# 4

## *Exploring the Need for a Statistical Collaboration Laboratory in a Kenyan University: Experiences, Challenges, and Opportunities*

**Thomas Mawora and Joyce Otieno**
*Maseno University*

**Eric A. Vance**
*University of Colorado Boulder*

### CONTENTS

## 4.1  Introduction

This paper explores the need for a statistical collaboration laboratory or "stat lab" (Vance and Pruitt 2022) at Maseno University in Kenya. It describes the experiences, challenges, and opportunities for statistics lecturers who established a statistical collaboration laboratory or "stat lab" called the Maseno University Laboratory for Interdisciplinary Statistical Analysis

(MU-LISA) and currently operate it. Section 4.1 introduces the Applied Statistics program at Maseno University and what statistical consultancy services are required in Kenya. Section 4.2 discusses nine challenges to operate a stat lab. Section 4.3 describes MU-LISA and how those involved in the stat lab are overcoming the challenges previously described. Section 4.4 concludes with lessons learned from creating and operating a stat lab in Kenya.

### 4.1.1  Context of Maseno University

The undergraduate degree program in statistics has been offered in Kenyan institutions of higher learning for more than 30 years. At the turn of the century, Maseno University pioneered the Applied Statistics degree program in the Department of Mathematics. As of now, more than five universities offer a similar program. Maseno further pioneered in introducing Information Technology (IT) to all its programs. This stretched the creativity of teaching at Maseno University and resulted in the then School of Mathematics opting to tailor their IT units to teach the skills their graduates should be expected to have (Stern et al. 2010). Maseno further pioneered being among the first universities in Kenya to offer fully online degree programs through the use of a Learning Management System (Musyoka et al. 2012).

The above activities pushed Maseno University's School of Mathematics, Statistics and Actuarial Science (SMSAS) to be in the limelight for innovations in student instruction. The innovations exposed the lecturers to several processes including working in projects through collaborations with members from the University of Reading's Statistical Service Centre (SSC). The SSC provided statistical consultancy services and had clients within the United Kingdom and internationally. In addition, they provided statistical support services to projects, some of them being in Africa. Apart from the statistical consultancy services, SSC was interested in modernization of teaching statistics. The SSC's collaboration with SMSAS resulted in the introduction of a variety of electronic resources to teach statistics, including Computer-Assisted Statistics Textbooks (CAST) (https://cast.idems. international/), and tailoring the IT units to be more relevant to the skills needed by students in the job market (Stern et al. 2014).

The IT courses were tailored to improve our undergraduate statistics curriculum. They were designed to enhance students' computing skills, needed for data analysis. The courses include basic concepts of IT, which introduces students to computers; web browsing and communication, which introduces students to the internet how to get information from the internet, and communication via the internet; descriptive data analysis and presentation; data management; statistical computing; and problem-based statistical analysis, among others. However, these courses typically run separately from the statistics courses. Only a fraction of students understand that they require these IT skills for statistics and data analysis. Most realize the importance of IT only after being exposed to the work environment. Usually, this is after leaving college and results in them posting on social media entreaties for current students to pay more attention to the IT courses. The few students who are engaged by the IT courses tend to be more interested in coding and not data management, visualization, or analytics. A current opportunity is to incorporate IT components into all statistics courses so that students will understand the importance of computing in statistics.

### 4.1.2  What Consultancies Are Required in Kenya?

The range of statistical consultancy services needed in a typical Kenyan university is vast. For instance, almost all schools in Kenyan universities offer postgraduate programs that

have a project/research component. Many students pay others to do their statistical analysis, many times after data has been collected. The analyst goes ahead to produce the graphs and tables the person wants for topic 4 of their thesis, which is usually included in the results and discussions sections. This is a thriving business in Kenya for statistics graduates out of school and still looking for jobs. They usually come to consult some of their lecturers and in several cases, the data has serious quality issues. These could include (1) the data has not responded to the objectives of the study, (2) the data is in a wrong format, hence the tests they want cannot be done—in some cases the researcher wants to replicate the report of another person, and (3) the data does not yield statistically significant results, much to the dismay of the researcher. As such, the students conducting their studies may need consultancy services not just after data collection but during proposal writing so they are adequately guided on what data to collect and how to collect the data to be of high quality (Seiss et al. 2014).

A good number of lecturers from many of the schools in the university have successfully applied for funded projects. All the projects contain some elements of qualitative and quantitative data collection and analysis, including monitoring and evaluation. Despite the projects being multi-disciplinary, the persons concerned get external collaborators who perform most of the analysis needs. However, sometimes they seek consultancy services from lecturers within the university to help with the analysis, though rarely from the SMSAS. The exclusion of lecturers particularly from SMSAS may be due to the impression the Principal Investigators have concerning the teaching methods in SMSAS being theoretical rather than applicable and suited to current needs in statistical computing.

Conducting short courses on statistical literacy is another area of skills development that has high demand in Kenya. The first online course that Maseno University offered was *e-Statistics Made Simple* (*eSMS*), offered in 2010 that was developed by the Statistical Services Centre. This was an 8-week course on basic statistical literacy. The course attracted 40 students despite short notice, most of whom were nonstatisticians (Stern et al. 2014). There is more need for face-to-face short courses on the same topic and on more advanced statistical computing skills.

Finally, statistical consultancy services are needed by the local government. The policymakers make many decisions touching on multiple areas without being well informed through data in the region. In addition, development partners are interested in developing local skills and capacity to help respond to the ever-needed interventions they sponsor. Statisticians working with development partners and local governments in the intersection of data production, data analysis, and data decision-making can help make positive impacts on society (Vance and Love 2021).

## 4.2  The Setbacks to Having a Stat Lab and Human Resources for It

Having worked with the SSC before, one might expect that the university's statistics lecturers should be ready to operate their own statistical collaboration laboratory or "stat lab." However, despite imparting some skills in consultation and collaboration and experience working on consultancy projects, working with the SSC did not prepare lecturers in SMSAS to have their own lab. Below are nine challenges for setting up a stat lab in the Kenyan context.

### 4.2.1  Minimal Time for Lecturers for Consultations and Collaboration

Lecturing is a prestigious job in Kenya. With the explosion of universities and satellite campuses after the turn of the millennium, many of the lectures are sought after to (1) take more administrative roles and (2) teach courses for both undergraduate and postgraduate students part-time at other universities. These commitments fill up 40 hours-a-week and more, leaving the lecturers with far much less available time for other ventures like offering consultancy services. This is especially so with lecturers who teach mathematics and statistics, because the demand for their teaching expertise is high.

### 4.2.2  Overly Theoretical Nature of Our Programs

The curriculum provided for the postgraduate students in statistics assumes they already have statistical computing skills, and hence the emphasis is on the theoretical and mathematical aspects of the discipline. The students are required to apply the theory to real-life problems to earn an Applied Statistics degree. The challenge the students experience is primarily related to the big data they collect during the research process. Students are not proficient enough to wrangle data using statistical software to achieve the objectives of the project because education in the theory of statistics does not prepare one to analyze a real dataset.

### 4.2.3  Limited Statistical Computing Skills in Lectures

In the current realm of statistics, one has to be up-to-speed with the ever-changing statistical computing software and packages, such as the *tidyverse* in R (Wickham et al. 2019). In addition to computing skills, another challenge is that a number of universities do not have the relevant equipment and software to support the students' learning and to provision classrooms for the teaching of statistical computing. For example, only a few universities have subscribed to the limited version of the Statistical Package for Social Sciences (SPSS). This makes it difficult for stat lab members to use the software for training and other consultation services.

### 4.2.4  University Setup Challenges

Setting up a stat lab can have multiple other administrative challenges. Currently, at Maseno University, there is no defined procedure for setting up a statistical collaboration laboratory and provisioning it with needed office space and an independent physical lab of computers for teaching statistical computing short courses and workshops.

### 4.2.5  The Challenge of Running Costs and Consistent Funding

Currently, the stat lab at Maseno University strictly runs on volunteers, hence there are no financial implications for the lab or university. However, for long-term sustainability, some essential services for the proper functionality of the stat lab, including accounting and secretarial services, require consistent funds. The lack of such services limits the effectiveness of the lab. Consistent funding would go a long way to attract persons who can offer specialized skills, including writing winning proposals and specialized statistical skills to expand the number and types of projects on which the stat lab can engage.

### 4.2.6 The Challenge of Collaboration

The perception of other scientists and data users is that statisticians are required only at the data analysis stage of a project, rather than during the design, data collection, interpretation, and transforming evidence into action stages (Vance and Pruitt 2022). In our experience, statistics graduates from Maseno University also tend to consider themselves as just data analysts rather than collaborators in all stages of the research process who have a stake in the outcome of the project and work toward creating shared understanding with domain experts (Vance et al. 2022). Further, the local government and many other partners use external collaborations to help with statistical analysis and other projects. We have observed that there is usually a tendency for most leaders in the African context to overlook their own who have similar skills and opt instead for expatriates. This, coupled with the indecision of the statistics lecturers in engaging in outreach to create working relationships with their leaders and their lack of collaboration skills, inhibits opportunities for consultancies to trickle to the lecturers.

### 4.2.7 Data Quality

For a long time, there has been a challenge of unavailable and untidy data. This is slowly changing with many organizations embracing the modern methods of data collection. However, many still lack the capacity and skills for ensuring the quality of data is of high standards. As already mentioned, most lecturers have used small and tidy datasets in their classes. In current times, we have big data, with a lot of it being unstructured. There is a need for the statistics lecturers who will work in stat labs to be conversant on how to mine and clean data before they conduct an analysis.

### 4.2.8 University Process for Accrediting a Course

All university programs in Kenya are regulated by the Commission for Higher Education. Before the programs can be submitted, there are many internal processes and hurdles to clear before they can finally be approved. One main function for a stat lab includes offering short courses that the university can credit. This will boost confidence among learners since many would use the certificates for career progression. Overcoming the hurdles of accreditation for short courses is a major challenge for any stat lab in Kenya.

### 4.2.9 Pandemic

The whole world felt the pang of the COVID-19 pandemic. This particularly affected physical meetings, which were canceled for over 6 months in 2020 and resulted in squeezed timelines for semesters in Kenyan universities. The pandemic also affected the regular scheduling for in-person stat lab training.

The stat lab is a relatively new concept for lecturers at Maseno University who have been used to conducting classes and training, but not running a stat lab. The skills required include administrative skills of entrepreneurship, management, and leadership and technical skills in statistics and data science. At the same time, the director/coordinator of the lab should be learning new skills to meet the demands in the market, including networking skills to help advertise the lab and its services and connect it to more researchers, local government officials, and nongovernmental organizations. Being pioneers, this is still a challenge which we are rapidly overcoming through the activities mentioned in the next section.

## 4.3 MU-LISA's Strategies for Success

The MU-LISA was initiated in August 2019 and became the 12th full member of the LISA 2020 Network in March 2020 (Vance et al. 2022). MU-LISA's mission is "To empower learners to use statistical tools and methods to conduct sound analysis for multi-disciplinary research projects." This is accomplished through cooperation within Maseno University and with other universities, corporate, and government institutions for research and informed policy development. One of the goals of MU-LISA is to emphasize statistical thinking in all of our students and all of our projects so that the analysis of data relies less on "recipes" and becomes an activity to foster active learning (Kasturiarachi 2000). The stat lab teaches statistical coding skills, data management, data visualization, and data analytics in its training. In the lab training, we aspire not to be restricted by the semester dates and therefore provide our trainees with a cyclic learning process that develops in four stages. First, the learner encounters concrete experiences, followed by reflections and observations, which direct the learner to abstract conceptualization, and finally ends with active experimentation (Kolb 1983).

Since its inception, MU-LISA has offered over ten training sessions for statistical students, focusing on statistical computing skills such as introduction to R, tidyverse (e.g., dplyr (Wickham et al. 2021), ggplot (Wickham 2016)), managing data in Microsoft Excel, and introduction to python. Despite the multiple challenges mentioned previously, MU-LISA is learning and adapting to overcome these challenges to become successful using several strategies described below. Some are yet to be actualized but are currently ongoing.

### 4.3.1 Administrative Strategies

First, the process of initiating MU-LISA involved fostering university support by seeking consent from the Vice Chancellor, who agreed for us to initiate the stat lab. Further, the stat lab opened a file with the Directorate of Research and Innovation, hence we submit periodic reports to them. This has caused MU-LISA to be acknowledged as an official entity in Maseno University. As a result of this, we have received some support from the university and SMSAS, including using the physical Maths Lab for MU-LISA training sessions.

MU-LISA members divided roles so that apart from training, we also manage administrative tasks collectively. Members include the coordinator (Dr. Edgar O. Otumba) and trainers (Joyce A. Otieno, Thomas M. Mawora, and James K. Musyoka). All of the time spent is voluntary because MU-LISA does not yet generate funds to pay members to do administrative tasks. MU-LISA members meet regularly to plan and follow up on matters arising from previous meetings. This helps members to keep track of what needs to be done and allocate time for it. Currently, MU-LISA members spend approximately 10 hours every week planning, strategizing, and conducting MU-LISA activities. For the office space, the trainers retained their personal offices as the lab offices and they used existing lecture halls and computer labs for training.

### 4.3.2 Attending Workshops

MU-LISA members resolved to continual improvement of their general statistical knowledge and statistical computing skills. As such, members have continuously attended workshops, mostly online, to help hone their skills. The workshops have included training on Epidemiology, Online teaching, Tidyverse, Machine Learning, Data Science for

Consultancy, and Statistics Education Software (e.g., Tinkerplots (Brodesky et al. 2008) and Fathom). One organization MU-LISA members have partnered with is the Pak Institute for Statistical Training and Research (PISTAR), which hosted the latter four training sessions. Members who attended the training formed group sessions where they read related books through weekly meetings and practice. The group is now focusing on writing joint publications on subjects of mutual interest.

### 4.3.3 Continuous Collaboration

When MU-LISA joined the LISA community, members were encouraged to submit proposals for a 1-year project through a Transforming Evidence to Action Capacity in Higher Education (TEACH) fund (Olubusoye et al. 2021). MU-LISA's proposal to build the capacity of staff at Lake Region Economic Bloc (LREB) was selected and awarded $11,000 from the United States Agency for International Development. The LREB Bloc (see https://lreb.or.ke/about/) is a consortium of 14 counties in Western Kenya that share common social interests. The interests are divided into 10 pillars, namely, financial services; infrastructure; information communication technology; education; health; youth, gender and people with disabilities (PWDs); water, environment and climate change; tourism; agriculture; and trade and industrialization. Despite all these interests, the LREB does not have a central repository for the data on these pillars from its 14 counties. MU-LISA's project sought to empower LREB staff so they can set up a data repository center to serve the bloc.

Being awarded the TEACH fund gave MU-LISA members visibility within the university. As a result, they have been involved in multiple collaborative project proposal writing trainings and activities. MU-LISA has received a lot of support and encouragement from the university's Directorate of Research and Innovation. Writing strong proposals requires engaging in a learning process, and even though MU-LISA has not yet secured additional funding, we are optimistic that a funding proposal will be successful as we continuously improve on our project proposals.

The TEACH fund collaboration between MU-LISA and LREB attracted the attention of County Governments. A second concept note was drafted on the need to create a data repository for LREB. This note was created during the regions' governors meeting (summit) in 2020. Currently, LREB and Maseno University are working on a Memorandum of Understanding to help jumpstart the process of designing a repository.

In addition to collaborations within the university and LREB, MU-LISA has partnered with the Society of Applied Statistics Maseno University (SASMU). This partnership helps advertise and market MU-LISA and its activities among students in SMSAS. As a result, MU-LISA offers regular training sessions to members of SASMU and others who come for the training. The practice of collaboration consists of skills that can be learned and improved with practice, especially by using the ASCCR Framework for collaboration of Vance and Smith (2019).

### 4.3.4 Attachment and Internship Opportunities

Most courses in Maseno University require students to go for a 2-month attachment, which is similar to an internship. This attachment is a requirement for graduation. One negative effect of the Covid-19 pandemic was that many people were losing jobs, hence companies were not in a position to take students for attachment. This need necessitated MU-LISA to take a total of 60 students for attachment in November–December 2020 and January–March 2021.

The attachment program was geared toward providing students with real-life job experiences. As such, MU-LISA created tasks to help students learn practical skills while applying what they learned in their classes. MU-LISA benefitted by tasking the students to create resources that MU-LISA would have otherwise paid for. The students worked on developing "cheat sheets" or study notes for different subjects that can be given to students in order to help them study or review for the courses. In addition, they were given tasks to download data, read about it, develop new study questions, analyze it, and produce reports. At the end of the attachment period, they were provided with letters of recommendation to help them as they source for greener pastures.

Through the attachment, we worked with a group of five students who participated in the International Statistical Literacy Program (ISLP) International Poster Competition. This was a first for Maseno University and Kenya.

MU-LISA also benefited from the African Institute of Mathematical Sciences (AIMS) Alumni Technical Assistance Sponsorship Program. This is a support program to enhance the transition of AIMS graduates to a professional career life (https://wil.nexteinstein.org/wilprograms.htm). In the program, an AIMS alumnus identifies an institution where he would like to go and work for a short period as an intern. The institution provides projects for the AIMS alumnus and reports to AIMS on a monthly basis. AIMS provides the student with a modest stipend. Through this, MU-LISA got its first intern who will work on the ongoing TEACH fund project and other MU-LISA training activities.

### 4.3.5  Online Career Talks

Upon discussing more with students, MU-LISA members learned that there was still a need to help students understand their place in the job market. However, due to unavailability of time, MU-LISA opted to initiate Saturday Night Career Talks. These 40-minute sessions are recorded via zoom and discuss a variety of topics relating to careers in statistics. Example topics include specializations in statistics, handling data, both structured and unstructured, social network analysis, geospatial analysis, qualitative data analysis, and the professional skills on how to market oneself. The sessions are recorded and uploaded to YouTube (https://www.youtube.com/channel/UCcDKadTekzSYheKrcQL0q7A) so anyone can watch them at their convenience.

The talks focus on professional skills rather than technical programming skills. This was informed by the fact that most of the attendees usually use their phones to view the sessions. Hence, coding might be difficult to follow. We have found that focusing on these important skills makes the career talks more meaningful to them.

## 4.4  Lessons Learned

Over the course of the 2 years in which MU-LISA has been in existence, we have learned multiple lessons. One lesson we have learned and have come to appreciate is the importance of having this initiative led locally. We believe it is as a result of our local efforts that we have gained massive support from Maseno University in terms of recognizing our accomplishments and supporting us by linking us to many other researchers in the university. The LISA 2020 Network and its mentors provided an impetus for creating the stat lab, but it was the local members who carried the vision to fruition.

Another important lesson we have learned is the need for being part of a consortium or network. The process of joining the LISA 2020 Network helped us understand our goals better, hence we are ever driven to their attainment. Regular meetings in the consortium help rejuvenate us every fortnight as we see more members join and learn from others about how they are overcoming the challenges of their stat labs.

We have learned that even though finance is really important, it is not the very first thing that one needs in order to start a stat lab. Rather, what is required is a few dedicated members and learning how to find different methods of getting in-kind solutions to some of the financial needs. This could include office space, internet, communication, and marketing oneself.

MU-LISA provides an opportunity to train statistical consultants and collaborators. The MU-LISA members intend to collaborate on more projects in the lab and encourage the engagement of statisticians at the problem statement stage, rather than after the data have already been collected. MU-LISA members are expected to collaborate to make decisions on the objectives of the study; help decide how and what data should be collected; assist in how the data is managed, analyzed, stored, and archived; analyze the data; help interpret the results; and finally to help transform the evidence collected and analyzed into action (Vance and Love 2021).

This chapter has shared opportunities that are there for a stat lab to exist in a university in Kenya. There are contextualized hurdles that the lab will have to overcome. However, the stat lab will become more solidly established and sustainable through patience, effort, and focus by its members.

## References

Brodesky, A., Doherty, A., and Stoddard, J. (2008), *Digging into Data with TinkerPlots*, Key Curriculum Press, 225 pp. ISBN 978-1-55953-885-5.

Kasturiarachi, A. B. (2000), *Interactive Labs for Introductory Statistics Courses*, Kent State University, Stark Campus, OH.

Kolb, D. A. (1983), *Experiential Learning: Experience as the Source of Learning*, Prentice Hall, Englewood Cliffs, NJ.

Musyoka, J., Otieno, J., and Stern, D. (2012), Using eLearning to Engage Mathematics and Statistics Students in a Kenyan University. *12th International Congress on Mathematical Education*, Seoul, Korea.

Olubusoye, O. E., Akintande, O. J., and Vance, E. A. (2021), "Transforming Evidence to Action: The Case of Election Participation in Nigeria," *CHANCE*, 34(3), 13–23.

Seiss, M., Vance, E., and Hall, R. (2014), "The Importance of Cleaning Data During Fieldwork: Evidence from Mozambique," *Survey Practice*, 7(4), 2864.

Stern, D., Ongati, O. N., Agure, J. O., and Ogange, B. (2010), Incremental Modernisation of Statistics Teaching and Curriculum at Maseno University. *International Conference on Teaching Statistics (ICOTS) 8*, Kenya.

Stern, D., Musyoka, J, and Stern, R. (2014), Building Strength from Compromise: A Case Study of Five Year Collaboration between the Statistical Services Centre of the University of Reading, UK and Maseno University. *International Conference on Teaching Statistics (ICOTS) 9*, Kenya.

Vance, E. A., Alzen, J. L., and Smith, H. S. (2022), "Creating Shared Understanding in Statistics and Data Science Collaborations," *Journal of Statistics and Data Science Education*. https://www.tandfonline.com/doi/full/10.1080/26939169.2022.2035286

Vance, E. A., Love, K., Awe, O. O., and Pruitt, T. R. (2022), "LISA 2020: Promoting Statistical Practice and Collaboration in Developing Countries." *Promoting Statistical Practice and Collaboration in Developing Countries*, eds. O.O. Awe, K. Love, and E.A. Vance, Boca Raton, FL: CRC Press. DOI: 10.1201/9781003261148

Vance, E. A., and Love, K. (2021), "Building Statistics and Data Science Capacity for Development," *CHANCE*, 34(3), 38–46.

Vance, E. A., and Pruitt, T. R. (2022), "Statistics and Data Science Collaboration Laboratories: Engines for Development." *Promoting Statistical Practice and Collaboration in Developing Countries*, eds. O.O. Awe, K. Love, and E.A. Vance, Boca Raton, FL: CRC Press. DOI: 10.1201/9781003261148

Vance, E. A., and Smith, H. S. (2019), "The ASCCR Frame for Learning Essential Collaboration Skills," *Journal of Statistics Education*, 27(3), 265–274. https://doi.org/10.1080/10691898.2019.1687370.

Wickham, H. (2016), *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H. (2019), "Welcome to the Tidyverse," *Journal of Open Source Software*, 4, 1686. https://doi.org/10.21105/joss.01686.

Wickham, H., François, R., Henry, L., and Müller, K. (2021). *dplyr: A Grammar of Data Manipulation*. R package version 1.0.7. https://CRAN.R-project.org/package=dplyr.