table 5: Logistic Regression Results for Predictors of Certificate Attainment (Statsmodels)

Variable	Coefficient	Standard Error	p-Value	Odds Ratio	95% Confidence Interval (Lower)	95% Confidence Interval (Upper)
Based in West Africa	-0.0030	0.176	0.986	0.9970	-0.349	0.343
Career stage	-0.0945	0.190	0.620	0.9098	-0.468	0.279
Education	-0.1495	0.169	0.375	0.8611	-0.480	0.181
Occupation: Others	-0.1616	0.300	0.590	0.8508	-0.750	0.427
Course Source: Colleagues	-0.1630	0.285	0.567	0.8496	-0.722	0.396
Occupation: Other health professional	-0.2838	0.247	0.251	0.7527	-0.768	0.201
Course Source: Social Media	-0.3215	0.202	0.112	0.7251	-0.718	0.075
Gender: Male	-0.3829	0.190	0.044	0.6818	-0.755	-0.011
Age	-0.4046	0.247	0.101	0.6673	-0.888	0.079
Occupation: Student	-0.5300	0.309	0.086		-1.136	0.076
Course Source: Professional Organizations	-0.5412	0.328	0.099	0.5883	-1.185	0.102
Course Source: Newsletter	-0.9202	0.589	0.118	0.3985	-2.074	0.233
Papers Submitted to Journals	0.0498	0.158	0.753	1.0510	-0.261	0.360
Peer Review Experience	0.0597	0.129	0.644	1.0616	-0.193	0.313
Previous Peer Review Training	0.1255	0.236	0.595	1.1336	-0.337	0.588
Uses Social Media	0.1376	0.218	0.527	1.1476	-0.289	0.564
Constant	0.6039	0.704	0.391	1.8297	-0.777	1.984

Sample size: N = 650; Coefficients represent log-odds of certificate attainment; Odds ratios provided for interpretability; Statistical significance assessed at $\alpha = 0.05$ (*$p < 0.05$).

Discussion

This study shows the broad interest in peer review training among researchers in Africa, with 1,086 individuals enrolled and 60% (n=651) accessing the course site. The demographic diversity of participants, ranging from early-career to senior researchers, points to the widespread need for structured peer review education. The majority of participants who enrolled in the course were early-career researchers (ECRs), accounting for 74% (n=790) of all registered participants. Among those who accessed the course, 75% (n=481) were ECRs, while they also made up 72% (n=309) of those who did not access the course. In contrast, mid-career researchers comprised 22% (n=237) of total registrants, and senior researchers represented only 4% (n=41). These findings suggest a strong interest in the course among ECRs, who were the dominant group across both active and non-active participants.

In terms of gender representation, 63% of participants identified as male while 37% identified as female. This reflects an underrepresentation of female participants and aligns with the broader trend of gender imbalance in MOOC participation (Macleod et al., 2014). Future course offerings should consider targeted strategies to improve female participation.

Among those who accessed the course, prior training in peer review was low; only 19% had received formal training, while 81% had none, demonstrating a need for a course such as this. However, a substantial proportion had experience with academic publishing, with 75% having submitted at least one paper to a journal (46% had submitted 1–5 papers, 14% had submitted 6–10, and 15% had submitted more than 10). Similarly, among the 435 who did not access the course, only 14% had prior training in peer review, while 85% had none. However, a substantial proportion had experience with academic publishing, with 71% having submitted at least one paper to a journal (41% had submitted 1–5 papers, 11% had submitted 6–10, and 19% had submitted more than 10). This shows that the majority of participants who accessed the course had some familiarity with academic publishing but lacked formal peer review training, depicting their need for the course.

Among those who accessed the course and disclosed their experience as peer reviewers (n=640), 46% had no prior experience in peer review, while 34% had limited experience (1–3 reviews), 16% had moderate experience (4–10 reviews), and only 4% had extensive experience (more than 10 reviews). Similarly, among those who did not access the course but disclosed their peer review experience (n=428), 54% had no experience, 26% had limited experience, 14% had moderate experience, and only 6% had extensive experience. The high proportion of participants, nearly half of course takers and over half of non-takers, with little to no peer review experience points to a critical training gap among African researchers.

The shortfall in training is compounded by the absence of formal peer review training for graduate students and ECRs at most higher institutions of learning and research in Africa and globally (Stupacher, 2025; Buser et al., 2023; Thomas, 2023; Isaacson et al., 2020; Xu et al., 2016; Mulder et al., 2014;; Freda et al., 2009). Without structured opportunities to develop peer review competency, future academics, many of whom will become the next generation of reviewers, may enter academia unprepared for reviewer roles to help strengthen research integrity (Isaacson et al., 2020). This lack of preparation could compromise the quality and consistency of peer review, reduce the pool of qualified reviewers, and perpetuate a cycle of reviewers with underdeveloped skills and poor reviews (Woldeamanuel, 2025; Willis et al., 2023b; Isaacson et al., 2020; Roediger, 2007). Beyond upholding research integrity, developing peer review skills offers significant benefits to ECRs and senior academics. Participating in peer review strengthens assessment skills, cultivates analytical thinking, establishing these competencies as vital components of an academic's skill set (Stupacher, 2025; Isaacson et al., 2020).

For African Diamond Open Access (OA) journals, which rely on volunteer reviewers and do not charge author fees (Kuchma & Ševkušić, 2024), the shortfall in training poses a particular challenge. Without a sufficiently trained and diverse pool of reviewers, these journals may struggle with prolonged peer review timelines, hindering the timely dissemination of important research findings.

While 651 participants accessed the course, completion rates for various activities indicate varying levels of commitment and engagement. The discussion forums were designed to encourage critical reflection and engagement with key topics in peer review. The highest participation rate (45%) was observed in the first discussion forum on the peer review process and its challenges. Participation declined in the second forum on AI in peer review (41%) and further in the forum on how to review a journal article (37%). Only 36% of participants contributed to all three forums required for gaining a certificate.

A similar pattern was observed in participants' engagement with course materials. The PREreview reviewer guide was the most downloaded resource (50%), followed by the PREreview bias reflection guide (45%), the PREreview review assessment rubric (44%), and the published public health preprint for practicing peer review (44%) Access to the model answer was not a requirement for earning a certificate of completion. 40% of participants downloaded all five required files for certification.

Varying levels of engagement with course materials reflect differences in participants' motivation and learning preferences (Kahan et al., 2017). Many MOOC learners enroll to explore topics or gain specific knowledge rather than complete the full course (Hew & Cheung, 2014). Kizilcec et al. (2013) identified four common engagement patterns: auditing, completing, disengaging, and sampling. A study of 24,412 participants in UNED MOOCs found that 1.2% intended only to consult materials and 2.4% planned limited activity without aiming to complete the course (Gil-Jaurena et al., 2017). Partial participation is valuable for learners with targeted goals, as many achieve their objectives through selective engagement (Kahan et al., 2017; Hew & Cheung, 2014; Lebar, 2014; Reich & Ho, 2014).

Furthermore, the requirement to submit a peer review assignment as a prerequisite for earning a certificate may have also filtered out less committed participants. Additionally, the choice of a public health preprint (Alabi et al., 2024) for the assignment may have posed a challenge, as not all participants were researchers or professionals in public health. Limited and unreliable internet access, which remains a common barrier across many regions in Africa, may have further affected participants' ability to fully engage with the course content and complete key requirements (Gunter, 2025; Mukuni, 2019). This combination of factors may have impacted the final course completion rate. Of the 651 participants who accessed the course, 203 submitted the peer review assignment, completed the course, and earned a certificate of completion, resulting in a 31% completion rate. However, the definition and calculation of completion can vary across different MOOCs. Jordan (2015) identified multiple definitions of MOOC completion, with earning a certificate being the most commonly used criterion. The 31% completion rate for the course should be viewed within the broader trends of MOOC participation, completion, and dropout rates (Reich & Ruipérez-Valiente, 2019; Goopio & Cheung, 2020; Zhang & Zhang, 2021; Huang et al., 2023). While completion rates in MOOCs are typically lower, often averaging between 7% and 10% (Gütl et al., 2014; Fu et al., 2021), a rate exceeding 15% is considered uncommon (Jordan, 2014; Hollands & Tirthali, 2014; Rai & Chunrao, 2016). Therefore, a 31% completion rate can be considered strong, reflecting the course's design, delivery method, weekly facilitation, and optional Zoom sessions that supported learner engagement.

The logistic regression model achieved 60% overall accuracy, correctly classifying 387 participants (119 true positives, 268 true negatives). However, 179 false positives (non-earners predicted as earners) and 84 false negatives (earners predicted as non-earners) highlight notable misclassification. With both sensitivity and specificity at 59 to 60%, the model demonstrated a weak ability to differentiate between certificate earners and non-earners. The positive predictive value (PPV) was only 40%, meaning that less than half of those predicted to earn certificates actually did. By contrast, the negative predictive value (NPV) was stronger at 76%, reflecting the model’s better performance in identifying non-earners.

The moderate 60% accuracy and low PPV in this model suggest that participant characteristics alone (such as age, gender, education level, course source, prior training, or experience with peer review) are not sufficient predictors of course completion. This aligns with broader literature indicating that in self-regulated learning, motivation, digital literacy, and internet access reliability often not captured in survey-based predictors are key determinants of MOOC success (Getenet et al., 2024; Alfayez, 2024; Badali et al., 2022; Revathy et al., 2022). To improve future predictive models, applying data-level approaches such as SMOTE or undersampling may help mitigate class imbalance (Mulyani et al., 2019), while incorporating behavioral data and explanatory variables such as login frequency, participant motivation, digital literacy, and internet access reliability could significantly enhance model performance.

Results from the statsmodels (Table 5) revealed gender as the only significant predictor, with other factors showing varied associations. With 16 variables entered into the logistic regression model there is a high probability of at least one being statistically significant, so we should not read too much into this observation, nor use this as a basis for any future enrolment criteria.

The course provides value beyond certificate attainment. All 651 participants who accessed the course had access to most of the course content and resources, including discussion forums. However, the model answer was only available to those who submitted the final peer review assignment. This approach allowed even non-completers to benefit from the course based on their individual learning needs. This shows the course’s capacity to offer value beyond completion or deadlines, aligning with the flexible learning objectives commonly associated with MOOCs. Additionally, the course is published under a Creative Commons (CC) license, enabling others to use or adapt it for their audiences (Peoples-Praxis, 2024).

Limitations

Participant recruitment was conducted primarily through social media, which may have introduced selection bias (Oudat & Bakas, 2023). This method likely favored individuals with regular internet access and online engagement, which may have limited diversity in the participant pool. The study sample was overrepresented by male participants and ECRs, suggesting that the recruitment strategy may not have effectively reached a balanced audience across gender and career stages.

The course design and engagement also presented areas for improvement. Declining participation in discussion forums and a completion rate of 31% indicate possible barriers such as unreliable internet access or misalignment between course content and participants' backgrounds. In particular, the peer review assignment, which focused on a public health preprint, may not have been suitable for participants from other disciplines. Future iterations of the course should align assignments with participants' backgrounds to increase relevance and engagement, which may improve the overall completion rate.

Certificates of completion were awarded based on the completion of key course activities and submission of the peer review proforma, without evaluating the quality of the reviews submitted. This approach limits the ability to assess actual skill acquisition. To better measure peer review competence, future courses should implement quality assessment and scoring of submitted reviews. A companion paper has been submitted to explore the reviews submitted as assignments in this course titled: Assessing the quality of peer reviews in Public Health: use of a standardised proforma among online course participants.

In addition, the study relied on self-reported demographic and experiential data, which may be subject to recall bias or inaccuracies (Koller et al., 2023). Future studies should consider implementing mechanisms to validate self-reported information to enhance data reliability.

The logistic regression model excluded key variables like internet reliability, digital literacy, and motivation, likely reducing its predictive power and limiting interpretation of certificate attainment.

Conclusion

The participation of 651 individuals in the course demonstrates the potential of an online course to train large numbers of African researchers and professionals to enhance peer review capacity. This is a critical need for all scientific publications, including those of our special interest of Diamond Open Access (OA) journals, which rely on volunteer reviewers. With a 31% certificate attainment rate (203/651 participants), the course outperformed typical MOOC completion rates, reflecting its relevance to researchers and professionals seeking to strengthen their peer review skills.

The model revealed only one weakly significant predictor of outcome, and combined with the model's moderate overall accuracy (60%) and low positive predictive value (40%) indicate limited predictive power. This, and the observation that approximately 83% of the registrants had no prior formal training in peer review suggests that course such as these have a place in peer review capacity building for a wide range of professionals.

Data Availability Statement:

The dataset used in this study has been deposited on Zenodo and is accessible at: https://doi.org/10.5281/zenodo.15733822.

References:

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. Frontiers in neuroinformatics, 8, 14. https://doi.org/10.3389/fninf.2014.00014

Agyei, D.D., Sangare, M., Anyiam, F.E., Ruredzo, P.I.M., Warnasekara, J. and Heller, R.F. (2023) 'Open access publication of public health research in African journals', <i>Insights: the UKSG journal</i>, 36(1), p. 6. Available at: https://doi.org/10.1629/uksg.605.

Akoglu H. (2018). User's guide to correlation coefficients. Turkish journal of emergency medicine, 18(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Alabi, M., Dougherty, L., Etim, E.-O., & Adedimeji, A. (2024). Quality of counselling, exposure to vaccination messages, and caregivers' knowledge on the uptake of Penta vaccine in six northern Nigerian states. medRxiv. https://doi.org/10.1101/2024.08.09.24311716

Alfayez, A. (2024). Effects of internet connection quality and device compatibility on learners' adoption of MOOCs. *Educational Technology & Society, 27*(2), 270–283. https://doi.org/10.30191/ETS.202404_27(2).RP12

Badali, M., Hatami, J., Banihashem, S. K., & others. (2022). The role of motivation in MOOCs' retention rates: A systematic literature review. *Research and Practice in Technology Enhanced Learning, 17*(5). https://doi.org/10.1186/s41039-022-00181-3

Baykal, P. I., Łabaj, P. P., Markowetz, F., Schriml, L. M., Stekhoven, D. J., Mangul, S., & Beerenwinkel, N. (2024). Genomic reproducibility in the bioinformatics era. Genome biology, 25(1), 213. https://doi.org/10.1186/s13059-024-03343-2

Beam, A. L., Manrai, A. K., & Ghassemi, M. (2020). Challenges to the Reproducibility of Machine Learning Models in Health Care. JAMA, 323(4), 305–306. https://doi.org/10.1001/jama.2019.20866

Buser, J. M., Morris, K. L., Dzomeku, V. M., Endale, T., Smith, Y. R., & August, E. (2023). Lessons learnt from a scientific peer-review training programme designed to support research capacity and professional development in a global community. *BMJ Global Health, 8*, e012224. https://doi.org/10.1136/bmjgh-2022-012224

Oudat, Q., & Bakas, T. (2023). *Merits and pitfalls of social media as a platform for recruitment of study participants*. *Journal of Medical Internet Research, 25*, e47705. https://doi.org/10.2196/47705

Freda, M. C., Kearney, M. H., Baggs, J. G., Broome, M. E., & Dougherty, M. (2009). Peer reviewer training and editor support: results from an international survey of nursing peer reviewers. Journal of professional nursing : official journal of the American Association of Colleges of Nursing, 25(2), 101–108. https://doi.org/10.1016/j.profnurs.2008.08.007

Fu, Q., Gao, Z., Zhou, J., & Zheng, Y. (2021). CLSA: A novel deep learning model for MOOC dropout prediction. *Computers & Electrical Engineering, 94*, 107315. https://doi.org/10.1016/j.compeleceng.2021.107315

Getenet, S., Cantle, R., Redmond, P., & others. (2024). Students' digital technology attitude, literacy and self-efficacy and their effect on online learning engagement. International Journal of Educational Technology in Higher Education, 21(3). https://doi.org/10.1186/s41239-023-00437-y

Gil-Jaurena, I., Callejo-Gallego, J., & Agudo, Y. (2017). Evaluation of the UNED MOOCs implementation: Demographics, learners' opinions and completion rates. *The International Review of Research in Open and Distributed Learning, 18*(7). https://doi.org/10.19173/irrodl.v18i7.3155

Goopio, J., & Cheung, C. (2020). The MOOC dropout phenomenon and retention strategies. Journal of Teaching in Travel & Tourism, 21(2), 177–197. https://doi.org/10.1080/15313220.2020.1809050

Gunter, A. (2025). The geography of distance education: Spatial disparities, accessibility, and impact across place. *South African Geographical Journal*, *1–19*. https://doi.org/10.1080/03736245.2025.2472653

Gupta, P., & Bagchi, A. (2024). *Introduction to Pandas*. In *Essentials of Python for Artificial Intelligence and Machine Learning* (Synthesis Lectures on Engineering, Science, and Technology). Springer. https://doi.org/10.1007/978-3-031-43725-0_5

Gütl, C., Rizzardini, R. H., Chang, V., & Morales, M. (2014). Attrition in MOOC: Lessons learned from drop-out students. In L. Uden, J. Sinclair, Y. H. Tao, & D. Liberona (Eds.), *Learning technology for education in cloud. MOOC and big data* (pp. 37–48). Springer. https://doi.org/10.1007/978-3-319-10671-7_4

Han, S., & Kwak, I. Y. (2023). Mastering data visualization with Python: practical tips for researchers. Journal of minimally invasive surgery, 26(4), 167–175. https://doi.org/10.7602/jmis.2023.26.4.167

Heller, R.F., Barrett, A., Oaiya, O., Heller, J. and Madhok, R. (2022) 'Final Report of a Novel and Successful Online Public Health Capacity Building Experiment – Peoples-uni', <i>Open Praxis</i>, 14(1), p. 83–92. Available at: https://doi.org/10.55982/openpraxis.14.1.150.

Hew, K. F., & Cheung, W. S. (2014). Students' and instructors' use of massive open online courses (MOOCs): Motivations and challenges. *Educational Research Review, 12*, 45–58. https://doi.org/10.1016/j.edurev.2014.05.001

Hollands, F. M., & Tirthali, D. (2014). MOOCs: Expectations and reality (Full report). Center for Benefit-Cost Studies of Education, Teachers College, Columbia University. http://cbcse.org/wordpress/wp-content/uploads/2014/05/MOOCs_Expectations_and_Reality.pdf

Huang, H., Jew, L., & Qi, D. (2023). Take a MOOC and then drop: A systematic review of MOOC engagement pattern and dropout factor. Heliyon, 9(4), e15220. https://doi.org/10.1016/j.heliyon.2023.e15220

Isaacson, R. A., Bay, S. N., & McCarty, M. M. (2020). Supporting the next generation of researchers: GENETICS peer review training program. *Science Editor, 43*(3), 77. https://doi.org/10.36591/SE-D-4303-77

Jordan, K. (2014). Initial trends in enrolment and completion of massive open online courses. The International Review of Research in Open and Distributed Learning, 15(1). https://doi.org/10.19173/irrodl.v15i1.1651

Jordan, K. (2015). Massive open online course completion rates revisited: Assessment, length and attrition. The International Review of Research in Open and Distributed Learning, 16(3). https://doi. org/10.19173/irrodl.v16i3.2112

Kahan, T., Soffer, T., & Nachmias, R. (2017). Types of Participant Behavior in a Massive Open Online Course. The International Review of Research in Open and Distributed Learning, 18(6). https://doi.org/10.19173/irrodl.v18i6.3087

Kelly, J., Sadeghieh, T., & Adeli, K. (2014). Peer Review in Scientific Publications: Benefits, Critiques, & A Survival Guide. EJIFCC, 25(3), 227–243.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13)* (pp. 170–179). Association for Computing Machinery. https://doi.org/10.1145/2460296.2460330

Koller, K., Pankowska, P. K., & Brick, C. (2023). Identifying bias in self-reported pro-environmental behavior. *Current Research in Ecological and Social Psychology, 4*, 100087. https://doi.org/10.1016/j.cresp.2022.100087

Kuchma, I., & Ševkušić, M. (2024). Landscape of no-fee open access publishing in Africa. Zenodo. https://doi.org/10.5281/zenodo.12792474

LeBar, M. (2014). MOOCs – Completion is not important. Forbes. Retrieved from http://www.forbes.com/sites/ccap/2014/09/16/moocs-finishing-is-not-the-important-part/

Macleod, H., Haywood, J., Woodgate, A., & Alkhatnai, M. (2014). Emerging patterns in MOOCs: Learners, course designs and directions. *TechTrends, 59*(1), 56–63. https://doi.org/10.1007/s11528-014-0821-y

Mukuni, J. (2019). Challenges of educational digital infrastructure in Africa: A tale of hope and disillusionment. *Journal of African Studies and Development*, *11*, 59–63. https://doi.org/10.5897/JASD2019.0539

Mulder, R. A., Pearce, J. M., & Baik, C. (2014). Peer review in higher education: Student perceptions before and after participation. *Active Learning in Higher Education, 15*(2), 157–171. https://doi.org/10.1177/1469787414527391

Mulyani, E., Hidayah, I., & Fauziati, S. (2019). Dropout prediction optimization through SMOTE and ensemble learning. In 2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI) (pp. 516–521). Yogyakarta, Indonesia: IEEE. https://doi.org/10.1109/ISRITI48646.2019.9034611

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

People-Praxis. (2024). *Peer reviewing – theory and practice*. https://courses.peoples-praxis.org/course/view.php?id=34

Publons. (2018). *Global state of peer review*. Clarivate Analytics. https://publons.com/static/Publons-Global-State-Of-Peer-Review-2018.pdf

Rai, L., & Chunrao, D. (2016). Influencing factors of success and failure in MOOC and general analysis of learner behavior. *International Journal of Information and Education Technology, 6*(4), 262-268. https://doi.org/10.7763/IJIET.2016.V6.697

Reich, J., & Ho, A. (2014, January 23). *The tricky task of figuring out what makes a MOOC successful*. The Atlantic. https://www.theatlantic.com/education/archive/2014/01/the-tricky-task-of-figuring-out-what-makes-a-mooc-successful/283274/

Revathy, M., Kamalakkannan, S., & Kavitha, P. (2022). Machine learning based prediction of dropout students from the education university using SMOTE. In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 1750–1758). Tirunelveli, India: IEEE. https://doi.org/10.1109/ICSSIT53264.2022.9716302

Roediger, H. L., III. (2007, April 1). *Twelve tips for reviewers*. *APS Observer*. https://www.psychologicalscience.org/observer/twelve-tips-for-reviewers

Ruredzo, P.I.M., Agyei, D.D., Sangare, M. and Heller, R.F. (2024) 'Open publishing of public health research in Africa: an exploratory investigation of the barriers and solutions', <i>Insights: the UKSG journal</i>, 37(1), p. 6. Available at: https://doi.org/10.1629/uksg.635.

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference* (pp. 92–96).

Stupacher, J. (2025). *Enhancing peer review skills in higher education: A mixed-methods study on challenges and training needs*. https://doi.org/10.31219/osf.io/89xju_v1

Tennant, J.P., Ross-Hellauer, T. The limitations to our understanding of peer review. Res Integr Peer Rev 5, 6 (2020). https://doi.org/10.1186/s41073-020-00092-1

Thomas, R. (2023, September 28). *Nurturing peer review excellence: The significance of early career researchers (ECRs)*. Enago Academy. https://www.enago.com/academy/ecr-in-peer-review/

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., … SciPy 1.0 Contributors (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature methods, 17(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

WACREN. (2024). *Establishing a Diamond OA publishing ecosystem for public health research in Africa: The Coalition for Open Access Publishing of Public Health in Africa (COPPHA)*. https://wacren.net/en/establishing-a-diamond-oa-publishing-ecosystem-for-public-health-research-in-africa-the-coalition-for-open-access-publishing-of-public-health-in-africa-coppha/

Willis, J. V., Cobey, K. D., Ramos, J., Chow, R., Ng, J. Y., Alayche, M., & Moher, D. (2023a). Limited online training opportunities exist for scholarly peer reviewers. *Journal of Clinical Epidemiology, 161*, 65–73. https://doi.org/10.1016/j.jclinepi.2023.06.023

Willis, J. V., Ramos, J., Cobey, K. D., Ng, J. Y., Khan, H., Albert, M. A., et al. (2023b). Knowledge and motivations of training in peer review: An international cross-sectional survey. PLOS ONE, 18(7), e0287660. https://doi.org/10.1371/journal.pone.0287660

Xu, J., Kim, K., Kurtz, M., & Nolan, M. T. (2016). Mentored peer reviewing for PhD faculty and students. Nurse education today, 37, 1–2. https://doi.org/10.1016/j.nedt.2015.11.031

Zhang, J., Gao, M., & Zhang, J. (2021). The learning behaviours of dropouts in MOOCs: A collective attention network perspective. Computers & Education, 167, 104189. https://doi.org/10.1016/j.compedu.2021.104189

Zimba, O., & Gasparyan, A. Y. (2021). Peer review guidance: A primer for researchers. Reumatologia, 59(1), 3–8. https://doi.org/10.5114/reum.2021.102709

Zhang, L., Geisler, T., Ray, H., & Xie, Y. (2021). Improving logistic regression on the imbalanced data by a novel penalized log-likelihood function. *Journal of applied statistics*, *49*(13), 3257–3277. https://doi.org/10.1080/02664763.2021.1939662