

Challenges and best practices in delivering remote hybrid bioinformatics training: The Experience of H3ABioNet University of Khartoum Node

Azza E. Ahmed¹, Ayah A. Awadallah², Mawada T. Elmahdi³, Maram Suliman⁴, Atheer⁵, Hassan Elsafi⁶, Basil D. Hamdel Nile⁷, mohamed.abdullateif⁸, Faisal M. Fadlelmola⁹

¹Centre for Bioinformatics and Systems Biology, Faculty of Science, University of Khartoum

²Ayah A. Awadallah ³Mawada T. Elmahdi ⁴Maram Suliman ⁵Atheer ⁶Hassan Elsafi ⁷Basil D. Hamdel Nile ⁸mohamed.abdullateif ⁹Faisal M. Fadlelmola

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXXX

ABSTRACT

Motivation: Delivering high quality distance-based bioinformatics training courses across the African continent is a challenging task. This often requires going beyond traditional teaching methods, using hands-on sessions, interactivity, and problem-based learning approaches. This will eventually lead to enhancing the course objectives and learning outcomes.

Results: In this article, we discuss the challenges and best practices related to delivering training in bioinformatics in lower-limited resource settings upon reflecting on our experience in hosting and running a multiple-delivery online course, Introduction to Bioinformatics (IBT), that was developed by the H3ABioNet Education and Training Working Group and delivered at the University of Khartoum Node. We believe that our local setting is similar to others in less developed countries, so we also reflect upon aspects like classroom environment and recruitment of students to maximize outcomes.

Contact: Faisal M. Fadlelmola (faisal.mohamed@hotmail.com)

Supplementary information: survey questions, survey data, figures, models and code

1 INTRODUCTION

The rapid advancements in genomics and molecular biology research and applications make adequate bioinformatics training not only needed, but also makes this need, for complimentary skills in biology and computer science, continuously evolving ??.

Bioinformatics training in the form of physical face-to-face workshops is one way to address this need, especially that it provides opportunities for networking and first hand understanding and discussions of concepts and ideas ?. However, when run in resource limited settings, like across Africa, where access to local bioinformatics expertise, funding and infrastructure to carry out research

is largely limited ?, this model soon becomes very expensive to run and therefore limited in the capacity of students' intake. More critically, the relevancy and applicability of skills acquired from training in these setting to the bioinformatician's own research in the long-term also becomes questionable ?.

To cope with scarcity of resources, distance learning methods have long been in use to deliver educational materials for physically segregated learners and educators, and have continuously evolved with technology from postal services, to radio and TV, and recently, to Massively Open Online Courses (MOOCs) ?. For bioinformatics training, both edX and Coursera, 2 popular MOOC providers, provide complete bioinformatics specializations that cover the necessary background for the biologist (<https://www.edx.org/micromasters/bioinformatics>) and computer scientist alike (<https://www.coursera.org/specializations/bioinformatics>). Despite the availability of these courses, some MOOCs may be inaccessible or less than ideal educational aids for developing countries' learners due to a complicated set of conditions like technological access, digital literacy, cultural relevance and social identity threats ??, not to mention the typical caveats associated with MOOCs, like other competing priorities and the overwhelming amount of information and options, even for a developed world learner ?.

H3ABioNet, the pan African Bioinformatics Network ?, is making strides towards bridging the bioinformatics training gap in Africa by designing and offering a 3 month multiple-delivery-mode training course, Introduction to Bioinformatics (IBT), across all its nodes including Sudan ?. The IBT model blends local in-person tutoring sessions with the online delivery of course materials from remote course instructors via the course website (https://training.h3abionet.org/IBT_2017/), the learning management system, Vula (<http://www.cilt.uct.ac.za/cilt/vula>), and the open source videoconferencing system, mconf (<https://mconf.sanren.ac.za/>).

The present study assesses the efficiency, effectiveness and relevance of the 2017 iteration of the IBT training model from

the learner’s and local teaching assistants’ perspectives in the H3ABioNet Node of Sudan based at the University of Khartoum. Through 3 surveys for learners at different points in the course, and another survey for the course local staff, we investigated the factors contributing to a successful training experience. Our results agree with empirical data suggesting that local group discussions improved the accessibility of the course material to the students ?, and also that these tutoring sessions were facilitated by volunteering previous IBT alumni ?. Also, the fact that the course was developed while keeping the African context in mind made the content relevant to the local learners, and hence the training aligned and satisfied their expectations ?. We believe that our local setting is similar to others in less developed countries, so we further reflect upon aspects like classroom environment and recruitment of participants to maximize outcomes.

2 RELEVANT LITERATURE

For MOOCs to fully realize their potential in democratizing education, a lot of effort is needed to overcome barriers in less developed countries in the form of technology and context ?. Efforts addressing these areas include +Acumen (<https://www.plusacumen.org>), which aims at empowering social change and provides MOOCs employing in-video transcripts, culturally-diverse case studies, and content that is viewable off-line and platform-agnostic . Consequently, +Acumen attracts participants from a diverse pool of countries, including Afghanistan, Botswana and Sri Lanka ?. Successful participation from those countries and other top Fragile States ? was also reported in AuthorAID’s offering on Scientific research writing, which utilized low-bandwidth friendly format via mainly text-based content, with occasional featured guest videos and voluntary course alumni as facilitators incentivized by certificates and badges ?. Though less obvious, Kizilcec et al have demonstrated that brief psychological interventions aimed at lessening social identity threats, like value affirmations and social belonging, significantly improved the persistence and completion rates of learners from less developed countries in MOOCs, at a negligible cost of only a few minutes from the course designer and learners ?. Furthermore, studies reporting on the application of the blended MOOC paradigm, which combines online MOOC components with in classroom interactions, have systematically shown positive educational indicators; even when applied in resource limited settings ???.

In the particular case of bioinformatics, there are huge, urgent, unmet training needs ? exacerbated by the breadth of the discipline, and its fast-pace of evolution ?. Contributing factors to this gap are the difficulty of curricula design due to the diversity of learners’ backgrounds ?, and also the shortage in experienced qualified trainers ?. Therefore, various international efforts have been exerted to bridge this skills gap, like GOBLET, ELIXIR, EMBL-ABR and BD2K TCC and H3ABioNet. These efforts have took many forms including short face-to-face and online courses which are preferable by researchers in the later carer stages ?.

For basic bioinformatics users, both Coursea and edX provide introductory level training aimed at the molecular biologist level.

Table 1 compares their offerings with H3ABioNet’s blended MOOC in terms of length, recognition and content. These three courses are comparable in terms of the content they provide, and the quality of these offerings is unquestionable. However, it remains to be investigated their accessibility to learners in less developed countries. In the rest of the article, we highlight and evaluate elements of the blended MOOC run by H3ABioNet, the IBT course in its 2017 iteration, that makes it particularly accessible to learners in less developed countries.

Course	IBT_2017
Website	https://training.h3abio.net/
Provider	H3ABioNet
Delivery	blended MOOC Instructor-paced (online) local tutor 13 weeks Statement of accomplishment mentoring
Number of weeks	13
Recognition	Statement of accomplishment
Molecular biology review	-
Databases and resources	x
Linux	x
Sequence Similarity Analysis	x
Genomics – Assembly and Comparison	x
Phylogenetics	x
Gene Expression and Regulation	Extra material
Protein Structure and Modeling	-
Selection Analysis	-
Metagenomics	-

** A verified certificate is offered at a cost, though financial aid is possible. Without a payment, edX offers a non-verified statement of accomplishment (i.e., honor code certificate), while Coursera restricts access to some parts of the course (and hence no statement of accomplishment).

Table 1. Comparison between On-line bioinformatics courses targeting basic bioinformatics users.

3 METHODS

3.1 The 2017 iteration of the IBT

In its second iteration of 2017, the IBT course started on May 9th 2017, with 2 days per week for in-person interactive sessions.

Building on its first iteration of 2016 ?, and aiming at building introductory level competence for basic bioinformatics users ?, the 2017 IBT was composed of the following 6 modules: Introduction to databases and resources, Linux, Sequence alignment theory and application, Multiple sequence alignment, Genomics, Molecular evolution and phylogenetics. The design, learning objectives and contents of these modules are already described in ?, so here we only comment on the local supporting set up of the classrooms.

In the H3BioNet Node of Sudan, 73 students registered in the 2017 iteration of the IBT course. The majority of these students (participants or learners hereafter) have been selected from a waiting list from the previous IBT run of 2016. The eligibility criteria was for them to have a basic understanding of the central dogma of molecular biology, and hence they came from a diverse set of specializations (Figure 1 and Supplementary Figure 3B) , at different stages in their education (6% were at the BSc level, 41% current MSc students, 11% current PhD students and 34% were MSc and PhD graduates not pursuing any degree) and career affiliations (66% in academic institutes, 4% in governmental ministries, 7% in Research centers, 3% in private companies and hospitals, with remarkable 12% unemployment rate) as can be seen in Figure 1 (C and A respectively). Figure 1 B further shows that despite these differences, the highest educational institute for the majority of those participants is the University of Khartoum (73% have graduated from the University of Khartoum, compared to 13% percent from the other Sudanese universities, and 4% who studied abroad).

To accommodate this large number, two classrooms were set up for physical interactions within the University of Khartoum main campus: the CBSB laboratory, equipped with 20 PCs and network ports to accommodate an additional 12 PCs/laptops; and the main Library of the University of Khartoum with its larger computer lab that can accommodate up to 70 participants. Collectively, these two locations accommodated the 73 participants who registered in the course, and were split to 33 and 40 between these locations respectively. Those two locations vary in their infrastructure as well: the PCs in the CBSB lab were chosen for bioinformatics training and research with larger screen sizes and more CPU and memory capacities, while those in the Main library needed more effort from the local IBT team and the University of Khartoum Information Technology Network Administration (ITNA) staff to set them up, and fix network issues at the beginning of the course.

This larger intake has been managed and facilitated by a local staff, which besides the CBSB node Principle Investigator (PI), was composed of 7 teaching assistants (TAs) who were themselves previous alumni of the first IBT iteration of 2016. Their prior IBT experience helped them provide actionable advice and guidance to the new course participants, and their facilitation job was tremendously eased with the on-line staff training sessions provided by the core IBT team. Only a single system administrator was available for the duration of the course, but it was manageable with both classrooms being physically close to each other. It is worth noting here that the recruited TAs have background in genetics & molecular biology (1 TA), and molecular medicine (the remaining 6 TAs), which matches the background of participants.

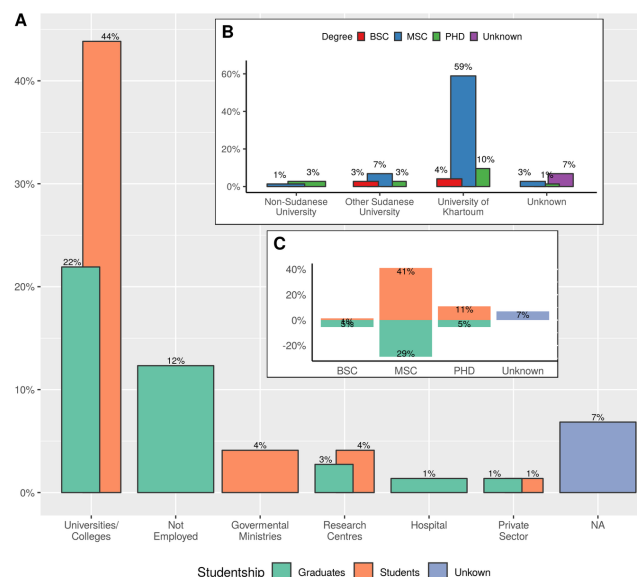


Figure 1. Demographics of the 2017 IBT participants in the H3ABioNet University of Khartoum Node, Sudan. A) Different affiliations of IBT participants stratified by their status as Current students or Graduates. Grouping into the shown categories was done by manual assignment of data to the appropriate category. B) The distribution of the institutes awarding the highest degree for the course participants. C) The distribution of the highest academic degree of the IBT participants stratified by their status as Current students or Graduates.

3.2 Measurement and evaluation

While we lack data on the exact performance of our course participants in the individual activities of the IBT (individual assessments and tests' results were collated and kept with the core IBT team in South Africa, and only shared directly with each participant), we have aggregate data from the following sources:

1. The initial registration database (i.e. waiting list from the previous IBT run of 2016) . We used this data to collect demographics on our course participants (**supplementary 1**)
2. Three surveys designed to monitor our learners experience throughout the course, and their satisfaction and progress. Those surveys were distributed at the start of the course (**supplementary 2**), the middle (**supplementary 3**) and end points (**supplementary 4**) as can be seen from Figure 2. In this figure, we also track this information against participants' attendance and withdrawal patterns (**supplementary 5**). Absence pattern is the difference between the total number of registered participants (73), and the present and withdrawn participants in each session, hence it is not explicitly shown.
3. There is also the data identifying participants who earned a certificate upon successfully satisfying the requirements of the course (**supplementary 6**). We corroborated the self-reported success status from our participants against aggregate class statistics shared by the core IBT team, because in addition to the official IBT certificate from the University of Cape

Town (South Africa), the local participants also appreciated a certificate of completion from the University of Khartoum.

- Another final source of data is the local course staff themselves, and the extent to which the course experience has been of value to them (**supplementary 8**).

Collectively, we use this data to investigate factors associated with a successful experience (or alternatively, failure to satisfy the course requirements or complete withdrawal from the IBT), so that we are better informed for future course runs, or similar training initiatives. In particular, our surveys (**also see supplementary 2, 3 and 4**) focus on the following aspects: participants' expectations, culture and interactions with each other and use of the online resources (Vula and mconf).

The results and discussions sections below collate and comment on all these data.

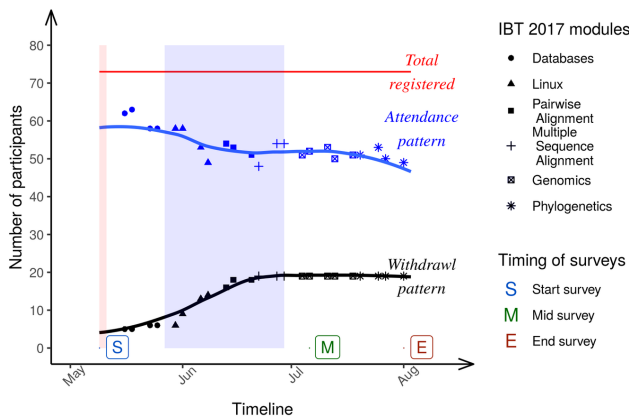


Figure 2. Patterns of attendance and withdrawals in the 6 IBT modules. Points of data collection (surveys) are also highlighted in the timeline. The red shaded area coincides with the orientation week, whereas the blue shaded area is the holy month of Ramadan.

4 RESULTS

4.1 Attendance & Withdrawal patterns

Students attendance and retention is a major concern contributing to a successful MOOC experience ?, especially that to many learners, the problem is about fitting the needed study hours into their busy schedules, and being on the same page as the course progresses.

In this 3 months course, we note that out of the 73 registered participants, the attendance rate was higher at the beginning of the course (~85%), then it dropped progressively towards mid-June and early July (Figure 2). This accompanied a steep increase of the withdrawal curve which can be related to 2 factors: 1) the Linux module, which is understandably challenging as the material naturally requires a bit different mode of thinking than what wet lab biologist are typically used to. 2) culturally, the IBT course started

just a few weeks before the holy month of Ramadan coinciding with end-of-year/semester holidays in many universities and colleges. As can be inferred from Figure 1A, during the few starting weeks it was easy for students (53% of the total IBT participants) both full-time (44%) and part time (9%) and other academic staff (22% of the total IBT participants) alike to follow up with the progress of material and activities in the IBT.

However, once these holidays were over, those IBT participants (collectively, 75% of the total IBT participants) needed to be back to full time working hours and classes in their respective institutes, making it harder for them to attend IBT sessions in person and timely submit their assessments or work towards their tests. It is also clear from Figure 2 that the withdrawal pattern plateaued after this point in time, such that in total 5 participants withdrew from the CBSB classroom and 14 from the Main Library classroom, making up a total of 19 participants (26% of the total intake of participants for the 2017 IBT).

A remarkable observation seen from Figure 1B is that the majority of the IBT participants (73%) have graduated from the University of Khartoum. Furthermore, all the local course staff have studied at the University of Khartoum. This could have been the reason why ~63.6% of the participants heard of the course through friends, 27.3% from their supervisors or mentors, and the remaining 14.5% through social media (Email, Facebook and Twitter) (**Supplementary 9**). This also suggests that local circles of friends/acquaintances were already in place before the course had actually started. This support system in place could explain why for those participants who stayed in the course to the end, 51 out of the 54 remaining participants (94%), they did actually satisfy the requirements of the course, and only 3 participants (6%) failed.

Given the high success rate of participants in both classes, it is more interesting to look for factors that contributed to a participants withdrawal from class (as opposed to failure). We investigate such predictive models based on the demographics of the class in the Discussion (section 5.1) (Figure 5, **supplementary 11,6**). We note however that there are unmoderated personal effects; like participants having to travel with their families, either for holidays or accompanying a close sick family member for medical treatment abroad; or for some work commitments that could not be waived.

4.2 Participants' perceptions & expectations

The data on the Sudan IBT 2017 participants' the Sudan IBT 2017 participants' perceptions and expectations were collected from the 3 surveys (**supplementary 2, 3, 4**) disseminated at the start of the course, the mid-point and the end of the course as shown at the times indicated by the S, E and M boxes in Figure 2, respectively. Out of the 73 participants, only 33 (45%) filled all the 3 surveys (**Supplementary 10**), while 15 (21%) never filled any.

The experience of the IBT course is both unique and new to the participants considering its blended multi-delivery learning model ?, and also its extended 3-months time span. Our surveys aimed to check the alignment between the participants' expectations and the course scope ?. On a labeled five-point Likert scale from *very uncomfortable* to *very comfortable*, we asked participants

about their perspectives in terms of their prior experience level in each IBT course module (start survey); the extent to which the content was appropriate, and the level it met their expectations (in the mid-course and end surveys, for each module taught until that point).

Participants' perceptions for each of the 6 modules of the IBT in its 2017 iteration largely followed the same trend (**Supplementary 17**), hence they are illustrated as an average of responses across all modules in Figure 3. Not surprisingly, most participants were largely unfamiliar with the various modules (with the 75th percentile of responses below neutral familiarity level), and this was especially true for the Linux module given their background (**Supplementary 6**). To the contrary, we see higher satisfaction levels (with the median of the averaged responses at a level above Comfortable in Figure 3) in terms of satisfaction with the appropriateness of the taught material and meeting participants expectations.

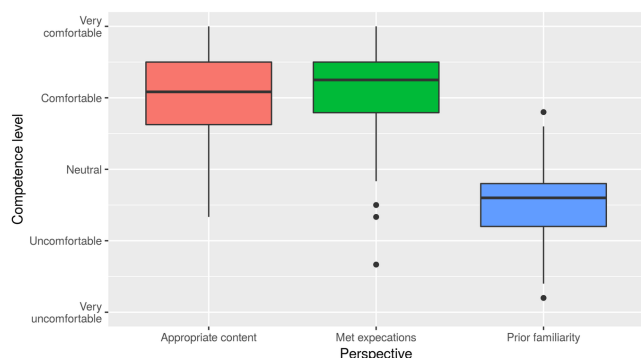


Figure 3. Average IBT participants' perceptions on the 6 modules taught in the 2017 iteration of the course, in terms of their level of competence, measured as the level of their prior familiarity with the modules, and their satisfaction upon completion according to the responses collected via the Start, Mid-course and End surveys

Another aspect investigated is the extent to which our IBT participants made use of the local and remote classroom elements to facilitate their learning experience. Namely, we were interested in the following aspects of the multi-delivery model: Networking with local and remote IBT participants, and use of the learning management system, Vula, for submission of assessments and asking/answering questions. We monitored the progression in these aspects via the mid-course and end surveys, via a five-level competency scale from *very uncomfortable* to *very comfortable*. These responses, as can be seen in Figure 4 show a positive trend as the course advanced which reflects that the participants were getting more familiar with the blended MOOC model. We remark however a modest amount of networking with other IBT classrooms, and also no change at the end of the course in the distribution of responses to this question (In total, 38 participants filled both the mid-course and end surveys, with 2 and 5 unique entries for each survey respectively (**Supplementary 10**)).

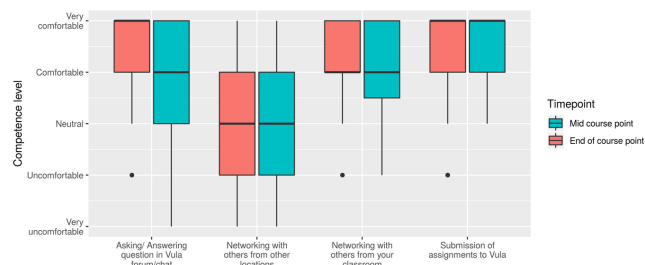


Figure 4. Students utilization levels of various elements of online learning: Vula and networking- measurements were taken from surveys in the middle of the course and at the end of the course

5 DISCUSSION

5.1 Classroom demographics & performance

Consistent with the largely female student's student's body composition in faculties related to natural sciences, health, and agriculture in Sudan ?, faculties from which most of the 2017 IBT participants hail, the majority of the IBT participants in both classes of the H3ABioNet node of Sudan were females (80%); and they were also more likely to satisfy the course requirements in comparison with their male peers (**Supplementary 11**). This remark, about gender as a strong demographic predictor of an IBT participant performance in the course can also be seen from the Recursive PARTitioning (rpart) tree ? of (**Supplementary 12, 13**), built by selecting splitting covariates that minimize the Gini coefficient as an information measure.

However, when changing the partitioning algorithm to a conditional inference tree ?, it is noted that the location of the local IBT classroom (the CBSB lab or the Main Library), is the most important covariate in predicting the performance of participants based on the complete dataset (**Supplementary 15**). This is also not surprising considering the inherent differences between the 2 locations in terms of infrastructure. The CBSB lab is designed to facilitate bioinformatics training and research in terms of infrastructure with stable internet connection and more powerful computers; whereas the Main Library classroom had logistic problems in terms of Internet connectivity at the beginning of the course, which was frustrating to some of the participants (and in occasions encouraged some of them to ultimately withdraw early on).

The difference in the trees constructed by these algorithms (**Supplementary 12, 15**), can be explained by the high degree of class imbalance in the entire dataset (70% Success, 26% Withdrawal and 4% Failure), and especially when seen in each classroom independently (85%, 15% and 0% for the CBSB class, and 58%, 35% and 8% for the Main Library respectively). We therefore built a multinomial classification model to further examine all demographic factors collectively (Figure 5 and **Supplementary 7**). While a potential problem in constructing this model is the assumption that each participant performance is independent and constant, here we see that both the physical location of the classroom and Gender are the only statistically significant demographic factors, yet none of the factors

has large effect in predicting a participants performance; except to note that MSc and PhD participants had higher odds of success than BSc level candidates; and so did unemployed participants, or those working in research centers and the private sector. Whether a participant is currently a student or has graduated from the said level had slight effect on their odds of success (also see **Supplementary 11**).

A remarkable observation of the IBT participants of 2017 (Figure 1A) is that about half the class (53%) are students, and 15% of those students are part timers with affiliations in either governmental ministries (the ministries of Agriculture and Animal resources in this case), or research laboratories. It is understandable for current students to have better odds of success, because they have a problem at hand and they wish to answer it; but we see no effect of interaction between being part or full time student in performance, hence we don't show it in the model of **Supplementary Table 2**. However, the fact that part timers were able to satisfy the requirements of the course (and have higher odds of success as predicted by our model in Figure 5 and **supplementary table 2**) suggests the appreciation of these institutes to equipping their researchers with modern and new techniques; and possibly suggests avenues for more sustainable development of research efforts.

Yet, there is the unemployment ratio of 12% among the participants. While we didn't explicitly investigate the employability of typical biological Sciences graduates in Sudan, by large, both MSc and PhD IBT participants indicated that they pursued higher education to pursue better opportunities. It is hard to conclusively say that enrolling in the IBT enhanced the employability of those participants. Yet, there is data from the 2016 IBT iteration alumni suggesting that some participants with MSc and PhD level education received teaching positions' offers in newly established faculties in Khartoum to teach bioinformatics-related courses or computational laboratories based on the skills they learned from the IBT in its 2016 iteration (**personal communications**).

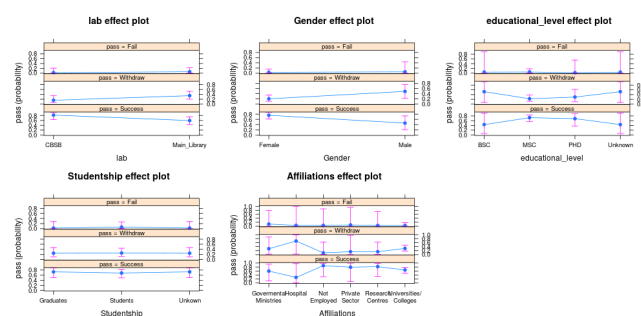


Figure 5. Effect of the logistics and demographics of the class (lab location, Gender, Educational level and Studentship) on IBT participants' performance (Success, Withdrawal or Failure) based on the 10-fold cross validated multinomial model of performance of Table 2.

5.2 Participants' perceptions & expectations

By design, the IBT course is taught over 3 months, to provide an introductory level knowledge and skills to basic bionformatics users. Besides providing relevant and high standard content, the core IBT team and various instructors have placed great emphasis on defining clear learning objectives and outcomes prior to each of the taught modules ?, and maintained the logical structure of the course despite removing the structural proteomics module in the 2017 iteration ?.

Therefore course participants had clear expectations out of each module, and we see largely positive indicators in their responses about all the taught 6 modules (Figure 3, and **supplementary 17**). This satisfaction about the course is also reflected in the high success rate of both classrooms in the University of Khartoum node(excluding early withdrawals, the success rate, i.e. percentage of participants satisfying the course requirements is 94%, see also section 5.1). Considering that the majority of the participants are in their MSc or PhD level (Figure 1C, supplementary 11) in a domain related to genetics or molecular biology (supplementary 6), it is expected that some of them had some exposure to concepts and uses of Databases; however, less familiarity with Genomics and related topics, because there is none or limited postgraduate degrees in bioinformatics currently in Sudan.

In terms of utilizing local and remote classroom resources, we make a few observations. While we see that our participants were comfortable networking with each other, we see less interactions with participants from other classrooms. In part, we can attribute this to the demographics of the course participants, as was elaborated in section 5.1, in terms of the highest awarding institute of each participant and also how the majority of participants heard of the course (supplementary 10). While a few minutes were spared at the beginning of each session for the scattered classrooms to introduce themselves via the videoconferencing system, mconf, this was ineffective in linking our participants with the other classrooms (Figure 4) because often times the connection would be too noisy to yield meaningful conversation. The same issue often arises during the discussion time at the end of each session where the instructor is typically available and answering questions live. The alternative our participants would employ in this case is discussing issues through the forum or the chat rooms available through the learning management system, Vula. We do indeed see more engagement in these avenues towards the end of the course in Figure 4.

Another element that could explain the high success rate of the 2 classrooms run in the University of Khartoum, is the local help provided to the participants through the local teaching assistants who were themselves alumni of the 2016 iteration of the IBT. From one perspective, testimonials from successful alumni have been shown to improve the sense of belonging in MOOCs, and their by their by learners' performance ?. This is especially true as the 2017 iteration of the IBT included training sessions for the local team in each classroom on how to best facilitate the course. These training sessions employed the Mental contrasting with implementation intentions (MCII) model ? in equipping the local staff with best strategies to facilitate the course by asking them to set goals for

the course, and then predict future challenges and devising solutions for them should they occur. While not all identified challenges were within direct control (see section 5.3), collectively looking at problems beforehand gave a sense of confidence to the local staff. It should also be mentioned that the majority of our volunteering staff already had prior teaching experience (7 out of the 8 TAs, in addition to the class PI). The responses from the TAs do indicate that they benefited from the previous IBT experience, and also that they were comfortable with facilitating such a class with different demographics (supplementary 18). On the personal level, we also see the TAs highly satisfied with the experience, even though they had other pressing commitments playing at the same time as the IBT (supplementary 19).

A final element in examining the IBT experience of 2017 is the non-cognitive factors pertinent to the course participants. These factors, unexamined in section 5.1, include the expected load incurred by participating in the IBT, which many found to be overwhelming when compared with commercial 1 week bioinformatics courses, that have no exams or assignments. This perception has exacerbated especially with the Linux module as can be seen in Figure 2. Another contributing factor may be the language of the course. Sudan is officially an Arabic speaking country, though university educated graduates have some competency in English. However, one teaching style commonly employed, especially at the undergraduate level, is to deliver teaching curricula in Arabic, while maintaining the English keywords and having English-based handouts (Supplementary 20). This means that for many of the participants, the experience of complete course delivery in English can be seen daunting. Collectively, these factors reasonably explain the large withdrawal of the course before the mid-point (26% of the total participants- also see Figure 2).

5.3 Reflections/ Local logistics and organization

The IBT used a blended MOOC (bMOOC) ?, or multi-delivery model ?, for learners to access, discuss and submit their assessments and take tests. The online resources used: mconf for example, were open source and were not network intensive; which made them appropriate to the local set up. Also, the fact that the instructors of the various modules come from different countries within Africa provided a context for the learners to relate to. These aspects, technological infrastructure and relevant context, are effective in making a MOOC accessible to local participants, and hence improving performance ?. We did not compare with students' performance on the offerings from edX and Coursera (table 1), so it is interesting to pursue this in the future.

A good article reflecting on challenges of bMOOCs is available here: <https://link.springer.com/article/10.1007/s12528-017-9133-5>. They are the only authors who call them H-MOOCs, but they focus on the institutional challenges with setting up the MOOC (as opposed to students' performance and perceptions).

By large, one should see that the design and running of the IBT followed the Ten simple rules for developing a short bioinformatics

training course of ?. In particular, we comment on the following rules (as they pertain to the local classroom organization):

- **Rule1: Set Practical and Realistic Expectations- achieved by specifying target audience:** The target audience for the IBT are molecular biologist who are familiar with the central dogma of molecular biology. We note that it was those participants were able to satisfy the requirements of the course (they did not fail nor withdraw early on). We note that graduates from certain faculties () withdrew the course early on. It did help in yielding a high success rate that we had a long waiting list in place.
- **Class diversity:** in terms of academic backgrounds, graduating universities... etc. Proper advertising, and conscious selection of participants (in lights of the drop out factors we highlighted) should help in a better outcome.
- **Rule 3: Ensure Computational Equipment Preparedness and Hands-On Support Availability:** Before and throughout the IBT, the local IBT team would meet weekly and update on resources needed and tasks assigned, report on issues faced, or make suggestions for improvement. A myriad of platforms were used for this: Trello boards for planning and follow up (<https://trello.com/>), a Google mailing list for emails, and authorea (<https://www.authorea.com>) for collaboratively working on this manuscript. For some of the members, a chatting app like Whatsapp (<https://www.whatsapp.com/>) was needed, but overall, experiencing other media was deeply appreciated and highly regarded.
- **Rule 9: Allow for Inter activity and Provide Time for Reflection, Individual Analysis, and Exploration:**
- This was achieved via the in-person class sessions. Those sessions assured that IBT participants had someone locally to turn to for help, and thereby reducing anxiety from participating- which is a major cause of withdrawal from online only delivery MOOCs ?. It helped in assuring this aspect that the **Teaching assistants, are alumni from the 2016 IBT iteration-** (In fact, in a typical MOOC, lack of local help is among the key reasons for course withdrawal ?, and for learners from less developed countries, success stories from previous course alumni have been shown to improve performance ?
- **Career progress and capacity building:** The IBT in all its iterations aims to equip African researchers with the skills and knowledge to launch their careers, and establish their science. While only 30 participants from the 2017 iteration responded to the follow up survey sent after 1 year, we see that many participants have moved from being students (53.3% at the time of the IBT) to being junior and middle staff (collectively 56.3% after 1 year)(Supplementary 20). For some of them, we even observe that the IBT helped them secure new job offerings (Supplementary 21).

6 BEST PRACTICES

The following are the top best practices that make the IBT 2017 a successful course in spite of all the reported challenges:

6.1 Online course design elements

Sessions and Modules learning outcomes: At the first session of each IBT module, trainers identified and emphasized on the learning outcomes to the participants. This gave the participants a clear idea of what to expect from each session. At the end of the course, they were able to evaluate to what level the course content met their expectations. Their satisfaction with the course was a good indicator of the course success. This was clearly reflected by the percentage of the successful participants, who were able to meet all the course requirements.

Trainers interactivity sessions with course participants: Trainers uploaded their session resources in the central panel of the Mconf interface should they have wanted to explain a concept on a particular lecture slide, for example. Trainers activated their webcams while answering questions.

Hands-on sessions & Teaching assistants: During this free 3-month introduction to bioinformatics, IBT2017, a large portion of course materials were dedicated to hands-on sessions, where participants are given the opportunity to practice what they are learning. Nevertheless, these sessions require a large number of teaching assistants, as they offer participants the opportunity to handle real data and run analysis tasks that implement the theory being illustrated in the lectures. This was found to be of great importance, as often trainees fail to appreciate how what is explained in the lectures can be directly applied to real data.

Video Conferencing System : IBT classrooms connected the trainer to all other classrooms via the Mconf open-source web conferencing platform (<https://mconf.sanren.ac.za/>), and classrooms either activated their microphones or entered text into a chat box to ask questions to the trainer. Considering it was a free resource, the offered features were to some extent satisfactory in some of the classrooms (real-time chat, screen sharing, file sharing, classroom mode). Issues with sound clarity and disconnection is motivating the core IBT team to consider alternative platform (personal communication).

Learning Management Systems: Throughout the IBT, Vula, the University of Cape Town's online learning and collaboration environment was utilized to send out announcements, manage participants, track their progress, and allow for live or delayed interaction amongst participants and with trainers and staff.

6.2 Local best practices & settings:

Registration database: We had so many people interested in the IBT, that at the time of writing, we have about 400 registered potential applicants in our bioinformatics training database, coming from different universities, career status, and educational backgrounds.

Teaching assistants - IBT alumni: During the planning for the IBT 2017 course, we came up with the idea of using IBT 2016 alumni best participants as teaching assistants for the IBT 2017. This turned out to be a great gesture as it improved help the IBT 2017 students as the IBT alumni gave them advice as they have gone through this phase.

Local Classes Logistics and Planning: Compare to IBT 2016 course intake of only 22 participants, we were successful to intake 73 participants for IBT 2017. There were many issues that support us to reach this target number. We talked to the librarian of the

Main Library to use the computer laboratory that accommodated 40 course participants, while 33 were accommodated at the CBSB.

Class diversity: course participants with different expertise and research interests.

7 CONCLUSIONS AND RECOMMENDATIONS

7.1 Challenges

1. **Location:** Finding computer labs to accommodate a larger intake of IBT participants is a challenge, because the duration of the IBT overlaps with parts of the academic semester in most of the relevant faculties, and hence they can't offer their labs for the entire 3 month duration.
2. **Timing:** Some of the course participants are full time MSc students, and some of the IBT sessions collided with the timing of their exams (~25 participant). The flexibility and sensitivity of the core IBT in giving them some grace period helped these participants make up for any missed IBT activity.

7.2 Recommendations/ lessons learnt

1. Working closely with concerned entities within the University could provide support in allocating more infrastructure and resources.
2. More active collaborations with other entities (Faculties/ research centers) within the university and/or the ministry of higher education.

7.3 Looking forward

1. Collaborations with other universities/ Research centers (in other states besides Khartoum).
2. Arranging similar courses (multi-delivery model) in other areas like data science and health informatics

8 KEY POINTS

- Learners in less developed countries are keen on to seizing educational opportunities. For MOOCs to be attractive, some effort is needed. Registration database...
- Blended learning employing multi-delivery models can be effective in bridging the achievement gap in MOOCs, even in bioinformatics courses.
- The training of local teaching assistants who are previous course alumni improves the sense of belonging to a MOOC, and hence improves learners' performance.
- Working in tandem with other bodies in the university to secure the needed infrastructural resources is also another factor contributing to a successful training.

9 AUTHOR CONTRIBUTIONS

- Conceptualization: AEA, FMF
- Formal analysis: AEA, FMF
- Funding acquisition: FMF
- **Investigation: all**
- **Methodology: all**
- Project administration: FMF
- Resources: FMF
- Software: AEA
- Supervision: FMF
- Writing/ Original draft: AEA, AAA, FMF, Maram, Mawada, Hassan
- Writing/ Review and editing: all? (all- add your initials)

10 SUPPLEMENTARY MATERIALS

1. Questionnaire_follow_up: questions from the follow up survey NOT the actual table data collected, because this data (even if names were deleted) can still identify participants
2. Survey1_Start (or, the actual results)
3. Survey2_Mid_course (or, the actual results)
4. Survey3_End_course (or, the actual results)
5. Table1_Attendance (just remove the names)
6. Table2_Graduates_demographics (with no names)
7. Table3_10-fold_cross_validated_performance_model_of_the_IBT_in_Sudan's_node
8. Survey4_Teaching_assistants (or, the actual results)
9. Figure1_Media_effect
10. Figure2_Surveys_filled_by_participants
11. Figure3_Class_demographics_and_performance_distribution
12. Figure4_rpart_performance_classification_tree
13. Text1_rpart_tree_specifications
14. Figure5_rpart_complexity_parameter_plot
15. Figure6_Conditional_Inference_Tree_for_Participants_Performance_in_the_IBT
16. Text2_10.fold.multinomial.model.summary
17. Figure7,8,9,10,11,12: Per module participants prior expectations and satisfaction levels
18. Figure13_TAs_perspective_teaching_experience
19. Figure14_TAs_perspectives_personal_reflections
20. Figure15_Teaching_languages

Table 2. 10-fold cross validated multinomial model of the IBT 2017 participants' performance in the course as measured in the H3ABioNet node of Sudan

	Dependent variable:	
	Withdraw (1)	Fail (2)
labMain_Library	1.103* (0.642)	1.206 (1.361)
GenderMale	1.374* (0.721)	0.883 (1.510)
educational_levelMSC	-1.362 (1.546)	-0.394 (2.914)
educational_levelPHD	-1.029 (1.676)	-1.243 (3.576)
StudentshipStudents	0.094 (0.697)	0.999 (1.767)
AffiliationsHospital	1.596 (2.772)	-0.190 (5.975)
AffiliationsNot_Employed	-1.580 (2.206)	-1.344 (3.421)
AffiliationsPrivate_Sector	-0.993 (2.825)	-0.808 (3.127)
AffiliationsResearch_Centres	-1.030 (1.938)	-1.386 (2.857)
AffiliationsUniversities_Colleges	-0.043 (1.534)	-1.149 (1.805)
Constant	-0.452 (2.304)	-2.624 (3.893)
Akaike Inf. Crit.	131.217	131.217

Note: *p<0.1; **p<0.05; ***p<0.01

100

Call:

```
rpart(formula = pass ~ lab + Gender + educational_level + Studentship + Affiliations, data = data_grads_demographics, method = "class", n = 73, complexity param=0.045, expected loss=0.3013699, P(no
```

```
CP nsplit rel error xerror xstd
1 0.04545455 0 1.0000000 1.000000 0.1782019
2 0.01000000 1 0.9545455 1.090909 0.1824398

Variable importance
Gender
lab
Affiliations
Studentship
```

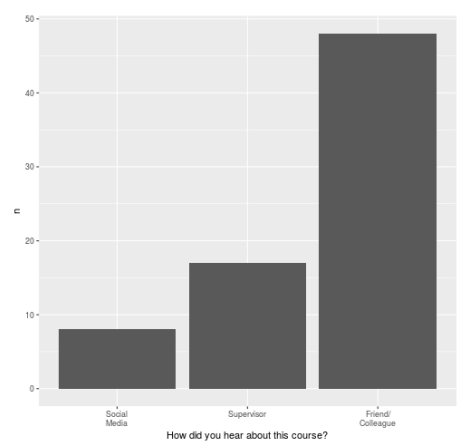


Figure 6. Figure1_Media.effect

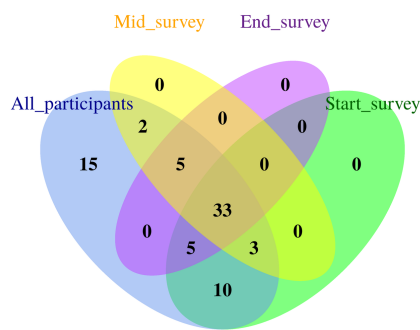


Figure 7. Consistency of filling the 3 evaluation surveys by the IBT participants in Sudan node: Main library & CBSB laboratory

```
educational_level splits as RLR, improved (10 fold)
Resampling results across tuning parameters:
Summary of sample sizes: 57, 58, 58, 57, 57, 58, ...
Resampling results across tuning parameters:
decay Accuracy Kappa
0e+00 0.6591667 0.07751040
1e-04 0.6591667 0.07751040
1e-01 0.7401190 0.08571429
Node number 2: 59 observations
predicted class=Success expected loss=0.2372881 P(node) =0.8082192
class counts: 45 12 2
probabilities: 0.763 0.203 0.034
Node number 3: 14 observations
predicted class=Withdraw expected loss=0.5 P(node) =0.1917808
class counts: 6 7 1
probabilities: 0.429 0.500 0.071
Accuracy was used to select the optimal model using the
The final value used for the model was decay = 0.1.

Penalized Multinomial Regression

64 samples
5 predictors
3 classes: 'Success', 'Withdraw', 'Fail'

No pre-processing
```

11 ACKNOWLEDGEMENT

- IBT 2017 core team: Kim T. Gurwitz (University of Cape Town), Shaun Aron(University of Witwatersrand), Sumir Panji (University of Cape Town), Suresh Maslamoney (University of Cape Town), Nicola Mulder(University of Cape Town)

- IBT 2017 trainers
- IBT 2017 Sudan Course participants
- University of Khartoum Faculty of Science
- University of Khartoum Main library
- University of Khartoum Information Technology Network Administration (ITNA)

12 FUNDING

This work was supported by the National Institute of Health Common Fund [grant number U41HG006941]. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

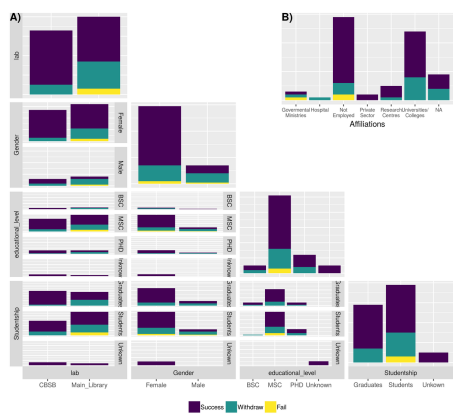


Figure 8. Figure3_Class_demographics_and_performance_distribution: A) Generalized pairs plot of the logistics and demographics of the class (lab location, Gender, Educational level and Studentship) against IBT participants' performance (Success, Withdrawal or Failure). B) Participants' Participants' affiliations distribution with respect to their performance

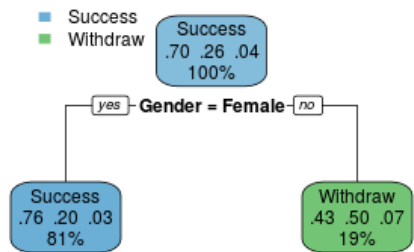


Figure 9. Figure4_rpart_performance_classification_tree: Recursive PARTitioning (rpart) classification model of the main covariates affecting participants performance in the IBT (Success, Withdrawal, or Failure). Each node shows the predicted learner performance, the probability of each performance category based on the node group, and the percentage of observations in the node. Here, we see that Gender is most important covariate for predicting the performance of an IBT course participant as per the 2017 data.

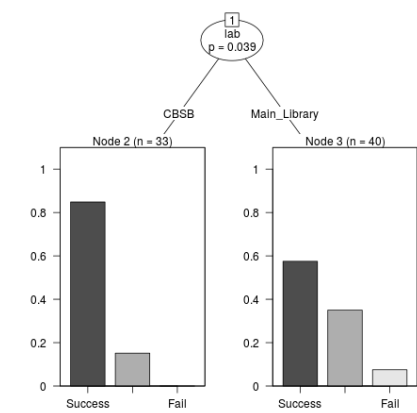


Figure 10. Figure6.Conditional_Inference_Tree_for.Participants_Performance.in.the.IBT. Here, we se that the physical classroom location is the most important covariate in predicting an IBT participant’s performance

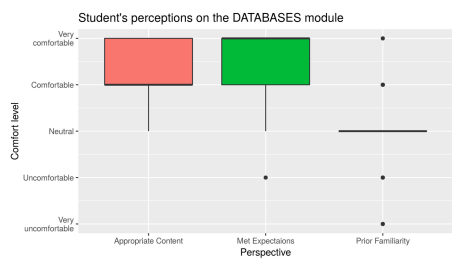


Figure 11. Figure_7_Databases_module_perceptions

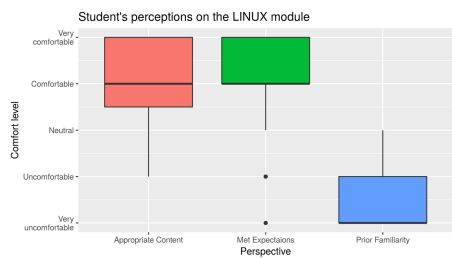


Figure 12. Figure_8_Linux_module_perceptions

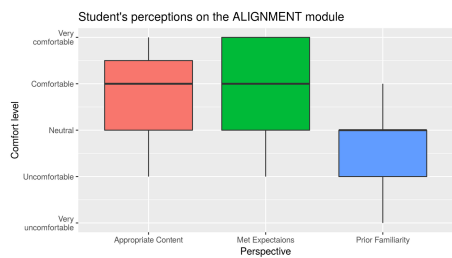


Figure 13. Figure_9_Pair_wise_alignment_module_perceptions

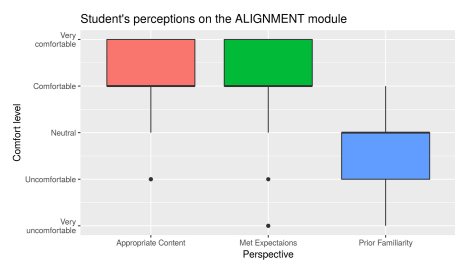


Figure 14. Figure_10_multiple_sequence_alignment_module_perceptions- For the prior familiarity, we used the same responses in producing figure_9 and figure_10, because participants were expected to have low familiarity with both types, which is confirmed by both figures

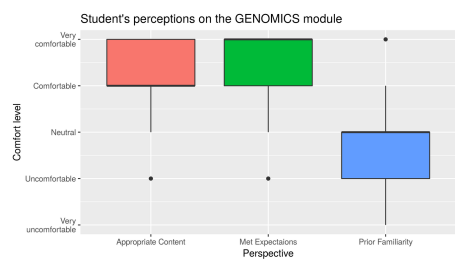


Figure 15. Figure_11_Genomics_module_pceptions

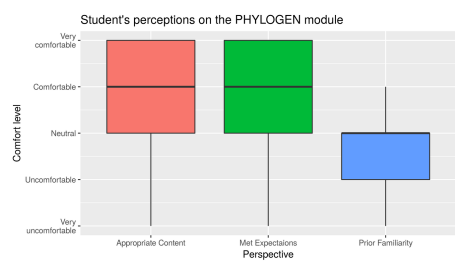


Figure 16. Figure_12_phylogenetics_perceptions

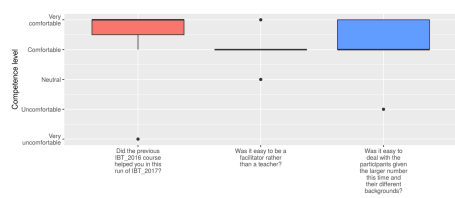


Figure 17. Figure_13_TAs_perspective_teaching_experience (n =7)

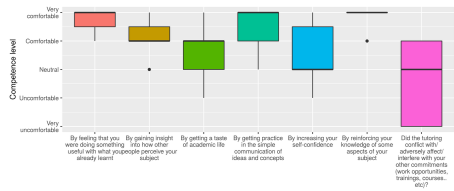


Figure 18. Figure14.TAs_perspectives_personal_reflections (n=7)

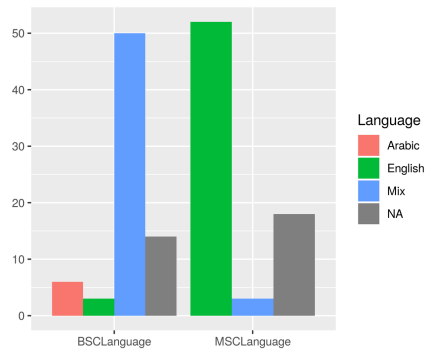


Figure 19. Figure15_teaching_language

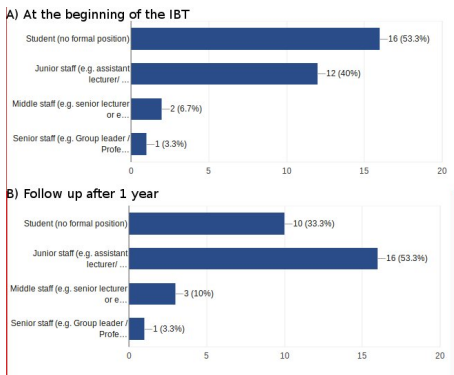


Figure 20. Progress of a 30 participant from the IBT 2017 iteration from their career status at the beginning of the course, and 1 year after that date.

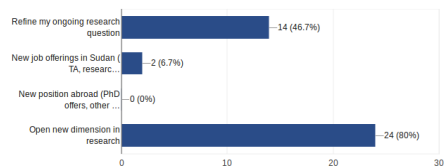


Figure 21. Responses from the 30 participants from the 2017 IBT run collected 1 year later, asking about which ways the IBT helped them with their career.