INTERNATIONAL PROGRAM IN SURVEY AND DATA SCIENCE

Frauke Kreuter[1,2], Florian Keusch[1], Evgenia Samoilova[1]
[1]University of Mannheim, Germany
[2]University of Maryland, USA
fkreuter@umd.edu

Working professionals at statistical agencies, local governments, companies and non-profit organizations alike see themselves increasingly in the need to collect and analyze data. For some these data collection efforts are related to surveys, while others use administrative data or scrape information from the web. To generate useful insights based on these data, it is essential that data users are familiar with the whole process – i.e., from formulating the proper (research) question to disseminating results and data alike. To fill this knowledge gap, the German government funded the International Program in Survey and Data Science. The paper describes the development of the modular, partially asynchronous learning environment of the program that allows students to fully engage in a learning experience without having to quit their jobs.

MOTIVATION

The volume of data including both big data (also referred to as "found" data) and data designed for particular purposes (e.g., survey data) keeps increasing and is expected to rise nearly tenfold by 2025 (Reinsel et al., 2017, p. 3). The value of data on human behavior and internalized states (e.g., attitudes or beliefs) across various sectors is undeniable. In the private sector, the new types of "big data" are in high demand, because data-driven decision making can help businesses become more efficient and gain substantive knowledge about their clients. Statistical agencies are also interested in integrating new digital sources of data in official statistics, as they hope that this will allow them to collect more timely data and to reduce response burden (National Academies of Sciences, Engineering, and Medicine, 2017). Given the recentness of change in the nature of data and a strong interest in data-driven decision-making across various areas, the demand for experts trained to work with different data sources is rising. It is important to note that found data are usually seen as a complimentary source not as a replacement for survey data, since there are still research questions where designed data is preferred. This also means that now an even larger skill-set is needed and it is not taught in traditional academic programs in social sciences and economics. The shortage in data expertise is especially challenging for small companies and organizations in the public sector, as big companies can afford to pay more and attract most of the talent from such a small group of potential employees.

To address this knowledge gap among working professionals through continuing education, the German Federal Ministry of Education and Research funded the International Program in Survey and Data Science (IPSDS). As newly graduated candidates often lack sufficient substantive knowledge in their respective work area and data training is often sought by people with a strong professional background in various areas (e.g., computer science, finance, consulting, official statistics, etc.), the program focuses on providing training to working professionals who already possess some skills and experiences. The program constitutes a consortium of Universities, led by the University of Maryland and the University of Mannheim, and aims to establish an online Master's degree targeting working professionals in the areas of opinion, market, and social surveys as well as employees of survey research enterprises, ministries, and statistical agencies worldwide. Since entering a regular graduate program for those who are full-time employed and have family responsibilities is not feasible, choosing an online format for the delivery was essential to make the program appealing and realistic to identified target groups. A 4-year funding (2014-2017) made it possible to select two cohorts of students who participated in the program free of charge and in turn helped evaluate and guide the program's development. In the following sections, we describe the development of the program's curriculum, the course design and online learning environment, the

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

program's evaluation strategies as well as discuss challenges and successes faced since the start of building the program.

ACADEMIC OVERVIEW AND CURRICULUM

While most data science programs focus their curricula on the areas of statistics and/or computer science, IPSDS adds to these key areas expertise in data collection and data quality given various types of data as well as their combination. In recent years, it became clear that the new digital sources of data will not replace surveys, but instead will be used in combination with survey data (Japec et al., 2015). Combining both new and traditional data sources is a tenable strategy already applied by leading organizations and business. Statistical agencies cannot rely solely on the new data sources, as they cannot capture all relevant aspects of a given research problem. One also cannot ignore the increasing popularity of "Do-it-Yourself" tools, which indirectly indicates the persistent reliance on survey data in the industry. For example, Survey Monkey reports that 90 million of their online questionnaires are filled out every month, while Qualtrics sends out about a billion survey invitations per year (Callegaro & Yang, in press). Even facebook, despite its access to a vast amount of user data, only within the last year collected survey data on 200 million people (both users and non-users of facebook) (F. Kreuter, personal conversation, May 26, 2017). As the latest report of the Committee of National Statistics of the U.S. National Academy of Science (2017) stresses, using multiple data sources and understanding the data generating process for each data type is of central importance for data quality.

The IPSDS curriculum draws on five areas identified by the Task Force Report of the American Association for Public Opinion Research: (re)defining research question(s), data generating process, data curation and storage, data analysis, as well as data output and access (Japec et al., 2015). The program consists of 75 ECTS points, that is equivalent to 15 months of full time studying (more information about the five core modules and example courses for each of the modules can be found here: https://survey-data-science.net/program/curriculum). Based on the data from the first two student cohorts, part-time studying will take on average three years. All of the students are required to attend the introductory course Fundamentals of Survey and Data Science. Following a successful completion of this foundation course, students can choose from a number of courses within each module. By allowing flexibility, we can be more responsive to students' career needs, their background, and interests.

TARGET AUDIENCES AND FORMAT

This program was designed to be particularly appealing to working professionals worldwide who have completed a bachelor's degree and would like to acquire profound knowledge in survey practice and data science to advance their current careers or to expand their career options. The German government funded two cohorts of students (16 students starting from 2016 and 15 from 2017) to test the program curriculum and format. As expected, the program attracted international working professionals with long work hours and, in some cases, additional family responsibilities. For all of the participants the online format was important in their decision to apply to the program at least to a certain degree. Nineteen students reported that online administration of the program was very important for their application decision. The first two selected cohorts (31 students in total) included a highly international body of students: 15 different countries of residence in total. Admission requirements included 12 ECTS/6 credits in mathematical/applied statistics. Successful applicants also had to take a statistical placement test. If the placement test identified knowledge gaps, students were expected to attend a bridge course covering topics ranging from fundamentals of probability and descriptive statistics to regression modeling.

While most students come from Europe, other countries include Chile, Brazil, Mexico, Oman, Qatar, and Kenya. In both cohorts, students are employed in the industry, public institutes, NGOs, or universities. Eight participants are employed in statistical agencies. Although the IPSDS participants differ in their professional biographies and backgrounds, they are very similar in their limited time due to long work hours, family responsibilities, and/or commuting time. Most students are full-time employed with only four students working less than 40 hours a week. Women are well

represented in the program (18 women and 13 men). Eight participants combine a full- or part-time job with family duties (underage children or elderly parents).

To address the needs of the program's target audience, the format entails the following features:

- course materials can be accessed any time and from anywhere asynchronously;
- in addition to asynchronous materials, students attend small live online classes moderated by the course instructor(s), in order to discuss questions and engage in collaborative problem-solving;
- IPSDS runs entirely in English, due to its international focus:
- once a year students are expected to participate in an immersion event (Connect@IPSDS) that takes place at the University of Mannheim, where they can meet their peers and faculty in-person, attend talks and workshops by leading experts in the field, network with international survey and data science professionals, as well as engage in additional learning and community-building activities.

IPSDS ONLINE ENVIRONMENT AND COURSE DESIGN

The goal of IPSDS is not just to create courses with high-quality content delivered by leading experts in the field, but to make courses interactive, engaging, academically rigid, and yet flexible. For this reason, the program adopts the flipped classroom model. In recent years, the flipped classroom design (also known as "inverted instruction" or "inverted classroom") received a lot of attention among researchers and practitioners (Prober & Khan, 2013). Although there is some disagreement on the exact definition of the flipped classroom (FC), literature agrees that it includes rotating between two phases: 1) (pre-class) interaction with course materials and 2) guided learning activities (He et al., 2016). The "flipping" designates the idea of moving what usually was done in the classroom (e.g., lecturing) out of the classroom and bringing traditionally out-of-classroom activities (e.g., work on problem sets or assignments) inside the class.

(Pre-class) Interaction with the course materials:

The course materials provided to the students on the course website (currently Moodle) include (depending on the course content) pre-recorded video lectures, required and recommended readings, examples for programing exercises, datasets, and other additional resources. Providing students with reading materials prior to the class is not a new phenomenon. In contrast to simply mediating this common practice with technology (upload documents online before class), the FC design stresses the quality of interaction with the provided resources. The integration of pre-recorded lecture videos on the learning platform allows for pausing, moving forward and backward in the video, re-watching (parts) of the video, as well as changing the video speed (both increasing and decreasing the speed is possible). The lectures are broken into several shorter videos to take into account the prototypical attention span as well as relatively short windows of free time among working professionals. Video materials include lectures, interviews and discussions with experts, as well as demonstrations of specific techniques and software tools.

Guided learning activities:

The in-class guided activities of the FC design are implemented with the help of online video conferencing and discussion forums. Each week, students join a 50-minute online session using the software Zoom moderated by the instructor. The online sessions are mandatory and serve multiple purposes: 1) discuss students' questions; 2) review problems with assignments; and 3) motivate students to persist in the course. Prior to coming to an online session, students are expected to go through the material of the respective unit. They are also encouraged to submit their questions via the discussion forum or email (to the instructor). The discussion forums are also used to provide opportunities for additional (often optional) communication as well as to refer students to relevant external resources (e.g., relevant news articles, websites, etc.).

Each course has its unique virtual room so that the same link could be used for all of the online meetings. The link is posted on the course website as well as communicated to the students in the welcome email. Zoom also allows splitting students into smaller groups, so that each group

is assigned a separate virtual room. The moderator (in our case the instructor) can move back and forth between these rooms. Small class sizes (up to 18 students) make it possible to see every person in the meeting simultaneously. The user can choose to use a speaker view or a full screen view. In the speaker view, the speaker will be shown in the main window (see Fig. 1a). In the full screen view, one can see all of the participants at the same time (see Fig. 1b).

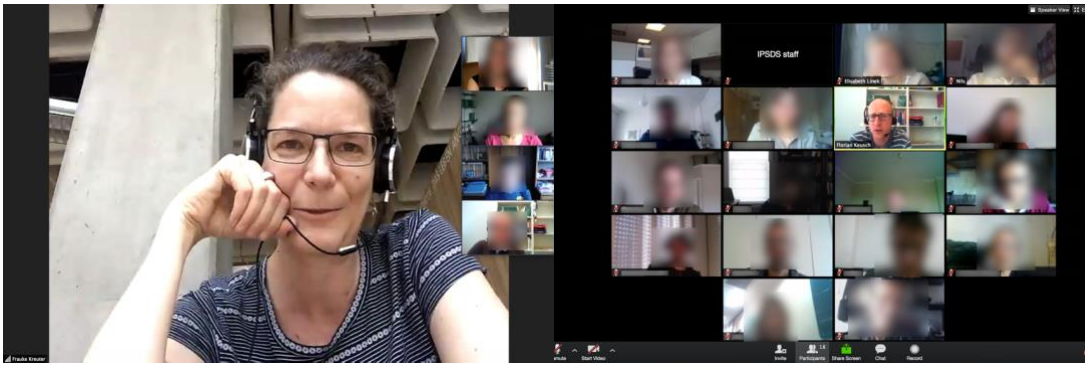

Figure 1a/b (from left to right): a. Zoom interface in speaker view; b. Zoom interface in the full screen view. Faces (other than the authors') are blurred here for privacy reasons.

IPSDS EVALUATION AND LESSONS LEARNED

Evaluation of the program involves systematic collection of data from various program stakeholders. Although program evaluation continues to evolve, currently the following data sources are used for guiding program changes:

- *Start-of-the-program/yearly milestone survey/qualitative interviews of students:* At the start of the program and after each academic year students are asked about their motivations, goals, expectations, employment, and family circumstance as well as achieved career outcomes.
- *Post-course student survey*: After each course students are invited to participate in the post-course survey that focuses on learning experience of students and their perceptions of teaching quality and various online course elements.
- *Learning analytics*: At the moment we are testing the use of anonymized learning analytics logs of online activities of the students. The results can be visualized and presented to instructors at the aggregate level. For example, logs capturing viewing of lecture videos can help identify which parts of the given video are viewed most frequently and which are skipped.
- *Qualitative interviews with instructors*: After having received student evaluation reports, instructors are invited to participate in semi-structured qualitative interviews conducted using the online conferencing software to discuss their teaching experience, quality of support they received from the program administration, and, if needed, suggestions for changes and/or additions to the course.
- *Pilot studies*: In order to bridge educational research and practice, regular course evaluation is accompanied by pilot studies implemented in single courses, which, if successful, can be scaled up later on.
- *Curriculum and courses review*: In addition to discussing further development of the curriculum with various partners, faculty, and the advisory board, we are currently working on introducing a peer review of the courses. Instructors will be assigned to review another course with respect to activities, structure, and workload. In addition to providing suggestions and comments, reviewers can themselves learn from reviewing other courses.

Given the identified target audiences with an eye on full-time work, family obligations and career development focus, we attempt to provide students with a versatile learning experience that is tailored to their current career stage and future goals. While in most aspects we have achieved positive outcomes, we have also faced some challenges that are likely to be common in the field of online adult education.

Quality of teaching: One of the core goals pursued by the IPSDS is recruiting, training, and retaining high quality faculty. According to the post-course evaluation surveys of 14 courses (n=73

with 84% response rate), students acknowledge the high quality of teaching at the IPSDS. When answering the question "How would you describe the quality of teaching by the course instructor(s)?" (1-very poor 2-poor 3-fair 4-good 5-excellent), none of the students perceived teaching as poor or very poor since the start of the program. Most students evaluate teaching as good or excellent (median=4, mean=4.3, sd=0.64). Furthermore, in most cases (81% of responses) students do not perceive differences in their learning and engagement in the IPSDS courses when compared to comparable onsite courses or rate the online format better.

Feasibility of combining studies with work and family: While the majority of students successfully adjusted to the program's pace and requirements, there has been a small number of students (four out of 31) who dropped out during the first two years. Two students have reported that they could not reconcile work and studies, for one participant the courses attended within the first year were enough to reach career goals, and one participant reported changes in personal circumstances. All 31 students successfully accomplished at least one course. On average students finished 6 ECTS of course work per term (12 weeks). Given the full-time employment of most students, this is a very positive outcome. While it is certain that combining work, family, and studies is not easy, experiences of the students during the test phase indicate that it is feasible.

Balancing flexibility and consistency: One of the main difficulties of online programs is finding the right balance between flexibility and consistency. While students often choose online course delivery mainly due to flexibility of time and place, it is the responsibility of the program to help students with difficulties that come hand-in-hand with flexibility such as greater need for time-management and self-discipline. Program evaluation indicated that having regular live interactions is very important for sustainable learning. The results from small-scale pilot studies where regular synchronous meetings were substituted with asynchronous discussion forums indicate that most students prefer weekly online meetings. Synchronous online meetings are associated with increased sense of community, with increased perceived learning outcomes, and, in some cases, also with improved self-reported time-management (IPSDS, 2017). As a next step the program will address the question of tailoring the live meetings to the course specificities, learning outcomes, and instructors' teaching styles. Post-course evaluations indicate that currently live-meetings are not used to their full potential in all of the courses, and perceived importance of online meetings for helping students to learn the material varies substantially.

Impact on work performance and career development: The first milestone survey conducted with the first cohort of students after the first year of studying demonstrated positive outcomes for students' work performance and career development. Among 14 surveyed participants, ten indicated that the program helped them improve their work performance. Two participants reported that participating in the program helped them find a better job position. Six participants received a pay raise since the start of the program.

Workload: One of the biggest challenges reported by students is successfully integrating studying into their everyday life. According to the first milestone survey (limited to the first cohort), 13 out of 14 surveyed students named workload as the main challenge for their learning. This is not surprising, since most students work full time and have family responsibilities. Moreover, seven students reported a very low level of satisfaction with work-life balance already at the start of the program. While the average workload for an IPSDS course based on the post-course self-reports constitutes 10 hours per week (median=10, mean=9.9, sd=3.5, min=3, max=18), workload varies greatly depending on students' background and individual study pace. To address this problem, as a next step the program will focus on exploring possibilities of micro- and mobile learning as well as on encouraging and supporting self-regulated learning strategies among students.

Workplace orientation: Although regular IPSDS courses include hands-on applications and working with data, they cannot replace experience of working on real-world projects. While most students work in the area of survey and data science, they do not always have an opportunity to have a relevant and demanding project to work on. Therefore, the program is currently focusing on creating project- and consulting-based courses via cooperating with external partners. Several statistical agencies already expressed interest in such collaboration. Such project work has to be well designed to fit the online and distributed work environment. The joint analysis work on (often

sensitive data) will likely need to be done in a cloud based (secure) environment. An example for such joint cloud based project work is set by the Coleridge Initiative (www.coleridgeinitiative.org).

In addition, we are currently exploring options for offering graduate certificates to participants who are not interested in or are not sure, if they can commit to the full Master's degree.

CONCLUSION

The project is currently at the stage of applying for the second phase of funding. The main goals for the second phase include further development of international cooperation with universities in South America, Australia, and China, as well as statistical offices worldwide, working on establishing tailored training for businesses and organizations, and strengthening real-world project work in the program's curriculum.

REFERENCES

Callegaro, M. & Yang, Y. (in press). The role of Surveys in the Area of "Big Data". In D. L. Vannette & J. A. Krosnick (Eds.), *The Palgrave Handbook of Survey Research*. New York: Palgrave.

He, W., Holton, A., Farkas, G., & Warschauer, M. (2016). The effects of flipped instruction on out-of-class study time, exam performance, and student perceptions. *Learning and Instruction, 45*, 61–71.

Japec, L., Kreuter, F., Berg, M., Biemer, P., Decker, P., Lampe, C., Lane, J., O'Neil, C. & Usher, A. (2015). Big Data in Survey Research. AAPOR Task Force Report. *Public Opinion Quarterly, 79*(4), 839–880.

National Academies of Sciences, Engineering, and Medicine. (2017). *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy*. Washington, DC: National Academies Press.

Prober, C. G., & Khan, S. (2013). Medical Education Reimagined: A Call to Action. *Academic Medicine, 10*(88), 1407–1410.

Reinsel, D, Gantz, J. & Rydning, J. (2017). Data Age 2025: The Evolution of Data to Life-Critical. Don't Focus on Big Data; Focus on the Data That's Big. IDC White Paper. Retrieved July 28, 2017, from: http://www.seagate.com/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.

International Program in Survey and Data Science. (2017). Program Assessment Report: Flipping classroom in online courses for working professionals: challenges and opportunities for student engagement. Retrieved November 1, 2017, from: https://survey-data-science.net/sites/default/files/ipsds_flipping_classroom_in_online_courses_for_working_professionals-_challenges_and_opportunities_for_student_engagement_0.pdf