

TRANSITION BETWEEN EDUCATION AND PROFESSION:
EXPERIENCES OF STATISTICIANS

by

Layla Guyot, B.S., M.S.

A dissertation submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
with a Major in Mathematics Education
August 2020

Committee Members:

Alexander White, Chair

M. Alejandra Sorto

Carlton Fong

Robert Sigley

COPYRIGHT

by

Layla Guyot

2020

FAIR USE AND AUTHOR'S PERMISSION STATEMENT

Fair Use

This work is protected by the Copyright Laws of the United States (Public Law 94-553, section 107). Consistent with fair use as defined in the Copyright Laws, brief quotations from this material are allowed with proper acknowledgement. Use of this material for financial gain without the author's express written permission is not allowed.

Duplication Permission

As the copyright holder of this work I, Layla Guyot, authorize duplication of this work, in whole or in part, for educational or scholarly purposes only.

ACKNOWLEDGEMENTS

While reflecting on my doctoral adventure, I am overwhelmed by the support I have had throughout these past five years. I would like to acknowledge those who have accompanied my work, directly or indirectly, but first and foremost, I would like to express my gratitude to the study participants who welcomed me at conferences, agreed to participate, and who trusted me to tell their stories. The adventure started when I applied to the program at Texas State University and I instantly knew when I met Dr. White and Dr. Sorto that they would be a great source of inspiration. I am very thankful they accepted to embark on the dissertation journey with Dr. Fong and Dr. Sigley.

I would like to thank Dr. White for sharing his genius and creativity about statistics and education to help me grow as a researcher and teacher. Thanks to him, I had incredible opportunities to get involved in research from the early beginning and he provided extensive support to help me develop instructional materials with fun activities to implement in my classroom. I also really appreciated our conversations about cultural differences between the U.S. and France which made me feel appreciated and integrated. It was a great honor to have had Dr. White's guidance throughout my graduate studies and sincerely hope that our collaboration continues.

Also, from the very start of my journey, I have been grateful to Dr. Sorto whose cheerfulness has been inspiring. She encouraged me to participate in conferences and gave me the amazing opportunity to be part of the editorial process of a renowned international conference that brought us to Japan. Thanks to Dr. Sorto, I met the most

influential researchers in statistics education because she knows everybody, or shall I say everybody knows her. I want to be like you when I grow up academically!

Advice given by Dr. Fong helped me shape my dissertation study and I very well recall the “Aha!” moment we had when we found the article that provided the theoretical framework and connected all the dots. It was a privilege to receive his guidance. My grateful thanks are also extended to Dr. Sigley who gave me the opportunity to experience qualitative research in statistics education and with whom I learned how to conduct interviews. I would also like to thank Dr. Melhuish who offered me the first chance to pursue educational research and included me on her team as a statistician.

This study could not have happened without the great support from the Department of Mathematics, the Graduate College, and the Student Government at Texas State University to enable me to travel to conferences to recruit participants. More than a financial aid, your support was a recognition of my work and aspirations. Conducting educational research would not make sense if it were not for students. I have learned so much from students at Texas State and I am very grateful for everything they taught me to grow as a teacher but also keeping me accountable for writing my dissertation on a daily basis.

A special thank you to my classmates Christina, Michael, Brittany, Mike, Suby, and Liz, with whom I was able to get through the coursework, preparing for presentations at a recurring headquarters, and who have shown constant support for my research and provided extremely valuable feedback.

I also would like to express my great appreciation to the persons that lead me to the topic of my dissertation. Philippe Bonneau, program coordinator of the Master's degree in Applied Probabilities and Statistics at l'Université de Lorraine as well as my classmate Yannick with whom I practiced the protocol for the interviews. Thank you to the team in the Department of Statistics at the Banque Centrale du Luxembourg for your warm welcome and for launching my career as a statistician.

To my parents, for instilling in me the passion for teaching and for mathematics ever since I can remember. A special thank you to my Mom who always believes in me and encourages me to follow my dreams. To Fifi whose strength is an inspiration to me and to my sisters Mélissa and Sonia for their unconditional love. To Naj, thank you for being by my side each and every day since day one of this adventure, bearing with me while I concluded this accomplishment, and for being quarantined with an introvert. Thank you to my brother at heart, Stéphane, for your constant love. Our adventures with Guillaume are full of lifelong memories and help me recharge.

This message acknowledges many people who have contributed to my success but I would not have been able to complete this work without the support of my music friends, in Austin and in France, for my horseback riding friends, especially Tanya with whom I could still go to play with when I had the chance, and to my friends who visited me, Christophe and Christine, Marie-Paule and Patrick. Finally, to Pilou (the cat) who stayed by my feet while I developed these ideas.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES.....	xiii
LIST OF FIGURES	xv
ABSTRACT.....	xvi
CHAPTER	
I. INTRODUCTION.....	1
Statement of the Problem.....	2
Purpose of the Study	5
Definitions of Terms	5
Statisticians	6
Mentors	6
Statistical Practices	6
Academic Settings	7
Workplace	7
Transition	8
Research Questions	8
Delimitations.....	8
Summary	9
II. LITERATURE REVIEW.....	10
Transition between Academic Settings and the Workplace	11
Transition for Statisticians	14
Theoretical Framework.....	20
Activity Theory	22
Boundary Crossing and Boundary Objects.....	25
Learning Mechanisms	26
Summary	27
Statistical Practices	28
Practices from the Perspective of Mentors at the Workplace	28
Practices from the Perspective of Statisticians	30

Practices from the Perspective of Mentors in Academic Settings	32
Promoting Authentic Statistical Practices in Academic Settings	33
Summary	36

III. METHODOLOGY 38

Research Design.....	39
Participant Selection	41
Members of the Statistical Community	42
Junior Statisticians and Mentors	43
Participants.....	44
Sample of Members of the Statistical Community	45
Sample of Junior Statisticians and Mentors.....	47
Description of Junior Statisticians	48
Gertrude.	49
Enid	50
Florence.....	52
Thomas.....	53
Stella	54
Janet	55
John	56
Edwards.....	57
Description of Mentors	58
Simone	58
David.....	59
Data Collection	61
First Phase of Data Collection: Sorting Task and Survey.....	61
Sorting Task.....	61
Paper collection.....	63
Digital collection.....	64
Survey	66
Second Phase of Data Collection: Interviews	66
Data Analysis	67
First Phase of Data Analysis: Identification	68
Identification of Important Practices through the Sorting Task....	68
Identification of Additional Practices on the Sorting Task.....	70
Second Phase of Data Analysis: Reflection.....	71
Third Phase of Data Analysis: Coordination	73
Fourth Phase of Data Analysis: Transformation.....	75
Survey	75
Interview	76

Triangulation.....	76
Researcher's Role	77
Institution Review Board	77
Limitations	78
Summary	78
IV. RESULTS	80
Perspectives from Members of the Statistical Community.....	81
Identification of Important Statistical Practices.....	82
Perspectives.....	83
Distinguishing and Consensus Practices.....	88
Additional Practices	91
Summary of identification	94
Reflection on Statistical Practices.....	95
Analysis.....	96
Communication.....	97
Traits	97
Design	98
Data Management	98
Domain Knowledge	99
Programming.....	99
Summary of Reflection	104
Perspectives from Junior Statisticians and Mentors	104
Identification and Reflection on Statistical Practices	107
Reflection and Coordination of Statistical Practices	110
Coordination of Design.....	110
Boundaries in the Division of Labor for Design.....	112
Boundaries in the Rules for Design	116
Boundaries in the Community for Design.	121
Boundaries in the Tools for Design.	123
Transition of Design.	125
Coordination of Programming	126
Boundaries in the Rules for Programming.....	128
Boundaries in the Division of Labor for Programming.....	134
Boundaries in the Tools for Programming.....	140
Boundaries in the Community for Programming.....	148
Transition of Programming.....	149
Coordination of Data Management.....	151
Boundaries in the Rules for Data Management.	151
Boundaries in the Division of Labor for Data Management.	155

Boundaries in the Tools for Data Management.	157
Boundaries in the Community for Data Management.	163
Transition of Data Management.	165
Coordination of Analysis	165
Boundaries in the Division of Labor for Analysis.	167
Boundaries in the Rules for Analysis.....	171
Boundaries in the Tools for Analysis.....	175
Boundaries in the Community for Analysis.....	182
Transition of Analysis.....	183
Coordination of Collaboration	184
Boundaries in the Community for Collaboration.....	186
Boundaries in the Division of Labor for Collaboration.	188
Boundaries in the Rules for Collaboration.....	191
Boundaries in the Tools for Analysis.....	194
Transition of Collaboration.....	195
Coordination of Communication	196
Boundaries in the Community for Communication.....	198
Boundaries in the Rules for Communication.....	199
Boundaries in the Division of Labor for Communication. ...	203
Boundaries in the Tools for Communication.....	208
Transition of Communication.	211
Coordination of Documentation	212
Boundaries in the Rules for Documentation.....	213
Boundaries in the Community for Documentation.....	214
Boundaries in the Division of Labor for Documentation.	215
Boundaries in the Tools for Documentation.....	217
Transition of Documentation.....	218
Coordination of Time Management.....	219
Boundaries in the Rules for Time Management.	219
Boundaries in the Community for Time Management.	221
Boundaries in the Division of Labor for Time Management.....	222
Boundaries in the Tools for Time Management.....	226
Transition of Time Management.....	228
Summary of Coordination.....	229
Triangulation and Transformation of Statistical Practices.....	231
Triangulation of Categories of Statistical Practices.....	231
Triangulation of Statistical Practices	234
Triangulation of Design Practices.....	234
Identification of Design	234
Reflection and Coordination of Design	235
Transformation of Design.....	236

Triangulation of Programming Practices	236
Identification of Programming.....	237
Reflection and Coordination of Programming.....	237
Transformation of Programming Practices.....	237
Triangulation of Data Management Practices.....	239
Identification of Data Management.....	239
Reflection and Coordination of Data Management.....	240
Transformation of Data Management.....	240
Triangulation of Analysis Practices	241
Identification of Analysis.....	242
Reflection and Coordination of Analysis.....	242
Transformation of Analysis.....	243
Triangulation of Collaboration Practices	244
Identification of Collaboration.....	244
Reflection and Coordination of Collaboration.....	245
Transformation of Collaboration.....	245
Triangulation of Communication Practices	246
Identification of Communication.....	247
Reflection and Coordination of Communication.....	247
Transformation of Communication.....	247
Triangulation of Documentation Practices	248
Identification of Documentation.....	249
Reflection and Coordination of the Documentation.....	249
Transformation of Documentation.....	250
Triangulation of Time Management Practices.....	251
Identification of Time Management.....	251
Reflection and Coordination of Time Management.....	251
Transformation of Time Management.....	252
Summary	252
V. DISCUSSION	254
Discussion of Findings.....	254
Important Statistical Practices.....	255
Transition of Statistical Practices.....	258
Discussion of Implications.....	261
Implications in academic settings	261
Implications at the Workplace	266
Implications for the Statistical Community	268
Recommendations for Future Research	269
Conclusion	272

APPENDIX SECTION.....	274
REFERENCES	285

LIST OF TABLES

Table	Page
1. Examples of Elements in the Activity Systems	24
2. List of Conferences For Recruiting Participants.....	43
3. Various Roles in the Statistical Community	45
4. Diverse Fields of Application for Statisticians	46
5. Description of Junior Statisticians	49
6. Description of Mentors	58
7. List of 24 Practices in Five Categories	83
8. Comparison of the Perspectives.....	84
9. Distinguishing and Consensus Practices.....	89
10. Additional Practices	92
11. Comparison of Programs Between the Systems	102
12. Percentage of Participants Learning New Programs	103
13. Percentage of Technology Users Across Fields of Application	103
14. Citations of Practices in Interviews	108
15. Important Elements for the Coordination of Design.....	111
16. Important Elements for the Coordination of Programming.....	127
17. Important Tools for Programming	141
18. Important Elements for the Coordination of Data Management	152
19. Important Elements for the Coordination of Analysis	166

20. Important Elements for the Coordination of Collaboration	185
21. Important Elements for the Coordination of Communication	197
22. Important Elements for the Coordination of Documentation	213
23. Important Elements for the Coordination of Time Management.....	220
24. Level of Continuity for Each Practice and Principal Boundaries	229
25. Statements on the Sorting Task Related to Design Practices	235
26. Statements on the Sorting Task Related to Data Management Practices	239
27. Statements on the Sorting Task Related to Analysis Practices.....	242
28. Statements on the Sorting Task Related to Collaboration Practices.....	244
29. Statements on the Sorting Task Related to Communication Practices	246
30. Statements on the Sorting Task Related to Documentation.....	249

LIST OF FIGURES

Figure	Page
1. The elements of an activity system.....	23
2. Statistical practices as boundary objects.....	25
3. Diagram for the sorting task.	63
4. Examples of tasks.	64
5. Webpage for data collection.	65
6. Scree plot for factor analysis.....	69
7. Comparison of categories of statistical practices between perspectives.....	85
8. Sorting task summarizing Perspective 1.....	86
9. Sorting task summarizing Perspective 2.....	87
10. Mean ranking of practices and variation.....	91
11. Comparison of practices learned in academic settings versus the workplace.	96
12. Comparing the number of programming tools used in each system.....	101
13. Lens for data analysis: Boundary crossing.	106
14. Distribution of elements for each practice.	109
15. Connections between categories of practices across findings.	232

ABSTRACT

The transition to the workplace is challenging across disciplines, but particularly decisive for statisticians since the demand for statistical skills is growing across many different fields and sectors. The employment of statisticians is expected to increase drastically, however, employers currently express difficulties in hiring. Indeed, a misalignment between practices developed by students in academic settings and practices expected of a statistician at the workplace has been identified. Despite some recent efforts to promote authentic statistical practices in academic settings, there is a need to understand how statistical practices transition to the workplace.

To study the phenomenon of the transition in detail, I considered a qualitative methodology with a quantitative component. Members of the statistical community were recruited at conferences around the world. They engaged in a sorting task to explore their inner perspective on the role of statistician at the workplace and reflected on their own experiences through a survey. Furthermore, a selection of participants who recently experienced the transition or accompanied statisticians in the transition were interviewed. Participants represented the diversity of the education and profession of statisticians. A cross-case analysis allowed for an in-depth description of the phenomenon of boundary crossing conceptualizing the transition, and identified the elements involved in learning statistical practices.

Findings revealed contributions and implications for statistical practices in academic settings and at the workplace. First, important statistical practices were

recognized by the members of the statistical community, contrasting different perspectives, and identifying practices that were previously overlooked. Second, participants confirmed that there was a misalignment between academic settings and the workplace by comparing practices developed in each system. Third, junior statisticians and their mentors identified boundaries occurring during the transition, uncovering what elements in academic settings or the workplace facilitate boundary crossing. Finally, the triangulation of findings formulated recommendations to promote the different statistical practices that emerged in this study. Implications addressed to the statistical community advocate for transformations in academic settings and at the workplace, promoting authentic statistical practices to facilitate the transition between education and profession for future statisticians.

I. INTRODUCTION

From actuarial science to zoology, the field of statistics has applications in many diverse disciplines. The amount of data being recorded across disciplines has exploded, making statistical skills increasingly valuable on the job market. The occupation of statistician was ranked as the best business job in 2020, and as the sixth best job across all job categories, according to online career guidance websites (American Statistical Association, 2020). The employment of statisticians is expected to grow by 31% compared to the projected 5% growth rate for all occupations between 2018 and 2028, as reported by the U.S. Bureau of Labor Statistics (2020). Yet, while the demand for data analysis skills has escalated worldwide, employers are experiencing difficulties in hiring statisticians across disciplines. In a report about the supply and demand of data analysis skills in the United States, it was found that 82% of 398 private, public, or nonprofit organizations, had positions in statistics available in 2016. However, as much as 78% of the organizations who hired data analysts indicated that the recruitment process was difficult (Society for Human Resource Management, 2016).

Even though the number of degrees in statistics and the number of students seeking degrees in statistics have been increasing over the past few years, the supply of graduates still does not match this thriving demand for statistical skills (American Statistical Association, 2015). But more importantly than the issue of the quantity of graduates in statistics is the issue of quality. Indeed, most employers are looking for experienced statisticians when hiring (Occupation Outlook, 2020) because graduates in statistics have not necessarily acquired the desirable qualifications for a statistician through their education. Desirable qualifications, as sought by employers, range from

interpersonal skills to more practical skills in statistics (Holmes, 1997; Ritter, Starbuck, & Hogg, 2001). The job market offers very few entry-level openings for statisticians and as a result graduates in statistics are encountering many challenges as they transition from their education to the profession.

Statement of the Problem

The transition from education to the profession is known to be challenging (Grosemans, Coertjens, & Kyndt, 2017), including for statisticians (Van der Berg, 2017). Educators, statisticians, and their mentors have uniformly identified a misalignment between the education of statistics in academic settings and the practice of statistics at the workplace. There are gaps in the training for specific methods and statistical packages (Harraway & Barker, 2005; Hijazi, Saeed, & Alfaki, 2019) as well as a lack of preparation for skills such as communication, writing, study design or data collection (Van der Berg, 2017). As much as we know about the misalignment between education and profession, we know little about how statisticians experience the transition and how they learn what practices they are required to perform at the workplace. My study aimed at exploring what practices statisticians develop as they transition to the workplace.

Indeed, statisticians develop practices, or practical skills, that grow out of experience and are relevant to their discipline and context (Pfannkuch & Wild, 2000). As a part of my master's program in Applied Probability and Statistics, I was taught courses led by professional statisticians from different disciplines such as biostatistics, data mining, or finance. They presented statistical methods related to their work as statisticians and offered the opportunity to experience statistical practices through projects involving real data and context. My educational background gave me insight into

what will be asked of me as a professional statistician and was of practical help during my transition to the workplace. In fact, Hoyles, Noss, Kent and Bakker (2010) recommended that knowledge and tools grounded in the context of specific professional settings should be implemented in the classroom. However, statistics education tends to focus on theory and concepts and lacks an authentic experience of statistical practices (Kenett & Thyregod, 2006). Training from professionals is rather atypical in educational contexts. Thus, one may wonder how authentic statistical practices can be promoted in academic settings. Before considering implementing authentic statistical practices in statistics education, these practices developed by statisticians across disciplines need to be identified.

Statistical practices can be interpreted from different perspectives. First, employers define what is required of a statistician at the workplace (Ritter et al., 2001). The perspective of mentors can help investigate in more detail the misalignment between the expectations of the workplace and academic knowledge. Van der Berg (2017) interviewed mentors to investigate what practices statisticians were lacking as they entered the workplace but did not specify in what ways mentors can support learning. Second, statisticians can reflect on their own experience and identify important practices. Pfannkuch and Wild (2000) interviewed statisticians from different disciplines and focused on statistical thinking. For example, they found that statisticians need to develop the ability to notice variation, navigate between various data representations, or build and reason from statistical models. However, Pfannkuch and Wild (2000) did not explicitly describe the role of a statistician at the workplace in terms of what practices matter the most and how statisticians have developed these practices. Third, perspectives from

educators reveal important goals for statistics education in promoting skills that reflect the role of statisticians. Anderson and Loynes (1987) described practices ranging from soft skills, such as communicating clearly and efficiently, to technical judgment such as recognizing the appropriateness of statistical methods. Cameron (2006) defined aspects of statistical practice as actions performed by statisticians, from preparing, analyzing, interpreting data to presenting statistical results and developing new methods. Anderson and Loynes (1987) and Cameron (2006) described a list of activities that statisticians do, based on their own experience or common understanding of what constitute the work of statisticians. We needed to involve the practitioners themselves. In addition, the practices need to be updated and grounded in context with concrete examples. Therefore, my study identified predominant practices and provided examples with extensive detail in diverse disciplines.

A misalignment between the training of statisticians in academic settings and their experience at the workplace has been identified, challenging the transition between the two systems. In order to understand how statisticians transition to the workplace, we need to give the practitioners the opportunity to share their experiences. In addition, other individuals are involved in the transition by acting as mentors for statisticians. Including their perspectives helped investigate how learning statistical practices is supported during the transition between academic settings and the workplace. A qualitative approach helped identify the challenges encountered by statisticians during the transition and allowed for an in-depth and rich description of the practices developed for the workplace. This study addresses gaps in the literature about the experiences of statisticians, combining the perspectives of statisticians themselves and their mentors. Investigating

the transition from academic settings to the workplace revealed significant details about statistical practices and what statisticians need to learn before, and as they enter, the workplace. The information collected also enlightened how to incorporate statistical practices in statistics education from the perspective of the practitioners.

Purpose of the Study

The purpose of this study was to examine the transition experienced by statisticians as they move from their education into their profession. Members of the statistical community from around the world were involved to identify and reflect on important statistical practices. Junior statisticians and their mentors were interviewed to give them the opportunity to share their experiences of the transition and depict a complete picture on how statisticians learn and develop practices to adapt to the workplace. A qualitative approach allowed for an in-depth and rich description of statisticians' experiences.

Findings from this study contribute to expand and update the existing research on what practices characterize the role of a statistician since practices are constantly evolving. Moreover, the information collected throughout this study enabled a discussion about promoting an authentic experience of statistical practices throughout statistics education from the perspective of the practitioners.

Definitions of Terms

Terms may have different interpretations, depending on the vision they represent. In order to distinguish between different interpretations, important terms need to be defined in the context of my study to designate the actors, with statisticians and mentors, and set the environment, with academic settings and the workplace.

Statisticians

Many positions require professionals to perform statistics as part of their job. In my study, statisticians designate professionals who self-identify as a statistician with a diversity of roles, including data analysts or data scientists, which involve performing statistical tasks as one of their main responsibilities, accounting for at least 50% of their work time.

In this study, there is a special focus on junior statisticians who are experiencing the transition at the workplace. I consider individuals in transition if they have graduated and have started working within the last five years.

Mentors

Statisticians sometimes transition to their professional environment under the supervision of a mentor. Mentors can have different roles and backgrounds. For example, a professor from the academic setting can facilitate the transition during a work placement or a coworker can share their knowledge with the statistician in transition to the workplace. The employer of a statistician in transition can also be considered to fill the role of a mentor by setting expectations at the workplace and guiding junior statisticians.

Statistical Practices

Practices that are required of a statistician at the workplace are considered as statistical practices. They range from practices involved in the process of data analysis, such as translating a real problem into a statistical form, managing data, and analyzing data, to practices denoting interpersonal and personal skills, such as communicating interpretations of statistics to non-statistical audiences, collaborating with domain

experts, or managing priorities.

Academic Settings

There are many different academic pathways that lead to becoming a statistician. In the context of this study, there is no restriction on the type of training a statistician has received, comprising of bachelor's, master's, and doctoral degrees within a variety of disciplines. For example, there are many statisticians who have graduated from programs in statistics or biostatistics as well as specific fields of application such as sociology or psychology.

Workplace

The workplace is defined as a professional environment in which statisticians are evolving and practicing statistics. As mentioned in the introduction, statisticians are hired across many different fields and sectors. In the context of my study, there is no restriction on the type of workplace. In fact, many different professional environments employ statisticians including government agencies, industries, private companies, and educational institutions. In addition, I also consider workplaces for statistical consultants who are even self-employed.

However, since the focus of this study is on the transition from academic settings to the workplace, work at educational institutions requires special attention. Often, statisticians are hired by educational institutions, solely for their work in statistics and not for instructional roles. For example, a biostatistician may work for research for a university hospital or a data analyst may work in institutional research. For the purpose of my study, these environments are considered as workplaces. On the other hand, if a statistician is also hired in an instructional capacity, as a professor for example, this work

environment is not considered as a workplace but still treated as part of academic settings, and therefore not illustrating the transition between academic settings and the workplace.

Transition

The transition can have multiple conceptualizations and I consider the transition to be a two-way interaction between education and profession. In the context of my study, statisticians are considered in transition from academic settings to the workplace if they have graduated from their highest degree within five years and if they consider they have been employed as statisticians for at least three months.

Research Questions

To give future statisticians the opportunity to have an authentic experience of statistical practices throughout their education and therefore better prepare them for the expectations of the profession, this research addressed the following research questions:

RQ1. What important statistical practices do members of the statistical community identify for the role of statisticians?

RQ2. How do statistical practices transition from academic settings to the workplace?

RQ3. In what ways should statistical practices be implemented in academic settings as recommended by members of the statistical community?

Delimitations

The choice of recruitment sites and the selection of participants are some delimitations for this study. First, recruiting member of the statistical community at conferences limited the number of research subjects who met the eligibility criteria.

However, the types of conferences where I recruited participants varied from focusing on statistical practice to teaching statistics, allowing for a diversity of perspectives. Second, the number of junior statisticians and mentors selected for interviews was restricted to ten participants. However, participants represented the diversity of the role of statisticians and allowed for an extensive description of statistical practices across disciplines.

Summary

Giving statisticians the opportunity to share their experience as they transition from academic settings to the workplace helps us understand and characterize their role. This study reveals how statistical practices were developed and how authentic practices should be promoted throughout statistics education in order to better prepare future statisticians.

The following chapter discusses in further detail research related to the experience of the transition to the workplace. The misalignment between learning statistics in academic settings and practicing statistics at the workplace was highlighted, including definitions and theories. The methodology for this study, with the research design, sampling procedures, data collection and data analysis, is described in the Chapter 3. I will also discuss my role as a researcher and possible limitations for my study in that chapter. In Chapter 4, I present the findings emerging from the members of the statistical community and from junior statisticians and their mentors. The last chapter outlines a discussion of the findings, providing recommendations for future research and implications for promoting authentic practices in statistics education.

II. LITERATURE REVIEW

As individuals transition from academic settings to the workplace, they encounter many challenges to become practitioners (Grosemans et al., 2017). An example of difficulties encountered by individuals in transition is the misalignment between what practices were provided in academic settings and the expectations of the workplace (Abrandt Dahlgren, Hult, Dahlgren, Hård af Segerstad, & Johansson, 2006). Previous research shows that there is a misalignment between the training of statisticians and what is required of them at the workplace (Van der Berg, 2017). Indeed, statistical practices differ substantially between academic settings and the workplace, mostly because statistics are taught within unrealistic and oversimplified conditions (Kenett & Thyregod, 2006). Prior studies have shown a misalignment between statistical practices, but research to date provides less insight into the experience of the transition for statisticians and how they learn practices (Hijazi et al., 2019). Therefore, since there is an urgent need for statistical skills (American Statistical Association, 2015; U.S. Bureau of Labor Statistics, 2020), and this need will keep growing (Kettenring, Koehler, and McKenzie Jr, 2015), a deeper understanding of statisticians' experiences as they transition to the workplace is critical to help statisticians overcome the challenges involved in the transition.

The literature review begins with a discussion of previous research on challenges encountered during the transition from academic settings to the workplace. I particularly discuss research that has identified a misalignment between statistics taught in academia and statistics performed at the workplace, noting some of the differences across the two systems. Next, I describe the theoretical framework, drawn on activity theory and more

precisely on the concepts of boundary crossing and boundary objects, framing this study in my conceptualization of the transition between academic settings and the workplace. The review of the literature also introduces different perspectives on statistical practices that are central to the role of statisticians, as identified by employers, statisticians, and educators. Finally, I review some efforts to implement and promote authentic statistical practices in the classroom and conclude with a discussion of gaps in the literature.

Transition between Academic Settings and the Workplace

The transition from academic settings to the workplace often presents challenges because it involves adapting to a new environment, with a new role, different rules, and learning new practices. In most cases, the knowledge that was acquired in academic settings does not perfectly match the knowledge required at the workplace because knowledge is contextually situated (Lave & Wenger, 1991). As a result, individuals in transition will have to engage in learning activities through practice. In a systematic review of the literature, Grosemans et al. (2017) focused on how learning that occurred during the transition from higher education to profession has been researched. They selected 45 empirical studies investigating the transition from different perspectives and disciplines. Overall, the studies revealed that individuals in transition to the workplace experience many challenges across diverse educational backgrounds and disciplines with discrepancies between the knowledge and skills acquired in academic settings and required at the workplace.

Individuals experiencing the transition need to apply and transform their knowledge and skills in order to adapt to their professional role. Although knowledge developed in academic settings is an important foundation for the preparation to the

workplace, the relative importance of this knowledge seems to vary across disciplines. Grosemans et al. (2017) identified that the transition between knowledge and skills learned in academic settings and putting them into practice at the workplace can be experienced in three different ways. First, individuals in transition can experience a relative continuity between their education and their profession. For example, Nilsson (2010) conducted a longitudinal qualitative study to follow physicians in their transition to the workplace to explore the meaning of their education for their profession. Physicians indicated that theoretical knowledge was necessary because every aspect of their work built on previous knowledge and they just needed to reprioritize their knowledge. Such transition is described as continuous. Similarly, Abrandt Dahlgren et al. (2006) followed graduates in psychology who reported that they experienced a certain continuity between their educational background and their work. Second, individuals in transition have reported some discrepancies between their education and their profession in other disciplines. Indeed, engineers reported that the knowledge they acquired through their education was too narrow and loosely coupled with what was required of them at the workplace, and thus considered their previous knowledge as mainly facilitating further learning at the workplace (Abrandt Dahlgren et al., 2006; Nilsson, 2010). In this case, the experiences of the individuals in transition are considered discontinuous. Third, individuals can have an experience that is in-between continuous and discontinuous, called detailing. Individuals in transition can transform their previous knowledge to develop more practical knowledge and skills. Indeed, graduates in political science described the process of their transition as transforming their general knowledge acquired through their education into specific knowledge applicable to the tasks they are required

to perform at the workplace (Abrandt Dahlgren et al., 2006). Thus, in all three cases, individuals in transition need to develop additional knowledge and skills but to a different degree, as continuous, discontinuous, or detailing, depending on how their previous knowledge applied in the context of the workplace.

Grosemans et al. (2017) also explored studies that investigated how professionals develop more practical knowledge and skills. Filstad and McManus (2011), for instance, conducted an ethnographic study to explore the learning experiences of new graduates from the same program in administration and economics who obtained office-based positions; and of new paramedics at the same workplace but having graduated from different programs. Participants were observed and interviewed about their work activities and in particular how they learned to adapt to their new role. Findings across the two case studies showed that learning during the transition was mainly facilitated by social support. Similarly, Lutz (2017) conducted interviews with recent engineering graduates and found that learning for engineers in transition to the workplace was principally described along social dimensions. Therefore, the roles of the mentors accompanying individuals in their transition was a crucial component of the transition, and Grosemans et al. (2017) recommended that in addition to the perspectives of individuals in transition, the perspectives of the mentors should be explored in order to identify a potential misalignment between experiences and expectations.

Thus, there is a need to research the transition between academic settings and the workplace (Hendricks, 2014; Lutz, 2017), and in particular since the experience of graduates may vary depending on the discipline, there is a need to research this transition in specific disciplines (Grosemans et al., 2017; Nilsson, 2010). For example, Hendricks

(2014) explored the transition for graduates in Family Sciences and, more recently, Lutz (2017) focused on the transition for graduates in Mechanical Engineering. In both studies, the authors discussed implications for the education of future workers based on their findings. Therefore, the aim of my study was to better understand the transition between academic settings and the workplace in the field of Statistics as perceived by junior statisticians and their mentors.

Transition for Statisticians

Previous research has investigated the transition between academic settings and the workplace for various levels of education and professional fields, but to date, there are few studies that have explored the transition experienced by statisticians (e.g., Van der Berg, 2017; Hijazi et al., 2019) and as viewed by their mentors (e.g., Van der Berg, 2017; Osman & Ismail, 2009). A misalignment was found between statistical training and statistical skills required at the workplace by the study conducted by Harraway and Barker (2005) which identified gaps in the transition between statistics in academic settings and what is expected at the workplace. In New Zealand, they followed 913 recent bachelor's, master's, and doctoral degree holders in five different disciplines (statistics, economics, business, biological sciences, and psychology) into their early career to identify discrepancies with their statistical training. Discrepancies in the techniques used at the workplace which were taught, or not, at the university were reported through a survey. They restricted their investigation to 46 common statistical techniques and research methods. The findings indicated that even though the nature of the methods lacking in the statistical training of graduates differed across disciplines, graduates from all disciplines agreed they needed more training in regression, multivariate methods,

research design, and power analysis. However, the quantitative study by Harraway and Barker (2005) only addressed specific methods as those were defined and restricted by the authors. Therefore, the misalignment identified does not entirely reflect what graduates may have experienced themselves as they were not prompted to share their own perspectives. In addition, the study describes statistical skills at the workplace in general, and not specifically for statisticians. For my study, I encouraged junior statisticians and their mentors to share their own perspectives in order to give details about their experience of the transition to the workplace beyond technical skills and in particular how they developed statistical practices.

By comparing the perspectives of junior statisticians and managers, Osman and Ismail (2009) explored what knowledge and skills acquired in academic settings met the expectations of the workplace for policy research centers in Egypt. They surveyed 24 junior statisticians and 12 mentors who rated a list of items reflecting statistical practices. To evaluate the preparedness of junior statisticians, each item of statistical knowledge and skills was given a score between 0 and 10 by participants. The median scores and interquartile range revealed the preparedness and the consensus on the items across participants. Both junior statisticians and mentors reported having gained the appropriate knowledge in academic settings to design graphs, prepare statistical tables, and conduct descriptive analysis. However, junior statisticians did not learn how to design databases and from the perspective of mentors they were less knowledgeable in simulation techniques. Junior statisticians felt mostly prepared for skills such as using statistical packages, conducting interviews for data collection, searching the internet, presenting results, and working with a team. Mentors did not refer to junior statisticians lacking any

skill in particular. The study by Osman and Ismail (2009) indicates to what degree junior statisticians transitioned to the workplace in terms of some specific practices. However, there is no indication to what degree these practices contributed to the role of junior statisticians. Indeed, my study intended to reveal important practices for the role of statisticians by ranking practices relatively to one another.

Van der Berg (2017) investigated the transition experienced by statisticians who just completed an internship at the end of their training in an institution for official statistics. In this study, situated in South Africa, 95 interns answered a survey to reflect on their experience of the transition to the workplace. Most statisticians (70.5%) agreed that they acquired the appropriate *statistical knowledge* needed at the workplace, however, most statisticians (71.8%) indicated that they did not acquire the appropriate *statistical skills* needed at the workplace. Therefore, there is a misalignment between statistical knowledge and statistical skills and what statisticians were taught through their education does not prepare them to perform the statistical tasks at the workplace. Furthermore, an open-ended question on the survey offered the opportunity for statisticians to list skills required at the workplace with the indication if that skill was taught or not. A list of 28 skills, mentioned by at least five respondents, was established. Out of these 28 skills, 21 were marked as not having been taught. For example, the most cited skills were data collection (mentioned by 54 statisticians), questionnaire design (mentioned by 54 statisticians), communication (mentioned by 49 statisticians), writing skills (mentioned by 45 statisticians), and using statistical software such as SAS (mentioned by 32 statisticians) or SPSS (mentioned by 21 statisticians). In her study, Van der Berg shows that the role of a statistician at the workplace requires many skills that

were not a part of the statistician's training. Nonetheless, because intern statisticians reported their perceptions of their transition through a quantitative instrument, we do not know *why* they felt underprepared for the statistical skills required at the workplace. And more importantly, we do not know *how* statisticians learn and develop these statistical skills. Thus, there is a need to investigate the transition experienced by statisticians thoroughly as was the aim of my study.

More recently, Hijazi et al. (2019) investigated the experiences of statisticians in the United Arab Emirates. Indeed, this country has a great need for statistical skills but experiences a lack of supply of statisticians. The main reason is because the country needs to improve their statistical studies for official statistics, but they restricted the employment of statisticians in the public sector to nationals. In this context, the authors surveyed 104 statisticians across approximately 40 different workplaces, with about 90% of the respondents worked in the public sector and 75% hold undergraduate degrees, mostly in statistics. Like Harraway and Barker (2005), the questionnaires investigated the use of statistics at the workplace, but respondents were also asked to share their experiences for professional development at the workplace. To keep learning, respondents used hard-copy and online materials, attended conferences, and took face-to-face training courses. They did not mention online courses as a common tool for professional development but alluded to mentoring. To meet their needs for professional development, respondents suggested to receive training in topics such as data management and statistical packages, survey design and analysis, and official statistics. They were less interested in advanced regression models or multivariate analysis but expressed interest with topics in simulation, programming, data mining and big data

analysis. The questionnaires also explored the preparedness of statisticians by having respondents rate the alignment between their education and their profession. Respondents indicated being satisfied with their education in terms of the traditional topics in statistics such as regression analysis or multivariate analysis. Statisticians with a degree in statistics ranked their preparedness for the workplace slightly above average compared to statisticians holding a degree in a related field who rated the alignment between education and profession a little under average. The study by Hijazi et al. (2019) indicates what practices statisticians needed to keep learning at the workplace and what tools they used for professional development. However, because statisticians mostly reported their experiences through close-ended questions on a limited list of practices, we need to research what tools statisticians use for professional development to learn a variety of practices. We also need to understand what aspect of their experiences in academic settings and the workplace affected the rating of their preparedness, for which I designed this qualitative study.

The study by Hijazi et al. (2019) also highlights the role of mentors at the workplace, since more than one-third of the statisticians reported that learning new statistical knowledge is done through mentoring. Previous research has shown that social support is crucial for facilitating the transition to the workplace (Filstad & McManus, 2011; Lutz, 2017). Investigating the perspectives of mentors can contrast the perspectives of individuals in transition and discrepancies between practices developed in academic settings and expectations of the workplace. In her study, Van der Berg interviewed ten mentors who accompanied statisticians during their transition to the workplace in order to contrast how prepared the interns were for performing statistical tasks. Most mentors

agreed that even though interns had some theoretical knowledge, they were lacking practical skills and did not know how to perform most of the statistical tasks. For example, from the perspective of eight of the mentors, statisticians should have primarily developed skills in statistical software packages, in data analysis, and in writing skills through their education. In addition, seven of the mentors agreed that statisticians should have acquired skills in communication, data collection, sampling, and survey methodology before entering the workplace. Previously, Denis, Dolson, Dufour, and Whitridge (2001) explored the perspectives of mentors and junior statisticians about the transition to the workplace for the statistical agency of Canada. New methodologists and supervisors were asked to discuss the transition in terms of preparedness for the knowledge, programming, communication skills, analytical skills, and teamwork. They also briefly discussed the factors affecting the transition. There was no misalignment found for analytical skills, however background knowledge in statistics seemed to be reinforced if recent graduates had courses in sampling theory or survey data, and some consulting experience in academic settings. While junior statisticians reported experiencing teamwork in academic settings, they found integrating larger scale projects challenging at the workplace. According to mentors, junior statisticians were well prepared for programming except in the context of big data. Mentors also suggested that the written communication skills such as summarizing and organizing ideas should be promoted in academic settings. They did not specify how the role of mentor supported learning for junior statisticians at the workplace, but some mentors had different strategies. For example, Anderson-Cook, Hamada, Moore, and Wendelberger (2017) suggest providing support for publications, written and oral presentations, learning

technical skills, interacting with clients, teams, and managers, taking part in networking, service, and socializing. Mentors also have the responsibilities to monitor the workload of junior statisticians so that they do not get overwhelmed by having too many projects or tasks assigned to them. Indeed, Love et al. (2017) recommended to look for mentors at the workplace to develop statistical practices, especially related to consulting or collaboration.

The transition between academic settings and the workplace is particularly challenging for statisticians because they develop practices out of experience (Pfannkuch & Wild, 2000). A misalignment between some aspects of statistical training and statistical tasks to perform at the workplace has been identified in general (Harraway & Barker, 2005) and the lack of preparation for statisticians has been examined in the specific context of official statistics (Van der Berg, 2017). A first inspection of how statisticians learn at the workplace (Hijazi et al., 2019) revealed some topics needed for further training. However, what is missing in the literature is how statisticians develop a variety of practices to adapt to the workplace. Thus, the goal of my study is to give the opportunity to statisticians and their mentors to share their experiences on how they overcome the challenges created by the misalignment between statistical practices learned in academic settings and statistical practices required at the workplace.

Theoretical Framework

As statisticians move from academic settings to the workplace, they encounter challenges, essentially because they need to apply what they have learned to specific contexts and situations. As discussed previously, there is a misalignment between statisticians' education and what is required of them at the workplace, making this

transition particularly challenging. Statisticians navigate the transition between academic settings and the workplace, going back and forth between what they have learned and what is required of them.

In order to investigate how individuals navigate the transition between different contexts, researchers have developed theories and ways of analyzing that transition (Beach, 1999; Konkola, Tuomi-Gröhn, Lambert, & Ludvigsen, 2007). In the context of this study, I chose to draw on activity theory (Vygotsky, 1978; Engeström, 1987) as opposed to the concept of transfer which tends to refer to one-way transitions and focuses on avoiding differences between contexts rather than embracing these differences. More precisely, I conceptualize the transition between academic settings and the workplace with boundary crossing and boundary objects (Tuomi-Gröhn & Engeström, 2003; Akkerman & Bakker, 2011; Star & Griesemer, 1989) because they emphasize a process of establishing continuity between two complex systems. Indeed, several studies have used this framework to study the phenomenon of the transition between academic settings and the workplace (Konkola et al., 2007; Bakker & Akkerman, 2014; Hoyles et al., 2010).

At the workplace, statisticians engage in activities that require them to have developed statistical practices. They may have acquired these practices in academic settings or may have grown practices out of experience. Academic settings and the workplace are complex environments, interacting with one another, where many different elements are involved. Each of these environments (academic and work settings) can be conceptualized as *activity systems* (Konkola et al., 2007; Engeström, 1987). Indeed, activity systems consider both individual-based actions and their interpretations in the

broader context of a complex structure.

Furthermore, the transition between the two systems is challenging for statisticians and the misalignment between statistics taught in academia and statistics practiced at the workplace creates challenges that statisticians encounter during their transition. Statisticians need to learn how to overcome these challenges to adapt to their professional environment. Many researchers have been studying these challenges using the term *boundaries* (Akkerman & Bakker, 2011). The concept of *boundary crossing* in activity theory represents the transition between at least two systems and the involved challenges, or boundaries (Suchman, 1994). For example, boundary crossing can occur when statisticians reflect on their practices that were developed in academic settings compared to practices that are required of them at the workplace. *Boundary objects* can function as bridges between practices from the different systems. For example, statistical practices can be interpreted as boundary objects, meaning that they can help bridge the gaps between the two systems if they are coordinated. Star and Griesemer (1989) introduced boundary objects as artifacts that are present in different interacting systems and that facilitate the communication between these systems by sharing a common identity. Thus, activity theory (Vygotsky, 1978) with the ideas of boundary crossing and boundary objects (Akkerman & Bakker, 2011; Suchman, 1994; Star & Griesemer, 1989) and the learning mechanisms involved at the boundaries (Akkerman & Bakker, 2011) define the theoretical framework for my study.

Activity Theory

Laying the foundations for activity theory, Vygotsky (1978) no longer considered individuals' actions as independent from their environment but rather defined actions as

embedded within a complex system. Engeström (1987) modeled and described the elements of an activity system where individual activities interact with collective activities (see Figure 1). Within an activity system, a *subject* reflects the point of view used to focus on the activity. The goal of the activity is the *object* mediated by *tools* and regulated by *rules* that restrain actions. The *community* refers to all the participants who share the same goal. Finally, the *division of labor* is the distribution of tasks and who hold authority among all the participants.

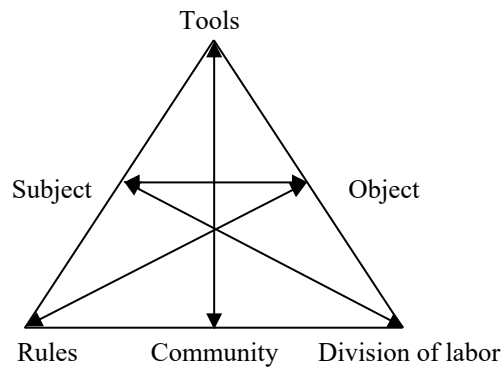


Figure 1. The elements of an activity system.

In addition, an important feature of an activity system is that the system is not fixed. The elements of an activity system are dynamic and often conflict or contradict themselves. These conflicts and contradictions are sources of change, development, and transformation of the activity system and its elements.

For example, in academic settings, the perspective of a student (*subject*) is the focus when investigating how statistical practices (*object*) are developed through teaching and learning strategies (*tools*). The *community* of students, teachers, and advisors work together ensuring that the student meets the degree requirements (*rules*). The teacher

usually defines the responsibilities (*division of labor*) of each participant. At the workplace, the point of view of a statistician (*subject*) is adopted to explore how statistical practices (*object*) are developed using previous knowledge or online courses (*tools*) for example. The collaboration between statisticians, managers, and clients forms a *community*, following regulations (*rules*) and a specific organization of the different tasks (*division of labor*). Table 1 summarizes how the elements of the two activity systems of academic settings and the workplace compare in the context of my study. The objects of the two activity systems are identical, namely statistical practices.

Table 1

Examples of Elements in the Activity Systems

Elements of an activity system	Examples in Academic settings	Examples at the Workplace
Object	Statistical Practices	Statistical Practices
Subject	Student	Statistician
Tools	Courses, Learning strategies	Computer, Knowledge
Community	Students, Teachers, Advisors	Statisticians, Managers, Clients
Division of labor	Teacher authority	Manager authority
Rules	Degree requirements	Best practices, Work ethics

Thus, statistical practices are considered from different perspectives and their interpretations are not necessarily aligned between the two activity systems of academic settings and the workplace. Investigating the boundaries between the statistical practices of each activity system is the goal of this study and that is why I used activity theory and the concepts of boundary crossing and boundary objects to frame this investigation.

Boundary Crossing and Boundary Objects

The transition from academic settings to the workplace occurs as the two activity systems interact with one another (see Figure 2) and can be characterized as boundary crossing (Konkola et al., 2007). Indeed, the term *boundary crossing* denotes how professionals may need to cross boundaries into unfamiliar situations for which they may be unqualified (Suchman, 1994). The misalignment between academic settings and the workplace creates boundaries lying between the two systems. By crossing the boundaries of their own activity system, participants can seek and bring information (the object) from the other activity system.

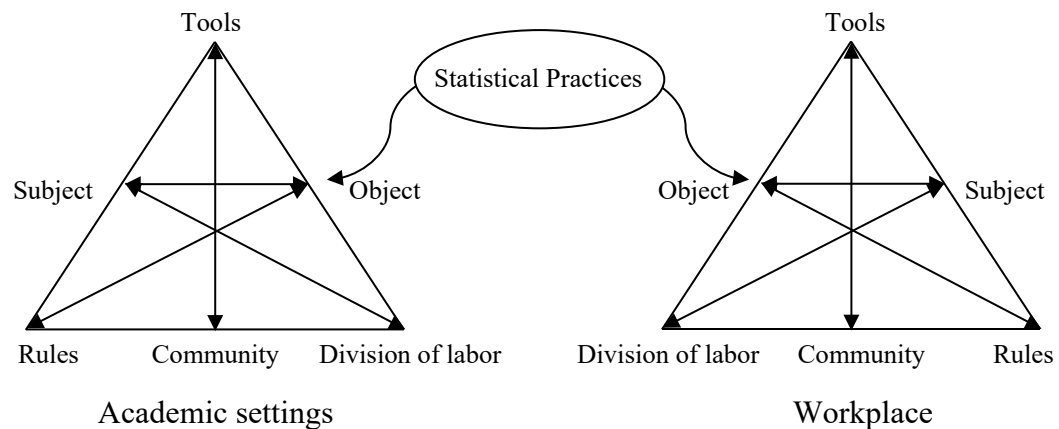


Figure 2. Statistical practices as boundary objects.

In the context of my study, subjects from the activity system of the workplace could cross the boundaries and bring back statistical practices developed by statisticians to the activity system of academic settings. Statistical practices are characterized as boundary objects because they are utilized to assist boundary crossing and should be coordinated between the two activity systems (Akkerman & Bakker, 2011; Star & Griesemer, 1989).

Learning Mechanisms

Akkerman and Bakker (2011) did a systematic review of the literature on boundary crossing, analyzing 181 studies from very diverse disciplines that focused on learning and were framed by the concepts of boundary crossing or boundary objects. Since boundary crossing and boundary objects have been emphasized to further opportunities for learning in all the studies they reviewed, Akkerman and Bakker researched how learning can be characterized at the boundaries. As a result, they distinguished four potential learning mechanisms that can occur at boundaries of activity systems: identification, coordination, reflection, and transformation. Next, I describe all four learning mechanisms and give examples of their meaning in the context of my study.

First, boundary crossing can reveal the identification of objects from each activity system and how they relate to one another. Statisticians and their mentors need to identify and make sense of the statistical practices that are crucial at the workplace and how these practices compare to the practices learned in academic settings. Second, expanding the perspectives from each activity system enables a reflection on the objects. By reflecting on their statistical training and on their experience at the workplace, statisticians can evaluate if there is a misalignment between the practices from the two interacting systems, and therefore they can inform future practice in how to better match practices. Third, the coordination of objects between the activity systems can be explored to establish continuity. In order to overcome boundaries, statistical practices can be coordinated between the reflection on academic settings and the workplace by matching their interpretations in each system. A better understanding of how the practices translate between the two systems is crucial to ease the transition. The role of the mentors strongly

supports the coordination of practices between the two systems. Finally, the activity systems can develop objects collaboratively by transformation. After reflection, statisticians can come to realize that they were lacking some important statistical practices in their training that need to be integrated in future training. From the perspective of statisticians and their mentors, the implementation of crucial statistical practices will conclude the findings.

Summary

The study by Bakker and Akkerman (2014) is an example of the application of activity theory and the concepts of boundary crossing to model the interaction between academic settings and the workplace. Indeed, Bakker and Akkerman developed a boundary object that took the form of a task to be implemented in the classroom but grounded in the context of the workplace. The goal of their study was to investigate to what extent the boundary object was facilitating the integration of statistical knowledge acquired in academic settings and more practical knowledge developed at the workplace during an internship. In particular, Bakker and Akkerman (2014) focused on the learning mechanisms of reflection, by taking different perspectives, and transformation, by integrating the different types of knowledge as a whole. They found that the three students they interviewed were able to increase their level of understanding on how statistics taught in academic settings and statistics performed at the workplace connected. Therefore, these results indicated that boundary objects based on statistical practices can support boundary crossing between the two activity systems and facilitate the transition.

Through the lens of activity theory and more precisely with the concepts of boundary crossing and boundary objects, my study explored the transition experienced by

statisticians as they move between the activity system of academic settings and the activity system of the workplace. In addition, the theoretical framework helped analyze how statistical practices transitioned between the two activity systems, and how we can eventually transform practices by utilizing the four learning mechanisms identified by Akkerman and Bakker (2011). Finally, the collaboration of statisticians, mentors, and educators built boundary objects, namely statistical practices, to facilitate the transition between academic settings and the workplace.

Statistical Practices

The transition from learning statistics in academic settings to practicing statistics at the workplace is challenging because statisticians develop practices that grow out of experience and are relevant to their discipline and context (Pfannkuch & Wild, 2000; Kenett & Thyregod, 2006). Loynes (1987) described the diversity of the training of statisticians with a review of twelve countries around the world, but more recent studies from different parts of the world have reported similar findings about statistical practices. The following sections outline existing research about defining statistical practices from different perspectives in the systems of academic settings and the workplace as well as some notable efforts to promote authentic statistical practices in statistics education.

Practices from the Perspective of Mentors at the Workplace

Employers are defining expectations at the workplace in terms of qualifications and skills. The Matching Education, Assessment, and Employment Needs in Statistics (MEANS) project (Holmes, 1997) was developed in the UK to bridge the gap and establish communication between employers' expectations and statisticians' experiences. The project mainly aimed at identifying statistical skills and knowledge needed at the

workplace and identifying examples of good statistical training. Reporting comments from employers, Holmes (1997) found that planning and designing experiments, communicating with non-statisticians, appreciating the practical importance of statistics were considered crucial skills for a statistician. In response to these comments, statisticians expressed experiencing a misalignment between their education and their employment needs because they felt they were not prepared to practice statistics in a real context throughout their education. Ritter et al. (2001) asked employers of statisticians met during the Joint Statistical Meetings, organized by the American Statistical Association, how they would describe the key elements of the role of statisticians in their organization. The contributors mentioned three main aspects of the role of statistician with statistical, technical, and nonstatistical responsibilities. For example, statistical activities included applying statistical methods and theory, collecting, analyzing, and interpreting data, preparing sample frames, or providing general consulting. Technical responsibilities regrouped using statistical software packages such as SAS, using databases, and conducting web-based search. Finally, nonstatistical activities were also mentioned with writing reports, making presentations, and participating in teams. As a result, Ritter et al. (2001) concluded that employers considered statisticians to be competent if they have developed theoretical knowledge, practical knowledge, and generic skills.

From the perspectives of mentors at the agency for official statistics in Canada, Denis et al. (2001) identified practices that statisticians are responsible for at the workplace. Practices were related to maintain current projects or design survey with emphasis on the methodology, data collection, and analysis. Mentors found junior

statisticians to be well prepared for analysis and programming, but less for communication, especially in writing. More recently, Van der Berg (2017) interviewed intern statisticians and managers who mentored them during the transition to the workplace in South Africa. Findings showed that mentors valued skills in statistical software packages, in data analysis, and in writing skills. In addition, seven of the mentors agreed that statisticians should possess skills in communication, data collection, sampling, and survey methodology. However, mentors noticed that intern statisticians were lacking the aforementioned skills when entering the workplace.

Overall mentors at the workplace with employers (Holmes, 1997; Ritter et al., 2001), managers, and coworkers (Van der Berg, 2017) seem to agree that statistical practices englobe not only the application of statistics theory learned through education, but also more generic skills such as communication. This result echoes what is found in the literature for qualifications of general employment with generic skills being increasingly valued (Grosemans et al., 2017).

Practices from the Perspective of Statisticians

Pfannkuch and Wild (2000) interviewed six statisticians from different disciplines such as biology, finance, medicine and quality control. They explored statistical practices developed by statisticians as the way they are reasoning. For example, they found that statisticians need to develop the ability to notice variation, navigate between various data representations, or build and reason from statistical models. These interviews grounded the theory of statistical reasoning as developed at the workplace and reflected the experience and practices of experts. However, the findings are restricted to few broad practices and do not indicate how statisticians have developed these practices.

Statistical activities at the workplace were assessed by Harraway and Barker (2005) across many disciplines and positions. Their findings indicate that performing data analysis was the most common activity, followed by reading published research. In addition, an important activity for statisticians is to engage in report writing. Similarly, Hijazi et al. (2019) investigated the use of statistics in different workplaces, mostly in the public sector which differed from the private sector in terms of the required practices. They found that, overall, statisticians focused on descriptive statistics more than inferences, modeling, or advanced techniques. Statisticians in the public sectors tended to work more on study and survey design than statisticians in the private sector who spend more time on data analysis. In both studies, the statistical activities were predetermined by the authors and respondents reported if they performed this activity or not at the workplace. We need to give the opportunity to statisticians to report their activities with their own perspectives. A qualitative approach to study their practices as proposed in my study will offer such opportunity.

Research has been conducted directly at the workplace in different fields to investigate statistical reasoning and practices (Kent, Bakker, Hoyles, & Noss, 2005; Bakker, Kent, Derry, Noss, & Hoyles, 2008). Kent et al. (2005) explored statistical reasoning in the workplace, specifically examining boundary objects called techno-mathematical literacies, a combination of mathematical, statistical and technological skills needed at the workplace. They identified a statistical practice during a workplace activity involving reasoning about the center and variation in data that employees needed to develop in the context of controlling a manufacturing process. As part of the same project investigating techno-mathematical literacies at the workplace, Bakker et al.

(2008) identified a statistical practice called statistical process control involving statistical inference. They compared statistical process control with a form of statistical inference presented in statistics education called hypotheses testing. They found that this practice developed at the workplace is different from the concept taught in academia primarily because it is more pragmatic.

Therefore, statistical practices were identified broadly from the perspectives of statisticians (Pfannkuch & Wild, 2000) or very specifically in one context by observing their activities (Kent et al., 2005; Bakker et al., 2008). Statisticians need to identify practices that reflect their role at the workplace in a way that is broad enough to be applicable across many disciplines, and specific enough to be illustrated in precise context.

Practices from the Perspective of Mentors in Academic Settings

Anderson and Loynes (1987) defined statistical practices as techniques of applying knowledge to a specific purpose and distinguishing a novice from an experienced statistician. The aims of teaching statistics should not only focus on the knowledge of statistical theory, methods or applications but also develop many abilities desirable for a statistician. Anderson and Loynes listed many abilities required for a statistician, from general skills, such as communicating clearly and efficiently, to technical judgment such as recognizing the appropriateness of statistical methods. They also recommended that a list of all the technical skills that most statisticians should master would help further the understanding of statistical practices.

Similarly, Cameron (2006) conceptualized aspects of statistical practices as actions performed by statisticians and identified categories of what statisticians do.

Indeed, he identified five components that form statistical practices. First, statisticians formulate a problem to be addressed through statistical means. Second, statisticians contribute to preparing data, including planning, collection, organization and validation. Third, they analyze data, by models or other summaries. Next, results are presented in written, graphical or other form. Finally, statisticians research the interplay of observation, experiment and theory.

Thus, statistical practices have been identified from different perspectives. They have been considered by employers, statisticians, and educators. However, the research on statistical practices did not report how statisticians have developed these practices and what practices are particularly challenging when statisticians transition from academic settings to the workplace. Statistical practices are constantly evolving so there is a need to update what has been found thus far. Next, I will describe some recommendations and efforts that have been developed to promote authentic statistical practices in academic settings.

Promoting Authentic Statistical Practices in Academic Settings

Van der Berg (2017) examined how the training of statisticians could be improved by investigating the perspective of statisticians themselves. They recommended that their training should focus on technical skills and have a practical component, applying theory to practice. Indeed, Kenett and Thyregod (2006) argued that there is a lack of authentic experience of statistical practice in the classroom. They claimed that the examples provided in textbooks are over-simplified and not connected to real life and that statistics education tends to focus solely on instructing students to perform statistical analyses and ignores other practices that are part of the role of a statistician. There is often little

connection between the context of a problem and the statistical methods used as statistical methods are often predetermined in solving statistical problems in education but the identification of the method is actually the issue in the context of the workplace. Thus, the training of statisticians should incorporate the ways statistical problems and applications are presented in a professional environment.

To include statistical practices in academic settings, there are diverse implementation ideas that have been developed. For example, after Anderson and Loynes (1987) reflected on the different approaches that could promote statistical practices in statistics education, they concluded that a mixture of methods should be implemented as a sequence of short assignments during a semester. Some of these methods are critical reading, statistical experiments, work placement, consulting courses, case histories, role-playing and case studies, seminar presentations, and discussion groups. They offered topics for mini projects, all involving articles and contexts from the 1970s and early 1980s that are not necessarily relevant to current statistical practices.

Similarly, after observing a misalignment between education and the workplace, Harraway and Barker (2005) concluded that support and collaboration should be established between departments at the university and between disciplines to include statistical practice. They offered multiple implementation ideas such as critical appraisal of articles, group projects, consulting, or work placement. They also mentioned that statistics courses could be taught by statisticians or practitioners who are aware of what statistical practices should be fostered throughout statistics education. Based on the perspectives of junior statisticians and mentors, Denis et al. (2001) proposed to collaborate with agencies and businesses to share real data but also to introduce research

issues they encountered in methodology, data collection, or analysis in academic settings. Members of the statistical community should be involved through seminars or publications, for instance.

To prepare qualified statisticians, Cameron (2006) recommends focusing on different aspects of the role of statistician. Statistics education should incorporate foundations of statistics and mathematical foundations that help understand statistical theory and applications, the theory and practice developed by professionals as well as collaboration and leadership. Both junior statisticians and mentors suggested to develop curricula materials to balance theory with practice in academic settings (Osman & Ismail, 2009). For example, students can be given the opportunity to complete a significant project in collaboration with an expert from another discipline. Also, Cameron (2006) suggested designing courses that review the history of applied statistics with its successes and pitfalls.

As an example of a successful implementation of a workplace situation with the introduction of a new statistical method, Bakker and Akkerman (2014) showed that students were able to integrate school-taught knowledge and work-related knowledge. Indeed, students successfully developed a sophisticated understanding of statistical practices from the perspective of the experts, grounded within the context. Likewise, Hahn (2011) developed a pedagogical situation to support students in linking their academic knowledge with their work experience and Gibbons and MacGillivray (2014) offered a developmental and mentored tutoring program to train statisticians grounded in a professional setting, inspired by the authors' experience. The program aimed at preparing graduate statisticians for employment needs by developing and enhancing

specific practices through an authentic experience.

For example, practices related to collaboration were found to be challenging for junior statisticians and should be developed with other disciplines (Kettenring et al., 2015). To teach collaboration, Vance and Smith (2019) break apart the practice into five different aspects: Attitude, Structure, Content, Communication, and Relationship (ASCCR). This framework can be used at the workplace by junior statisticians in transition or can be implemented in academic settings. Indeed, Vance and Smith (2019) recommended to use the ASCCR framework across courses and at their institutions, they integrated the framework into consulting centers to give students an experience of collaborating with clients.

Therefore, recommendations and efforts have been developed to promote an authentic experience of statistical practices in academic settings. However, in order to implement statistical practices throughout education we need 1) to identify important statistical practices from the perspectives of statisticians and their mentors, and 2) to provide illustrative examples of statistical practices so that educators can develop activities, tasks, or curricula to integrate work-related knowledge.

Summary

The review of the literature has shown that statistical practices can be conceptualized as boundary objects facilitating the communication between the activity system of academic settings and the activity system of the workplace. The misalignment between education of statistics and the application of statistics at the workplace is primarily caused by the lack of authentic experience of statistical practices in academic settings. Before designing curricula and programs that incorporate statistical practices to

ease the transition to the workplace, it is crucial to recognize important practices as identified by the experts of the field and better understand the elements involved in the transition for statisticians to learn, including the role of mentors.

A detailed description of the transition between academic settings and the workplace experienced by junior statisticians and from the perspective of their mentors will reveal important statistical practices and how they were developed. The methodology including the research design, sampling procedures, data collection and data analysis, will be described in the following chapter.

III. METHODOLOGY

The purpose of this study was to describe how statisticians learn practices in the transition between academic settings and the workplace. Based on their experiences, important statistical practices performed by statisticians at the workplace and the elements that supported learning these practices were revealed. The implications of my study give future statisticians the opportunity to gain an authentic experience of statistical practices throughout their education and therefore better prepare them for the workplace's requirements. I addressed the following research questions:

RQ1. What important statistical practices do members of the statistical community identify for the role of statisticians?

RQ2. How do statistical practices transition from academic settings to the workplace?

RQ3. In what ways should statistical practices be implemented in academic settings as recommended by members of the statistical community?

To answer these questions, I developed a qualitative study that involved 154 members of the statistical community, with statisticians, educators, and students of statistics. By engaging members of the statistical community in an activity, I gathered perspectives on important practices for the role of statisticians. The results of a survey indicated what practices statisticians developed in academic settings and at the workplace and reflected on what they wished they had focused on in academic settings. To learn about how statistical practices transition between academic settings and the workplace, I conducted interviews with eight junior statisticians and two mentors. Via interviews, I obtained rich

and in-depth descriptions of how junior statisticians learned during the transition to the workplace and how mentors supported their learning.

This chapter describes the research design for the study, sampling procedures, information about the participants, data collection, data analysis as well as a discussion of my role as a researcher and possible limitations for the study.

Research Design

While transitioning from academic settings to the workplace, statisticians encountered boundaries arising from the lack of development of some practices (Harraway & Barker, 2005; Van der Berg, 2017). I interpreted the transition as an interaction between two activity systems, academic settings and the workplace, which both involve statistical practices but from different perspectives. The concept of boundary crossing represents this interaction (Akkerman & Bakker, 2011) and occurred as statisticians reflected on practices that were developed in academic settings compared to practices that were developed at the workplace. The four learning mechanisms (identification, coordination, reflection, and transformation) involved at the boundaries of academic settings and the workplace, operationalized the design of my study to answer the research questions.

First, the *identification* of important statistical practices at the workplace revealed the perspectives of the statistical community on the role of statisticians (RQ1) through a sorting task. Second, participants engaged in a *reflection* about their experiences through a survey and described what practices they developed as they crossed boundaries between academic settings and the workplace. This reflection was the basis for interviews of junior statisticians and their mentors that showed the *coordination* of statistical practices

between academic settings and the workplace (RQ2). Third, participants suggested *transformation* of statistical practices to promote an authentic experience in academic settings in the survey and during interviews as well. Finally, I confronted the perspectives of junior statisticians and their mentors to discuss better coordination in the last chapter.

Since the purpose of this study was to investigate the perspectives of statisticians on their role at the workplace and how they experienced the transition from academic settings to the workplace, a qualitative approach was considered. Indeed, the phenomenon of statisticians transitioning to the workplace is not well-understood and need further exploration (Bakker & Akkerman, 2014). As recommended by Grosemans et al. (2017), research on the transition to the workplace has been conducted from a qualitative approach, and I added a quantitative component to identify important statistical practices (Brown, 1980). A qualitative study allows for a better understanding of the phenomenon of the transition (e.g. Hendricks, 2014; Lutz, 2017). Although findings from qualitative research can hardly be generalized, providing thick descriptions and maximum variation in the selection of participants strengthened the results (Merriam, 1998). A phenomenological approach was considered as this study focused on interpretations of lived experiences (Creswell, 2018). Indeed, the results gave a rich and in-depth description of the experiences of junior statisticians while crossing boundaries between academic settings and the workplace, and how mentors provided support. Each participant constituted a case for learning practices in the transition to the workplace and a cross-case analysis lead to “a unified description across cases” (Merriam, 2009, p. 204), depicting statistical practices across participants. I added a quantitative component to the study to investigate the perspectives of the statistical community on the role of

statisticians, identifying important practices. My study intended to uncover a complete picture on how statisticians transitioned from academic settings to the workplace and, in particular, how they learned important statistical practices.

Thus, in order to explore the perspectives of junior statisticians and their mentors, multiple sources of data were gathered. Data included a sorting task, open-ended survey, and interviews, to provide a rounded picture of the transition from academic settings to the workplace as experienced by statisticians.

Participant Selection

In order to learn about practices performed by statisticians and how practices are developed at the workplace the data collection involved two phases in participant selection: first, a sample of convenience of members of the statistical community, and second, a purposeful sample of junior statisticians and mentors. I recruited participants through the following four steps. First, I met members of the statistical community at conferences focusing on statistical practice or statistics education worldwide. The role of a statistician is universal and the transition between academic settings and the workplace for statisticians is challenging in many parts of the world, for example in Oceania (Harraway & Barker, 2005; Gibbons & MacGillivray, 2014), in Europe (Bakker & Akkerman, 2014; Hahn, 2011), in Africa (Van der Berg, 2017; Osman & Ismail, 2019), or in America (American Statistical Association, 2015). Conferences offer professional developments where attendees are willing to improve, learn, and share their practices. Second, attendees engaged in a task that explored what practices they value at the workplace. This task enabled attendees to start reflecting on their experience and raised their awareness about the issue of the transition to the workplace. They realized they

could participate in my study and contribute to better understand this phenomenon. As they decided to take part in my study, participants completed a survey that collected information about their educational background, their professional experience, and how they transitioned from academic settings to the workplace. The survey also served as a tool for checking eligibility requirements and recruiting participants for interviews by asking them to share their contact information for further research. Finally, a purposeful selection of junior statisticians and mentors who were willing to participate and represented the diversity in educational backgrounds and fields of application, were contacted to be recruited for interviews. The goal of the interviews was to investigate the transition between academic settings and the workplace in detail.

Members of the Statistical Community

A convenient sample of members of the statistical community was involved at conferences. The themes of the conferences focused on the practice or on the education of statistics. Overall, 154 members of the community were recruited at seven conferences across three continents (see Table 2). I engaged the audience in my study through various formats: from workshops, to speed presentation, to poster sessions. For the last conference, the Joint Statistical Meetings, I sent invitations prior to the conference by posting information on the conference's online forum. All attendees were provided the opportunity to participate and those who were interested returned completed materials with a signed consent form (Appendix A). Six attendees submitted materials without the consent form signed therefore they were not included in the analysis. In the first phase of the participant selection for members of the statistical community, participants who wished to be contacted for further research shared their contact information.

Table 2*List of Conferences for Recruiting Participants*

Name	Location / Date	Type of presentation	Number of participants
Conference of Texas Statisticians	San Antonio, TX April 2018	45-minute workshop	30
50èmes Journées de Statistique	Paris, France June 2018	15-minute speed presentation	19
International Conference on Teaching Statistics	Kyoto, Japan July 2018	Poster	8
Conference on Statistical Practice	New Orleans, LA February 2019	35-minute workshop	39
US Conference on Teaching Statistics	State College, PA May 2019	Two 60- minute workshop	31
51èmes Journées de Statistique	Nancy, France June 2019	15-minute speed presentation	3
Joint Statistical Meetings	Denver, CO July 2019	4-minute speed presentation, poster, and forum	24

Junior Statisticians and Mentors

To investigate how statisticians developed practices at the workplace, I invited junior statisticians and mentors to share their experiences of the transition. The goal of the selection of participants for the interviews was purposeful sampling with the following criteria to recruit participants who (1) self-identified as statisticians or similar roles that involve performing statistical tasks in at least 50% of their scope of work, (2) started working at the current workplace for more than three months, and (3) graduated from their highest degree within the past five years. These criteria defined who I considered to be junior statisticians. In addition, participants were recruited if they considered themselves as mentors of junior statisticians, satisfying the eligibility criteria (1), (2) and (3). Indeed, Grosemans et al. (2017) recommended to include the

perspectives of the mentors accompanying the individuals in transition to better capture the experience of the transition and confront expectations.

By answering a survey in the first phase of participant selection, members of the statistical community reported information that I used to determine their eligibility for being selected for interviews. I started the selection of participants for interviews after the third conference. I analyzed the survey responses and identified eligible participants. Among eligible participants, I looked for participants who maximized the variation between fields of applications and types of degrees for junior statisticians. I also prioritized participants who had lengthy responses on the survey, indicating that they were eager to share their experience. Concerning mentors, I selected participants who supported junior statisticians at the workplace or students in academic settings, in various fields of application. Overall, I contacted 13 potential participants for an interview, of which 10 responded positively, with eight junior statisticians and two mentors. An informed consent form was sent to them by email prior the interview and described the procedures, risks and benefits as well as protection of confidentiality. The ten interviews were conducted between May and October 2019. Recruitment for interview participants concluded when the information collected started to become redundant (Merriam, 1998). I believe the size of the sample was adequate because it allowed for a variety of fields of applications and educational backgrounds.

Participants

In this section, I describe characteristics of members of the statistical community who took part in the sorting task and the survey. I also summarize background information for junior statisticians and mentors who participated in interviews.

Sample of Members of the Statistical Community

Participants self-identified with diverse roles within the statistical community (see Table 3). The role of statistician is diverse itself and participants may not have self-identified as statisticians while they were involved in performing statistical tasks on a regular basis. For example, some professionals may primarily self-identify as actuaries or zoologists but dedicate a large amount of their work time to performing statistical tasks. Therefore, the category for the role of statistician was not restricted only to professionals who self-identified as statisticians but required that statistics was one of the primary focus of the professional's position, accounting for at least 50% of their work time. For example, participants in that category also considered themselves as analysts, data scientists, biostatistician, consultant, or business intelligence developer. I distinguished between two categories of statisticians depending on the type of workplace, academia versus the industry. I classified participants as *statistician in academia* if they identified as a statistician and having other responsibilities in academic settings. For example, researchers in the theory of statistics fall in that category as well as professors.

Table 3

Various Roles in the Statistical Community

Role	Count	%
Statistician	49	33
Educator	46	31
Graduate student	26	17
Statistician in academia	15	10
Undergraduate student	14	9

Note. Two roles were missing.

Table 4*Diverse Fields of Application for Statisticians*

Field	Count	%
Health / Pharmacy	27	57
Government / Policy	8	17
Economics / Finance	6	13
Engineering	3	6
Consulting	3	6

Note. Two fields of application were missing.

Participants who identified themselves as educators mainly mentioned teaching mathematics or statistics. They also reported some subject matters such as educational sciences, psychology, biology, or computer science. Graduate students reported studying statistics, biostatistics, or machine learning while undergraduate students mentioned pursuing mathematics, statistics, and data science. Participants who identified themselves as statisticians represented the diversity of the role in the fields of application of statistics (see Table 4), with more than half worked in healthcare or in the pharmaceutical industry.

Statisticians reported professional experiences ranging between one month to 40 years at the time of their recruitment. On average a statistician had 10.1 years of experience with a standard deviation of 10.7 years. Participants who identified as statisticians had various levels of educational backgrounds as shown by the highest degree earned with three bachelor's, 26 master's and 15 doctoral degrees. They also had their degree in various disciplines with 18 statisticians acquired degrees in statistics and 14 in biostatistics, two in mathematics and seven others in specific disciplines such as engineering, psychology or public health.

Overall, participants represented different roles in the community of statistics with statisticians, educators, and students. They represented the diversity of fields of application of statistics with various educational backgrounds and professional experiences.

Sample of Junior Statisticians and Mentors

Participants who identified as statisticians from the first phase of participant selection were considered for the second phase. Statisticians were invited to participate if they met the additional eligibility criteria (2) and (3) mentioned earlier and represented the diversity of fields of applications shown in Table 4, with various educational backgrounds. In order to share their experience about the transition between academic settings and the workplace, participants were required to have graduated from their highest degree within the past five years and they should have been working at their current workplace for at least three months. Participants meeting these eligibility criteria are called junior statisticians. In addition, statisticians who identified as mentoring junior statisticians were also invited to participate. Indeed, the experience of the transition should be reported by the person who is in transition as well as direct observers of the transition as suggested by Grosemans et al. (2017). To select participants, there was no restriction on the type and level of degrees obtained by the participants because statisticians have very diverse pathways. The highest degree earned by the junior statisticians was at the level of a master's degree in various disciplines. There also was no restriction on the field of application at the workplace.

Therefore, the junior statisticians and mentors selected for this study represented diverse educational backgrounds as well as professional experiences (see Table 5 and 6).

Previous studies have considered workers in transition who either came from similar academic settings (Filstad & McManus, 2011; Nilsson, 2010; Hendricks, 2014; Lutz, 2017) or were transitioning to the same workplace (Filstad & McManus, 2011; Kent et al., 2005; Van der Berg, 2017). However, since the misalignment between academic settings and the workplace has been identified worldwide and at different levels of education, the selection of participants provided meaningful results. Indeed, the analyses of the experiences of statisticians within diverse settings revealed common themes, indicating that junior statisticians identified similar challenges during their transition from academic settings to the workplace with no regard to specific disciplines. However, the results showed that there may be some differences induced by the educational backgrounds. Next, I summarize the educational background and the context of the professional experiences of the eight junior statisticians and two mentors I interviewed for this study.

Description of Junior Statisticians

An overview of salient characteristics for each participant is displayed in Table 5, reporting their different degrees in various disciplines and the year they graduated, leaving academic settings. At the workplace, the fields of application and the number of years of experience are mentioned for each participant, characterizing their professional experience. The names given to the participants are pseudonyms, named after famous statisticians. I included more detail about each participant by summarizing their overall experience in academic settings and at the workplace with information about their motivations to study statistics and pursuing their degrees as well as their search of a position in statistics and a description of their professional environment.

Table 5*Description of Junior Statisticians*

Name	Degree	Discipline	Graduation Year	Field of Application	Years of Experience
Gertrude	MS	Applied Statistics for Social Science	2016	Book publishing in retail	3
	BS	Mathematics Economics	2014		
Enid	MS	Applied Statistics	2016	Public Policy	2.5
	MS	Mathematics	2014		
	BS	Mathematics	2011		
	AA	Mathematics	2009		
Florence	MS	Biostatistics	2017	Genetics	1.5
	BS	Biostatistics			
Thomas	MS	Biostatistics	2019	Government	2
	BA	Quantitative Economics and Community Health			
Stella	MS	Applied Statistics	2017	College Access	2
	MS	Physics			
	BS	Earth			
	AS	Mathematics			
Janet	MPH	Public Health	2019	Omics	1
	BS	Global and Community Health			
John	MS	Biostatistics	2018	Pharmaceutical	1
	BA	Mathematics	2017		
Edwards	MS	Applied Statistics	2015	Biomedical	3
	BS	Mathematics			

Note. AA = Associate in Arts; AS = Associate in Science; BA = Bachelor of Arts; BS = Bachelor of Science; MPH = Master of Public Health; MS = Master of Science.

Gertrude. Gertrude started studying psychology, but she decided to pursue a degree that was more interdisciplinary and majored in Mathematics and Economics. She became especially interested in statistics because she could play “in everybody's backyard.” After graduating, she had a visit from the Census Bureau to complete a survey and through this experience she realized how people could lie on a survey and wondered if there are techniques to account for such bias in the analysis. She came across an ad for

a master's program in Applied Statistics in Social Science at her university and decided to enroll: "I'm going to go study statistics. Like all signs are pointing here."

After graduating in 2016, Gertrude first looked at positions with nonprofit organizations, but she was advised to start her career in the private sector, to first get the experience and money. She interviewed and found her job at the Joint Statistical Meetings organized by the American Statistical Association in a midsize book publishing company. At the time of the interview, she had been working there for about two years. To navigate the transition to the workplace, she identified a coworker "as a mentor now, not as a statistician but someone who is like a, AI and machine learning professional."

Gertrude had an opinion about what data science is and what data scientists do: "they've written a model. It has very good accuracy. Therefore, it is a good model and it takes some, you know, five minutes to write this model." She noticed that data scientists tend to overlook data cleaning: "they're not examining what they're looking at, they're just building a model and saying the accuracy is very high." She thought that data scientists tend to be less rigorous with the model assumptions and validity than statisticians.

When reflecting on academic settings, Gertrude recommended to include more programming, especially starting in undergraduate courses. She was speaking on panels at the university she graduated from to share her experience with students. Gertrude was looking for a new position at the time of the interview.

Enid. Mathematics was a passion for Enid growing up. She started a degree in Mathematics with a minor in Education to become a teacher. She was not able to find a job as a teacher in local schools even with 8 through 12 and K through 6 generalist

certifications. She decided to continue her education with a master's degree in Mathematics and the university she attended offered a focus on Statistics. She graduated in 2014 and started a doctoral program in Statistics at another university. The requirements of the degree were different from what she had expected and Enid struggled to adapt to the program even though she built a "strong support system [...] with students that I was going to school with as well as the faculty." After the first year, she decided to switch to a master's in Applied Statistics. She was disappointed to abandon her pursuit of a doctoral degree, but she was still thinking about going back at the time of the interview.

Enid graduated in December 2016 and started her job search the summer before. She attended the Joint Statistical Meetings and had a dozen interviews with a variety of companies and institutions in different fields: biostatistics, pharmaceutical, agriculture, demographics, economics, public policy, and retail. She "just wanted somebody to let me put my foot in the door so I could get the experience." It "took me the entire semester to really even get close to a job offer. It was frustrating."

Enid had been working for a government institution for a year and a half at the time of the interview. She "felt like [her job] was a continuation" of her education. In particular, for the skill of programming she had "had all of that education to kind of bolster it, to be like, well, I know enough and if I don't know it, I can Google it, which is what I do." She mentioned receiving encouragement from a mentor in academic settings through her job search and her transition to the workplace, as well as being supported by a mentor at the workplace through a mentoring program organized by the institution. The situation of Enid was unique because she was a teen and single mom and had to take care of her kids, which motivated her even more to complete her degrees and find a stable job.

Florence. After taking many engineering courses, Florence took some biology courses and “fell into” statistics. As a result, she decided to pursue a bachelor’s degree in Biostatistics. She did an internship at a diabetes clinic to do statistical analysis and this experience reinforced her desire to study biostatistics: “the first week I was there, we made a significant discovery in Alzheimer’s genetics about a certain gene that no one had discovered before. And it had this amazing feeling of being able to help people through doing mathematics.” She also completed a thesis as part of her undergraduate studies. She enrolled in a master’s program in Biostatistics with a genetics track, for which she took more courses in advanced mathematics than she expected.

Florence started to apply for jobs a few months before graduating, in December of 2017. She targeted a wide range of fields but she prioritized the location to be closer to her family. She was recommended for a position at a research institution by a person she worked with during her internship. She started to work almost immediately after graduation even though “they were looking for someone a little more experienced than me. And that has definitely caused some issues just in terms of their expectation versus my skill level. That took us awhile to figure out.” Indeed, Florence’s transition to the workplace was challenging because she did not have a mentor until eight months after she started. The role of the mentor was extremely important to Florence who struggled with adjusting to the expectations of the workplace and the intervention of a senior statistician helped bridge the gap between the employers’ expectations and Florence’s skills. Florence recommended to include courses on data cleaning across courses and to practice communication skills with interdisciplinary collaboration with students across different disciplines within the college.

Thomas. Thomas initially wanted to go to medical school, but he decided to study Quantitative Economics with a minor in Community Health. As part of his undergraduate studies, he did an internship in a healthcare market research firm.

After graduating, Thomas decided to apply to jobs to strengthen his application to master's degrees with some hands-on experience and publications. He found a research assistant position at a career fair organized by his university. The recruiters were looking for someone with a background in health, mathematics, and economics. He worked at this health policy organization for two years doing “the dirty work of research, the literature reviews, the dataset cleaning and manipulation and some of the more standard analyses and a little bit of manuscript writing, kind of the organization of the research.”

Thomas knew he wanted to go back to further his education to do more advanced research and he enrolled in an accelerated master's degree in Biostatistics and graduated within three semesters. He focused on data science, machine learning, and deep learning, and he did a master's thesis that was “more applied than theoretical.”

Thomas had his graduate studies sponsored by a stipend and he owed the sponsor a one to one match for employment. At the time of the interview, Thomas had been working for the sponsor, a government institution, for six months. He identified a mentor who helped him navigate the government work and proposed topics of research.

Thomas had a second job, working for a Law school as a remote clinical research associate which he said had a “little more independent work and then a little bit more of a continuation of what I learned on both undergrad and my master's.” He balanced his work with 40 hours a week onsite for the government job and between 5 to 15 hours remotely

for the Law school. Next, Thomas was thinking about becoming a data scientist or a consultant, in the field of healthcare.

Stella. Stella started her academic journey with Earth Science and a minor in Mathematics. She took a course in statistics and “fell in love with it” but it was too late to change her major. She wanted to be a geologist but there was no opportunity where she was located, and she was not willing to move. She worked as a chemist for a few years. After running into her high school physics teacher, she decided to apply to a master’s degree in Physics. Her thesis used statistics and she “enjoyed the statistics part more than anything else. And the funny thing is my physics courses really didn’t have a lot of statistics in them, but my research did.” She worked for a scientific government agency for six years and then did underwriting at a health insurance company. She felt she was not challenged enough by her job, so she decided to pursue a master’s degree in Statistics. She completed one course per semester online while working fulltime except for the last year, which involved a capstone project that was time consuming.

After graduating, Stella sought the support of the career services at the university she attended for her undergraduate studies to look for a job. They made her realize that “nobody knows everything ... you learn over time” which helped her deal with the imposter syndrome and worrying about being a good statistician. She found that her job search was very challenging, it was “feast or famine. Either I am not qualified enough or I’m overqualified.” She had many degrees and years of experience, but not as a statistician. Overall, she noticed that employers were looking for experienced statisticians who had published in journals. She was also worried that being older than other graduates could be held against her and she restricted her job search in a specific location.

Stella found a job within a year after graduation, with a nonprofit organization in college access for first generation college students. She had been working there for two years at the time of the interview. She loved her job and enjoyed the flexibility in making an impact, she said: “even though I'm behind the scenes crunching the numbers I help make a difference in the lives of our students. And it's amazing to hear stories about students that have gone through our program.”

Janet. Janet decided to pursue a degree in Global and Community Health because she wanted to work in the field of healthcare with a “practical application of preventing disease, rather than going into the medical field where I would have to be treating it.” After graduating, she joined a volunteer program run by the government for two years, to provide social and economic development abroad. She did survey development and qualitative research, mostly collecting data and she realized she was more interested in conducting research.

Janet decided to pursue a master’s degree in Public Health with a focus on epidemiology and biostatistics, expanding what she had learned abroad. She took many courses from the biostatistics department and had the opportunity to do research at a government health institute with which she partnered to do a thesis. Overall, she thought that it “was a very practical masters, like you definitely leave knowing how to do something when you walk out.”

Janet started applying to jobs six months before she graduated and applied to maybe 100 jobs. At the time of the interview, she was still receiving calls for interviews from the jobs she applied to six months earlier. She was mostly interested in doing research at the university she graduated from and started working there right after

graduation. She felt “like it's been a really good transition, but it's expanded my knowledge and expertise in ways that I never expected.”

Janet had been working as a biostatistician for “an internal professional services organization” for one year at the time of the interview. She worked for “genomics projects or just omics projects in general.” As she collaborated with clients, she learned to be patient and having experience in customer service and retail helped her develop that skill. She kept learning, by reading, taking more courses at the university she worked for, especially in biostatistics or programming, or attending conferences and workshops nationwide. She also had an extra job to teach one laboratory section for a biostatistics course and started a journal club.

John. John started college with the idea of pursuing Chemistry and Physics but realized he was not interested in careers in these domains so he took “random classes in undergrad, trying to figure out what I wanted to pursue.” After he took an introductory statistics course in psychology, he was inspired by his instructor and decided to major in Mathematics with a minor in Education. There was no degree in statistics at his university at the time. During the last year of his undergraduate degree, he included master’s coursework and then completed an accelerated master’s program in biostatistics within a year. He thought that having a degree in biostatistics would give him “a strong foundation, just statistical theory in general that even if I didn't want to work with biological applications, I could apply it elsewhere.” As part of the program he did an internship over the summer with a research lab and he did a one-year master’s thesis.

John started to apply for jobs about five months before he graduated. He was initially looking for a position in business analytics. He had several interviews and started

getting offers before he graduated. To decide between different offers, he considered the support of other statisticians at the workplace, how he could develop his career, and the salary negotiation. He chose a pharmaceutical company for its team spirit and had been working there for a year at the time of the interview. He was looking for a mentor by scheduling one on one meetings with everybody to learn about their background, experience, and career growth. He found a mentor after two or three months who he considered as a role model. In the future, John might consider pursuing a doctoral degree because he observed that doctoral level statisticians had different responsibilities that he would be interested in having, instead of focusing on programming.

Edwards. Edwards was first interested in business, then engineering but had more affinities with mathematics so he pursued a degree in Mathematics with a minor in Mathematical Statistics. He started to take courses in statistics to have “a nice backup in case I couldn't get into a doctoral program for mathematics.” Indeed, he was initially interested in set theory but decided to pursue a master’s degree in Applied Statistics. He thought that there were not as many prospects with set theory whereas he found that machine learning was great “in terms of marketability.”

After graduating with his master’s, Edwards applied to teach at the college level. He primarily taught statistics but soon realized that “I didn't want to be a teacher that hadn't done it before. I'm teaching statistics but had never actually done it in practice.” He decided to apply to statistician’s positions in the field of medicine. He applied to one job found on an online job search engine and had been there for three years at the time of the interview. Overall, he said that his transition was not entirely continuous and “some of that has to do with logistics.”

Edwards worked for a not-for-profit organization and academic medical center and valued the relationship he had with his coworkers. Indeed, he was assigned a mentor who helped him transition to the workplace and motivated him to get the job in the first place during the interview process. Edwards also coordinated educational opportunities and served on committees. He recommended to teach future statisticians how to search for resources and digest papers.

Description of Mentors

Similarly, characteristics in academic settings and the workplace is displayed for each mentor in Table 6. I described their own educational background and experience as statisticians as well as their role of mentors for junior statisticians and future statisticians.

Table 6

Description of Mentors

Name	Degree	Discipline	Graduation Year	Field of Application	Years of Experience
Simone	PhD MS BS	Epidemiology Statistics Economics	2010	Statistician in Health Care Adjunct Professor in Biostatistics	14
David	MBA MS BS	Strategy and Management Statistics Biometry	2011	Statistician in Higher Education	8

Note. BS = Bachelor of Science; MBA = Master of Business Administration; MS = Master of Science.

Simone. Simone was first interested in going to Law school and completed a bachelor's degree in Economics. Since she had a full scholarship and completed her coursework in three years, she took additional courses in her last year that were electives for a master's program in Statistics. She then decided to finish this master's degree and enrolled in a PhD in Epidemiology and Biostatistics. Simone started working fulltime

while finishing her master's in the field of psychometrics, continued in that field for four years and then she switched to biostatistics. At the time of the interview, she worked as a statistician for a nonprofit agency and she was also an adjunct professor at a university.

As a mentor at the workplace, Simone was in the managerial track and she participated in the recruitment of junior statisticians. She thought it was easier to recruit junior statisticians rather than senior statisticians because there were many more applicants for entry level positions and she considered that senior statisticians had already been trained a certain way while junior statisticians can be trained at the company's standards. She said: "we invest in people, so, um, that's part of their development." She thought junior statisticians take about six months to adapt to the workplace, especially because they have a lot of internal training for the first two months.

As a mentor in academic settings, Simone taught one graduate course per semester, either advanced epidemiology or an independent study to prepare students to work in the industry. She assigned randomly projects to students because she said that in the real-world context, statisticians do not always pick what they are interested to work on. She also assigned two projects at the same time and gave short deadlines, so students learned how to balance their time. She collaborated with another university to discuss how to give students authentic opportunities to learn.

David. David became interested in statistics as a child by collecting baseball cards and reading the statistics on the back. He started his educational journey with a bachelor's degree in Statistics and Biometry with a minor in Communication. He took advantage of some research and internship opportunities, for example, he "worked with a professor and we analyzed tweets from Twitter." He found that the job market did not

have many offers for him with an undergraduate degree, so he decided to pursue a master's degree in Statistics. Overall, he said that "the big thing my education was, it gave me a nice toolbox of techniques and theories to really prime myself off of when I actually had to use real world data. I think that, that was a big thing."

David applied for jobs in different fields and with different types of position, had several interviews, and right after graduation, he received a job offer and "got to live my dream of becoming a baseball person." He worked for a professional baseball team as a statistician to perform analyses, define metrics, and find "better strategies so that you can perform better on the field." After a couple of years on this job, he worked as a consultant in marketing for about three years, and then he looked for a new position with the opportunity to grow. He found an opportunity to work in administration at a university and developed a team of statisticians. He had been there for five years at the time of the interview and had hired three statisticians.

When looking to hire, David struggled to find a good candidate. It took him three months to hire two statisticians and six or seven months to hire a third one. The reason he struggled was partially because "a lot of people that I talked to were more interested in showing off coding and the techniques they're using as opposed to actually be an analyst." He described their transition as "a mixed bag" he provided support by asking them a lot of questions and encouraged them to attend conferences and get experience in consulting.

While working at the university, David completed an Master of Business Administration in Management and Strategy which helped him use the "language of talking statistics in a more business sense and really kind of drive decisions that way." He

recommended to include longer term projects in academic settings, with emphasis on interpretation and decision rather than techniques.

All participants had unique experiences although the goal of this study was to look for similarities and differences in the elements that sustained their transition. These elements are identified and described in detail in the next chapter.

Data Collection

There were two phases for data collection. First, participants recruited at conferences completed two sources of data, a sorting task and a survey, described in the following section. These two data sources were also used to check for eligibility criteria to select participants for the second phase of data collection with interviews. A tailored interview protocol was developed for each participant based on previous knowledge about the participant gathered in the first phase. The interview allowed for rich and in-depth descriptions of the experiences of the transition.

First Phase of Data Collection: Sorting Task and Survey

I lead workshops at conferences worldwide and engaged the members of the statistical community in a sorting task and in answering a survey. They completed these instruments on paper or online.

Sorting Task

The sorting task was derived from Q-methodology. The purpose of the task was to prompt participants to share their perspectives on what important practices are crucial to the role of statistician. Q-methodology aims at investigating inner perspectives and the possible differences and similarities of perspectives within a community (Brown, 1996). In a study using Q-methodology, each participant is asked to sort a sample of statements

from their own perspective. The statements relate to the topic to be investigated and even though there is no rule of thumb for the minimum or maximum number of statements to be used, the sample should contain enough statements to cover all aspects of the topic (McKeown & Thomas, 2013). I created a list of 24 practices based on a review of previous research as mentioned in the literature review and refined with my personal experience as well as a review of information from career guidance companies and job postings (see Appendix B).

Members of the statistical community met at conferences shared their inner perspectives on what important statistical practices constitute the role of statistician. To do so, they sorted the 24 statistical practices in order of importance from their own experience. They attributed a score of importance between 1 as least important and 9 as most important, following a specific distribution represented in Figure 3. In addition to the given practices, participants were also able to add up to six practices to complement the list. If participants did not have any practice to add, they were directed to leave the middle column blank, considered neutral. Traditional Q-methodology does not allow participants to add statements but giving participants this opportunity enabled them to share what practices were truly valuable to them and added to the literature. I will discuss briefly in the next section how having additional practices did not affect the analysis.

Data was first collected on paper and later I had a website made to collect data digitally. For the first three conferences, data was collected on paper, the fourth included both paper and digital collection, and the last three conferences only had the digital option.

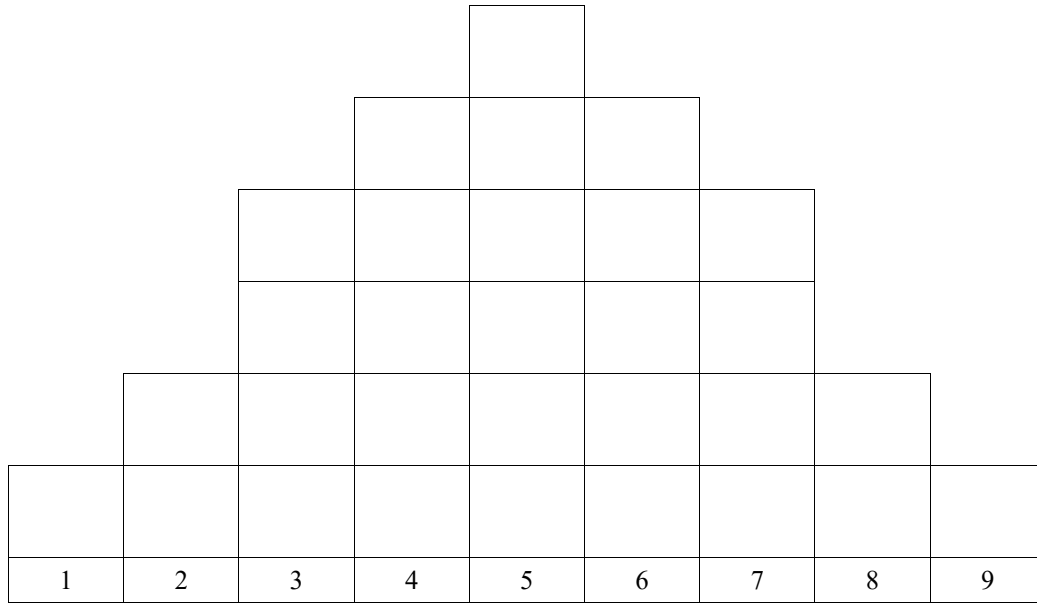


Figure 3. Diagram for the sorting task.

Paper collection. To conduct the sorting task on paper, the 24 statistical practices were marked on slips of paper, together with a unique code that identify each practice. Six blank slips of paper were also provided to add practices. Participants were asked to arrange the list of statistical practices marked on slips of paper into a diagram of nine columns with a specific distribution (see Figure 3). Each column was assigned a value of importance for the practices present in that column, from 1 (least important) to 9 (most important). If participants did not have additional practices to write on the blank slips of paper, they would place the blank slips of paper in the middle of the diagram, which is considered neutral.

The data for the sorting task on paper was recorded on a handout representing the grid on Figure 3 and participants reported the codes for each practice present in each column (see Appendix B and Figure 4). I also added a space for participants to elaborate

on the additional practices they considered important. The back of the handout asked the participants to self-identify to a list of roles, selecting all that applied (see Appendix C).

Task Handout

Complete the following grid with the codes for the corresponding practices:

			LL27			HD64		
		HS21	TL97			FD11	GT37	
		LH16	FH22			YN31	IO44	
	UA40	RT71	EH28			TX27	XC59	IK84
HI57	GI15	OV47	AL13			SK97	WD86	WA74
1	2	3	4	5	6	7	8	9

Please select your role(s) and circle your main role:

☐ Undergraduate student (Major: _____)

☐ Graduate student (Degree in: _____)

☐ Lecturer (Department: _____)

☐ Professor (Department: _____)

☐ Researcher (Field: _____)

☒ Statistician (Field: data science ; engineering company)

☐ Other, please specify: _____

Task Handout

Complete the following grid with the codes for the corresponding practices:

				GJ37				
			YN31	XC59	IR84			
		B3	B5	AL13	GI15	HD64		
		B6	UA40	WA74	SK97	WD86		
		CH16	OV47	HI57	BH28	FH22	PM34	
			IO44	IT521	TL97	FD11	LL27	RT71
1	2	3	4	5	6	7	8	9

Please select your role(s) and circle your main role:

☐ Undergraduate student (Major: _____)

☐ Graduate student (Degree in: _____)

☐ Lecturer (Department: _____)

☐ Professor (Department: _____)

☐ Researcher (Field: _____)

☐ Statistician (Field: Public Health)

☐ Other, please specify: _____

incorrect
 • misplaced 3 blanks
 • incomplete
 missing TH27

Figure 4. Examples of tasks.

Note. The upper picture shows a correct sorting task. The lower picture shows an incorrect sorting task with misplacements of the blanks and missing statements.

Digital collection. I decided to create a website to collect data to control for incorrect sorting tasks. Doubling statements, missing statements, or not leaving the

middle column blank (see Figure 4) resulted in 32 incorrect sorting tasks which were unusable for the analysis and produced an error rate of nearly 44%. To conduct the sorting task online, I created a website: <http://statistician.intelligentedge.com/>. The 24 statistical practices were represented as blocks on the webpage and could be moved with the mouse and placed on the diagram (see Figure 5). Participants could type a new practice which created a new block to be placed on the diagram. They could not add up more than six practices. Practices in the same column were attributed the same score of importance, ranging from 1 to 9. If there was no additional practice, the middle of the diagram should remain empty. A process automatically checked to validate the answers of participants before they submitted their answers.

statistician.intelligentedge.com/step1/

SORTING TASK

From your perspective, what important practices are statisticians performing at the workplace?

Instructions

- Read the list of 24 statistical practices.
- Think about up to 6 additional practices that you believe are part of the role of statisticians and add them to the list.
- Arrange the practices in order of importance, from the least important (1) to the most important (9), into the shape. If you have no additional practice, leave the middle column (5) blank, considered neutral.

Note : Practices are assigned a value of importance depending on which column they are placed. The vertical order has no meaning.

1	2	3	4	5	6	7	8	9	
Least Important				Neutral				Most Important	

Drag these practices the grid above

- Preparing sampling frames / Drawing samples
- Cleaning data / Managing missing data
- Communicating in writing / Writing reports
- Being curious / willing to learn
- Communicating interpretations of statistics to non-statistical audiences
- Consulting / Working with a client to solve a problem

You May Enter Up To 6 New Practices

Click The Button Below After Completing The Grid

Figure 5. Webpage for data collection.

In total, I collected 152 sorting tasks, including 73 on paper and 79 online. Out of the 73 sorting tasks on paper, 32 were incorrect. As part of the presentation and if time

allowed, participants were encouraged to discuss the differences and similarities of the right-end tail, representing the most important practices. In particular, statisticians were urged to identify how and discuss why the right-end tail for all the participants of the workshop session might be different. Then I invited them to reflect on how well their education had prepared them to perform these most important practices which lead to the next source of data collection.

Survey

After engaging in the sorting task, statisticians were offered to answer a survey. The survey consisted of mainly open-ended questions to collect information about educational background, professional experience, and the transition between academic settings and the workplace (see Appendix D for a list of questions on the survey). Details about their experience helped examine specific practices developed in academic settings and at the workplace. The results from the survey were also used to determine the eligibility of participants for the second phase of data collection.

At first, questionnaires were handed out on paper but after I created a webpage for the sorting task, I had one webpage for the survey as well. In total, I collected 117 questionnaires, including 61 on paper and 56 online. Identification numbers helped match the results of the sorting task to the survey.

Second Phase of Data Collection: Interviews

After collecting the sorting task and responses for the survey, I analyzed the data to identify potential participants for the interviews. I describe in detail the process of the analysis in the next section. I invited 13 participants to take part in interviews and 10 responded positively. I conducted semi-structured, one-on-one interviews with

participants to investigate in further detail their experience as junior statisticians or mentors, and their perspectives on the transition from academic settings to the workplace. I wrote a generic interview protocol for junior statisticians (Appendix E) and for mentors (Appendix F) and I used the sorting task and the responses on the survey to develop a tailored protocol for the interview of each participant. Indeed, previous knowledge acquired about the participant collected before the interview through the sorting task and the survey helped refine the protocol. The protocols were developed so that the questions asked to the participant were aligned with the research questions using the theoretical framework.

The interviews were conducted online via Zoom except for one interview that was conducted on the phone. The duration of an interview varied between 45 and 115 minutes. Each interview was audio recorded and transcribed. I took notes during the interview and summarized my impressions right after the interview to reflect on the conversation and identifying briefly the practices mentioned by the participant.

Data Analysis

In this section, I describe how I analyzed the data to answer the research questions. The data analysis was conducted in four parts, guided by the four learning mechanisms at the boundaries (Akkerman & Bakker, 2011): identification, reflection, coordination, and transformation. First, I analyzed the sorting tasks completed by the members of the statistical community to *identify* important statistical practices from different perspectives on the role of statisticians. Second, the analysis of the survey *reflected* on what statistical practices were developed in academic settings in comparison to the workplace. Third, I *coordinated* the interpretations of statistical practices shared by

participants from each activity system. Finally, I summarized how participants suggested to *transform* statistical practices in academic settings.

First Phase of Data Analysis: Identification

The identification of important practices principally occurred with the analysis of the sorting task and was refined by the analysis of additional practices. The analysis of the survey responses and of the interviews refined the identification of practices as well.

Identification of Important Practices through the Sorting Task

The results of the 120 sorting tasks helped identify perspectives among the statistical community. In Q-methodology, the analysis of the sorting tasks intends to differentiate different perspectives among the participants and therefore participants are correlated and grouped based on their sorting tasks to identify patterns of views, or perspectives (Brown, 1980). The factor analysis performed in Q-methodology considers participants as variables and practices as individuals so it can be considered as reversed compared to a traditional factor analysis. The analysis involved three statistical procedures: correlation, factor analysis, and computation of factor scores.

First, a correlation matrix was built by calculating Pearson's correlation coefficients for each pair of sorting tasks resulting in a 120 by 120 correlation matrix. Second, a factor analysis was performed on the correlation matrix to search for sorting tasks that share similar points of view and were grouped into a perspective. To conduct the analysis, I used an R-package called *qmethod* developed to perform the type of analysis specific to Q-methodology (Zabala, 2014) with principal component analysis (PCA). To determine the number of factors, I originally anticipated three perspectives on the importance of statistical practices: from the perspective of practitioners, educators,

and students of statistics. However, by analyzing the proportion of explained variance given the eigenvalues against the number of factors, it seemed that the two first factors conveyed most of the variation. Indeed, the scree plot (Figure 6) indicates that the proportion of explained variance level off after the second factor.

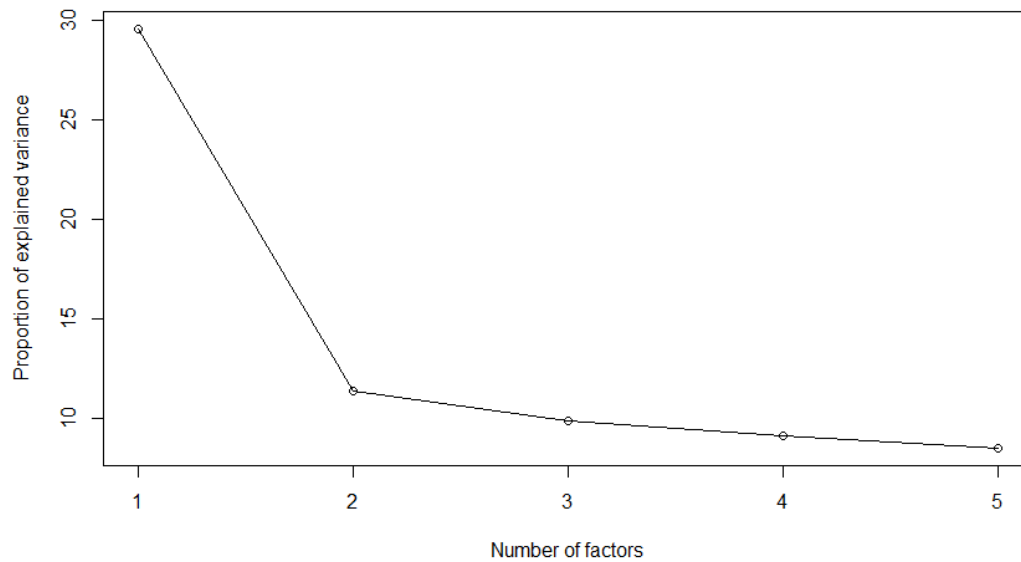


Figure 6. Scree plot for factor analysis.

A varimax rotation was applied to maximize the amount of variance explained by the factors. Such a procedure shows the range of points of view that are favored by the participants in a factor. The participants contributing to the factors were flagged automatically by the program with the condition that their factor loading was greater than $\frac{1.96}{\sqrt{120}}$ and greater than the loading for the other factor.

Finally, to interpret the factors as perspectives on the role of statisticians, I summarized in a sorting task the points of views of participants contributing significantly to that factor and computed the scores of each practice. First, the weights of each participant flagged on the factor were calculated based on their loadings to measure their

contribution. The combination of weights and the scores of each participant on a practice gave a total score for that practice. This process was repeated for each practice and then the mean and standard deviation of the total scores were calculated over all practices. To find the score attributed to a practice, the practices were positioned relatively to each other by calculating the z-score of each practice. The practices are sorted from the lowest z-score to the highest z-score to reconstruct the distribution. I argue that this process allowed for the flexibility to add practices since the scores of the practices in a perspective are calculated relatively to one another and not many additional practices were added.

The output produced by the package also provided information about distinguishing practices. The absolute difference between the z-scores of a practice between the two factors was calculated and compared to a significance threshold for .05 and .01 *p*-value levels. If the difference in z-scores between the two factors was not significant, the practice was qualified as a consensus.

In addition to the traditional analysis of Q-methodology, I looked for agreement for the importance of statistical practices across all participants by representing the deviations from the mean score of a practice using a caterpillar plot. This plot revealed what practices were scored higher on average and what practices displayed the most and the least variation, which I interpreted as disagreement or agreement, respectively.

Identification of Additional Practices on the Sorting Task

Participants could add up to six practices that were not included in the list of 24 statistical practices. In total, 106 practices were added by 41 participants. First, I identified 11 practices that were related to teaching statistics in academic settings which I

eliminated. Second, I considered if practices fell into one of the 24 statistical practices already listed. If there were more than three participants citing a subcategory of existing practices, I created a new practice. For example, calculating sample size can be viewed as part of the existing practice “Preparing sampling frames / Drawing samples” but since three participants specifically referred to calculating sample sizes as an additional practice, I considered it as a new practice, refining the original list of practices. However, if a duplicate was not mentioned by more than three participants, it did not bring new information to qualify as a new practice. In total, I coded 72 additional practices which I grouped in common themes and identified seven new statistical practices.

Second Phase of Data Analysis: Reflection

Among the 117 survey responses collected, only one participant had not completed the sorting task. The goal of the survey was to give participants the opportunity to reflect on how they learned statistical practices as well as generating their own categories of practices. One question on the survey asked participants to share what practices they had learned in academic settings and another question prompted for what practices they had developed at the workplace (Appendix D). Since the questions were open-ended, important quotes from the responses of participants were highlighted for each question. There were 202 quotes for practices learned in academic settings and 130 quotes for practices developed at the workplace. I started with open coding the quotes to look for common themes within each system. After coding practices in academic settings and at the workplace, I compared the categories to look for similarities and differences between the two activity systems and adjusted the names of the categories to match if needed. I also compared the categories with the practices that emerged in the sorting task

to align the category names. For each category, I refined the list of practices with subcategories. For example, I found that most practices in academic settings referred to analysis and subcategories for analysis included the interpretation of models or descriptive statistics, inferences, or modeling. Another analysis subcategory contained specific techniques in advanced mathematics, modeling, inference, or probability theory, such as convex optimization, linear mixed models, randomization tests, or stochastic processes.

All practices fell into one of the same seven categories that emerged in both academic settings and the workplace: Communication, Context knowledge, Data management, Design, Analysis, and Traits. Comparing these seven categories to the five categories of practices I used for presenting the results of the sorting task, namely Design, Data, Analysis, Interpersonal skills, and Personal skills, there are two categories in common and five derived ones. The categories in common are: Design and Analysis. The four additional categories were contained in the other three categories but have become their own. For example, the practice of programming was classified under Data for the sorting task as “Using statistical software.” Because it had many occurrences in the survey, Programming was created as its own category. Similarly, practices related to communication appeared under Interpersonal skills, but many participants mentioned specifically this practice in the survey to represent its own category of Communication. I applied a similar process for context knowledge and data management.

Questions about specific programming tools helped identify what statistical packages were commonly used in academic settings and at the workplace. I determined which programming tools were learned by participants by finding the differences

between the statistical packages mentioned in academic settings and the ones cited at the workplace. Finally, I looked for differences across the disciplines.

The analysis of the survey enabled me to prepare individual protocols for the interviews and anticipate what practices each interviewee considered as important. However, I did not solely focus on these practices in the interview and allowed for new practices to emerge.

Third Phase of Data Analysis: Coordination

The coordination of statistical practices was the third phase of the data analysis and focused on the interviews of junior statisticians and mentors. To coordinate the interpretations of participants for each practice, I did a cross-case analysis across participants with the lens of boundary crossing between the two activity systems of academic settings and the workplace.

I transcribed the audio recordings for the interviews and the analysis process included multiple steps as advocated by Merriam (2009). To code the transcripts, I used Quirkos, a qualitative data analysis software. With a first read of the transcript and field notes for each interview, I identified statistical practices, creating a category or *quirk*, and highlighted quotes for each practice. I also took notes on how I identified each practice in the transcript, and why I highlighted quotes. After I read all the transcripts and identified practices, I recognized differences and commonalities between the meanings of the practices across participants and identified main categories of statistical practices, arranging the quirks accordingly. For example, one of the main categories was related to programming practices. Third, I read the transcripts again to refine with evidence from the data for each main category of statistical practices. Next, within each statistical

practice, I created a quirk for each element in the framework: *tools*, *rules*, *division of labor*, and *community*, for each activity system, *academic settings* and the *workplace*. Therefore, for each statistical practice there were eight quirks representing each element of the two activity systems. I coded the quotes for a statistical practice with these eight quirks to identify what elements of which activity system were mentioned by participants. For example, in the quotes that mentioned programming, I identified if the quote was related to tools in academic settings for learning programming and would include the quote in the quirk of tools in academic settings. This step of the analysis resulted in a 4-by-2 table with quotes for each element of both activity systems. After reading quotes in each cell of the table representing each element of an activity system, I looked for differences and commonalities across participants. For example, in the cells for tools for programming in academic settings I identified that some participants took computer science courses that fostered the logic of programming. Finally, I compared the differences and commonalities across participants in an element between the two activity systems to identify boundaries in the tools, rules, division of labor, and community between academic settings and the workplace and the level of boundary crossing. I used the levels described by Grosemans et al. (2017) to categorize the transition as continuous, detailing, or discontinuous. For example, for the practice of programming, participants found that diverse programs in academic settings facilitated the transition to learn new tools at the workplace. By analyzing the boundaries between academic settings and the workplace along with the differences between the experiences of participants within each element, I was able to classify the transition of the practice for each participant as continuous, detailing, or discontinuous. For example, for participants who had computer

science courses in academic settings the transition was considered continuous compared to discontinuous for participants who did not have these types of courses. The example in this paragraph was simplified to illustrate the analysis of the interviews and the transition of the practice of programming will be described thoroughly in the next chapter. Each practice is illustrated with concrete examples provided by junior statisticians and their mentors.

As suggested by Merriam (2009), the analysis of the interviews started during data collection, right after the first interview had been conducted. There was no adjustment that needed to be made to the general interview protocol.

Fourth Phase of Data Analysis: Transformation

Members of the statistical community were solicited to provide recommendations to transform statistical practices in academic settings through the survey and during the interviews. The triangulation of statistical practices across the three data sources summarized the identification, reflection, and coordination of practices, and resulted in suggestions for transformation.

Survey

At the end of the survey and after reflecting on what statistical practices they had learned in academic settings and at the workplace, participants were asked to reflect on how well their education had prepared them for their profession and what they would like to change in academic settings. As a result, they suggested transformations of statistical practices in academic settings. I first identified what statistical practices they recommended to transform in academic settings. Second, I open coded their answers and found categories for the ways to support the transformation. For example, one participant

recommended to “learn many software” which referred to programming practices and including different programming tools in academic settings.

Interview

Through the interviews, junior statisticians were asked what they wished they had learned in academic settings and mentors reported what they wished junior statisticians knew before entering the workplace. Participants not only mentioned which statistical practices should be transformed, they also suggested how to transform them. I identified the different elements such as tools, rules, division of labor, or community that should be promoted in academic settings as described by participants. Some participants shared examples and ideas to build curriculum materials such as case studies or projects.

Triangulation

After identifying important practices with the sorting task, I refined the identification of practices with the analysis of the survey and of the interviews. With the triangulation of the three data sources, I highlight the connections between categories of practices that emerged with the analysis of each instrument, retaining especially the eight categories that emerged from the interviews. The triangulation is conducted for each category of statistical practices to compare interpretations across the three instruments. I also triangulate the reflection and coordination of practices across the survey and interviews. Finally, I summarized recommendations from members of the statistical community expressed through the survey and during interviews to formulate transformations to ease the transition of practices between academic settings and the workplace.

Researcher's Role

Through my role as a researcher, I was the primary instrument for data collection and analysis (Merriam, 2009). Therefore, it is important to recognize my personal assumptions, biases, and values that may influence my interpretations. After I graduated with a master's degree in statistics, I experienced myself the transition from academic settings to the workplace. My degree emphasized on applied statistics and offered many opportunities to develop practices that were valued when I entered the workplace as a statistician. This degree program inspired me to pursue this research study but also framed my own views about statistics education. In addition, as a teacher of statistics, I have my own opinions about what should and what should not be included in the curriculum. I shared my background and experiences as well as my research goals to the participants throughout the research process.

The triangulation of the different sources of data increased the validity of the findings by providing enough evidence to support the analysis. I also included extensive and relevant quotes from participants to substantiate my findings (Merriam, 2009).

Institution Review Board

Prior to beginning the study, I received approval from the Institution Review Board (IRB) at Texas State University to conduct the study. Participants of the first phase of data collection were given an informed consent on paper or consented online for data collection, analysis, and publications. Participants of the second phase of data collection with the interviews were given a consent form online for audio recording. Pseudonyms were used in the findings to ensure that participants stayed anonymous and any references to specific location or people were deleted.

Limitations

The nature of qualitative studies may appear as a limitation for the generalization of the findings however, extensive detail and rich descriptions captured the experiences of statisticians as accurately as possible. Even though the representativeness of the sample was maximized by including participants with diverse educational backgrounds and fields of application, the role of a statistician is so broad that all practitioners who identify as a statistician might not be represented. Moreover, participants were recruited at conferences which indicated that they were willing to learn and share their knowledge about statistics and in such a way they might differ from the general population of statisticians.

Other limitations specifically related to participants for the interviews included the geographical location of statisticians who were all working in the United States. However, participants were located in the four regions of the US: the Northeast, the Midwest, the South, and the West. Junior statisticians who took part in interviews all had the same educational level with master's degrees, but they obtained degrees in a variety of fields, both at the undergraduate and graduate level.

Summary

Qualitative methodology with a phenomenological approach and a quantitative component was used to investigate the transition of statistical practices between academic settings and the workplace in detail. A large sample of convenience of the members of the statistical community gave access to select a purposeful sample of participants for the interviews. Participants for the interviews were selected to represent the diversity of the role of junior statisticians and the mentors accompanying them in the transition. The

approach for this study best reflected the perspectives of the members of the statistical community, including from junior statisticians and their mentors. A cross-case analysis allowed for a rich description of the phenomenon of boundary crossing and revealed the elements involved in learning statistical practices in the transition between academic settings and the workplace.

IV. RESULTS

The purpose of this study was to explore the experiences of statisticians concerning the transition between academic settings and the workplace. First, I engaged members of the statistical community to reflect on the role of statisticians at the workplace by identifying important statistical practices. Second, participants shared their educational background, professional experience and reflected on their transition to the workplace by answering open-ended questions on a survey. They identified practices they developed in academic settings and practices they developed at the workplace. Based on their answers and to represent the diversity of the role of statisticians, a selection of participants were invited to take part in interviews to better understand how they managed the transition to the workplace and to reflect on how we can improve the transition from academic settings. This chapter discusses the results of the three sources of data for this study and answers the following research questions:

RQ1. What important statistical practices do members of the statistical community identify for the role of statisticians?

RQ2. How do statistical practices transition from academic settings to the workplace?

RQ3. In what ways should statistical practices be implemented in academic settings as recommended by members of the statistical community?

Results are presented in three sections, each section addressing a research question and following the learning mechanisms suggested by Akkerman and Bakker (2011): identification, reflection, coordination, and transformation. The first research questions *identified* the importance of statistical practices relatively to one another for the

role of statisticians. The second research question focused on the transition of these practices from academic settings to the workplace, involving a *reflection* and *coordination* between the two systems. The third research question investigated the transition from the workplace to academic settings, by *transforming* statistical practices. In the first section, I present different perspectives on the role of statisticians that emerged from the members of the statistical community through the sorting task and survey. I report important statistical practices identified by the participants who also reflected on practices being mostly developed in academic settings or at the workplace. Second, I incorporate the concepts of continuity, detailing, and discontinuity (Abrandt Dahlgren et al., 2006) to describe the transition of practices across participants in lines with the framework of boundary crossing between the systems of academic settings and the workplace (Akkerman & Bakker, 2011; Star & Griesemer, 1989). With cross-case analysis across participants to explore trends and patterns for statistical practices, I compare differences and similarities, and identify salient characteristics between the experiences of participants and the development of the statistical practices. In the last section, I build statistical practices as boundary objects to facilitate boundary crossing with the triangulation of the three sources of data (sorting tasks, survey, and interviews) and summarize recommendations given by members of the statistical community to transform statistical practices in academic settings.

Perspectives from Members of the Statistical Community

To learn about the important practices defined by members of the statistical community, I met statisticians, educators and students of statistics at conferences. They completed a sorting task and a survey which I described in the previous chapter. The

analysis of the sorting tasks and of the questionnaires revealed the identification of important practices for the role of statistician and a first indication of the coordination of these practices between academic settings and the workplace.

Identification of Important Statistical Practices

I engaged members of the statistical community to identify important practices for the role of statisticians with the sorting task. Participants sorted a list of 24 statistical practices that was based on a review of the literature and refined with an overview of qualifications required in job offers and my own experience. To ease the interpretation of the results, I divided the practices into five categories. Practices focused on either *Design*, *Data*, *Analysis*, *Interpersonal skills* and *Personal skills* (see Table 7). Participants sorted these 24 practices in order of importance: from least important, attributing a score of 1, to most important, attributing a score of 9. Participants could write up to six additional practices that were not included in the list, if they noticed that a practice that was important according to their own experience was missing. The rankings were recorded as a sorting task for each participant, assigning a score to each practice.

Overall, I collected 120 sorting tasks, 41 on paper and 79 online. As mentioned in the previous chapter in the section describing participants, participants represented the diversity of the roles in the statistical community with statisticians, educators, graduate and undergraduate students as well as the distinction with statisticians in academia. Participants had various educational backgrounds with bachelor's, master's and doctoral degrees obtained in different disciplines. Furthermore, statisticians worked within a variety of fields of application including healthcare, policy, or economics. To identify important statistical practices, I discerned different perspectives across participants.

Table 7*List of 24 Practices in Five Categories*

Category of practices	Statistical practices
Design	Translating a real problem into a statistical form Designing studies / Aligning design with research goals Preparing sampling frames / Drawing samples Using knowledge of the context
Data	Collecting / Gathering data Creating / Maintaining databases Cleaning data / Managing missing data Using statistical software package / Writing computer programs Producing visual representations of data Interpreting data (Limitations of methods / Bias)
Analysis	Researching appropriate statistical methods and techniques Developing new statistical methods and techniques Applying statistical methods and techniques Using advanced mathematics
Interpersonal skills	Participating in teams / Collaborating Communicating in writing / Writing reports Communicating orally / Making presentations Consulting / Working with a client to solve a problem Communicating interpretations of statistics to non-statistical audiences
Personal skills	Being curious / willing to learn Being skeptical / critical Meeting deadlines Considering ethical issues Working independently

Perspectives

The sorting task was derived from Q-methodology with the possibility for participants to add their own practices. The process of analysis was explained in the previous chapter with finding correlations between participants in terms of their sorting tasks and performing a by-participant factor analysis to reduce individual opinions summarized into perspectives. The analysis of the sorting tasks revealed two main factors, or perspectives, with 100 out of the 120 participants contributing significantly to either factor (see Table 8).

Table 8*Comparison of the Perspectives*

Characteristics	Perspective 1	Perspective 2
Number of loading sorting tasks		
Positive	60	37
Negative	2.0	1.0
Eigenvalue	28.7	17.3
Percentage of explained variance	24.0	14.4
Mean score per category of practices		
Design	5.0	5.3
Data	4.7	4.8
Analysis	3.3	6.3
Interpersonal skills	7.6	5.4
Personal skills	4.8	3.6

The first perspective, Perspective 1, summarized the individual points of view of 62 participants, 60 of them contributing positively to the perspective meaning that they share the same subjectivity. Two participants weighted negatively, rejecting the ranking of the practices expressed by Perspective 1. The amount of variance explained by the first factor was 24% of the total variance between the 120 sorting tasks while the second factor explained slightly over 14% of the total variance. The second perspective, Perspective 2, summarized the opinions of 38 participants, of which 37 contributed positively and one contributed negatively. Both perspectives explained 38% of the total variance in the scores of statistical practices.

The interpretation of each factor was based on the reconstruction of the scores for each practice. To ease the interpretation of each perspective, the five categories are illustrated by different shades of grey and practices related to interpersonal skills and personal skills are represented with a white font (Figure 7, 8, and 9). The mean score of practices was calculated for each category and reported in Table 8 for each perspective.

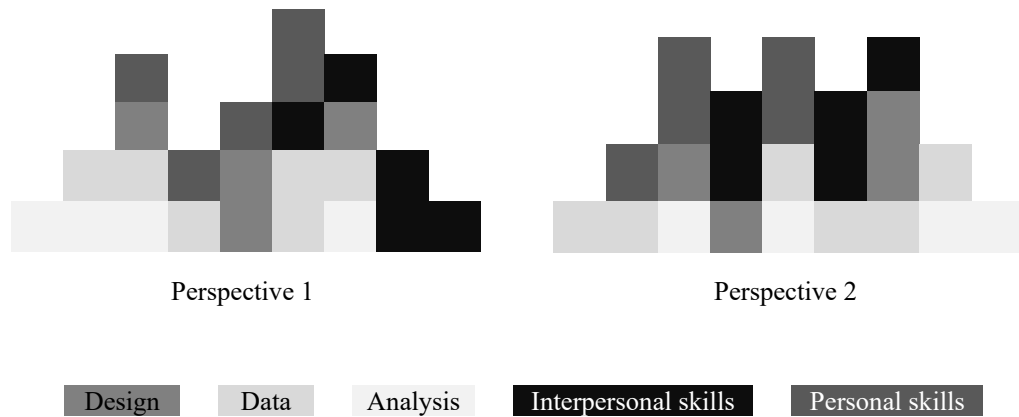


Figure 7. Comparison of categories of statistical practices between perspectives.

The dark blocks on the right of the diagram for Perspective 1 indicate that participants who contributed positively to this perspective focused on interpersonal skills with a mean score of 7.6, higher than other categories of practices. In comparison, interpersonal skills were found at the center for Perspective 2 averaging to a score of 5.4, which suggests that participants who contributed significantly to the second perspective had mixed views about these practices. The light blocks on the left of the diagram for Perspective 1 showed that participants considered practices related to analysis as less important attributing the least mean score. On the contrary, analysis practices were placed on the right by participants contributing positively to Perspective 2, receiving the highest mean score of 6.3 and therefore considering these practices as important. Perspective 2 also differed from Perspective 1 by attributing lower scores to personal skills. Both perspectives ranked practices related to design and data similarly. There was no significant difference between the two perspectives in terms of the roles of participants, their educational background, or the fields of application.

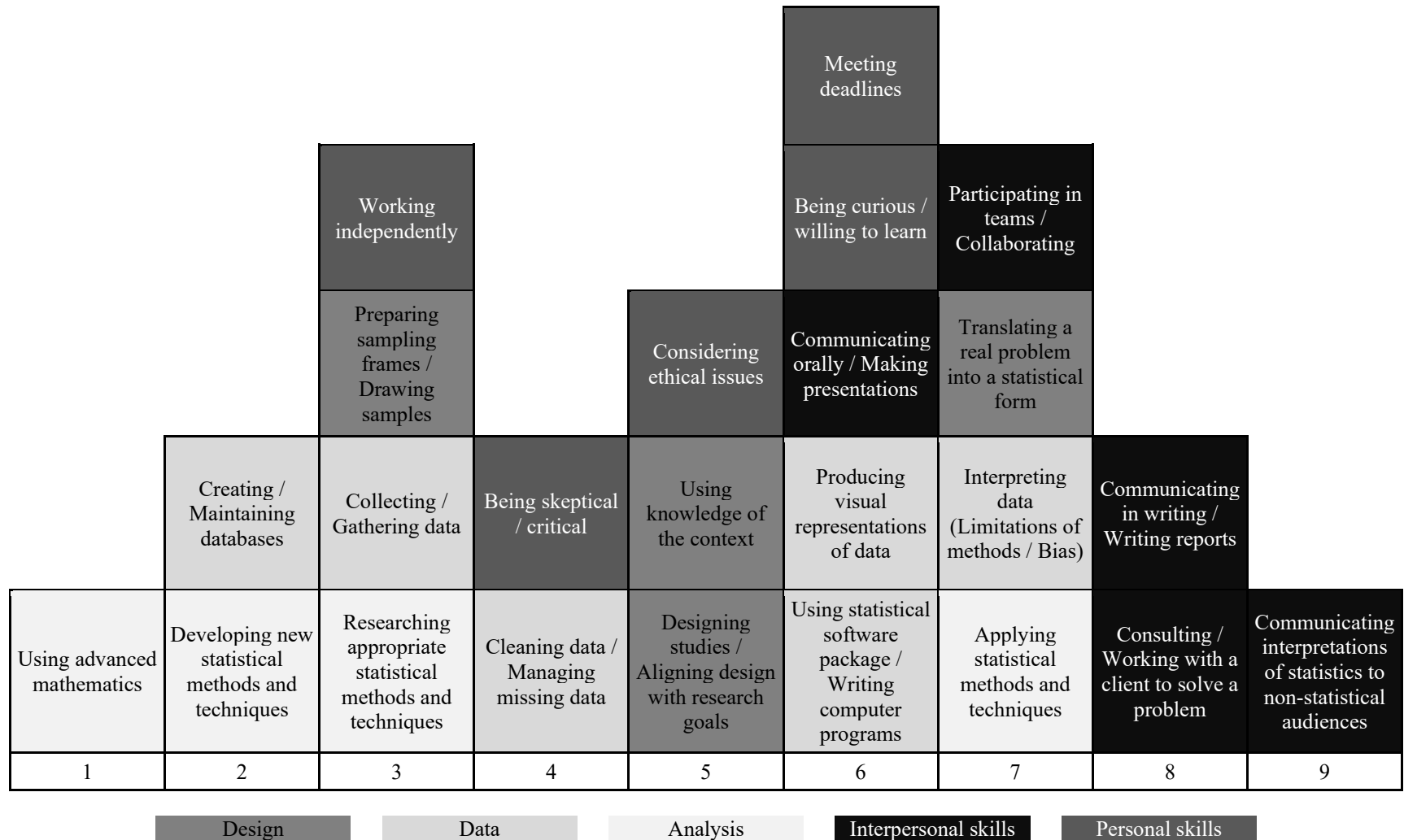


Figure 8. Sorting task summarizing Perspective 1.

		Meeting deadlines		Being skeptical / critical		Communicating interpretations of statistics to non-statistical audiences		
		Considering ethical issues	Communicating in writing / Writing reports	Being curious / willing to learn	Participating in teams / Collaborating	Translating a real problem into a statistical form		
	Working independently	Preparing sampling frames / Drawing samples	Consulting / Working with a client to solve a problem	Cleaning data / Managing missing data	Communicating orally / Making presentations	Designing studies / Aligning design with research goals	Interpreting data (Limitations of methods / Bias)	
Creating / Maintaining databases	Collecting / Gathering data	Using advanced mathematics	Using knowledge of the context	Developing new statistical methods and techniques	Producing visual representations of data	Using statistical software package / Writing computer programs	Researching appropriate statistical methods and techniques	Applying statistical methods and techniques
1	2	3	4	5	6	7	8	9
Design		Data		Analysis		Interpersonal skills	Personal skills	

Figure 9. Sorting task summarizing Perspective 2.

By taking a closer look at the arrangement of specific practices on the diagram in Figure 8, I noticed that Perspective 1 emphasized the importance of interpersonal skills with the practice of “Communicating interpretations of statistics to non-statisticians” ranked as the most important with a score of 9. Next, the practices of “Consulting/Working with a client to solve a problem” and “Communicating in writing” had a score of 8, highlighting the importance of interpersonal skills as well. On the opposite, practices focusing on analysis such as “Using advanced mathematics,” “Researching [...]” and “Developing new statistical methods and techniques” were ranked as less important, respectively scored 1, 2, and 3. However, the practice of “Applying statistical methods and techniques” was ranked as one of the most important practices with a score of 7.

Perspective 2 summarized the points of view of participants who mainly focused on practices of analysis at the right of the diagram in Figure 9. with the practice of “Applying statistical methods and techniques” being scored as 9 and “Researching appropriate statistical methods and techniques” scored as 8. The less important practices were part of the category about data with “Creating / Maintaining databases” and “Collecting / Gathering data,” respectively scored 1 and 2. Practices related to personal skills such as “Working independently,” “Meeting deadlines,” and “Considering ethical issues” were also identified as less important, with scores of 2 and 3.

Distinguishing and Consensus Practices

The perspectives can also be compared in terms of distinguishing and consensus practices between the two groups of participants (see Table 9).

Table 9*Distinguishing and Consensus Practices*

	Difference in z-scores	Practice	Category	Score on P1	Score on P2
Distinguishing	-1.9716	Researching appropriate statistical methods and techniques	Analysis	3	8
Distinguishing	-1.7952	Developing new statistical methods and techniques	Analysis	2	5
Distinguishing	-1.4964	Using advanced mathematics (calculus, algebra, differential equations...)	Analysis	1	3
Distinguishing	+1.4009	Communicating in writing / Writing reports	Interpersonal traits	8	4
Distinguishing	+1.4052	Meeting deadlines	Personal traits	6	3
Distinguishing	+2.0173	Consulting / Working with a client to solve a problem	Interpersonal traits	8	4
Consensus	+0.0054	Being skeptical / critical	Personal traits	4	5
Consensus	-0.0545	Working independently	Personal traits	3	2
Consensus	-0.1165	Creating / Maintaining databases	Data	2	1
Consensus	-0.1189	Producing visual representations of data	Data	6	6
Consensus	-0.1198	Preparing sampling frames / Drawing samples	Design	3	3

Note. P1 = Perspective 1; P2 = Perspective 2.

The difference between the z-scores of a practice on each factor is calculated and compared to a significance threshold to determine which practices are distinguishing or showing consensus between the two perspectives. Practices related to theoretical background such as “Researching [...]” and “Developing new statistical methods and techniques” and “Using advanced mathematics [...]” were ranked significantly lower on Perspective 1 than Perspective 2. Participants who contributed to Perspective 1 ranked practices related to the traits of a statistician with “Consulting / Working with a client to solve a problem,” “Meeting deadlines” and “Communicating in writing / Writing reports”

significantly higher than participants who contributed to Perspective 2 with a difference of at least three points between the two scores.

Both perspectives showed consensus of the importance of “Producing visual representations of data,” scored at 6. The personal skill of “Being skeptical /critical” was considered relatively important with scores of 4 or 5. The practice “Preparing sampling frames / Drawing samples” for the design of a study was not considered as important with a score of 3 for both perspectives. Next, the practice of “Creating / Maintaining databases” was considered as one the least important practices for the role of statistician, with scores of 1 and 2. Finally, “Working independently” was scored 2 and 3, therefore considered not so important and contrasting with the practices that involved working with others which were considered more important especially for Perspective 1.

Considering the sorting task from all 120 participants, the caterpillar plot in Figure 10 shows the mean score attributed to each practice and compared the variation for the scoring of each practice by representing an interval for the score within one standard deviation from the mean. The most important practices in terms of their mean scores were “Communicating interpretations of statistics to non-statistical audiences,” “Applying statistical methods and techniques” and “Interpreting data.” These practices also had less variation than most practices meaning that not only they were the most important on average, but participants tended to agree that these practices were important. Practices related to more theoretical practices such as “Developing new statistical methods and techniques” and “Using advanced mathematics” were the least important in terms of the mean but also displayed some variation. The practices of “Collecting / Gathering data” and “Cleaning data / Managing missing data” indicated disagreement

among the statistical community with the largest standard deviations. In addition, the practice of “Preparing sampling frames” was on average in the lowest half of the scores and therefore considered as less important but demonstrated agreement among all participants with the lowest standard deviation.

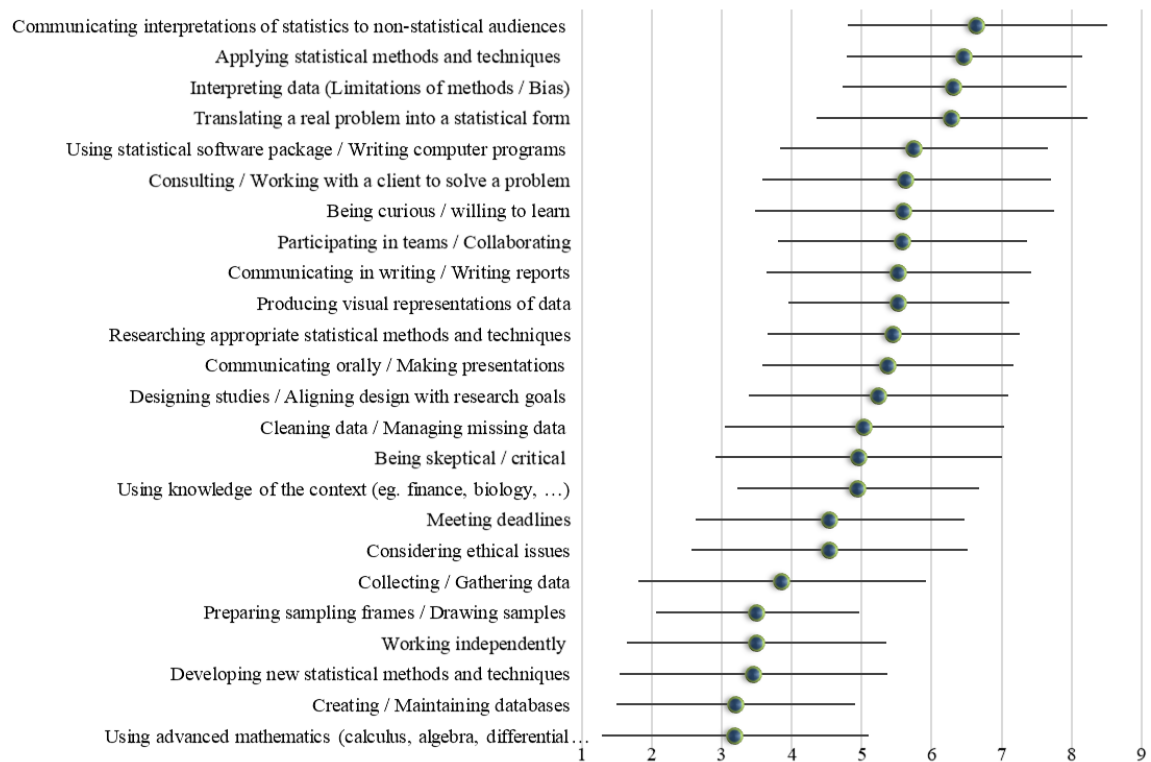


Figure 10. Mean ranking of practices and variation.

Note. The dots represent the mean score of a practice and the bars represent the interval of scores within one standard deviation from the mean.

Additional Practices

As part of the sorting task, participants could provide up to six additional practices that they believed were missing from the list of 24 practices and included them in the sorting task. A total of 41 participants provided 106 new practices which I first compared to existing practices. I identified 72 practices from 33 participants that did not fall directly into any practice on the list and had at least three occurrences to create their

own category. To summarize the 72 practices, I identified common themes and coded them into seven statistical practices (see Table 10).

Table 10

Additional Practices

Practices	Occurrences	Mean score
Training constantly	19	4.5
Training others / Mentoring	19	4.6
Managing projects / budget / people	14	5
Establishing a relationship with a client	9	5.3
Accessing and storing data	5	6
Documenting work	3	4.3
Using power analysis for sample sizes	3	5.7

First, 19 participants pointed out that they had to keep learning and seek training constantly. For example, they mentioned “attending conferences, workshops, webinars, to stay relevant in the field.” They also valued “reading peer reviewed journal articles to advance knowledge” and contributing to the statistical community. The topics they needed to learn were tools such as “new programming languages,” also “learning new methods” and “keeping up to date about latest developments in statistical methodology.”

Besides training themselves, 19 participants shared that they needed to train others by providing statistical support and “mentoring junior statisticians.” While “acting as a leader on a team,” some of their responsibilities included “developing and recommending organization-wide strategies” as well as “advising and consulting with other departments in the company.”

Participants mentioned being involved across teams and 14 of them also reported managing projects, budget, or people with “other statisticians and data professionals.” They considered “applying for grants,” “structuring effective meetings” and “tracking time spent on projects” being part of their role.

I believed the existing practice of “Consulting / Working with a client to solve a problem” did not take into account the aspect of establishing a relationship with a client which was specifically designated by participants. Therefore, I created a new practice representing nine occurrences. Some participants cited “addressing statistical anxiety of clients” and “building strong relationships with clients.” For example, one participant noted that they needed to establish “a reputation among non-statisticians in order to obtain their trust.”

Another practice that emerged concerned the responsibility to access and store data. Even though this additional practice could be related to the existing practice of “Creating / Maintaining databases,” there were five participants who specifically referred to data storage and a practice was created to reflect quotes such as “assuring data integrity (from data creation to storage of final, cleaned data),” “evaluating data quality,” or dealing with “data security / privacy.” In addition, this practice tended to be ranked higher than other additional practices with a mean score of 6.

Three participants mentioned documenting their work as an important practice by keeping “organized files, accessible to others.” Indeed, they thought it was the responsibility of statisticians to “write code or conduct statistics with a focus on reproducibility.”

Finally, a practice related to “generating sample sizes for research studies” was suggested by three participants, referring to power analysis in particular. I distinguished this practice from the existing practice of “Preparing sample frames / Drawing samples” not only because it was mentioned by three participants, but more importantly because these three participants ranked the additional practice high with a mean score of 5.7 while the existing practice had a mean score of 3.5 with the lowest standard deviation, demonstrating agreement among participants (see Figure 10). Therefore, it was necessary to distinguish the additional practice of calculating sample sizes.

Summary of identification

The analysis of the sorting task addresses the first research question. Members of the statistical community shared two perspectives on the role of statisticians at the workplace. The perspective summarizing the greatest number of participants identified interpersonal skills as the most important practices while the other perspective focused on more technical skills. Overall, interpersonal skills were still considered important, for example with communicating with non-statisticians being ranked the highest on average compared to other practices, and across all participants. In addition, seven new practices were identified by approximately one fourth of participants. One of the two most cited practices was related to learning, having to train constantly. Even though participants referred to learning as a practice, I will consider learning as embedded in each practice in the rest of this chapter. Indeed, one of the goals of my study was to understand how junior statisticians learn specific practices, which will be described in detail with the analysis of the interviews. However, the fact that many participants indicated to keep learning at the workplace shows that the phenomenon of learning in the transition to the

workplace was worth investigating. Next, participants reflected on their experiences to identify a possible misalignment that created boundaries and therefore, opportunities to learn.

Reflection on Statistical Practices

A total of 117 participants completed the survey, with 61 paper questionnaires and 56 online questionnaires. Open-ended questions engaged participants in a reflection on statistical practices learned in academic settings and at the workplace. The goal was not to get an exhaustive list of practices but to reflect on what were the most important practices learned in each activity system. Participants were asked to share what were the most useful concepts or practices they learned in academic settings and the most important practices they developed at the workplace. The survey questions are included in Appendix A. I highlighted quotes in participants' responses and created a list of 202 practices learned in academic settings and 130 practices developed at the workplace. To identify common themes across the practices generated by participants, I coded these practices in emerging categories that I constructed. The percentages of practices representing each category were calculated within each system (see Figure 11), indicating prevailing practices developed in each system. The categories were the same across the two activity systems but had different interpretations and were cited more or less often. For example, practices of analysis were the most common practices in academic settings while at the workplace participants cited communication practices more often. The other practices that emerged, namely traits, design, data management, domain knowledge, and programming were equivalently learned in each system.

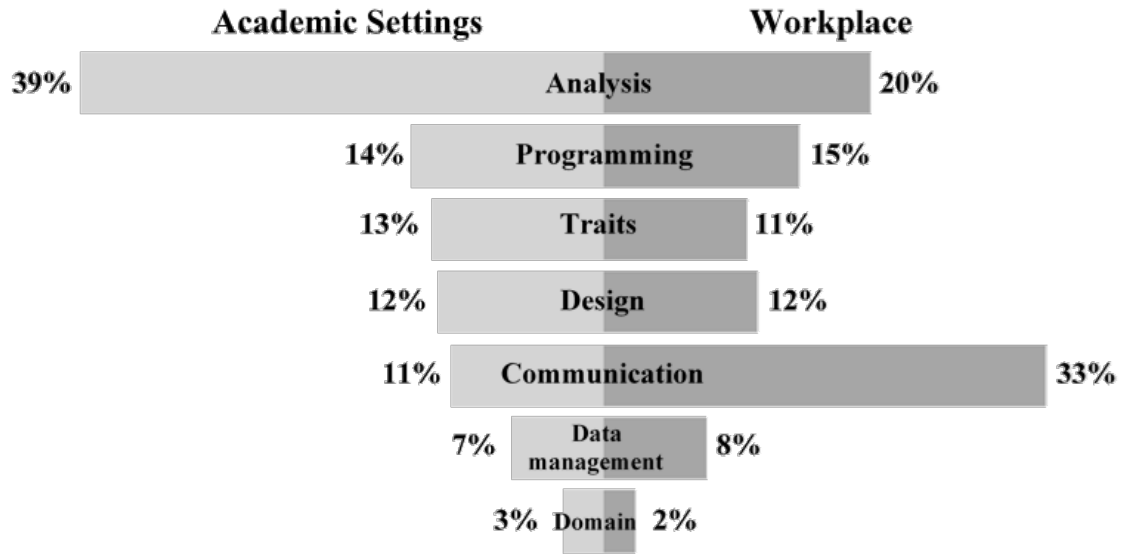


Figure 11. Comparison of practices learned in academic settings versus the workplace.

Analysis

Practices coded as *Analysis* were mentioned the most in academic settings, representing 39% of the practices in that system. Some specific analysis techniques were cited, including *advanced mathematics* topics such as “convex optimization” and topics in *probability theory* such as “random variable” and “stochastic processes.” Some participants also highlighted the importance of *interpretations*, how “to understand the interpretations of coefficients in models and how to take the mathematical form of a model and make sense of it.” At the workplace, practices coded as analysis still represented 20% of the 130 practices, indicating that statisticians needed to keep developing these practices. The interpretations of the practices focused on different *branches of statistics* with “survival analysis,” “time series analytics,” or “Bayesian statistics,” and referred to topics in *data science* with “machine learning” and “unsupervised learning.”

Communication

Practices of *Communication* were the most frequent at the workplace, representing 33% of the practices in that system. Participants reported that they had to learn how to *collaborate* and *communicate with non-statisticians* at the workplace. For example, one participant shared “The most useful practices are those related to communication and collaboration: being able to meet someone new, learn about the person, their studies or research, goals, and the particulars of their current concern.” Other communication practices included *presenting* with “report writing” or “communicating orally” and *translating* “statistics into English” by “learning to talk to people in their language, not the language of a statistician.” Whereas in academic settings, the practices coded as communication were less frequent, accounting for only 11% of the practices. Communication in academic settings focused on *presenting* and particularly by “writing reports.” *Teaching* was also mentioned as a useful practice learned in academic settings, probably as a teaching assistant.

Traits

Participants reported learning *Traits* in academic settings, representing 13% of the practices, compared to 11% at the workplace. In academic settings, practices included “learning how to learn,” “how to use online resources for help,” or “that it is always ok to ask question or search out more resources.” Participants developed *time management* skills with “work organization” to “submit report on time.” They also learned “how to think critically” and doing “independent research.” At the workplace, participants also mentioned that they had to be “constantly learning new techniques” and manage their time as well as teams, projects and workflow. Under traits, I also found *attitudes*, for

which participants added that they had “to be more understanding and patient with others,” learning “diplomacy” and “leading without authority.” One participant also shared that it was very important to “find a mentor within your organization,” highlighting the role of the mentor in the transition to the workplace.

Design

Design was equally represented in both system, accounting for 12% of practices. In academic settings, practices provided “an understanding of experimental design” and “how to create a statistical analysis plan for a collaborative research project.” Participants specified that *choosing the appropriate method* was an important practice learned in academic settings:

Regardless of the statistical tools one has, because it is relatively easy to learn new ones on the fly, it is just as important to know when it is appropriate to apply or not apply a tool to a given situation.

However, this practice was mentioned more often at the workplace, indicating that the practice tended to be developed on the job. In particular, time might influence this practice at the workplace, as explained by:

Academic frameworks are great, but you must keep in mind that sometimes you have to adapt the frameworks to your organization’s needs. You might not be able to do the perfect analysis if time is a factor. It is better to have some data driving decision making than nothing at all.

The practice of *translating* a problem into a statistical form was common to both systems, with “formulating questions” in academic settings versus “helping investigators to formulate research questions” at the workplace.

Data Management

Practices related to *Data management* were mentioned in both systems, representing 7% of the practices in academic settings and 8% at the workplace. They had

similar interpretations across systems with *data cleaning*, dealing with *missing data* and *big data*, *building databases*, or ensuring *data validation*. In academic settings, participants learned “cleaning methods,” “coding to work with messy datasets,” or “how to handle missing data.” One participant explained the steps for “data validation. Using plots, descriptive tables, diagnostic checks, etc., to check the assumptions and quality of data before going forward with analysis.” At the workplace, participants developed skills to be “working with large databases” or “interpreting unclean and missing data.” They also realized “the importance of data quality and how to build infrastructure and/or design surveys to improve it.”

Domain Knowledge

Participants also valued developing *Domain knowledge* in both systems, accounting for 3% of the practices in academic settings and 2% at the workplace. In academic settings, the fields of “epidemiology,” “clinical trial” and “bio” were all related to the medical field while at the workplace participants also mentioned “business knowledge” and “specific domain knowledge” in general.

This practice was included in the list of practices to rank for the sorting task with “using knowledge of the context” and was identified again as a practice by participants in the survey. However, I will consider domain knowledge to be integrated into other practices, such as design or analysis, in the next section with the analysis of the interviews.

Programming

Practices related to *Programming* had a similar share in both systems, accounting for 14% of the practices in academic settings and 15% at the workplace. They mostly

referred to programming in general, using terms such as “coding” or “statistical programming.” Some participants reported specific programming languages or software which differed between the two systems. In academic settings, participants mentioned using R and SAS as the most important programming practices while at the workplace participants mentioned a variety of tools such as SQL, Python, Tableau or GitHub. Indeed, one participant pointed out that in academic settings: “The most useful practice I learned was computer programming. I have to pull my own data and learning SQL was vital to do this. Learning to program in multiple languages makes you more marketable as a statistician.” Aside from a variety of tools, the practice of *documentation* related to programming emerged. Indeed, participants mentioned “data and code documentation” by “creating code and analyses that are reproducible.” One participant explained the importance of this practice:

Being very diligent with organization of datasets and code. While working on many projects concurrently, some of which span multiple years, it is necessary to have a system of labeling and commenting that allows work to be carried out efficiently, accurately, and thoroughly. This organization is also extremely helpful (and even necessary) for regulatory auditing purposes.

Documentation was already mentioned in additional practices through the sorting task. It will be presented as its own category in the next section because documentation did not only refer to programming but was also expressed as a mean to foster collaboration.

Specific questions on the survey investigated the variety of programming tools used in each system. Out of the 117 questionnaires collected, 115 participants reported using statistical programs in academic settings and 60 participants reported what programs they used to perform statistics at the workplace. Participants could select all tools that applied to them from a list of seven tools or add some tools that were not listed.

The total numbers of programs learned per participant are displayed in Figure 12. In academic settings, participants learned between one and seven different programs with a mean of 3.1 programs, while at the workplace they used one to five programs with a mean of 2.7 programs, indicating a diversity of programming in both systems with a tendency to have exposure to more programs in academic settings.

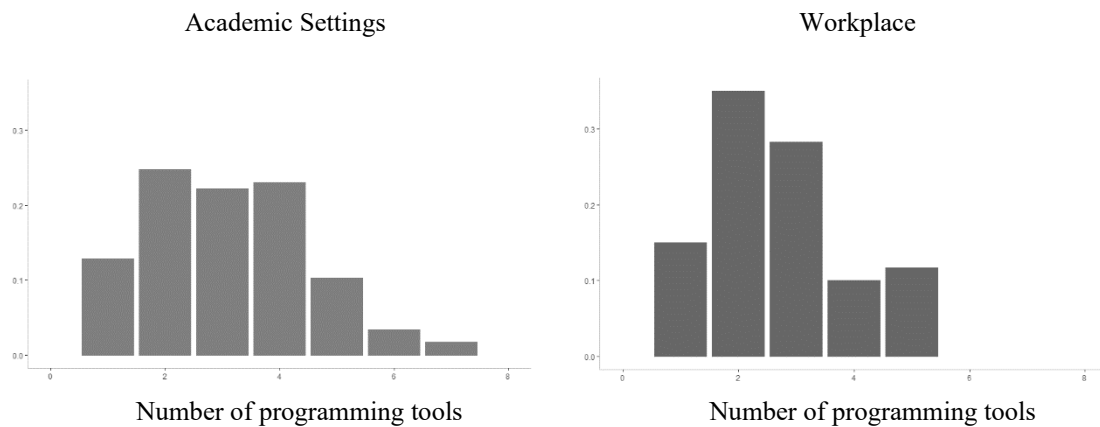


Figure 12. Comparing the number of programming tools used in each system.

The most common programs were the same in both systems with R, SAS, and Excel (see Table 11). R was used by 75% of the participants in academic settings and by 82% at the workplace. The second most used tool was SAS which was less represented at the workplace with 50% of participants versus 68% in academic settings. Excel was almost equally used in both systems, with 44% and 45% of participants reporting using this tool. In academic settings, over 20% of participants had also learned SPSS, Minitab, or Stata. Over 20% of participants also reported using SPSS, Stata, or Python at the workplace. Less participants used Python in academic settings. There were also other programs that were only mentioned once in academic settings: BMDP, Java, Mathematica, Statistica,

SYSTAT, Scilab, and Tensorflow. Added programs at the workplace were: EAST, BUGS, PASS, Minitab, StatCrunch and Tensorflow.

Table 11

Comparison of Programs Between the Systems

Percentage of users in academic settings (n = 115)		Percentage of users at the workplace (n = 60)	
Program	%	Program	%
R	75	R	82
SAS	68	SAS	50
Excel	44	Excel	45
SPSS	33	Stata	33
Minitab	27	SPSS	20
Stata	20	Python	20
JMP	13	JMP	5
Python	12	SQL	3
MATLAB	7	Minitab	3
S+	3	Tableau	3
SQL	2		
C++	2		

To measure which programs statisticians had to learn at the workplace, the differences between having learned a program in academic settings and using the software at the workplace were found for the 60 participants who reported both (see Table 12). The most common tool that had to be learned by participants was Excel however, participants may not have reported it in academic settings because they did not use it for the purpose of statistics but still knew how to use Excel. Over 10% of participants also had to learn Stata, R or Python at the workplace.

Table 12*Percentage of Participants Learning New Programs*

Program	Excel	Stata	R	Python	SPSS	SAS	JMP
%	23	18	12	10	7	5	5

Table 13*Percentage of Technology Users Across Fields of Application*

Field	R	SAS	Excel	STATA	SPSS	Python
Consulting	100	100	50	50	100	0
Economics	100	25	50	50	0	50
Education	67	33	75	25	50	0
Engineering	100	20	40	60	20	60
Healthcare	95	71	24	19	5	10
Math and Stats	64	45	36	64	18	36
Policy	50	50	75	0	0	25

All fields of applications at the workplace required diverse tools. In Table 13, I represented the percentages of the participants who reported using a programming tool in each field of application, highlighting the cells for which the tool was used by at least 75% of the participants in dark grey and between 50 and 75% in light grey. For example, R was used by at least 50% of participants across all fields with 100% of participants using R in the fields of Consulting, Economics, and Engineering. In addition, 95% of the participants in healthcare reported using R. Healthcare professions also reported using SAS, representing 71% of participants. SAS and Excel were used across all disciplines at different levels while Python was mostly used in Engineering and Economics.

Summary of Reflection

The analysis of the survey complemented the findings from the sorting task to identify important statistical practices for the role of statisticians. Findings from the survey identified practices that were classified into seven categories, making programming, communication, data management, and domain knowledge, their own categories. Details about the use of programming tools revealed that members of the statistical community tended to be exposed to a variety of programs in academic settings. In addition, participants reflected on the transition of the statistical practices between academic settings and the workplace. The leading practices at the workplace were related to communication while practices associated with analysis were prevailing in academic settings. The contribution of other categories of practices was balanced out between the two systems. These findings started to address the second research question, which focused on the transition of the statistical practices. Thorough explanations of the transition of each practice are given in the next section with the analysis of interviews. Junior statisticians and their mentors were recruited to reflect on their experiences and describe how they coordinated statistical practices between academic settings and the workplace.

Perspectives from Junior Statisticians and Mentors

After the analysis of the survey responses, I considered members of the statistical community who met the eligibility criteria to qualify as junior statisticians, namely statisticians who had been working at their current workplace for more than three months and who had graduated within the past five years. I also recognized mentors accompanying junior statisticians in the transition from both academic settings and the

workplace. The selection of participants was further discussed in Chapter 3 but to summarize I operated purposeful sampling to maximize the variation in terms of educational background and field of application, representing the diversity of the education and profession of statisticians observed within the members of the statistical community.

Overall, I invited 13 selected participants to take part in an interview and 10 responded positively, therefore interviews were conducted with eight junior statisticians and two mentors. The eight junior statisticians all earned a master's degree but with different majors in applied statistics, in biostatistics, or in public health. They also applied statistics in various fields: in healthcare, policy, retail, or higher education. In addition, I interviewed two mentors. One mentor was a senior statistician for administrative research at a private research university. He hired, supervised, and provided support for a team of junior statisticians. The other mentor functioned both at the workplace and in academic settings. She was a senior statistician in the management track within the healthcare industry and contributed to the recruitment and training of junior statisticians. She was also an adjunct professor for a master's degree program in biostatistics and provided opportunities to learn for future statisticians. More detail about the background of junior statisticians and mentors can be found in the previous chapter under the section describing participants.

Besides checking for eligibility criteria, responses from participants on the sorting task and survey helped me tailor an individual protocol for each participant. Half of the participants interviewed represented Perspective 1 in the sorting task, and the other half represented Perspective 2. They indicated a wide range of practices learned in academic

settings and learned at the workplace in the survey. In order to explain how practices were learned and developed, the focus of the interview was to explore how important statistical practices transitioned between academic settings and the workplace.

During the interview, participants discussed how their educational background had prepared them for the expectations of the workplace in terms of statistical practices. I interpreted the transition between academic settings and the workplace through the lens of boundary crossing (Akkerman & Bakker, 2011). The principle of boundary crossing made participants consider different perspectives on statistical practices, first, as a student in academic settings, and second, as a statistician at the workplace. Academic settings and the workplace were perceived as activity systems where the learning of a practice was mediated by different elements: tools, rules, division of labor, and community (Figure 13).

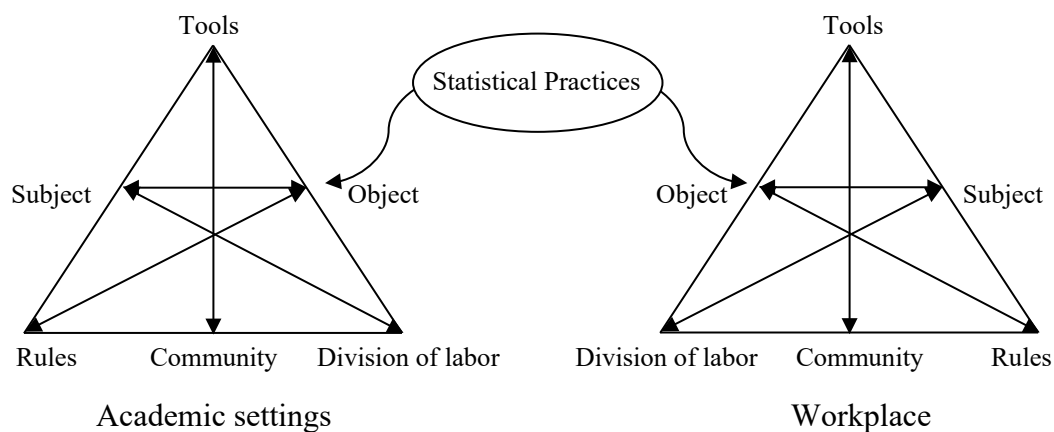


Figure 13. Lens for data analysis: Boundary crossing.

Within an activity system, individuals used tools to learn and perform a statistical practice, following rules. I identified rules constraining learning as conventions or best practices set by the activity system. The division of labor designated who was in charge

of learning and performing a statistical practice while the community specified the different actors involved. To describe boundary crossing, I characterized the experience of junior statisticians as either continuous, detailing, or discontinuous, which was illustrated by Grosemans et al. (2017) and conducted a cross-case analysis between participants to summarize statistical practices. The theory of boundary crossing was explained in the literature review with the theoretical framework in Chapter 2, and the process of data analysis for the interviews was explicitly expressed in Chapter 3.

First, I present the practices that emerged in the analysis of the interviews, with some salient characteristics such as the number of quotes, the percentage of quotes pertaining to the workplace, and the number of participants who cited each practice. Second, each statistical practice is described in light of boundaries encountered by participants and including extensive quotes to depict the transition. Within each practice, I also portray the role of the mentor to support learning statistical practices, as mentioned by the junior statisticians or by the mentors themselves.

Identification and Reflection on Statistical Practices

Across the eight junior statisticians and the two mentors interviewed, eight categories of practices were created (see Table 14), of which, six were mentioned by all participants. Among these, practices related to *Programming* were cited the most with 232 quotes, followed by 183 quotes about practices associated with *Communication*, and 166 quotes for practices linked to *Analysis*. Fewer participants alluded to practices of *Documentation* and *Time management*, also less extensively described, with only 53 and 87 quotes, respectively. However, these practices were mostly mentioned at the workplace with 70% of the quotes for *Documentation* referring to this system. Nearly all

quotes tied to practices of *Time management* were accounted by the workplace, representing 91% of the quotes from eight participants. The third most cited category of practices at the workplace was *Data Management*, followed by *Communication*. Overall, all practices predominantly pertained to the workplace.

Table 14

Citations of Practices in Interviews

	Total number of quotes	Workplace % of quotes	Number of participants
Design	110	60	10
Programming	232	57	10
Data management	142	67	10
Analysis	166	58	10
Collaboration	150	63	10
Communication	183	64	10
Documentation	53	70	5
Time management	87	91	8

In Figure 14, the elements of each activity system, namely the tools, rules, division of labor, and the community, were represented in accordance with their contribution to quotes in academic settings and at the workplace. Increasing contributions of elements between the two systems unveiled boundaries. As illustrated by practices related to *Design*, the percentage of quotes in the division of labor almost tripled from academic settings to the workplace, meaning that junior statisticians tended to hold authority for designing studies more so at the workplace than in academic settings. The share of the division of labor across the elements increased at the workplace for all the practices, except for *Collaboration*, which indicated that junior statisticians were responsible for developing practices at the workplace and created boundaries for the transition. On the other hand, the

share of tools in academic settings was consistently greater than or equal to the share of tools at the workplace, except for *Time Management*. Indeed, in academic settings, participants mainly focused on developing tools rather than experiencing the division of labor or working with the community, and therefore, tools in academic settings often distinguished between continuous, detailing, and discontinuous experiences of boundary crossing, across participants.

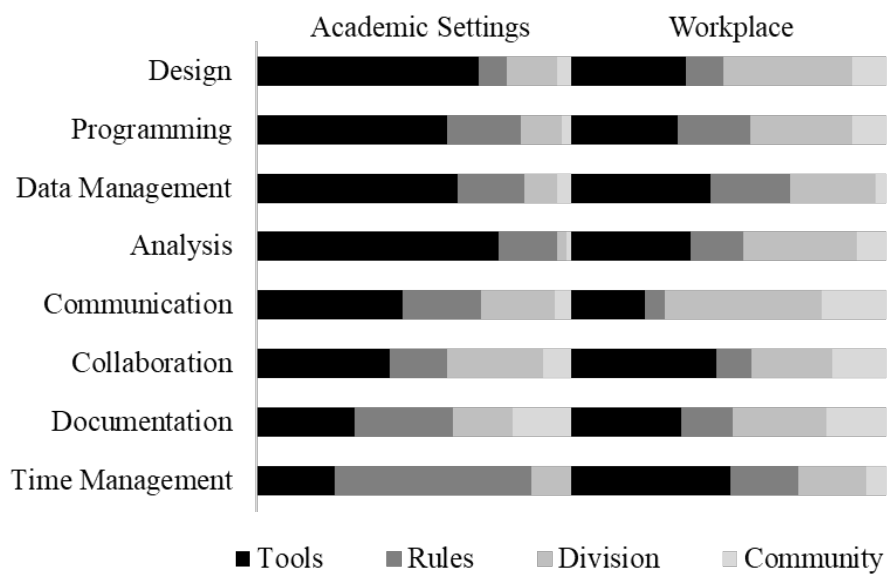


Figure 14. Distribution of elements for each practice.

Next, by reflecting on the experiences of participants, I sought to highlight the coordination of statistical practices between academic settings and the workplace. Each of the nine categories of statistical practices is described by unfolding boundaries encountered by participants in terms of the elements (tools, rules, division of labor, community) and classifying the experiences of the transition as continuous, detailing, and discontinuous. I will keep referring to Table 14 and Figure 14 in the description of each practice.

Reflection and Coordination of Statistical Practices

A cross-case analysis enabled for a rich description of the statistical practices, revealing differences and similarities across participants with diversified degrees and working within various fields of application. I start the description of each practice with a summary of participants who mentioned the practice and identify important elements from academic settings and the workplace that either facilitated or challenged the transition.

Coordination of Design

The design of a study defines the procedures of data collection and planning for statistical analysis to answer a research problem. The practice of designing studies included formulating statistical questions, choosing the appropriate measures, calculating sample sizes, or planning for analysis. Participants also engaged in activities such as communicating with domain experts, mentoring others, writing protocols, and applying for grants. All participants mentioned this practice and both mentors indicated ways they support learning in academic settings and at the workplace (see Table 15). The principal boundaries encountered by junior statisticians in the transition lay in the division of labor as most participants reported holding the authority at the workplace. Indeed, Figure 14 indicated that the proportion of division of labor increased compared to academic settings, also representing the most cited element at the workplace. John was the only participant to experience discontinuity in design practices because he had not acquired tools in academic settings and reported that learning about design exclusively occurred at the workplace. Tools in academic settings facilitated the transition of participants with courses in experimental design, clinical trials, survey data or sampling, for instance. As shown in Figure 14, tools were the most cited elements in academic settings. The

boundaries in the tools across participants revealed the distinctions between detailing and continuous experiences, reinforced by boundaries in the division of labor and in the community at the workplace. For example, Florence and Janet reported similar tools in academic settings but Florence experienced a boundary in the division of labor as she was the only statistician at the workplace while Janet was part of a supportive team. The community at the workplace played an important part in the transition with the support of mentors but also with interactions with teams and clients. Boundaries in the rules occurring as junior statisticians transitioned to the workplace included dealing with small sample sizes, specific types of methods, different implications of the results and considering the feasibility of studies.

Table 15

Important Elements for the Coordination of Design

Participant	Academic Settings	Workplace	Transition
John	Consulting course	Plan for analysis	Discontinuous
Stella	Course in design of clinical trials, course in sampling	Using textbook and meet with domain experts	Detailing
Thomas	Master's thesis	Online course on experimental design and analysis	Detailing
Edwards	Course in sampling	Half of the projects are on design and support from mentor	Detailing
Florence	Course in study design Projects	Working with principal investigators with mentor	Detailing
Gertrude	Course in survey data	Setting up experiments	Continuous
Janet	Course in introductory biostatistics and projects	Working with principal investigators with team	Continuous
Enid	Course in experimental design and consulting experience	No responsibility at the workplace	Continuous
Simone (mentor)	Independent study course	Write protocol but not the responsibility of junior statistician	Observed discontinuous
David (mentor)		Let statisticians create their own studies	Observed detailing

Boundaries in the Division of Labor for Design. Participants mentioned different responsibilities for designing studies at the workplace which they did not assume in academic settings, and thus, created boundaries. For example, they had to formulate statistical questions, choose the appropriate measures, calculate sample sizes, or determine the appropriate techniques. They also reported communicating with the domain experts as well as mentoring others to design studies.

Formulating a Statistical Question. When conducting interviews to hire junior statisticians, David was interested about how candidates thought about design and he would ask them to share “a project that the candidate worked on from start to finish and really try to dig in deep. In every single part from how, from the beginning, to how the question came about.” As illustrated by the experience of Florence, junior statisticians were responsible for formulating a statistical question at the workplace since she helped principal investigators with “creating a good study design for the question that they want to answer.” John wished he was taught how to “formulate a statistical question of interest and developing a statistical plan and getting that operational experience, would have been nice to have. Because I feel like that's something I had to learn on the job as well.” Indeed, even though he had a biostatistical consulting course in academic settings that John hoped would have given practical opportunities, the course was more about theory than practice and he did not have any experience in working on an actual study. His experience was thus considered discontinuous because designing studies was a practice he was lacking when he started his job. On the contrary, Enid experienced continuity since she had a consulting course for which she was accountable for helping the client design their study, together with her group members and without the input from the

professor: “we'd schedule the meeting, sit down with the client and go over, you know, what are your needs?” She had developed this practice in academic settings however, she did not mention if she contributed to the design of studies at the workplace.

Edwards contributed to the design on some projects at the workplace, but after three years of experience he wanted to have more responsibility and relied on his mentor to help him develop the skill of study design. He said that:

I haven't had to like really work on designs for that often, but next year is, my goal is to remedy that. So I think next year I'll be asking for more from [my mentor] in terms of questions and resources.

As Edwards needed his mentor to develop the practice of design, his experience of the transition was detailing. Similarly, Thomas worked with his mentor at the workplace to propose research that he wanted to do. In academic settings, he only had the opportunity to choose his own question for his master's thesis, otherwise there was no flexibility. He had to “decide what kind of approach I'd like to take, what question I'd like to answer with this data set that I had.” He was in charge of the division of labor and he “implemented on my own, interpret my own and kind of answer a comprehensive beginning to end.” Because design became his own responsibility at the workplace, the experience of Thomas was qualified as detailing.

Choosing the Appropriate Measures. Gertrude considered that defining metrics were important to design experiments and evaluate performance but apparently it was not supposed to be her responsibility. She shared that:

For many of our ongoing projects, we don't have metrics that are defined and unfortunately I'm not in charge of my team so I, it's hard for me to take the time to like define the metrics we've been running and building and someone has been building models without defining metrics to evaluate them.

On the other hand, Stella was in charge of the division of labor for defining metrics. Talking about educational program evaluation, she did “get involved a lot in the development of the program, the curriculum of the program and then at the end measuring and analyzing the program.” As a statistician for a baseball team, David learned how to “come up with metrics on how to judge a player.” In academic settings, Janet reported having some experience with this aspect of design practices by doing a project in an introductory level course in biostatistics, for which she had “to give the appropriate measure of such and such, like would you use an odds ratio, would you use a relative risk.” She held the division of labor because her professors “they made us learn it. They made us like take it in and then apply it immediately. It was very good.” The tools Janet acquired in academic settings helped her experience a continuous transition.

Calculating Sample Sizes. During consultations with principal investigators, Janet focused on power analysis and sample size calculations. Sometimes she had to refuse to analyze the data because the sample size was too small to provide meaningful results. As a statistician, Simone said that this aspect of the practice is not well defined but could be learned with experience over time:

One thing that we always struggle with is sample size calculation. It doesn't seem to be that definitive answer in many domains. We're always struggling with that. Even now. Um, I think ... it's gradual, over time, you learn that.

Simone thought that in general, study design was the most challenging practice for junior statisticians, by working on protocols which included sampling:

I think the most difficult, really, is the protocol. Only for the fact that, they... I think that therapeutic area or domain, they really would have to work with the medical director, the clinician, to figure out, you know, what samples, so what type of people, with what diseases, need to be collected and what are the appropriate statistical method that should be performed? Analysis plan.

Planning the Appropriate Analyses. As just mentioned by Simone, statisticians need to choose the appropriate analysis as “they’ll be part of finding the statistical analysis plan for the protocol.” She found that junior statisticians were actually better prepared for this aspect of the practice than other aspects mentioned for design. As a mentor at the workplace, she made sure to have them “mostly focus, entry level, focus more on just the analytical stage, for quite some time.” Indeed, Thomas thought that designing studies consisted in “searching, researching and implementing the solutions with the appropriate statistical approach for the task at hand.”

As a mentor in academic settings, Simone found that there were not many opportunities to develop the practice. Therefore, she created a new course to offer opportunities to students to hold the division of labor to work on design. She explained:

That's why we started this new independent study this year, um, for students that is, they want to go into industry, that I'm working with them to give them projects to see, you know, are they able to, well, they're only working with me, but I'll formulate the problem, they're doing analysis, they're figuring out what is needed.

Even though Simone was still holding the authority for design by defining the problem, she believed these projects better aligned with the expectations of the workplace.

Writing Protocols. As a mentor at the workplace, Simone believed that the contribution of statisticians for the practice of design was through writing a protocol. In academic settings, she thought that even doing a thesis was not necessarily sufficient because:

A lot of people with a thesis are not involved in creating a protocol or a project, there's sort of obtaining information that is going off to study. Only if they were part of a PhD program and if they were required to actually design a study, um, would they have, would they have previous experience on, um, something similar to what we would need them to do for a protocol.

Similar to protocols, Florence mentioned writing proposals for grants at the workplace.

Communicating with Domain Experts. Florence sometimes needed her mentor to communicate with the investigators about design since she did not feel she had enough experience or credibility. She shared an anecdote where investigators had to repeat “the experiment three more times before we were satisfied because they didn't consult us in the study design the second time they ran it. So we had them sit down with us and deal with the study design first.” More detail about this aspect of the practice are given within practices of collaboration and communication.

Mentoring. At the workplace, Thomas provided support to his coworkers who “are interested in getting into the field, and me, I might not work explicitly on their projects, but they have a question about how they should analyze different data and kind of answer the questions that they're hoping to answer.” Janet was mentoring principal investigators about the importance of calculating sample sizes by letting them know “ ‘No, you can't do this for these samples because you don't have sufficient power.’ What does that mean, so I walk them through that.” To explain, she had to “be the math person and like the sample size person and the kind of, sometimes I feel like the downer cause I’m like ‘No, no, you can't do that.’ ” She wanted to focus on getting researchers “in the habit of doing a power analysis. So I've been doing that a lot lately.”

Boundaries in the Rules for Design. Rules set by the workplace differed from the rules in academic settings with the issue of small samples, restrictions on the types of methods, the implications, ethics, or the feasibility of a study.

Small Sample Size. Participants had to deal with small sample size and adapt other practices such as analysis accordingly. Thomas mentioned that he had to design studies to “maximize the amount of information we can glean from the limited resources

we have.” Similarly, Florence talked about the impact of sample size on the design of the study for the analysis method: “We don't quite have enough data to do like testing samples versus independent samples. So that's the unfortunate downside of working with small sample sizes.”

Janet wanted to establish calculating sample size before collecting data as a rule for design at the workplace. Indeed, she said: “I'm working to do some frontend stuff as far as like catching people early and so we can get them the appropriate sample size techniques and do their thing.” For example, “I have a meeting after [the interview] with them to go over the results of their power and sample size analysis. Which is never fun, cause its always way more than they want.”

Specific Method. Some workplaces restricted the types of analysis or design methodology to be used. For example, Gertrude designed experiments to test if a strategy would “bring in more revenue or not” and to compare book prices. “Comparing is like: did this group of books perform better than the control group when we left the prices the same or did they not?” She had to find the appropriate analyses to investigate the differences but without using *t*-tests for example because “they seem at my company, not to like the idea of *t*-tests or *p*-values.”

Janet was looking forward to implementing at the workplace what she had learned about machine learning in academic settings but first she had to organize and classify the different methodologies used at her workplace. She mentioned “some very obscure metric that we use that nobody really knew how to interpret or use appropriately.” She did some research to learn more about this particular technique, “read the original like seventies paper that was scanned PDF of whatever from back in the day” and “30 years’

worth of debates about this one mathematical formula.” Finally, “I had to learn to zoom and pull back and like, okay, how do I explain this to my supervisor. And what is relevant to them about like, what is the executive summary of this 30-year long debate.” Before offering to use new methods, Janet had to learn the methods used by convention at the workplace.

Implications. The importance and implications of designing studies shifted between academic settings and the workplace in terms of interests from the perspective of statisticians. Gertrude found that the rules have changed from academic settings because working in retail, the implications, or importance, of the experiments were not the same: “the data in a business setting you might end up being a little bit cleaner and, and finding the exact cause of something might be a little bit less important.” Indeed, Simone thought that it was important for statisticians to learn how to work on projects for which they saw less value, meaning that the implications did not seem that important to them:

How to, for your real work to do that you're not really interested in, but how'd you become interested in it? So like, how do you, you know, get excited about a project, research it, do the best job on it.... And you're working in this domain that you don't know in theory, don't know and you have to learn about it.

As a mentor in academic settings, Simone assigned students projects that they were not specifically interested in to give them opportunities to learn:

I also started to give them things that they don't necessarily want to work on. So many courses, they'll say, okay, you know, I want you to do a presentation, design a study that you're interested in, blah, blah, blah. And the real world is, you're not always working on stuff you're interested in.... In my class, have to pick up, pick up out of a hat and what you have to research to do. Because in the real world you were young, your boss say you need to do this. So it's like, how do you experience, you know, trying to figure out something that you're not interested in or it's not your specialty, and how do you go about that? And I'm, so that's another aspect. I think it's different between academia and industry. In academia people tend to follow into their research passion. Um, but in industry, you know, is, is, what is the company's passion, not necessarily yours.

On the contrary, David supported the learning of his junior statisticians at the workplace by offering to design their own studies:

We had a process of them coming up with projects and it wasn't the easiest to process, but I really wanted them to think, okay, this isn't just about being a cool analysis, you know, we, our time is pretty valuable here. Why would this analysis be important? Who would it be important for? What kind of things are you looking to get out of it and you know these are questions that may not necessarily that they would have thought of, even though they seem pretty obvious. I don't think they're doing something interesting versus doing something important are often two different things. But I was trying to really try to bridge the gap between those two.

David highlighted that there are two main implications for a study: the interest with being curious about the results, and the significance, with implications for the field. This distinction created boundaries in the rule for the design of studies.

Ethics. Statistics help make informed decisions and statisticians need to ensure that the design of statistical studies complies with ethics. Stella and Florence took courses in academic settings which explicitly introduced ethical issues in the context of healthcare studies. Stella “talk[ed] a lot about ethics” in a course about clinical trials and she “didn't realize how much goes into drug testing” before she took that course:

I was really amazed at the design of the trials for cancer medications, because obviously you can't treat one group with, you know, a new type of chemotherapy and then give one group a placebo. That doesn't do anything. I mean, that's very unethical. So I have learned a lot about those types of situations in that class and also just random studies how you can breach ethics, if you know a participant needs a particular drug and if you know you willingly break the random trial, random blind trial and give them the drug, that's unethical as well because you could be hurting another person. So I didn't realize that, this, all these things go into just drug research.

Even though Stella did not give an example of how she might have dealt with ethics at the workplace, she considered ethical issues as crucial for setting up an experiment and with sampling procedures. Similarly, Florence took a course about ethics in biostatistics

“which was utterly valuable. It was specifically catered to our particular master's program. So it's all about the ethics of statistics and of medical studies and pharmaceuticals, which was really important.” In order to give some experience with these concepts, for the ethics class:

[The professor] cobbled together various case studies and just plain ethical writings, like medical ethics writings that weren't stat specific. So every week she'd have us read three or four different articles about a series of topics that we were going to be discussing and she'd expect that we'd have them read before we walked in. And then most of the class time would be kind of a discussion about that and feeling them out. And sometimes she'd throw us some theoretical cases and put us on different sides in various different positions as the PI or the statistician or a data manager to see what we would do in that situation. And that was really valuable to have her both make us look into past cases that went really bad.

Not only this course introduced the rules of ethics for medical studies, but it provided tools for Florence with some knowledge of ethics throughout reading articles and reviewing case studies as well as experiencing different roles for the division of labor.

Feasibility. Another boundary that statisticians encountered as transitioning to the workplace was about the feasibility of a study, related to expenses, time and implications. Edwards shared spending about half of his time on “design, budgeting.” While describing the difference about academic settings and the workplace in terms of designing studies, Simone referred to the restrictions to take into consideration:

I don't think we, we spend enough time on actually designing studies and um, breaking the design of the study, and choose the correct analysis. And then knowing what I had, what an output would be, that we would interpret, um, I don't think we have enough practicing. Okay, this is my research question or this is the objective. Well, what would be the ideal study design, let's plan that out and then let's look to see if that's realistic in terms of the real world. Like you know, cost-wise, timewise? And if we have to make some adjustments to that and what would be the plan B? We are, can we sacrifice, you know, maybe what time points or less people, but will we still get a valid result that would mean, you know, good sound statistical practice. I don't think they get enough practice, you know, building that and then kind of real, really look at that to see if it's feasible

going forward and getting that feedback from others, things you know, and they'll say, this doesn't make sense and I think you need to adjust it. So, um, that's what I would like to focus on.

As a mentor in academic settings, Simone was planning on giving her students opportunities to learn this aspect of design practices.

Boundaries in the Community for Design. Working on design often implied working with domain experts. Further explanations about establishing a working relationship with domain experts are described for the collaboration and communication practices. Here I highlighted the specific aspect of collaborating with domain experts for design.

Domain Experts. Florence had a study design course for which she learned how to align the research goals with the analysis with “A needs to match A and B needs to match B and here's how you help develop, help an investigator develop towards that.” However, at the workplace, principal investigators:

Sometimes they come after you've done already some work in or you're looking at it together and they're like, this doesn't look right, or this is not what I want. So it does take experience to get that, to realize that, oh, there's a lot of pre prep work before we even start the statistical modeling.

Similarly, Janet worked with principal investigators who did not usually involve statisticians in the design of their study, and she had to adapt to studies that may not have been well designed. However, she chose to share a more positive anecdote with one specific collaborator:

This has been the best meeting ever, and he was really interested, he kind of knew what his experimental outcome of interest was, for example, and he had an idea of like the relevant interaction variables and the relevant things that might affect his outcome. And I was like, [gasps], it was so great. So we had that meeting and I was just really excited about doing that experiment and moving forward, and I would be very happy to work with him going forward. Provided he can pick, he's not too put off by the number of samples that he'll need.

Simone described that the different aspects of design have to be done in conjunction with domain experts and other team members:

It's really working to try to understand, you know, um, all the data collection, because that would go into the cost and the time and the duration and working with those cross functional teams to build all that out. But that really is a more of a senior level staff position, not entry level.

For example, Enid developed a series of questions to interact with a client through her consulting course. She would ask: “What are your needs? Do you need help setting up a design? Do you need help doing any sort of sampling? Do you need help creating a survey?”

Domain Knowledge. Working with domain experts involved learning domain knowledge. According to Simone: “a lot of it is working, um, with the most professional to learn their domain and then being able to say, ‘Okay, how will I apply the correct statistical methods to answer the question that they want?’ ” As a student, Simone noticed that “in my PhD, we integrated so much with uh, you know, clinical or medical world, we had to, we were required to take, um, there's a level classes in that domain itself” which had prepared her to work in the medical domain at the workplace. Janet also had courses that were specific to some medical domains such as a methods course for analyzing tuberculosis cases or a course about clinical trials.

To set up the design of her study and conduct an evaluation of an educational program, Stella met with the program director to understand the context, “how the program is administered, the dosage of the program which is really important.” She said “when I evaluate, I need to find out what exactly is going on. And from there I try to develop a tool and for a program like that I would do a survey.”

To have experience about this practice in academic settings, Enid thought that attending an agricultural school was an advantage to learn how to design studies in different domains. She considered that “setting up design of experiments was super easy ... doing experiments on wheat. Awesome. Or snails or something.”

Boundaries in the Tools for Design. Most participants mentioned learning tools for design in academic settings with specific courses in design, dealing with experiments, sampling, or survey.

Experimental Design. Gertrude mentioned that in one of her classes about data science for education and survey data, she talked about designing experiments and dealing with other types of studies:

Causal inference become much easier when you've already like set up an experiment and done that, whereas we were taught to deal with like okay, what about all the cases where you don't have an experiment that was set up.

Gertrude was given the opportunity to learn about different types of studies in academic settings which facilitated her transition to the workplace. Thus, her transition was found to be continuous. Janet found that she had a very practical experience in her masters overall but in particular for design which she practiced through projects:

We learned what - when would you do a case control study, what kinds of stats can you apply to a case control study, what's appropriate, what's not. What's a valid measure of change in a given study or what's something that you would report with a study like this, or one of the typical measures of such and such a thing. When do you use which, what are the pros and cons of using each one.

Florence had a study design course which pointed out “where logical fallacies were” for aligning research questions with the design. To practice:

I did a study design or a mockup for a study design where we were putting forth our own proposal for whatever it was. And that was really helpful to go back when I actually ended up needed to write that piece of it.

Florence experienced continuity for design because she referred to her study design course as “that particular piece of my education very much continued on when I started this job where I’m in, even now I get to help people with creating a good study design for the question that they want to answer.”

On the contrary, Thomas had to design experiments at the workplace and contrasting with what he learned in academic settings, he said that “really required a lot of little more advanced thinking of just like let's get a population, set up an experiment, collect the data and analyze it.” To learn the practice, he took an online course on experimental design and analysis within a few weeks at the workplace. Indeed:

It wasn't really something that I was explicitly exposed to in my masters, but it was something that I kind of brushed up on through online courses and I frequently use online courses to kind of filling the gaps from my more formal education.

His thesis gave him a first opportunity to design his own study in academic settings, and the workplace, Thomas would research proposals of past projects to put together future research.

Survey Design. Gertrude mentioned a course about survey data, but Stella had to learn how to design questionnaires at the workplace, qualifying her transition as detailing. She had “to kind of teach myself some things too. I had to buy a textbook. And I also looked at some of the previous examples that I had found from the previous person that had held my job.” She developed questionnaires to evaluate the programs of education she was assigned to monitor.

Enid had an opportunity to design a survey in academic settings with her consulting course as “the client that my group had needed help setting up a survey.” Janet

learned on a previous job before her master's degree how to do survey development by creating questionnaires with pre- and post-assessments.

Sampling. Stella mentioned taking a course in design of clinical trials and one in sampling methods while Enid had a course about the design of experiments where she “learned how to set up sample frames and how to set up designs, how to take that data from the design and apply analysis to it.” Edwards also had a course on sampling for which he had some opportunities to learn how to determine “power and sample size calculations. We did very few. I think I could count the number of power analyses I did in grad school on one hand.” Thomas chose not to take the course offered about clinical trials in his degree because he was prioritizing another area of study but overall he “had an idea about experimental design and power analysis and different ways to, to analyze data that results of an experiment.”

Grant Applications. One of the tools Florence was missing at the workplace was how to write grant applications which characterized her experience as detailing. While she had learned how to write a paper, “no one taught us how to write grant applications, let alone figure out what information is needed so someone else could write them for you.” Indeed, she had to figure out by herself how to write grant applications and proposals for new research.

Transition of Design. Participants identified various activities while engaging in the practice of design. For example, they needed to formulate a statistical problem, define appropriate measures and the sampling frame by calculating sample sizes. Part of their responsibilities also included communicating with domain experts which highlighted the importance of involving statisticians in the design of a study. Indeed, participants

contributed to the writing of protocols and proposals for grants, and in turn, after getting some experience, provided mentoring to coworkers or collaborators. The transition of the practice was qualified as detailing for most participants, mainly because tools developed in academic settings facilitated boundary crossing. Tools such as courses in experimental design, clinical trials, or sampling, as well as experience in consulting were particularly beneficial to support the transition. Indeed, the lack of tools and division of labor in academic settings rendered a discontinuous experience. In the transition, participants learned design practices, using tools provided by academic settings, with the support of a mentor, or by researching new resources with online courses or textbooks. Mentors provided opportunities both in academic settings and at the workplace to support learning. They recommended that students should be able to practice how to design their own studies in academic settings, yet they should not only pursue their own interests. Similarly, at the workplace, mentors should progressively introduce junior statisticians to design, starting with the plan for analysis, which they were the most prepared for.

Coordination of Programming

With the amount of data statisticians need to handle, they develop the practice of programming to complete different types of tasks. For example, programming was used to perform data analysis, pull data out of databases, or clean data. All participants focused on this practice and was mentioned by both mentors. In fact, the practice of programming was the most quoted practice in the interviews, with most quotes referring to elements at the workplace as shown previously in Table 14. Junior statisticians all expressed the need to diversify their programming tools at the workplace, however, they had acquired

programming tools in academic settings that made the transition to the workplace either detailing or continuous (see Table 16).

Table 16

Important Elements for the Coordination of Programming

Participant	Academic Settings	Workplace	Transition
Florence	2 undergraduate courses in statistical and mathematical software, and 1 graduate course in statistical software	Learn a new programming language for data management	Detailing
Gertrude	1 graduate course in statistical software	Learn many new programs with summer course, online courses and online resources	Detailing
Edwards	Many undergraduate and graduate courses in statistical software	Use a combination of statistical software Learn new programming language	Continuous
Enid	Many undergraduate and graduate courses in statistical software and 1 course in data management	Use a combination of statistical software Learn new software at workshops	Continuous
Janet	Undergraduate course in computer science and graduate courses in statistical software	Learn new programming language within same software	Continuous
John	Undergraduate course in computer science and statistical software Graduate course in statistical software Internship with new programming language	Use a combination of statistical software Learn internal software	Continuous
Stella	Graduate courses in statistical software Experience with mathematical software and programming languages in Physics	Learn a new programming language for data management	Continuous
Thomas	Many undergraduate and graduate courses in statistical software and programming languages	Use a combination of statistical software Learn new programming language with online courses	Continuous
Simone (mentor)	Students do not meet expectations in advanced graduate courses	Most important skill for hiring, statisticians need to know various programs	Observed detailing
David (mentor)		Most candidates show off their code during interviews	Observed continuous

The need to keep learning originated in boundaries created by different rules between academic settings and the workplace. The routines in which junior statisticians engaged had different goals at the workplace. In academic settings, programming focused on applying specific tools to a given task while at the workplace participants needed to find the appropriate tools to perform a specific task, which was also a boundary in the division of labor. Other rules such as the limitations of technology to handle big data or the time allocated to complete a task constrained the learning of the practice at the workplace. Differences between detailing and continuous experiences were caused by a variation in tools offered in academic settings. As an example, undergraduate courses in computer science instilled the logic of programming for Janet and John and supported their learning for new programming tools. Florence wished she had such a course in academic settings and thus her experience was considered as detailing. The community at the workplace also contributed to the practice of programming, by requiring specific programs, by including members with different roles on a team, or by offering training.

Boundaries in the Rules for Programming. Participants experienced boundaries in the rules for programming while they transitioned from academic settings to the workplace. They engaged in different routines and had to share their code. Their practice was also constrained by the development, the cost or the limitations of the technology, and by time.

Routines. The practice of programming tended to be developed for the purpose of performing analysis in academic settings while at the workplace it involved different routines, creating boundaries for participants and the need to learn new tools. As a mentor at the workplace, Simone noticed that junior statisticians had not acquired programming

tools applicable to perform just any routine. She believed they “need stronger programmable skills” in order to do “data cleaning and data manipulation and data re-coding and all that.” Indeed, Florence found that programming focused mainly on conducting analyses during her master’s degree while she actually needed programming to mostly do data management at the workplace: “I got a lot of experience in grad school on how to work with code for the purposes of doing statistics, but very little experience in doing data cleaning.” Florence had to learn how to do data management which involved learning a new programming language. By contrast, the experiences of Enid and one of her coworkers showed how boundaries in the rules, specifically related to data management affected the transition of programming. Enid started working at the same time as a coworker at the workplace. They worked on similar projects, but the main difference between their experiences of the transition was about programming. In fact, there was a difference in their educational background. Their degrees covered a similar coursework, but at a different level, with Enid’s coworker having completed an undergraduate degree while Enid had two graduate degrees. Enid felt that she had:

Such an advantage over my coworker because I had like, at least with the data management SAS class, I had this whole semester's worth at least where, you know, I was doing this day in, day out, like quizzes once, twice a week where, you know, I have somebody who can give me like almost immediate feedback.

Enid compared her experienced to her coworker’s, who still had to take classes and go to workshops “to actually teach her the basics of SAS before she could start doing anything.” The rules between the workshops at the workplace and the rules of courses in academic settings are different:

It's just not the same as in like a university class setting where you're going three days a week to this class. You're having to like constantly show your comprehension through projects, papers, quizzes, tests.... whereas like in the

workshops and like the online trainings that we have at work, you don't get that feedback.

Thomas encountered a similar boundary as Enid's coworker when he started working after his undergraduate studies in quantitative economics. The tool required at the workplace was not adapted to the routine of data management and Thomas had to learn a new programming language:

We're doing everything in Microsoft Excel and kind of like, I was kind of fed up with Microsoft Excel and I had only use Stata in my undergraduate studies. So I started learning R just for data cleaning and found that it was kind of very useful for a lot of the work I was doing.

After this first experience, Thomas decided to pursue a master's degree in biostatistics, combining the field of healthcare with quantitative skills. The courses required "mostly coding exercises where you're getting a model to a dataset that's provided, interpreting results." However, the emphasis was on performing analyses as Thomas found that "the exercise wasn't really the programming, it was more the model validation and interpretation and write up." The routine seemed to determine what tool would be used for programming. Indeed, John would choose a different tool for programming depending on what task he intended to perform: "on my more simulation projects or data visualization projects, I do them in R" or "sometimes the data that's given is already a SAS data set and by convenience it's just easier to work with SAS."

Collaboration. Another boundary in the rules as participants transitioned to the workplace was that they shared their programming at the workplace for collaboration or for an audit, for instance. As a mentor at the workplace, Simone expected:

The highest level of rigor on programming, because we have to have, everything's doubled their size because we know that there are for audit, um, and those kinds of things, and that we have to read through those down, you know, 20 years from now.

Simone thought that there was a shift from academic settings “when people learn programming, they don't typically have that rigor in terms of people typing their programming.” The lack of programming rigor created boundaries for Gertrude at the workplace, making her experience detailing. She mentioned having issues with programming because “people were not documenting their code” and they only shared the GitHub page for the company. Since she did not understand the tools they were using for their programming, Gertrude had a hard time understanding the purpose of the coding and “it took me literally like a year and a half to understand what the code was trying to do, after begging people to explain it to me, like in English.”

To be able to communicate with the community and collaborate with others, David learned a tool. Indeed, he found “Python to be useful, just because I know a lot of other people use it. And so for me it's more useful in being able to understand other people's work as opposed to what I actually use myself.” The analysis of documentation practices will further focus on the aspect of collaborating with programming.

Technology Requirements. The technology is constantly evolving, with new statistical programs being developed and impacting the way programming is done. Some participants reported having to learn new tools to meet the technological demand of the workplace. While looking for a job, Gertrude noticed that employers were interested in programming tools that were not included in her master's degree. Not knowing how to use GitHub “was seen as like a huge red flag and all these jobs [told her] like ‘what do you mean you don't have experience with Git?’ ” As she started her career, she had to learn GitHub because it was required by the workplace.

In addition, Florence thought that her master's degree "wasn't technologically advanced enough to help me." In particular, she was not taught tools to retrieve data from the cloud system which she had to do at the workplace. She also wished she had "been picking up another coding language. We only had to learn two in grad school." She mentioned the programming language Python because it "is now being used more and more." Indeed, at the time of our interview, Gertrude was looking for new career opportunities and she was teaching herself Python. She realized that in the area she was living "basically all the jobs require Python," because "no one uses R for data science, they all use Python." She added that the job announcements "are very much like: you have to use Python, you have to use scikit-learn." On the contrary, Simone favored candidates who could program with R for applications in data science: "we are currently looking for, you know, statisticians with the big data or data science, machine learning, um, background or experience and, uh, particularly, you know, programming experience with the R language." Even though, SAS was the principal statistical program at Simone's workplace, her company was looking for "people who are very sound with R but also with some SAS," but not Python. Gertrude also taught herself TensorFlow because she wanted to find a job in artificial intelligence for which "I guess there are other things they use also but TensorFlow is sort of where most beginners start."

Technology Cost. The access to technology could be restricted by the cost of the tools themselves, especially in academic settings. Enid thought that being exposed to different software through homework was very beneficial, but she understood that "some universities are restricted on, like, of financial, you know, like financial restraints. R is open source, it's free, you know, and there's so many packages out there that like

universities can download.” Edwards commented on the difference between R and SAS in terms of cost:

You have to pay SAS for the software that allows you for, SAS Enterprise miner, in order to do [machine learning or data mining] stuff in SAS, whereas in R it's open source, if someone made the software available through an R package, go for it.

On the opposite, the cost of technology commanded the choice of program for Stella who was working for a nonprofit organization. She had the option to choose which program she wanted to use, and she picked “R because it's free” and “SAS is also really, really expensive.”

Technology Limitations. The choice for programming tools could be constrained by computation power or the ability to work with big data. For example, Florence had to learn how to program with SQL “to pull out a huge chunk of data from what was the SQL database.” In addition, she thought that parallelization will become essential “in a future job that includes just how much data we all collect now, that's the way we need to be moving.” Enid was also dealing with large datasets and at the workplace, her programs took “a long time to run. And sometimes I'll make mistakes on my code and then I'll have to go back and run it. And then it takes another like two hours to go through.” The amount of data impacted the time for running programs and using different programming tools could resolve this issue. John encountered limitations for computation and had to learn a new programming language during an internship he did as part of his master's degree:

We ran into computational issues with R cause it was taking so long to do our analysis. So we, when we moved to Python, we found that our optimization of our programs are running so much faster and so much more efficiently. That, I had to learn on the job.

Edwards also noticed limitations between tools in terms of performing analyses: “if you use an R function that's using Newton Raphson and it's not going to converge, SAS will switch to another method. It has some protections built in. R ain't gonna do anything.” Indeed, Edwards talked about R being more flexible for techniques, optimization and visualizations:

You can write functions that are optimized and stuff like that. In SAS, I'm not sure how you go about doing that. I guess you'd have to use SAS procedures that they optimize on the back end. So it's, it's, it's just limiting.

Time. At the workplace, shorter deadlines constrained the practice of programming and participants had to balance between different programs to optimize their time. Enid explained: “Being in like nonacademic world, you don't necessarily have the luxury of like, ‘Oh, well, you know, I'll just find my time. I'll, I'll make some time to make this work in SAS.’ ” She thought that “SAS and Excel kind of go hand in hand to an extent.” She gave an example of using a combination of SAS with Excel because “if I need to do just like a quick row summation or a quick column summation, you know, sometimes it's just easier just to do those quick calculations in Excel.” Or on the contrary she would use SAS to “read in the Excel file, made the cumulative with the do array with the do loop and the arrays. Then I exported it to Excel.” Enid balanced her use of different tools to perform the task faster.

Boundaries in the Division of Labor for Programming. As a mentor at the workplace, Simone valued the practice of programming. During the hiring process of junior statisticians, she would especially pay attention if applicants had learned tools in academic settings: “first and foremost is their programming, um, level, that's what we're looking for.” At the workplace, participants realized that programming became one of

their main responsibilities, sometimes assuming the role of programmer or data manager. They also had to choose the appropriate tool to program for specific tasks and they were responsible for their learning.

Assuming Different Roles. Participants mentioned the practice of programming as being one of their main responsibilities at the workplace. After checking her emails every morning, Florence would usually open “R studio and [be] writing code all day.”

Indeed, she found that:

Most of my job is really just doing code and a lot of that is yes, data cleaning, some of it's exploratory data analysis and some of it is actual analysis. Most of my job day to day is just code.

Enid had a similar experience, in fact, she said: “right after [answering emails] then I jump pretty straight into my SAS code.” Enid always tried to improve and said that “sometimes I might spend an afternoon researching better ways to code my program better too, because, you know, I want it to be more efficient, I want it to run faster.” John found that there were differences between the responsibilities of a master’s level and doctoral level statistician. Indeed, he thought that master’s level statisticians focus more on programming to support the work of doctoral level statisticians. His responsibilities at the workplace were mainly centered around programming and he said that “a more appropriate title would be statistical programmer” for his job. Indeed, he had the feeling that master’s level statisticians, “we're just programming like 24/7.” On a typical day, he estimated that:

At least half of my day can be spent doing programming and working on the projects that I need to get done. So often I have a lot of deliverables that need, that require statistical programming and output that they need to see. So half of my day will be spent on a computer using R and SAS to get those completed.

The time taken up by programming came as a little surprising to some participants. John

was “kind of sad to say it, but it's the programming courses, so the SAS programming and the R programming courses are probably what I use the most in my work right now.” Gertrude also mentioned that “for my job I'm in now, the only useful thing I got out of my degree was that, that I could write code in R and it's mostly, you know, data cleaning mode.”

In addition to taking most of their time on a daily basis, programming seemed to appear under different forms and make them assume various roles at the workplace. Parts of Gertrude's responsibilities were to “mak[e] sure all of the steps that go into a process are automatic and are working every single week.” At the workplace, John described his role similar to one of programmer:

Using R and SAS, I do that constantly in my work and I utilize those skills daily to the point where I feel like I am growing as a programmer and I am definitely more comfortable in R and SAS than I have ever been in my grad school career and I've learned programming techniques that I didn't even know existed or didn't think I would be able to know, but just by using it every day and my work requiring it. I think my programming skills are probably at the highest they have ever been. So that has been continuous.

Similarly, Florence also experienced a boundary in the division of labor between academic settings and the workplace as she portrayed her job as:

It was a lot more of how to think like a programmer because I'm trained as a mathematician, not as programmer. How do you automate something to run multiple times in a time efficient manner and you get it to pull up something in the particular way you want it without using a million lines of code?

In academic settings the professors were providing the programming tools with the type of programs and commands to use. Gertrude mentioned a programming course for which:

We were given SAS code to run. We didn't have to write any of our own SAS code. It's just like copy and paste it in, look at the results and like, okay. So I didn't really learn how to write anything.

Besides assuming the role of programmer, participants also improvised themselves as data managers. Stella experienced a boundary in the division of labor at the workplace when the “database administrator just up and quit one day and I was the only one that had access to the database.” She became responsible for managing the database and “to keep the organization going, I had to, you know, teach myself more SQL.”

Florence also learned SQL because “I was going to be asked to be the, the backup database manager for our database manager while she was gone.” David explained that a tool such as “SQL was always useful” for statisticians because “often most companies don't have people that their role is to actually just pull [data] for other people to analyze.”

In academic settings, Enid had a course in data management and it “really made [programming] concrete for me.” Indeed, she felt like “someone's holding my hand, baby step through the way” and “it was nice to be able to just sit down and go through the logic of everything piece by piece.”

Choosing the Appropriate Program. Participants experienced boundaries in the division of labor in terms of choosing the appropriate programming tool to complete a task. In academic settings, the tools were often specified by the professors and the requirements of the course. In fact, Enid had a variety of courses for her two master's degrees and they “either only specifically taught a statistical software or it taught the, the stats concepts and you had to do your homework in the statistics software.” Similarly, Thomas found that he learned different programs which were determined by the topic of the course:

General data science using R and Python, and the other one was more with a deep learning focus, in using, it was kind of like an introduction to using Keras for, in Python, for deep learning models. And then a machine learning course that was taught primarily in R.

Thomas also learned “a lot of Python more focused on machine learning, deep learning, data science.” Programming in academic settings was structured by the course requirements and as a student, David did “apply R to the problem that's being asked. But I would say, actually, just, I really became good at programming and just on the job learning.” Indeed, David learned more programming tools at the workplace.

There was a boundary in the division of labor to determine the appropriate tools for John and Edwards who balanced different tools depending on what task they were working on. For example, John considered either using R or SAS:

SAS is like kind of my more secondary, like if I need to do a project in SAS, I will do it in SAS, but, my preference is definitely for R and I would say like 80% of my work is in R and 20% it is in SAS. So it depends what's needed to get done.

Likewise, Edwards balanced his programming between R and SAS with “60/40 R to SAS.” He was actually hoping to “build up the software I need. The functions I need, the packages I need, stuff like that to get the stuff I do, done efficiently.”

The division of labor for programming at the workplace seemed to rely on statisticians, while in academic settings the professors were in charge of choosing the tools, with the exception of Thomas and Enid, to some extent. Thomas had the option to choose the program with which to complete assignments during his master's program. There was no restriction, only suggestions for the choice of programming tools. He decided to use the programming language he was the most familiar with, but still choose to practice others:

[The professors] kind of let it up to you and encouraged either R, Stata or SAS and I mostly use R.... I did use SAS just because it was kind of, I really wanted to diversify the different programming languages that I knew, I had a pretty firm grasp on how to use R and so I use SAS, didn't use a lot of Stata.

Because Thomas was able to choose the appropriate tools in academic settings, his experience of the transition was considered continuous. Similarly, the choice of programming tools was flexible to complete homework for one of Enid's courses. She used the programming language R but it was not mandatory and she could use SAS if she preferred. Thus, she also experienced continuity for programming in the transition to the workplace.

Self-teaching. To diversify their programming tools or adapt to the tools at the workplace, participants referred to the learning process mostly as “teaching myself.” Even though Gertrude was not using Python at the workplace, she decided to learn it to strengthen her application to future jobs. Indeed, she had “to spend an hour each morning just teaching myself Python.” She considered the task of learning new programming tools as her own responsibility and did “teach myself SQL” and “teach myself TensorFlow also.” Likewise, Thomas used online courses to start learning Python and did “follow out that along and then started using it in my everyday life and kind of like learned through doing.” If he encountered coding issues, he would “hop on the internet and Google like ‘How do I do this?’ ” The reason he kept learning was:

To broaden the different programming languages for newest so I kind of forced myself to struggle with Python. I still do. So definitely not the best in Python, probably close second but it's kind of something I'm just learning to, like learn through using it at work.

Learning with online courses is a tool Thomas started using at his first workplace, between his undergraduate and graduate degree. He had to learn R and he “kind of self-taught myself. I can't totally say self-taught because it, it was mostly through online courses.” He got a three-month membership for a website specialized in statistical programming and “having economics background, I thought I only have three months,

here you can maximize the amount of usage I can get out of this.” The online courses helped him learn:

I probably put in maybe 100 hours of their trainings and really got up to speed on it and started using that and almost by the time I left that, that job as a research assistant, I was using R every day.

Mentoring. Participants reported providing support for programming at the workplace. At John’s company, many statisticians in his team created R shiny apps to improve workflows and which they “make accessible across the company.” He did not create apps himself, but he said:

I've been more playing a supportive role if someone needs me to enhance their shiny program or if they need me to validate their shiny program. That's kind of where I come in. But we have a lot of statisticians in our company who do R shiny, who create R shiny apps to kind of help improve workflows in our company.

After getting more comfortable with programming, Thomas started to mentor others and for example he helped a colleague “figure out like how to implement a machine learning model in Python.”

Boundaries in the Tools for Programming. All participants mentioned learning programming tools in academic settings but they also all reported that they had to keep learning at the workplace. The differences between detailing and continuous experiences were related to the diversity of programming tools acquired in academic settings, as well as specific tools and learning about the logic behind programming.

Diverse Programs. Table 17 shows the programming tools learned by participants in academic settings and the tools used at the workplace. All junior statisticians learned SAS in academic settings but only half of them use this programming tool at the workplace whereas all participants, except Enid, used R.

Table 17*Important Tools for Programming*

Tools	Academic settings					Workplace
	SAS	R	Python	Stata	MAT LAB	List of programming tools
Gertrude	X	X		X		R, SQL, Bash, Python, Tensorflow, GitHub
Florence	X	X			X	R, SQL
Thomas	X	X	X	X		R, SAS, Python, Stata
Edwards	X	X		X		R, SAS, Python*
John	X	X	X			R, SAS, Internal software
Enid	X	X				SAS, Tableau*
Janet	X		X		X	R*, Python
Stella	X		X		X	R*, SQL*

Note. * shows which programming tools were learned at the workplace

As a mentor in academic settings, Simone noticed that her students did not have the necessary programming tools she had hoped for an advanced course at the end of their degree. She said that “they're not at the skill level really needed to do what I'm trying to execute in my class. I spent some time, you know, still focusing more on the programming and stuff like that.” As a result, Simone did not want junior statisticians to start too early with programming at the workplace as “it takes a good, a while for them to, you know, um, to give them R tasks.” Indeed, Gertrude noticed that she was missing programming tools after she graduated, while interviewing for jobs and transitioning to the workplace. She found that Git and SQL “were not things that we were taught in school, that in the field of data science, whatever data science still means, those end up being really key. Um, Git, SQL and Bash and really to a large extent, Python also.” She had used GitHub in academic settings but not enough to know how to use it: “Git, I had

heard of it, had used once or twice in school as part of one of the projects I'd been working on but really didn't have experience with that.” Referring to SQL she said:

That I learned the summer after I graduated, from the masters. There were a lot of courses online you can take to learn it and you're like one of those boot camps, like the one-day SQL bootcamp. But I think I just took one of those like mass open online courses. It was one of those things that is not super difficult to learn. It looks really intimidating at first and you're using it. It's like, oh, I can do this, you know, in a day and we really competent enough in it that I could then go to job interviews, can say, yes, I could use this. That was nice. I was afraid I was going to have to like go through the steep learning curve of R.

Stella also had to learn new programming tools, such as R and SQL, when she entered the workplace, however, she felt that the tools she learned in academic settings made her transition continuous since “SAS calls a lot of SQL”:

Learning how to program in SAS has helped me become a better programmer in R, how to pick languages a lot faster, especially with the SQL, like I was amazed at how fast I was able to pick that up. To think, like the computer programming skills that I learned through the majority of my classes at [university] really helped me a lot.

The programming tools from academic settings differentiated the experience of Stella and Gertrude. Stella added that “once you can program in one language, picking up another one really isn't as bad as like learning one for the very first time.” She found R comparable to MATLAB and Python, which she had experience with while working for a scientific agency.

Similarly, tools from academic settings supported the learning of new tools at the workplace for Janet, John, and Enid. Through a course on machine learning, Janet had learned Python but she particularly appreciated learning about Jupyter Notebook which compile code from various programs: “You can use a different coding programs to run your code and see your results printed in line with the code itself, you would run like a

SAS kernel, and R kernel, or a Python kernel.” Janet had to learn R since it was the tool required by the workplace and she found it “easier [to learn] with the Jupyter Notebook framework because I could walk through the code and see what it does. A little bit faster than R Studio. I don't really use R Studio.” As for John, he used tools from academic settings to learn his company’s internal specific software. Even though it was something “that you cannot learn in school cause your school does not have access to particular software or tools that a company has,” he mentioned that having “my computer science and programming backgrounds, from both undergrad and grad school provided a good foundation to make that learning a lot easier.” In addition, Enid felt that she had “an advantage when I got into the workforce because I'd had that kind of formal classroom training.” She said to “know enough and if I don't know it, I can Google it, which is what I do.”

In addition, the support of mentors at the workplace eased the transition for John during an internship required by his master’s degree. He learned Python “on the job. And that was kind of a challenge.” His colleagues at the workplace helped him find resources and “after my internship was done for the day, I could kind of do some tutorials on my own to learn Python.” Gertrude also worked on Python with her mentor who had a background in software development, and he helped her “write senior code that doesn't break as often.” Indeed, she said “my issue is I don't know how to write [code] in a way that, you know, it's clean and reproducible, and doesn't break.” However, Gertrude thought that she would have had a more continuous transition “if someone had told me [...] how to use Git and we'll just do a quick Git walk-through that then I would have started looking at Git a little bit earlier.”

In academic settings, the mentors guided students in the choice of the programming tools but also gave support and feedback. In her SAS courses, Janet had “coding assignments, largely.... we would submit our code for review.” While at the workplace, even workshops for learning programming tools did not provide that level of feedback according to Enid:

Whereas like in the workshops and like the online trainings that we have at work, you don't get that feedback. You don't... If you have questions you may or may not be able to ask, Hey, why are you doing that? Why can't you do it this way? Or, you know, you might not have the time or the availability to like ask whoever's teaching that little two day training workshop, you know, ‘I don't understand.’

Learning new tools at the workplace seemed to be facilitated by the diversity of tools provided in academic settings. In addition, undergraduate courses seemed to create differences between the experiences of participants. John, Edwards, and Janet mentioned learning programming through their undergraduate degrees which made their transition continuous. Edwards learned a variety of tools: “The first one was SAS Enterprise Guide. For software. Really elementary stuff. The second one, we use SAS, full fat, we didn't use SAS Enterprise Guide. So we had to actually code.” He was also exposed to JMP and Stata which he used for classes in econometrics. Janet was introduced to programming with “a computer science course that was in MATLAB.” John started using R “through my introductory statistics course and I'm really glad that I was exposed to R in undergraduate, because that really made the transition so much smoother when I was doing more advanced R in graduate school.” Indeed:

Once I took an R programming class in grad school that got a lot more complicated. With like data visualizations and how do we optimize simulations in R and, let's see, how do we make, how do we incorporate functional programming in R and so I felt like undergrad gave me a good foundation and then learning the more advanced stuff later on was a lot more easier because of it.

On the contrary, Gertrude felt she was at a disadvantage because she did not have any experience of programming in her undergraduate studies:

In my bachelor's program though, I, you know, I took statistics classes at like six different times and then in none of those classes did I learn any code. I had never written Stata before I graduated college and I was an econ major. I had used SPSS like once, as a very near psych major.

She added: “We didn't even do any programming in Excel in my entire math degree. Just like no programming whatsoever.”

Logic of Programming. As mentioned for the division of labor, participants felt that part of their responsibilities was to be a programmer. The summer before Gertrude enrolled in her master's degree, she “took like C++ class at my community college just to know what [programming] is, how code works.” Indeed, she thought that offering opportunities for programming in undergraduate degrees is crucial: “even if it's SAS or R or Stata or something at the undergrad level.” Florence started to learn programming during her undergraduate studies, mainly with R and had some experience with MATLAB. Her graduate degree “started with R but quickly left it behind for SAS.” However, she still felt that something was missing and at the workplace she would “go to my little brother, he's very helpful because he thinks like a programmer.” In general, she suggested that instead of redundant courses in probability theory, what “would have been more useful is to have like a CS 101 type course that took us through the basics of programming” with “specific things that are more infrastructure like.” She concluded with:

The more technology driven world that we live in, having a more solid coding background would've been more helpful than having additional probability. In a perfect world you'd have both. But unfortunately with an 18 month program that wasn't going to happen.

Stella also wished she had been taught the logic of programming because “I was never taught that in school and I, it would be great that, you know, there would be classes like that for students like specifically just on programming practices.” To help organize her programming, she “use[d] a lot of flow charts and I, you know, I put out a plan. You know, I ask myself, what do I need to do, what do I want as the outcome.” As an example, John had experience programming with a “computer science class in undergrad.” This course in Java was very valuable when he “learned R because a lot of the logic that you learn in computer science IF statements and loops and all of that. The transition to learn R is so much easier because I knew the foundational principles of programming.”

To understand the logic of programming, Enid kept referring to the course she took in data management. She said:

That was absolutely the best thing that ever happened to me because it really gave me an understanding of basics, the logic behind the coding program, you know, because like us being math, mathy, people in general, like we're very logically minded anyways. So I feel like it just makes sense that we'd kind of naturally gravitate to understanding, you know, these coding programs.

She suggested “to introduce SAS or R as like an introductory kind of data management, just here's how you do the basics with these programs, I think would really set students up for like success.”

Specific Programs. Participants mentioned specific tools and referred to specific courses that supported their learning in academic settings. The course about data management that Enid just mentioned was to prepare the programmer certification for SAS Base. Janet mentioned a similar course in her master’s degree: “I took a specific SAS class part two, because there was part one and part two. And that geared us up for

taking the SAS exam.” They both found this course to be valuable for their understanding of programming.

Edwards found that R was a better programming tool to learn because “it forces you to interact more, know more about what's going on than SAS does, at least in my opinion.” Comparing R to SAS, Edwards concluded that “R is way better” mainly because “you have to understand more. You have to think about stuff more. It's not as easy as, as SAS in the sense that it's case sensitive.” And that is why he believed that “using R has made me better. Also because R is capable of more than SAS. You can do more in R.”

As a statistician, David said that “R has been by far the most useful” tool he learned from academic settings. David thought that the opportunity to take courses on specific tools was a recent opportunity offered by degrees in statistics because at the time he graduated there was no course focusing on specific tools: “As opposed to, I know a lot of students now actually get courses and how to use R and I remember there was no such thing when I was in school.” Stella had chosen to prioritize taking courses on theory and application rather than programming in R or SAS, but she wished she had taken more classes about programming:

But I, now looking back, I do wish that I took a semester to learn how to, you know, better program in R because that you know, I'm still I'm learning, you know, something new about R every day and they, I just think that, you know, once you learn one programming language, it's beneficial in learning other languages, but also it helps you think logically and di-, differently.

Enid also mentioned a sequence of classes covering SPSS to do statistical analyses. She really appreciated that tool because “it was very point and click and I was like, I can do this. I just need to understand the statistical analyses behind it.” Indeed, she suggested

that having “more educational, like statistical software is helpful. Um, even if it's something more touch-and-go based like JMP, Minitab, Tableau, um, or SPSS.” Gertrude added that doing calculations by hand help understand the concept and she appreciated to have the opportunity to do so, “but after a certain point, the code is helpful.”

Boundaries in the Community for Programming. Between academic settings and the workplace, the community was setting up the rules, with the different types of programming tools to use, and required statisticians to assume different roles like data manager if none were included on the team.

Rules. At the workplace, Florence had to use R for her projects because her collaborators “only code in R, they don't code in SQL and sharing project files with them, it would have been incomprehensible to them in SQL.” To be able to collaborate with her coworkers, Gertrude needed to be fluent in several programming languages. She worked with “five people on my team, two of us use R, and 3 use Python. So I have to read their code.” Similarly, Janet had to use R because that was the program required “coming into this job and they were like, yeah, we use R.” On the contrary, Stella was given the option to choose which program she wanted to use and she picked R because it was free.

In academic settings, Edwards thought that he didn't have any experience in Python because the community was not using this tool: “none of our professors used Python at the time. Half are pretty heavy R users and the rest use nothing but SAS.”

Training. Enid and John received training opportunities by attending workshops and forums organized by the community at the workplace. John described the forums as “one thing about my company that I really appreciate.” At Enid's workplace they offered “a very large scope of classes and you can pick them” and for example “they have SAS

skills kind of broken down into different components.” She described them as “two- or three-day long training sessions. Um, but yeah, so since this does provide these training workshops on things like Python, R.” The workplace is “very encouraging of, if you want to take trainings that are off campus,” supporting their employees to keep learning.

Cross-functional Teams. Some teams at the workplace included data managers which facilitated the transition since statisticians did not have to assume that role for programming. For example, Stella did not have to program for data management before the data administrator quit unexpectedly nor did Florence unless the database manager was away. Janet also mentioned having a coworker who was specialized in data management and took care of programming for these tasks on her projects.

Florence was challenged by the practice of programming because she was the only onsite, fulltime statistician at the workplace, with the other two statisticians in her team working remotely. In addition, “both my senior statisticians have PhDs in statistics. Um, very little coding background.” She said that “no one else worked on site and also coded.” Even though she had learned R in academic settings and continued to use R at the workplace, as she “ran into a coding issue, it was not easy to get assistance quickly.” She used online resources such as Stack Overflow or Google, or even contacted her mentors in academic settings, going “back to my professors once or twice.” Her experience of the transition was considered as detailing since she had acquired the appropriate tools but still encountered some boundaries with the community, feeling isolated.

Transition of Programming. Participants reported spending more time on programming than they expected before the transition. The practice of programming was

expanded from focusing solely on data analysis in academic settings to completing different tasks at the workplace, such as data management. Besides a boundary in the focus of the activity, junior statisticians had to deal with technology requirements, cost and time constraints, or limitations, which affected the types of program to use. As a result, all participants reported to diversify their programming tools at the workplace, through training courses offered by the community or online, or using tools acquired in academic settings. The transition of the practice of programming was considered either as continuous or detailing because all participants had acquired some tools in academic settings. The distinction between detailing and continuous experiences was related to courses in statistical programming at the undergraduate level and courses in computer science, both facilitating the boundary crossing because participants had developed a logic of programming. Boundaries in the division of labor also challenged the practice of programming with choosing the appropriate programming tools which was only included in academic settings for two participants. Junior statisticians did not recognize the role of the mentor at the workplace as particularly encouraging for this practice as they mentioned self-teaching new tools. However, the community had a greater impact on boundary crossing in terms of setting the rules for which program to use at the workplace, providing support with members in the team who had a programming background, or by offering training onsite. Mentors recommended to offer opportunities to perform different tasks with programming in academic settings, such as data management. Indeed, as discussed in this section, participants used programming for data management which is the next practice I will discuss.

Coordination of Data Management

Collecting, storing, and processing data before analysis define the practice of data management which englobe important steps for a statistical study. Participants reported spending a lot of time engaging in activities such as extracting data out of databases, cleaning data, or handling missing data. All participants mentioned this practice as well as both mentors, with 67% of the quotes referring to the workplace (see Table 14). Indeed, the practice of data management was developed at the workplace for most participants. The rules differed from academic settings in terms of dealing with messy data, big data, and data security at the workplace. Time also constrained the learning of the practice and added to the boundaries in the rules. To hold the division of labor and address these issues, participants used tools developed in academic settings, principally programming (see Table 18). Gertrude, Enid, and John mentioned courses in academic settings specifically geared towards data management which facilitated the transition. Indeed, the main difference between continuous and not continuous boundary crossing pertained to specific tools acquired in academic settings. Furthermore, the degree to which participants assumed the division of labor also made the distinction between continuous and discontinuous experiences. The mentors both considered that junior statisticians experience discontinuity with the practice of data management and offered opportunities to learn at the workplace with training on the database system.

Boundaries in the Rules for Data Management. Students are usually given clean data in academic settings whereas at the workplace statisticians deal with messy data or need to manipulate big data. They also ensure the security and ethics of the data and are constrained by time.

Table 18*Important Elements for the Coordination of Data Management*

Participant	Academic Settings	Workplace	Transition
Edwards	General programming in R	Handle missing data	Discontinuous
Florence	General programming in R	Clean data, deal with big data, extract data from databases, sub for data manager	Discontinuous
Stella	Programming with SAS and SQL queries	Clean data, extract data from databases, sub for data manager	Detailing
Gertrude	Course about missing data	Handle missing data and extract data from cloud databases	Detailing
Enid	Course in data management with SAS	Clean data and deal with big data	Continuous
John	Introductory course in statistics with data management assignments	Clean data	Continuous
Thomas	Further his skills in programming based on past experience	Clean data	Continuous
Janet		Data manager on team	Continuous
Simone (mentor)	Offer opportunities by including real data	Offer training on the database system and include data manager on team	Observed discontinuous
David (mentor)		Offer training on the database system	Observed discontinuous

Messy data. As Florence entered the workplace, she encountered a boundary in the rules as she noticed she was dealing with data that had not been cleaned:

Almost always in school, they just kind of give you these perfect data sets and that's great because it means that you learn the statistical topic pretty fast. I mean without having to go through four levels of [nonsense]. But it's really unrealistic once you get to the real world.

David confirmed that in academic settings, his experience was that:

I never had to clean up data, you know, the data is always generally pretty clean, ready to be used, ready to input into the analysis, even if there is cleaning it's almost structured so well that you just have to do one or two steps and it's completely cleaned up. I never really had to think about what the data meant.

Indeed, Gertrude noticed that her master's program in Applied Statistics intended to give her experience in "dealing with like real data in an applied setting [because] you basically never have data that's already cleaned." The rules differ from academic settings in terms of the quality of the data for Enid as well, since at the workplace "sometimes I'll get really nice datasets that I don't have to do any cleaning up or a maintenance with and sometimes they're just nasty, ugly and all kinds of messy." For example, Janet recalled her experience consulting with some clients:

[The clients] didn't clean any of the data. So people were coded as male and female differently throughout the data set, they didn't know how many people they actually had because there were four different ID variables that varied across data sets, people would have different dates of birth across data sets. There were different, like it was terrible, just a messy data set.

Stella said that "data collection can just be very frustrating." For one of her projects, she had to gather data "from so many different sources and it's in multiple formats." She even "sometimes [had] to physically drive out with a thumb drive and download data on it and bring it back to my office." Stella learned to be patient because "when you have to get data from outside sources, it can take a long time."

In academic settings, Enid and John followed the rules by completing assignments to practice data management. In an introductory statistics course for his undergraduate degree, John "had weekly lab projects where we did data cleaning." At the graduate level, Enid had a course in data management and she "had this whole semester's worth at least where, you know, I was doing this day in, day out, like quizzes once, twice a week."

Indeed, as a mentor in academic settings, Simone noticed that students only "get to clean up data sometimes" and they "don't really have to do a lot of data management to clean data, and just focus on the analytical piece." She believed it to be an issue and said

that “one problem is the data [students] get in the class is too clean and they have no experience on how to use raw data, or then dirty data and things like that.” However, she mentioned a boundary in establishing such rules because students “need stronger programmable skills to be able to do all that.”

Big Data. The amount of data impacted the way to perform data management. For example, Enid was “dealing with like really big data sets. So [the program] takes a long time to run.” Florence also talked about how the amount of data constrained the way to perform computations:

Big data are specifically gonna go forward like genetics, data, terabytes, large. And at this point in time, the way that my group does it is that most of us are doing the computation on our personal computers and personal computers are great but don't quite have the muscle to chuck through terabytes of data, nor should they be stored on there. Cause I would just take up so much space and the computers the [workplace] provides also don't have the capability. The best way to do that would be some sort of cloud. And my education didn't at all prepare me to figure out how to do coding within the cloud environment.

Dealing with big data created boundaries and participants needed to learn new tools to store data in a cloud environment for instance.

Data Security. Florence switched to cloud databases to deal with big data but also to manage data storage securely. At her workplace, data was “literally stored on hard drives in people's desks, which there are no backups and that should just concern everyone cause how expensive genetics data is.”

Ethics. Boundaries introduced by the confidentiality of data also presented ethical issues. In the context of storing data, Florence referred to an anecdote where she researched how to address ethical issues about unprotected data and the rules that could guarantee the confidentiality of the data:

A guy lost his computer at the [workplace] and it had a ton of patient information on it. It wasn't locked and a lot of that patient information ended up being stolen. So we kind of studied that case and the problems that he had in terms of delaying recording it and the issues there, and also just the ethics of having unprotected data, and online and identifiable data on your personal computer.

Having an ethics course in academic settings gave Florence some experience in being attentive to ethical issues.

Time. Participants were constrained by time to perform data management at the workplace. Enid needed to optimize her time and “Sometimes you're just like, ‘Nope, I've got to do whatever is quickest and easiest to get this data ready so that I can get it done for the project that I need to do.’ ” She mentioned balancing between tools such as Excel and SAS to complete her tasks faster.

Boundaries in the Division of Labor for Data Management. As statisticians transitioned to the workplace, some participants realized that data management became one of their main responsibilities, taking a considerable portion of their time. The role of the mentor in academic settings and at the workplace supported participants in crossing the boundaries in the division of labor.

Cleaning Data. Since most participants reported not being taught data management in academic settings, it was surprising to them how much time they had to dedicate to it at the workplace. Florence realized that she “spent much more time data cleaning than anyone told me.” At first, she did not “realize that it was acceptable for me to take my time to do [data cleaning], and not that I was being too slow about it. That was a big thing.” After she had a mentor helping her navigate the transition at the workplace, she learned from her that “75% of your time will just be cleaning data. Expect that, that is what most of your time you'll be doing.” The role of the mentor at the workplace helped

Florence manage the expectations for the division of labor concerning data cleaning. Indeed, her mentor made her realize that it was part of her responsibilities and that she had to dedicate time to it. Had her mentor intervened earlier in the transition, Florence may not have had a discontinuous experience. The role of the mentor in academic settings was also important for Enid who appreciated to “have somebody who can give me like almost immediate feedback.”

Gertrude also noticed that data cleaning was an important part of her role, she said: “This is the reality. It's 90% data cleaning. I don't think anybody does a very good job of explaining that. There are so many disappointed statisticians in their first year on the job.” However, Gertrude expected to hold the division of labor for data management because she had some experience with missing data in academic settings. She was getting frustrated at some point because she was spending so much time on data management, but her professors wanted her to learn that it was necessary: “I'm spending all my time cleaning the data, no time doing stats, and they were like: ‘that gets the point.’ ” John also considered being in charge of data cleaning at the workplace as “Sometimes the data that I get isn't clean so I do have to manually clean it.” However, he mentioned that data cleaning take “maybe like 10% of my job.” His estimates for spending time on data management were lower than for Gertrude and Florence but he still considered it as one of his main responsibilities.

Gertrude found that at the beginning of her master's program, professors were providing data that “was clean and perfect.” However, “they did make an effort to take the training wheels off as we went through, which I was frustrated with at the time.” With the support of her mentors in academic settings, Gertrude was able to learn progressively

how to clean data. Yet, she wished she had learned how to deal with cloud databases in academic settings, because it would have been “a little bit faster with the help of an instructor versus just try and do it, you know, on your own.” In fact, Gertrude was in charge of handling missing data at the workplace and “getting the right data out of the database.”

Extracting Data. For some participants, performing data management went as far as assuming the role of a data manager. After the database administrator quit at her workplace, the division of labor was passed down to Stella and she had to figure out how to get the data out of the database. She taught herself SQL and “had to do that for a while but fortunately now we have a DBA so I handed everything over to him.” Similarly, Florence used SQL “to pull out a huge chunk of data from what was the SQL database.” She had to research how to move the databases to the cloud because people at her workplace “want[ed] to just switch to the cloud.” She shared the division of labor with a database manager but said that:

Not every statistician who goes out, will have a database manager or a data engineer who's storing the data and learning a little bit more about how to store it and how to store it securely and have access to it easily would have been helpful.

Florence was “asked to be the, the backup database manager for our database manager while she was gone.” As a mentor at the workplace, David made sure to have his junior statisticians “started on really learning our database and our systems.”

Boundaries in the Tools for Data Management. As participants learned how to perform data management, they acquired new tools to complete various tasks. Indeed, David thought that statisticians should be trained in “understanding the data. Knowing how it's collected, how it's gathered, how it's manipulated. How it goes from being out in

the open to actually going on your computer screen” because “you really can't do any sort of real interpretation as to what's going on without that.”

Programs. The main tools mentioned by participants to perform data management involved programming tools. This practice was described in depth in the previous section, but I would like to highlight some aspects of programming directly related to data management. Florence thought that programming for “the purpose of doing statistics,” with which she had experience in academic settings, was different than programming for “doing data cleaning.” Indeed, Enid really appreciated having a course about data management in SAS with “proc means, proc, proc univariate” because this course “gave me an understanding of basics, the logic behind the coding program” even though she had already learned SAS for statistical analysis. She said this course was an advantage compared to the experience of one of her coworkers who had no prior experience in data management and “I attribute a lot of my success I think to that class where we just talked about data management.”

According to Gertrude, “for my job I'm in now, the only useful thing I got out of my degree was that, that I could write code in R and it's mostly, you know, data cleaning mode.” John started to learn about data cleaning with R an “introductory statistics when we did our programming.” On the contrary, Thomas encountered a boundary during his transition to the workplace after his undergraduate degree. He was doing “the dirty work of research” and had to start “learning R just for data cleaning.... data wrangling, data mining.” In fact, the rule at this workplace was to use Excel but he found it to not be adapted for data management and that is why he switched from Excel to R. At the time of the interview, he was working at two different workplaces, and he did the “research dirty

work” at both places. However, considering the workplace he was principally working at, Thomas had previously acquired the proper tools for data management, therefore his boundary crossing was regarded as continuous.

Databases. As mentioned in the division of labor, participants found themselves responsible to pull data out of databases and do some data manipulation. They had to learn some tools and strategies to cross that boundary. Florence clearly said that she was not prepared to deal with cloud environments and large databases: “my education didn't at all prepare me to figure out how to do coding within the cloud environment,” to store data securely and access it easily. In fact, she revealed that “my education wasn't technologically advanced enough to help me with stuff like that.” In particular, she wished she had learned “specific things that are more infrastructure like.” She mentioned learning SQL to:

Pull out a huge chunk of data from what was the SQL database and that I don't use for data cleaning. Um, it's pretty inflexible. What it's really useful for, for me at least, is to identify the chunk of data that's relevant to what I want to do.

Florence also learned about cloud computing and “spent the first couple of months doing a bunch of research on how to move things to the cloud and which of the various platforms is going to be the best.” She talked about flat files versus functional databases and having only experience with the former. She described:

Flat data is like an excel spreadsheet. There's only one Excel spreadsheet and all the data is right there. Even if someone is consistently male through all 17 visits that that person has, they will still have 17 lines where it says that they are male versus a functional database where there's one table that says visible information in a different table that says demographic type information and then some sort of linkage system across the two of them.

Performing data management included thinking logically for Florence and figuring out the way that is “computationally less intensive ... since we're talking [about] a couple

millions of patients was so incredibly difficult.” She used “A really big whiteboard” and “fake patients are written down on my board to see if, if they went through this path, then sort of logic, how would they, where would they end up and did they end up in the right spot?”

Gertrude did not learn about cloud databases in academic settings either and she learned how to “make sure that the data pipelines are working” and “pull data out of the database to make a one-time report” using Bash, a command processor. Indeed, she said that cloud databases are more and more common “since databases are becoming more accessible, thus working only with like Excel files and CSV files, it doesn't make sense.” As databases are getting bigger, that creates a boundary in the tools between academic settings and the workplace as, “you're not going to be working with Excel files, you are going to be pulling out of cloud databases.” She characterized the skills of “data pulling” as being important and being part of “the things that go into a job that’s not like the statistics part of the job.”

Similar to Florence, Stella relied on the programming tools she acquired through her master’s in statistics to learn SQL, mastering “how to append tables, how to delete tables, how to join tables together, how to extract the data.” Because she had experience using SQL queries within her SAS code, her transition to SQL made her experience detailing.

As mentors at the workplace, Simone and David supported junior statisticians to learn how to access the database at the very beginning of their transition. Simone offered training to junior statisticians “on the data management system. You know how to get access to it, how to find data within there. Um, and they would have an understanding of

how data is edited and cleaned within that system.” However, at Simone’s workplace, statisticians were not in charge of data management, they just needed to understand the structure of the data. As his employees entered the workplace, David started by assigning them “a simple [project]. It was really more so they could learn the structure of the data” and “get them started on really learning our database and our systems.”

Data Cleaning. As we just mentioned the importance of extracting data, Gertrude said that “it doesn't matter if you can build all these models and your data is really good if you can't get it out of the database in the first place” and added “when you're a statistician it doesn't matter what kind of model you have, if your data is trash.” As a mentor in academic settings, Simone gave the opportunity to her students to experience “data cleaning and data manipulation and data re-coding” with real data, using SQL functions. However, not all participants reported to have these opportunities in academic settings. Florence had some experience with data cleaning while she worked on her thesis: “we're a little more prepared because we have to deal with it then, but we didn't have to deal with it in classes.” Indeed, Florence wished she had more experience in data cleaning before she transitioned to the workplace. Participants shared some strategies they learned at the workplace to cross boundaries for data cleaning, checking for abnormal values, for example.

As “data set collection, data cleaning, data mining” and “manipulating, constructing data sets for analysis” were Thomas’ responsibilities, he had to get “the raw data to the dataset where analysis could be done. There was a number of steps that had to happen there and kind of issues along the way.” To perform data cleaning, Florence described her strategy:

Start with taking the data from whatever format it is and putting it in R, I save an original copy of the data in its original form and then I don't touch it again just in case I mess up the data somehow I don't want to corrupt the original. And then I go through a series of, of taking it, um, your original data, once you read it in, taking it to data that's... I'm aware of where the missingness is and then kind of standardizing it so that anything that says it's a character is, you know, actually a character, anything that should be between 0 and 100 is actually between 0 and 100. And I call it, taking it to technically correct in terms of that, that fits within the parameters. And then step back, and I actually start looking at the form of the data itself and see what sort of trends are there, which side the skew is on for anything that's continuous and which levels at which for anything that's dichotomist. So that's kind of, it starts to go into exploratory data analysis near the end of it. But that's also where I'm doing a bunch of transformation of the data if it requires it. And so it's important to know the shape of the data before I commit to transforming out. And that I do any sort of transformation that kind of tails into what the analysis plan is if that's what we actually needed to do. So that last part it's always done in conjunction with people who are more familiar with the statistical analysis plan and what's more common in the field cause sometimes you just can't transform it. It, it won't be interpretable to anyone in the field if you do that.

To summarize, Florence's strategy was first to save a copy of the original data, check for abnormal values and that the variables were defined as they intended to be, and apply transformations if needed. For the last step, she mentioned working with domain experts to meet the requirements of the field. Part of Enid's strategy was also to check for data values that "look abnormal like at a place or like really high or really too low, then we want to be able to let upper management know about it" while she did data cleaning in real time during data collection to sustain data quality.

Messy Data. Edwards handled messy data and he referred to missing data in particular. He noticed that "the sample size might vary from model to model based on who has at least one missing data point for at least one variable." At the workplace, he was "learning, making up my mind in terms of how I handle certain things. How do I handle missing data." Indeed, he wished he had learned "how to handle missing data" in academic settings because his program "didn't put that much effort in that." To deal with

missing data, he learned by reading papers he found on the internet and he “used mice. An R package. But it's complicated.” He realized that “variable Selection, data imputation and validation, these are things that, never touched on again, never thought about and most of my colleagues here have never thought about that.” He also mentioned that there is no definitive correct answer and he was “still making up my mind on that stuff. But then again, so is rest of the world.” Lacking tools from academic settings, Edwards was considered to experience a discontinuous transition to the workplace.

Gertrude also concluded that “there's no one right answer” for how to deal with missing data and she actually took a half-semester course about it as part of her master's degree. She learned “what happens if you just decide to ignore the missing data” for example. She explored techniques with “mean imputation or median imputation” and “regression analysis.” She also learned to question the reason why “your data might be missing, if you have any insight like ... if it's a survey like either someone just choose not to report, it is more likely than not um, and then you have to account for that.” Simone confirmed that her role as a mentor in academic settings was to offer her students the opportunity to learn data management by giving them real data in her course for advanced epidemiology.

Boundaries in the Community for Data Management. The boundaries between the community in each activity system were created by the lack of rules for data management in academic settings and boundary crossing was facilitated in terms of the division of labor depending on the background of other members on the team.

Rules. Professors and students did not typically get involved with the practice of data management in academic settings, except for Gertrude, Enid, and John, while all

participants mentioned the practice at the workplace. As part of her degree requirements, Gertrude had to take a course specifically designed to learn about missing data. She found this course to be “pretty unique” and “very important,” in fact being “the most valuable class.” In addition, data management was also included across her degree program. On the other hand, Enid focused on data management through one specific course that made her think about the structure of the data while developing her skills in programming. John also learned the importance of data management but from an early stage of his educational journey, starting to clean data in his first undergraduate course in statistics. These types of courses made Enid’s and John’s experiences continuous, however, Gertrude had to learn more to extract data from databases and therefore, she had a detailing experience as she transitioned to the workplace.

Cross-functional Teams. Some participants referred to having a database or data management specialist on their team which facilitated the boundary crossing by taking off the responsibility from the statistician, as it was the case for Janet and junior statisticians on Simone’s team. However, not all workplaces considered that statisticians should assume the role of data manager. Gertrude did not seem to think that “people working in industry don’t know or care to handle [missing data].” Concerning one of the projects Janet worked on, she had a “colleague who worked primarily with database management” take over building the database “because my supervisor decided that it was not an effective use of my time if they weren’t going to pay for me to do the cleaning.” Although Simone had “a data manager in my team who’s got the primary responsibility” and thought that statisticians are “not to do a data management,” she made sure that junior statisticians were trained to access the database in case they needed to.

Transition of Data Management. As junior statisticians enter the workplace, they realized that data management accounted for a large share of their responsibilities. Indeed, this is the only practice for which participants estimated the amount of time they spent on this practice compared to other activities. Some participants reported engaging principally in data cleaning, but other participants also reported being in charge of handling messy data and pulling data out of databases. Similar to the practice of programming, the rules created boundaries at the workplace and tools in academic settings facilitated boundary crossing. To cross boundaries between academic settings and the workplace, participants experienced different levels of transition, and the division of labor within the community at the workplace and the tools acquired in academic settings distinguished between the experiences to cross boundaries. Tools in academic settings, such as data management courses or opportunities to deal with messy data, facilitated a continuous transition to the workplace, while the inclusion of data managers on the team at the workplace relieved the division of labor for junior statisticians. Indeed, mentors offered support on the database system as an introduction to the practice of data management since they observed that junior statisticians were not familiar with how to work with databases as they entered the workplace. Participants were not particularly explicit on how they learned the practice of data management at the workplace, but they rather focused on how challenging learning this practice was, yet, it was necessary to prepare the ground for data analysis, which I will discuss next.

Coordination of Analysis

Unlike data management, some participants reported spending less time on analysis than they expected from the perspective in academic settings, before entering the

workplace. For the practice of analysis, participants engaged in activities such as performing analysis, choosing the appropriate technique, interpreting results, and checking assumptions for validation. All junior statisticians mentioned this practice as well as both mentors, with 58% of the quotes referring to the workplace (see Table 14), which was one of the lowest percentages compared to other practices. Figure 14 also emphasized that the proportion of quotes referring to tools for analysis in academic settings was the greatest in comparison to other practices. Indeed, participants indicated developing many tools in academic settings with various courses in theory, applications, and data science, for example (see Table 19).

Table 19

Important Elements for the Coordination of Analysis

Participant	Academic settings	Workplace	Transition
Florence	Advanced mathematics machine learning	Manage validity issues without a mentor	Detailing
John	Mathematics, survival analysis, machine learning	Learn Bayesian statistics	Detailing
Stella	Mathematics, introduction to probability	Use trial and error to choose the appropriate technique	Detailing
Edwards	Mathematics, theoretical statistics, machine learning	Develop machine learning and focus on theory for interpretation	Continuous
Gertrude	Mathematics, statistical inference, data science	Wish to spend more time on analysis	Continuous
Janet	Machine learning	Review currently used methods	Continuous
Thomas	Statistical inference, Bayesian statistics, machine learning	Use machine learning with mentor	Continuous
Enid	Mathematics	Choose the appropriate technique according to time	Continuous
Simone (mentor)	Encourage students to focus on interpretation	Assign junior statisticians analysis tasks	Observed continuous
David (mentor)		Encourage junior statisticians to focus on interpretation	Observed detailing

Because participants reported having some experience of this practice in academic settings, the transition to the workplace was either qualified as continuous or detailing. Similar to design practices, boundaries encountered by junior statisticians concerned the division of labor. Stella, John, and Florence confronted boundaries such as the choice of the technique for analysis, using a new field of theory, or managing the validity of the model with the supervision of a mentor. To support learning, Simone encouraged students to focus on the rules, by interpreting results with implications for the domain, while David echoed that support at the workplace with interpretations for decision making. The community facilitated the transition by offering training or having members with different backgrounds on a team, but also introduced some boundaries by setting the rules and requiring specific techniques.

Boundaries in the Division of Labor for Analysis. At the workplace, participants encountered boundaries in the division of labor by focusing less on performing analyses and more on choosing the appropriate technique, interpreting results, or checking assumptions.

Performing Analyses. As a mentor at the workplace, Simone thought that statisticians had learned the practice of doing statistical analysis in academic settings best compared to other practices since “they focus on the analytical piece.” Indeed, Florence’s mentor determined that “typical expectations for recent biostats graduates was competence and exploratory data analysis and like standard forms of regression.” One of her responsibilities at the workplace was to do analysis, “some of it is exploratory data analysis and some of it is actual analysis.” Thomas shared the division of labor with his mentor whose role was to be “helping me with like actual analyses and projects and

statistical applications that I'm working on.” At the workplace, Stella was in charge of setting up analyses to keep track of test scores over time and Edwards shared that about half of the projects he worked on were to do analysis. Gertrude said:

What I mostly do is these models they break every week, something goes wrong, so I spent a lot of my time just trying to make sure that every week they are working and they're fixed and there a little bit better.

Even though Gertrude was “always trying to improve these models or find new methods,” she considered she was not spending enough time doing analysis. For example, she mentioned a new project for which “I'm trying to add one new model, the pricing thing and it's very exciting working with you know, network analysis. So I'm thrilled about that. I can just do some kind of analysis again.” Similarly, John encountered a boundary in the division of labor between academic settings and the workplace as he realized that “I don't even do statistics at my job anymore. Why did we study so hard for a statistics graduate degree when we're really not using our statistics knowledge that we've gained?”

Choosing Appropriate Techniques. Participants were responsible to choose the appropriate technique to fit a specific data at the workplace while in academic settings the data was chosen to illustrate a specific technique. For example, David said that the structure of courses was usually “based off of a technique and not really based off of understanding the data and then trying to pick the technique to fit the data.” Indeed, as a mentor at the workplace, David started to give different projects to his junior statisticians guiding them with which technique to use and given them exposure for professional development and recognition:

Even though some of the simpler projects, there was one where they were all just doing t-test, was highly seen by higher up. So it had as much value as one of the more technical projects that was only really seen by me and one other person.

He guided junior statisticians on what appropriate techniques to use and for what audience to emphasize on the methodology with analysis techniques.

After one year at the workplace, Janet also shared that she had “learned a lot of knowledge and skills around applying those techniques and when and how the typical sort of analyses that we do for a given type of experiment.” Indeed, she learned different techniques by being in charge of:

Triaging our current methods. So that's taking up most of my time but ideally once we get past that I'd like to apply some of the methods to the questions and they can give our clients, the opportunity to see that, that's possible, and use those kind of techniques to answer some of their research questions.

Janet was also responsible for doing research, reading about particular techniques to report to her supervisor who understood the statistics, and presenting to her colleagues or to the principal investigators. As Stella transitioned to the workplace, she also encountered a boundary to choose the appropriate method and her experience was considered as detailing because she still needed more practice after being at the workplace for two years:

It's still early in my career that I'm still doing trial and error. I am always amazed at, you know, people have been statisticians for years and you talk about a project and oh yeah you can use blah, blah, blah, you know, and say, I wouldn't even, I wouldn't think that off the top of my head.

However, Thomas experienced how to choose the appropriate technique in academic settings with his “master's where it was like you're given a data set and given some questions to answer. And you had to know how to apply the right tool.”

Interpreting Results. Both mentors gave a great deal of importance on the ability for statisticians to interpret results. David thought that at the workplace, statistics inform making decisions and that has usually “been a more of a learning on the job type of thing.” As he was recruiting junior statisticians to build a team, he was looking for skills to interpret results:

I wanted people that knew the techniques and knew how to do statistical analysis. But I was mostly looking, when I was asking a lot of those questions, you know, I was less interested in the technical aspect, but more of the interpretive aspect.

He added that he “wanted people that were creative and could understand what was going on.” David encountered boundaries himself as a statistician during the transition since he said “I mean to be honest, I feel like most of the interpretation that I do and use, let's say I learned on the job.” To support the need to interpret the results from junior statisticians, David “would always ask them ten different questions that really, even if I knew what the answer was going to be, I would make them go in and really understand what was going on in the data.” Simone did not identify many opportunities in academic settings to develop this aspect of the practice, except if students did a thesis for which “they have to interpret the results.”

Checking Assumptions. At the workplace, Simone had junior statisticians on her team “do the statistical analysis and verify other people statistical analysis.” Gertrude also had to “spent a lot of time in this, learning like check your model assumptions, check all the things, like make sure your model is valid, your insights are valid.”

Florence had a detailing experience, mainly because she did not have a mentor at the workplace at the beginning of her experience to help her with validity. She had to

deal with researchers who were “pushing so hard to just get that analysis done right away” and she was tangled in a situation where:

I ended up catching one of those validity mistakes and I didn't know who to report it to cause I didn't have a senior statistician on the project to report it to and I didn't yet know how to speak up to the researchers.

Boundaries in the Rules for Analysis. Different aspects of analysis constrained learning, such as time, handling big data, ethics, and laying the ground for decision making. Another important boundary in the rules was that some workplaces required specific techniques.

Specific Technique. According to David, one of the boundaries between the rules in academic settings and the workplace was that in academia you “find a data source, and how does it fit into the technique and then at the end, okay, what does it mean.” The courses were focused on the implementation of the technique, while at the workplace participants focused on the interpretations of the results. In academic settings, a dataset was chosen to illustrate a technique while at the workplace a technique was chosen to explain the data. Indeed, Thomas mentioned that learning techniques for analysis in academic settings was “basically learning how to apply different models to specific data.”

Some workplaces required specific techniques for analysis. For example, Simone was particularly interested in recruiting junior statisticians “with the big data or data science, machine learning, um, background or experience.” John was required to do an internship as part of his master’s degree, which gave him an authentic experience of applying machine learning in context. However, he said that “I left understanding that it's really important and I could see the value in it. But it just wasn't the most interesting to me, naturally.” Contrastingly, he wished he had learned Bayesian statistics in academic

settings but that “was something that our program kind of pushed for PhD students to learn and the master’s students they didn’t really prioritize that a lot.” Because his team at the workplace required techniques involving Bayesian statistics, but this branches of statistics was not required in academic settings, John shared that “the theoretical statistics and applying those to the modeling questions have been kind of not continuous.”

Decision Making. David pointed at another boundary in the rules between academic settings and the workplace. In academic settings, doing an analysis was “more of putting an answer on paper so you can finish the homework assignment or finish the project.” However, at the workplace, David used results of an analysis to make informed decision: “in professional life, you know, I really have to figure out what it actually means because potentially millions and millions of dollars are depending on my decision as opposed to the difference between an A and an A minus.” In academic settings, he wished he had “been pounded with just how do you like... what does it mean, what does it mean, what does it mean, what are the implications?” While the rules in academic settings do not focus on the implications, he thought that “a lot of it I don’t know if you can actually do until you actually work because you actually don’t have the repercussions of saying something wrong.” As a mentor at the workplace, David provided support to junior statisticians by making them focus on the results and not on the analysis when they presented their work:

I think that was a little bit annoying for them. Because you put in potentially weeks or months of work into analysis into five bullet points. But in the end, that’s what matters, that’s what people are making decisions off.

To prepare future statisticians on this aspect of analysis, Simone required her students to “spend the majority of their time looking at the data, questioning data, understanding the

data and not going straight to modeling, which I think a lot of people focus on.”

Domain Knowledge. Simone recommended that teachers give “real data that they can get from industry, use and, give it to the students to use and teach the statistical methods on that. But give them, you know, a real opportunity to see, you know, a real-life problem.” The reason why she recommended to have real-life scenarios was for students to practice “on interpreting the models. You know, not just interpret what's the point I was going to make, but how does this relate, you know, to, to clinical, to clinical practice.” Indeed, the interpretations should be related to the domain knowledge in order to provide meaningful results. In academic settings, Gertrude felt that the analysis techniques she learned from social science applied to other domains:

In statistics the principles generally tend to apply in a similar way and I've actually found the social science to be sort of a nice framework to teach all kinds of things because I feel like dealing with, at least in my experience like experimental data in say a medical setting or biology setting, it's like a subset of social science.

On the contrary, Thomas developed tools in the domain of healthcare in academic settings while he worked with military strategy at the workplace. However, he said: “it's definitely a different domain than I'm used to, but it's still the same approach from those statistical analysis perspectives.” For his second position with a Law school, Thomas recognized that “some of the research is in like again domains that I'm not typically exposed to.” He gave one example for which he did:

Help someone with an event study based on stock data and he wanted to construct a portfolio and just it was a, is one of those projects where I needed to read a lot on the like, the underlying domain and like I've never really done a lot of financial analysis.

Overall, Thomas said that “when it gets down to statistical analyses, I feel a lot more comfortable, but the domains are very different.”

Similarly, Florence mentioned that she “needed to ask questions to help me understand the context of what I was doing so it was more helpful” when she was doing consultations with principal investigators. She sometimes looked for readings to learn about the context:

When I'm beginning a study so that I understand the why they're doing it. And also we like to pick up previous articles on whatever they're doing to see that their techniques are similar and what previous statisticians who have done those ones, what they've done for the statistical analysis.

However, Florence explained that “I don't love that because it means sometimes, we rely on techniques that are older fashioned rather than using newer ones.”

Ethics. At the workplace, Florence mentioned ethical issues encountered during the analysis when she worked with clients who asked to retrieve patients from the analysis. She noticed that she “already had two or three patients in my professional career where that's not an okay thing to do.” In particular, Florence learned in academic settings that dropping individuals from a study “cause they don't fit the trend” was not ethical. She said: “we were pointed out very explicitly in both our math classes but mostly in our ethics class in college that you can't do that because there are ethical implications to doing something like that.”

Time. Participants encountered boundaries in the rules with time at the workplace. For example, the time given to Florence to work on a project impacted the analysis process with validation. She referred to an anecdote:

We've had at least one case where someone came and said, I need this thing really fast. We need it, get it down for us right away. And I kind of just did it because they asked me to, and the results were really weird. So, I ended up going back and backtracking and it turned out that they were, the controls that they were using as controls, there was no possible way they could be controls. It was so far out of the sphere. So, we ended up, because of wanting to go back and validate the results, we ended up going back to them and saying, you have to repeat the experiment.

If Florence had had more time to work on the analysis, she would have been able to check for the assumptions at the beginning of the analysis.

As mentioned previously, the techniques for analysis mattered less than the implications at the workplace. David shared that one of his junior statisticians struggled in terms of picking the technique that was quickest rather than the best analysis technique:

He was very caught up in trying to do the best analysis possible, and with deadlines you can't always do that and sometimes even the best analysis, you don't really get the best results. But if you get some results, it's better than no results, otherwise the people that are making decisions are making bad decisions based on no evidence as opposed to at least we can give them a little guidance.

Big Data. The large amount of data required Florence to apply machine learning: with “a really large cohort of patients, just impossible to use anything that wasn't machine learning.” Therefore, she had to adapt the technique to be able to analyze big data.

Boundaries in the Tools for Analysis. All participants reported gaining tools in academic settings, even though they still had to keep learning. Indeed, David considered that he learned tools for analysis in academic settings:

The big thing my education was, it gave me a nice toolbox of techniques and theories to really prime myself off of when I actually had to use real world data. I think that, that was a big thing.

The tools mentioned by participants included a solid background in theory from academic settings while at the workplace, participants used new techniques, included topics in data science, domain knowledge, and validity.

Theory. Enid said that she wanted to “get a job as a statistician where I can apply all the math and all the statistics that I learned in like a real-world setting.” When prompted what prepared her for her role, she gave a list of courses, including “one of the

courses was called introduction to mathematical statistics. And so that was more of like the theoretical side of statistics.” However, she did not mention explicitly performing analysis at the workplace. Her experience was probably continuous since she did not refer to any challenge concerning that practice.

More specifically, Gertrude learned standards techniques of analysis such as “regression, multilevel modeling, GLM” as well as data science with “feature engineering,” “natural language processing and dealing with text” but she wished she had more “mathematically speaking, I guess a little bit of knowledge about how time series worked and um, a little bit of practice just on like forecasting kinds of problems might have been useful.” Similarly, John learned many tools in academic settings “but one thing I wish I kind of had that I didn't know that I really needed, leaving graduate school, was Bayesian statistics. My training is very frequentist.” As he transitioned to the workplace, “learning terms like priors and posterior distributions... was a steep jump initially and I'm still learning things about Bayesian statistics till this day.” He talked about his first exposure to Bayesian statistics with one project about “simulating a clinical trial that utilized a Bayesian model.” It took him some time to understand the model and program it and John concluded that it was “definitely one project that I really did not understand very well.” To learn about Bayesian statistics, John did some independent research and met with the doctoral level statistician in charge of the project.

Stella talked about a course in “introduction to probability and to date, that is the most challenging class I've ever taken into my life.” Even though “I’m gonna be honest, no I'm not using it at all,” she greatly appreciated this course for interpreting the results:

I don't use any type of probability methods in my career, but that course taught me how to think about a problem differently and how to approach it, and how to

really interpret and understand the results. Instead of just putting them in an R program, just using an R command and just looking at the results like this class really taught me how to thoroughly think about what the input says, what does it mean and how I can apply this and maybe use it for a different test. So just, just learning how to approach a problem differently. To me, that had the greatest impact on my career so far.

Theory was also valuable to Thomas. He gained a solid background in mathematics in his undergraduate studies as he took calculus, “up through calculus III” and he said that “the general tool belt that you get from taking a calculus course was very useful down the line.” He also took “a course of linear algebra, a statistics course in econometrics, and a few other quantitatively focused economics courses that they kind of prove useful later when I was studying statistics for a masters.” For his master’s degree, he mentioned courses such as “applied Bayesian analysis and probability” or “statistical inference.” The most useful concept he learned was “convex optimization” which “kind of helped me get from the real world to a mathematical understanding of how what we're really doing.” He explained:

When it came up to kind of training supervised learning classification and regression models that we, you know, is actually a convex optimization problem, it kind of came full circle and it really helped me understand, kind of making the leap between different machine learning algorithms and, and realize that from a high level they're all doing the same thing where they're trying to optimize some kind of objective function. And it also helped me understand some of the decision theory methods.

By taking a variety of courses, Edwards had a broad exposure to different techniques which gave him background to “learn something new and complex”:

From elementary stat courses that went over, you know, degrees of freedom, basic designs, what a chi square test is, normality all that kind of stuff and went to regression, advanced modeling, so modeling with splines, so machine learning stuff.

Edwards also referred to:

A two-course sequence in theoretical statistics. Everything else is more applied based, but we had a full year of applied - of a theoretical statistical inference using the book that I'm sure almost every university uses, the Casella & Berger book.

Even though Edwards valued theory, he appreciated having a degree in Applied Statistics. He illustrated the difference between his degree in Applied Statistics and his degree in Mathematics, with the example of a nonparametric course answering questions such as:

What do you do about confidence intervals? What do you do about your effect size? What's your effect size with a Wilcoxon rank sum? Things like that. Some questions, I would never have thought about coming out of the math department.

During the sorting task, Edwards remembered putting the practice of advanced mathematics “a lot higher than I think everybody else. Because if you don't know a technique and you now need to know a technique, you better have a foundation.”

To build his “toolbox,” David shared his experience in academic settings where he “took a whole slew of courses. Everything from basic probability and statistics theory to stochastic processes and time series and different types of regression model hierarchical modeling and... Bayesian analysis.” He also referred to a “stochastic processes course” as being unique and challenging but “a lot more rewarding when you have to work a little bit harder.”

Diverse Techniques. Having a background in theory from academic settings facilitated learning new techniques at the workplace. However, most participants mentioned they had to keep learning. Indeed, Stella kept “learning new methods, there's so many of them.” She took a variety of courses with “mathematical statistics, multivariate statistical analysis, time series analysis.” She said that for her position:

The most useful [techniques] I learned was just how to do a simple student t-test just testing between two means. Chi-squared analysis. I work with a lot of

categorical data. So I like to use that first to see a test of association and then from there I decide if I should use a little you know, logistic regression model or if I should pursue a linear regression. So like just the simple test set for us, but the t-test and the chi-square definitely ones that I use the most.

John learned new analysis techniques by attending forums at his workplace and in fact, that is “one thing about my company that I really appreciate is that we have a lot of statistics forums.” When starting on new projects, Florence needed to find information about techniques and she would look for resources online focusing on materials shared by “university names that I recognize just under the assumption that a, a school that has good public standing is more likely to have slides that are accurate.” Even with more experience, David reported that he still had to keep learning. For example, he mentioned the topic of machine learning which he learned “just on my own through reading papers” and from junior statisticians who had just learned about it in academic settings. When learning on his own, he searched for papers online and reached out to the statistical community:

I am also a member of the ASA. So I just was at the JSM conference this past July and, you know, just talking to a lot of people there and writing down everything that I see in the pamphlet and everything I look at just so I can take that home and just kind of go a little bit more in depth and know what to look at.

David also supported his junior statisticians to learn new techniques by attending conferences. “For example, one of them really wanted to go to this conference on data science so we sent them there.” Indeed, an important approach for the analysis that was mentioned by participant included topics from data science.

Data Science. Janet said that “the machine learning class was super beneficial as you might imagine, trendy” and Edwards thought that his machine learning course was in

the top 3 of most important courses, especially “in terms of marketability.” As Edwards entered the workplace, he realized that:

I wanted to learn more about machine learning and stuff like that. I found it very intriguing. It was something different than dividing sums of squares by other sums of squares. So, but right now I'm the go to person. And I've been here three years.

Janet had not yet used machine learning tools that she learned in academic because she was “triaging our current methods” at the time of the interview. However, she was hoping to implement machine learning in her projects after reviewing the types of techniques traditionally used at her workplace. Florence did not realize that she had learned about machine learning in academic settings. As she continued to learn about analysis techniques at the workplace, she took “summer courses for biostatistics” offered by a close-by university. Referring to a course she took about supervised learning, she said:

Having picked up a lot of the basics in grad school, I'm not sure the class was helpful. We just learned most of it in my master's degree, but no one had ever told us it was machine learning. Okay. I just assumed it was another statistical technique and not its own category.

Indeed, data science topics were included in some of the master’s degrees that participants obtained. For example, Thomas took “machine learning and deep learning classes” and learned tools to conduct analysis with “a lot of Python more focused on machine learning, deep learning, data science [courses].” John also used Python for machine learning, which he was first exposed to during an internship he did with a research lab over the summer between his undergraduate and graduate degree. The internship “was really devoted to machine learning, AI and specifically working on face recognition at the time and optimizing it.” Then he “took a machine learning course in

grad school that required Python as well.” However, through the internship, John had realized that he was not interested in implementing machine learning because:

It seemed like the field of machine learning, to be successful, you had to really like keep up to date with all the new algorithms that are constantly popping up with machine learning and that seemed kind of - not for me.

Domain Knowledge. On another hand, John mentioned a course about survival analysis that was particularly helpful to work for a pharmaceutical company. It helped him learn common terminologies “like hazard ratios and Cox proportional hazard models or Kaplan Meier plots.” Indeed, he thought that having a degree in biostatistics would give him “a strong foundation just statistical theory in general that even if I didn't want to work with biological applications, I could apply it elsewhere.”

Florence had to do some research to learn about techniques that may have been developed for a specific domain. She said:

Papers that I read like during a project will be on various statistical techniques. I'm still trying to keep them in a similar field, at least in the medical research. So they'll almost always be applied or at least somewhere nearby. I spent time on a lot of graduate school websites, reading slides that have been posted and some of them are a couple of years old, going back to like 2002, but they'll still have really good examples of like, here's the technique and here's a case study on it so you can learn about it. Just kind of a lot of time on publicly posted, and on university slides that are just out there on the internet somewhere.

Janet also did some research about the domain knowledge and she wished she had taken courses “more on specific for clinical trial data, specific for high dimensional data, specific for time series experiments, survival analysis.” She considered researching new techniques as an important aspect of the practice of analysis because “typically, I'm at my desk doing the analysis and reading about different techniques and things and then I'm presenting those results to my colleagues or the PIs.”

Validity. Florence mentioned knowing about “validity” of the models as part of the practice of analysis, to check for assumptions for example. However, she said that she was “not sure how much I learned about that in school versus learning about it kind of more on the job.” Thomas had some experience in academic settings because the emphasis of his degree was on “the model validation and interpretation and write up” results. Indeed:

Written out explicit math really helped me get a firmer grasp on statistics, not just applying models and interpreting results, but actually understanding what's going on underneath the hood. And it really helped me kind of like I said grasp kind of the theoretical underpinnings of different statistical methods and different analyses of data.

Therefore, understanding the mathematics helped Thomas interpret the results. David suggested steps to ensure validity by first understanding the data and “knowing how it's collected, how it's gathered, how it's manipulated. How it goes from being out in the open to actually going on your computer screen,” which is really important to “do any sort of real interpretation.”

Programming. Participants used their practices in programming to perform analysis. I already discussed these aspects pertaining to analysis in the rules for programming routines but David made a connection between programming and analysis as he experienced the transition to the workplace himself: “I knew the theories that I wanted to use and I knew, like the techniques that I wanted to use and I was able to translate that into how do I translate that into R on the job.”

Boundaries in the Community for Analysis. The rules set by the community, opportunities for training, or having cross-functional teams introduced or inhibited boundaries in the transition to the workplace.

Rules. John was not familiar with the theory that his team stated, and he found it difficult to be part of a team with a background in Bayesian statistics. He said that it “was a challenge for me to kind of grasp in the beginning of my career here, and I've only been here for about maybe a year now.” Therefore, his experience was considered detailing. On the contrary, Thomas had acquired techniques in academic settings that were advocated by his workplace. The community at the workplace made “a big push to kind of implement artificial intelligence, machine learning, deep learning all those buzzwords in everything they do.” Rightly, his educational background had provided tools to perform these specific techniques.

Training. As mentioned previously for the practice of programming, John really appreciated the opportunity to take part in statistics forums at the workplace, where he also learned about new techniques. Other participants found opportunities in going to conferences to keep up to date with newly develop techniques for analysis.

Cross-functional Teams. At Simone’s workplace, teams included members with different backgrounds and specialties, from whom statisticians could learn more about the domain knowledge for example. Even though, Edwards had colleagues with backgrounds in mathematics, he thought that he was the only one to be interested in making sense of the theory.

Transition of Analysis. The practice of analysis involved activities beyond simply performing analysis. For example, participants had to research new techniques to determine which ones were appropriate to apply to their datasets. Boundaries in the rules such as dealing with big data and needing to complete task on short deadlines restrained the choice of the technique. However, tools developed in academic settings facilitated

boundary crossing for the practice of analysis, with all participants having acquired tools such as a solid background in theory and applications of statistics, with topics in data science and especially machine learning. The main differences between detailing and continuous experiences resided in the division of labor and the community, in particular to choose the appropriate technique, to learn a new field of theory, or to manage validity issues. The community at the workplace played an important role in the transition by setting the rules for which field of statistics to focus on, and provided opportunities to cross boundaries by offering forums, encouraging junior statisticians to pursue professional development, or building teams with members of different backgrounds and roles. To keep learning and overcome boundaries, participants mentioned going to conferences, researching resources online, doing trial and error, or seeking the support of a mentor. Indeed, mentors recognized assisting junior statisticians in the transition, both in academic settings and at the workplace. They reinforced the priority of getting the analysis done over performing the best analysis, and the importance of interpreting results to provide information to domain experts for them to make informed decisions. Indeed, statisticians collaborate with domain experts to answer questions of interest, which I present next.

Coordination of Collaboration

As discussed with the design practices, statisticians contribute to answer research problems arising from different fields of application. To provide answers, participants mentioned collaborating with their team or with domain experts which involved understanding their question of interest and sharing expertise. All junior statisticians

mentioned this practice as well as both mentors (see Table 20) with 63% of the quotes referring to the workplace (see Table 14).

Table 20

Important Elements for the Coordination of Collaboration

Participant	Academic Settings	Workplace	Transition
Florence		Collaboration with many domain experts but isolated statistician	Discontinuous
Gertrude		No collaboration on programming	Discontinuous
Janet		Collaboration with many domain experts Experience in customer service	Detailing
John	Consulting course with no domain expert	Collaboration with domain experts through PhD level statisticians	Detailing
Stella	Consulting course with no domain expert	Collaboration with domain experts	Detailing
Thomas		Collaboration with coworkers Collaboration domain experts at his second job which he started during his master's degree	Continuous
Enid	Consulting course with domain experts	Collaboration with coworkers	Continuous
Edwards	Consulting course with domain experts	Collaboration with coworkers and domain experts	Continuous
Simone and David (mentors)		Collaboration with domain experts is not the responsibility of junior statisticians	Observed Discontinuous

Participants had a variety of experiences, from discontinuous to continuous, and emphasized on boundaries in the community and the division of labor. Indeed, Enid and Edwards referred to consulting courses in academic settings that promoted the division of labor for collaboration as they were providing statistical support to domain experts. However, Stella and John did not have the opportunity to work with domain experts through their consulting courses, marking the difference between detailing and

continuous experiences. Rules at the workplace mediated the relationships between domain experts and statisticians as well as their attitudes. For example, Janet applied her past experience in customer service to interact with domain experts by being diplomatic and understanding. The absence of explanation for programming practices at the workplace caused the transition of Gertrude to be discontinuous, while Florence experienced discontinuity because she first worked with many different domain experts with no support from a mentor or fellow statisticians. As a matter of fact, the role of the mentor facilitated boundary crossing by mitigating the meetings with domain experts for instance.

Boundaries in the Community for Collaboration. Between academic settings and the workplace, the community introduced boundaries because participants collaborated within interdisciplinary teams or with domain experts.

Team. As junior statisticians transitioned to the workplace, they encountered boundaries when meeting people in the community who did not necessarily shared the same interests or background. Thomas worked for a large government organization and by participating in seminars or interest groups, he managed to meet a collaborator with similar interests in machine learning. In academic settings, he used to work with classmates who shared the same background and he “had a study group and the people to work with” during his graduate studies. However, working in a team at the workplace introduced boundaries if the members of the community had different background. For example, Gertrude encountered boundaries as she entered the workplace because her coworkers used a programming language that she was not familiar with and they did not document their code to explain their work. As a matter of fact, Florence was told by her

university that working “in a team with people who don't speak the same technical language as you is really hard” but she was not provided specific support to develop that practice.

Florence experienced discontinuity because she had to deal with isolation at the workplace, being the only statistician physically present. She was “the only one who works on site. The other two statisticians both work off site.” As a result, she joined a peer group to connect with the broader community of statisticians. Because the senior statisticians working remotely had very little programming background, she also connected to the community at the workplace to find coworkers to help her with the logic of programming. Within her team at the workplace, she collaborated with her coworkers “who are not mathematicians or statisticians” and helped her with her projects by “coming over and telling me if things made logical sense.” After an incident that happened on one of her project, she started to work closely with one of the senior statistician who she considered as her mentor and met with her to work “on a project at least once a week.”

Finding coworkers to collaborate with within a team also helped Edwards improve skills in consulting with domain experts, which he had previously acquired in academic settings with a consulting course. Indeed, John valued the team spirit at his company and in particular to be working with doctoral level statisticians to collaborate with domain experts. Doctoral level statisticians were “kind of like the middleman between us so there's like an indirect connection between us and the researcher.”

Domain Experts. At the workplace, participants worked with domain experts to solve a statistical problem. For example, Janet and Florence consulted with principal

investigators on many different projects in the field of healthcare, Thomas provided statistical support to graduate students and professors at a Law school at his second job, and Stella met with stakeholders in higher education. She said she had “to meet with clients on a regular basis and walk them through what I think is appropriate and why, for their particular projects.”

Some participants identified consulting courses in academic settings to facilitate boundary crossing for working with domain experts. Enid mentioned a consulting and collaboration course “where we've basically had a real-life client who said, here's my need.” She provided statistical support for study design and analysis for experts in agriculture, in particular. Edwards also had a consulting course and he worked with many consultees under the supervision of a professor. As opposed to Enid and Edwards who had a continuous experience working with domain experts, Stella and John took a consulting course but did not interact with real clients, resulting in detailing transitions. In addition, Janet and John found that consulting courses offered in academic settings seemed to be proposed to doctoral students in priority. Indeed, John was one of the few master's level students enrolled in the consulting course and Janet knew of a consulting course offered in her department, but it was only for doctoral students.

Boundaries in the Division of Labor for Collaboration. The boundaries encountered with the community about working with domain experts had some repercussions on the boundaries in the division of labor.

Working with Domain Experts. Opportunities in academic settings to experience how to work with domain experts facilitated the transition to the workplace. Indeed, Edwards noticed that “in terms of what to put on your CV, having consult, consulting

experience out of school really is, is a pretty big deal.” In particular, participants identified consulting courses and emphasized on the important of experiencing the division of labor. For example, about his consulting course, John noted that “even though it sounds really like a lot of application base, it was actually more theory of biostatistical consulting.” Indeed, there was no opportunity to work with domain experts:

We were reading a lot of papers and we had good discussions and we did have one project that we worked with a consulting center but my project just wasn't the best experience like, they kind of pulled out halfway through, so I didn't get the full experience of like working with someone that requested statistical support on their work.

Similarly, in Stella’s consulting course, the professor would provide a research problem, and paired up two students, one to be the client and the other to be the statistician.

Students were in charge of planning a conversation and determining the appropriate analysis, but they did not experience how to build relationships with a client.

On the other hand, Edwards and Enid were both in charge of the division of labor in their consulting courses. Edwards had a one-year sequence in consulting, for which he worked on approximately ten projects with different consultees while Enid provided support for at least two projects.

Understanding the Problem. To start collaborating on a project, the objectives needed to be clearly defined as well as the problem in question. As a mentor at the workplace, David focused his meetings on “what should we be looking at, what are some of the strategies we should be thinking about.” However, he was not expecting from his junior statisticians to be in charge of formulating a problem but progressively contributing to it.

Janet mentioned the challenge to “understand what people really want and how to answer that question.” Reflecting on her experience in academic settings, she said: “I don't think anything in my Masters of Education Program, particularly had prepared me for that at all. I just already had that skill.” Indeed, her experience in customer service helped her understand researchers’ needs and made for a detailing transition. Enid had developed a set of questions to ask researchers while taking a consulting course. In this course, the professor determined which students were working with which client, but students were responsible for setting up meetings with the domain experts. Enid would ask them:

What are you working with? What do you have to start with? What do you need at the end? Like what are you hoping to find out or attain knowledge wise and then we can figure out the in between and how to get from A to B.

Asking these questions helped determine what researchers expect from the collaboration. For Florence, the role of her mentor was to facilitate meetings with principal investigators and made sure that “they've said that they want to get done.” Indeed, as a mentor at the workplace, Simone explained that collaboration begins with understanding the design of the study with the different steps such as data collection, analytical plan, or dissemination of results. She identified boundaries because domain experts did not necessarily formulate properly what they needed:

It's a lot of working hand and hand with people because of what they may ask is not really what they, they need. They just don't know how to ask for it the correct way. So you really have to work with them to tease out what they need. And sometimes they come after you've done already some work in or you're looking at it together and they're like, this doesn't look right or this is not what I want. So it does take experience to get that, to realize that.

Indeed, Simone considered that junior statisticians were better prepared for preparing the analysis plan while collaborating with cross-functional teams was “more of a senior level

staff position.” She noticed that the reason junior statisticians usually experienced discontinuity with collaboration was about sharing knowledge with the domain experts.

Sharing Knowledge. Janet viewed collaboration as a way to share her statistical expertise with domain experts who brought domain expertise. After getting some experience during a research opportunity, she realized that it was not reasonable to expect domain experts to be fluent in statistical language. Similarly, Florence had to meet the expectations of her collaborators in terms of the programming languages because they were more familiar with another language than the one she was using.

Explaining Results. Participants were also in charge of explaining results which is described within communication practices by communicating with non-statisticians.

Boundaries in the Rules for Collaboration. The relationships with domain experts and the rules moderating the attitudes between statisticians and their coworkers or with clients created boundaries at the workplace.

Relationships. In academic settings, Enid described the interactions with consultees as “rigid” and “very strict where it was like start, do, finish” whereas she noticed that interactions while consulting with her coworkers were more fluid at the workplace. Indeed, she collaborated with people from other divisions on several projects and after a while she would find herself working with the same people. To ensure fluidity in communication, Thomas shared an anecdote about how he learned how important it was to define the objectives and assumptions at the beginning of a collaboration. He was working with domain experts on a project and the “message was fragmented and we're just kind of playing a game of telephone across a kind of a complex project and one that was a lot more complicated than we thought.” How he resolved this issue was by:

Making that message clear, concise, because a lot of people don't read long emails, as you probably know, and kind of getting everyone on the same page. And it actually took like a phone call, a conference call, with all of us to kind of lay out what's been [problematic].

To establish relationships with domain experts, Janet used her experience in customer service because it taught her how “to listen and trying to understand what people really want and how to answer that question as opposed to what they're saying on the service.” Indeed, she made the analogy between this past experience and her experience as a statistician at the workplace for which she had to understand a problem and translate it into a statistical form. She struggled with this aspect of building relationships during her internship while she was interacting with principal investigators who did not necessarily have a background in statistics. She realized that she had “an unreasonable expectation for PIs because they don't usually know all that stuff.”

Attitude. Statisticians learned how to be diplomatic, understanding, and patient to work with collaborators. For example, Janet learned how to interact with principal investigators and “share expertise in a way that wasn't me going, ‘how do you not know this?’ ” Once again, her experience in customer service as she worked in retail or at restaurants, provided insights on how to interact with domain experts:

I have to say I have a fair amount of experience and just dealing with a lot of different kinds of people. Conduct myself with people like in a different language and having to, like, just be quiet all the time, is a skill that you have to learn.

However, being diplomatic was also a skill developed in academic settings, as Enid shared her experience doing group projects. Indeed, she found that group projects were challenging because of how accountable, or not accountable, other students in the group may have been. She emphasized how she had to deal with others not putting enough effort into completing the project and found that the rules were “not fair then because

then I get the good grade that I want” but “these two people that contributed nothing to the group project are going to reap the benefits of my hard work.” On the contrary, at the workplace, Enid collaborated with coworkers on projects who had very different attitudes because she felt “like I can depend on my coworkers. So if I need something done, you know, I'm as invested in this project as they are, I want to see it succeed, they want to see it succeed.”

According to John, interactions with his coworkers at the workplace differ from interactions between professors and students in academic settings “because there's that professor and student expectation” while he found that collaborating with doctoral level statisticians was difficult because they “are not as understanding of what level you're at.” Similarly, Florence had to manage the expectations of the principal investigators she was working with because they were “pushing so hard to just get that analysis done right away.” To help her in the transition and learn “how to speak up,” she eventually had a senior statistician intervene and did not meet alone with the researchers.

Furthermore, Stella felt discriminated against as she dealt with different expectations related to her gender, noticing that “some men do not like working with women.” She had experiences in the past where she “wasn't treated professionally” and she “felt a lot of it was because I was a female. I was called emotional, I - when I tried to express my opinion about something. I've been called stupid because I didn't know the answer to something.” She realized that these attitudes were not only towards her while sitting at a panel about equality for women in statistics at a conference. She “was amazed because I thought I was the only one that this was happening to and it wasn't.” To learn

how to overcome this boundary in the rules, she wished she “had learned more communication skills when dealing with a difficult person or a client.”

Boundaries in the Tools for Analysis. Participants referred to tools with means of communication, meetings, or documentation. They also referred to consulting courses as demonstrating collaboration in academic settings.

Communication. The practice of communication is further explained in the next section but here I emphasize how Stella and Enid learned about communication to foster collaboration at the workplace. As just mentioned, because Stella dealt with difficult people at the workplace, she took a course offered at a conference to learn about communication and “more specific[ally] towards female statisticians or minority statisticians.” To learn how to better communicate with others, Enid referred to workshops offered at her workplace with “classes on collaboration, but it's really more of like diversity collaboration where it's like we're going to learn how to talk with people that are different than us.”

For most participants, important communication was conveyed by emails, and John learned at the workplace “what it takes, in terms of like the operational side of like the email, email communication and how much time you have to invest.” Edwards, Enid, and Florence mentioned starting their day by checking and answering emails from principal investigators and sometimes setting quick meetings for clarifications and discussions. Thomas also used an instant messaging chat to collaborate with his coworkers because “you can see what other people are working on and get involved and get connected that way.” Communication by email was even more important for his second job since he worked remotely for a Law school:

My communication is done, I'd say 95% of it is done via email. And so it's, it's important to kind of be, have attention to detail and to communicate clearly and effectively and to kind of get feedback and move projects along, especially when you're not at the office every day.

Emails helped Thomas staying in the loop for projects he worked on remotely, but at the workplace other participants used meetings for staying up to date.

Meetings. Janet scheduled regular meetings with domain experts to understand their needs or to present preliminary results. Meetings were an important part of Enid's day since she had about two each a day with "other branches, other divisions, other non-statisticians" who needed help with their projects. She used the meetings to "know their context or maybe updates or changes or extra information that might affect the project that I'm doing for them." Similarly, John organized or attended meetings to get updated or to ask for help. He would meet with his supervisors to clarify what is expected of him. To prepare for the meetings, John would do some research, prepare questions about the research, and know his collaborators' "working style."

Documentation. Participants reported using documentation to collaborate on projects which was actually extensively described in the interviews and defined an entire practice. For example, Gertrude identified the lack of documentation as a boundary for collaboration since it was "so hard to collaborate" without having explanations in the code created by her coworkers. Instead she recommended to define standardized ways to write code: "You should write your code like this. Because it's easier to collaborate with people."

Transition of Collaboration. To solve statistical problems arising from different fields of application, participants needed to collaborate with domain experts or within their team, which introduced boundaries in the community. To collaborate with domain

experts, they needed to understand the problem, share their expertise, and explain results. Rules for collaboration defined the norms for establishing a working relationship between statisticians and domain experts, with attitudes to adopt. Most experiences of the transition were qualified as detailing or discontinuous, with courses in consulting causing the distinction with continuous experiences. However, to be effective, consulting courses had to include domain experts and let students be in charge of the division of labor. Indeed, they would gain experience in contacting domain experts and organizing meetings. Mentors at the workplace considered that junior statisticians should generally not be responsible for collaboration because they first need to develop more collaborative skills such as communication, which I describe next.

Coordination of Communication

To foster collaboration, statisticians developed the practice of communication, by explaining statistics to non-statisticians, presenting to stakeholders, writing for publications, listening, and asking great questions. All junior statisticians mentioned this practice as well as both mentors (see Table 21), communication practices being the second most cited category with a total of 183 quotes (see Table 14). Communication had the greatest proportion of quotes associated with the division of labor as well as an important share of the community at the workplace (see Figure 14). Similar to collaboration, boundaries in the community challenged the transition of participants the most, with the audience of non-statisticians. Communicating with non-statisticians affected the rules, requiring statisticians to consider the background knowledge of their audience and their expectations. Opportunities to practice communication in academic settings made the distinction between continuous and detailing experiences of the

Table 21*Important Elements for the Coordination of Communication*

Participant	Academic Settings	Workplace	Transition
Florence	Programming	Create visualizations Communicate with non-statisticians	Discontinuous
Gertrude	Programming	Create visualizations Explain machine learning to non-statisticians	Detailing
Thomas	Writing assignments	Communicate with non-statisticians at two workplaces	Detailing
Stella	Consulting course with no domain expert	Present to non-statisticians	Detailing
Enid	Consulting course with domain experts	Communicate with non-statisticians	Detailing
Janet	Documentation	Communicate with non-statisticians Experience in customer service	Detailing
John	Teaching experience Consulting course with no domain expert	Communicate with other statisticians Communicate with non-statisticians	Continuous
Edwards	Writing assignments Consulting course with domain experts	Experience teaching Communicate with non-statisticians	Continuous
Simone (mentor)	Encourage students to ask questions to each other	Encourage junior statisticians to rehearse presentations and practice	Observed discontinuous
David (mentor)		Encourage junior statisticians to focus on decision making rather than analysis	Observed discontinuous

transition. For example, Edwards practiced writing for publications in academic settings.

In addition, John regarded his teaching experience as providing the tools for his

continuous transition by explaining statistics to non-statisticians, using layman's terms.

Mentors identified the experiences of junior statisticians to be generally discontinuous

and they progressively supported learning how to communicate at the workplace. Indeed,

without the support of a mentor, the experience of Florence was discontinuous. Overall,

participants recognized the importance of communication, because as David stated: "you

can do all you can do the greatest analysis in the world. But if nobody understands what's going on it's just going to go to waste. You really have to be able to learn how to communicate.”

Boundaries in the Community for Communication. Participants encountered boundaries in the community by communicating with non-statisticians or with other statisticians.

Non-statisticians. In academic settings, the community of students and teachers shared common knowledge whereas, at the workplace, statisticians communicated with collaborators who did not necessarily have a background in statistics. Indeed, Thomas explained how he interacted with his coworkers who did not have a background in statistics:

Communicating with them what results are from a statistical point of view and what the implications are, has been a lot more challenging because based on just previous jobs and being in graduate school where everyone's studying the same thing I'm studying. So kind of putting, not really in layman's terms, but kind of making statistics and analyses and different approaches more accessible to kind of foster collaboration.

All participants except Gertrude reported communicating with non-statisticians. For example, Thomas was the only statistician at his workplace:

I'm the only person on base with a statistics background. They're all civil engineers, mechanical engineers, electrical engineers, so they, they're, they understand math and some statistics but communicating with them what results are from a statistical point of view and what the implications are, has been a lot more challenging.

Similarly, Enid principally worked with coworkers who had some notions about statistics, but she avoided using statistical jargon:

I'm like a small group [of statisticians] among a ton of non-statisticians and, and, and you have to kind of also think in context, like they're not completely dumb to statistical concepts, like they know some basics, but if I had to try and explain a

logistic transformation to them, their eyes would glaze over and they would probably pass out from a coma.

Enid had to learn how to “translate statistics to non-statisticians in a way that they understand it.” In academic settings, she “got a bit of a taste of it in like that consulting and collaboration class, but that's something I think I can say I wish I had more experience of back in my college experience.” She also wished there were workshops offered at her workplace that focused on communication, depicting her experience of the transition as detailing.

Other Statisticians. John experienced boundaries in the community by communicating with other statisticians and in particular doctoral level statisticians. Indeed, he explained the difference between communicating with professors in academic settings and with other statisticians at the workplace:

Communicating statistics with other statisticians, I think I have a lot more experience with that as well. That is different than speaking from someone who doesn't have a statistics background and especially talking to a PhD level statistician. I think I have a lot more skills with that because that was tricky at first because you have your professors that you talk to you in school, but I think sometimes they are more clear. And there's a little more understanding of where you are because there's that professor and student expectation. But when you're at the workforce, sometimes the PhD level statisticians are not as understanding of what level you're at. So kind of navigating to communicate with them has been something I had to pick up.

Simone mentioned cross-functional teams with statisticians and non-statisticians, with backgrounds in “all specialties from different areas: statistics, regulatory, quality, R&D, medical directors.” Therefore, she considered that junior statisticians needed to learn how to present to all types of audience.

Boundaries in the Rules for Communication. Most participants identified opportunities in academic settings to communicate through presentations, however,

presenting to audiences with different domain knowledge, with the intention of making decisions, and ensuring ethics created boundaries in the rules.

Domain Knowledge. As the community differed between academic settings and the workplace, participants realized they needed to consider their audience's domain knowledge and their familiarity with statistics. At the workplace, Janet developed "her presentation skills, like presenting things for different audiences has been a skill set that I've worked on a lot." Indeed, participants established rules to tailor their presentation to their audience. For example, Enid considered that data visualizations "get the point across more to like non-statisticians than I think like anybody else." However, Stella added that she needed to be careful about what types of visualizations to use:

I don't use real fancy graphics either because I find most, a lot of people are afraid of those as well. So I'll use like simple you know bar charts or a line plot or something like that. And I, I did do a couple of box plots. But that didn't go over very well. Like, I spent a lot of my time in the meeting explaining what the box chart actually was and the story it tells, and I just made a mental note, don't use those anymore.

Similarly, Enid thought that some visualizations are better than others, using traditional graphs for non-statisticians such as bar graphs or pie charts, and definitely avoided using summary tables. In addition, Stella decided not to use equations and technical terms, because "just over time from doing that, I've learned people are afraid of equations." She also did not use the word *p*-value either:

Cause If I said well the *p*-value is less than .05 so it's statistically significant, and usually I get with, you know, the blank stare that says, what are you talking about. So I try to make sure that I don't use terminology like that.

However, Stella needed to adapt the terminology to the audience because sometimes she presented results to people who had a background in statistics:

I have to present to our like outside evaluators, that's when I geek it up so to speak and that's when I can really start throwing out statistically significance and p -values and you know, Bayesian methodology, you know, all those big words because, because they understand it.

Since Enid and Stella shared specific strategies they have developed at the workplace, their experiences of the transition were considered as detailing. Indeed, no participant referred to rules in academic settings for presenting to non-statisticians, and as Florence said, what “they never talked about in school was how to communicate statistics to a non-statistics-based audience.” Although, Gertrude recognized efforts in promoting such skills because her degree program had evolved since she graduated: “communicating concepts to practitioners is a pretty important thing and I think they've sort of added that into their curriculum now.”

To support junior statisticians in crossing boundaries encountered with the rules at the workplace, David encouraged them to focus on the implications rather than on the analysis techniques for an audience of non-statisticians:

I think really the big thing for me has been transitioning them from thinking about, you know, they like to present often how they did things. And when you're presenting to people that don't know anything about statistics, it means absolutely nothing. You can say you did this amazing analysis, with this technique. Nobody that is of interest that you're presenting to actually really cares that much. I need a little bit just to keep a little bit of validity.

David also offered the junior statisticians on his team to take a data visualization course and a communication course at a conference because they had to learn how to communicate and present results.

On the other hand, Simone mentioned boundaries in the rules by communicating with non-statisticians who believed they understood statistics: “I find that the biggest skill sometimes it's not to explain what we're doing, but really, sometimes, I think they don't

understand it and they don't know it.” She especially noticed these boundaries occurred during presentations with cross-functional teams, when nobody asked questions about results that needed further discussion. Simone suggested that it was the responsibility to ensure everybody was on the same page:

The statistician has to ask to confirm or do something. So we've had issues like that in the past where it's like making sure that other people don't misrepresent, um, and have a team, you know, correct back. Because I even feel like they know, know a lot about statistics, yet they're still not fully qualified for those nuances and that has always been a down for us.

Making sure non-statisticians understood the results presented by statisticians was crucial, because results were often shared to provide insights to inform decisions.

Decision Making. In academic settings, Edwards was required to write papers and the rule set by the professor was to provide some insights about the interpretations of the results, “saying what the effects were.” Similarly, Thomas had writing assignments for which he was required to explain the problem of interest and the results, but the professor expected students to tie the results to implications, and “even if your result was right, if you didn't type it up correctly and coherently, we will be taking, it would be penalized for that.” These opportunities gave authentic opportunities to learn for communication at the workplace because as discussed for analysis practices, the rules shifted to focus on implications. As a mentor at the workplace, David found that junior statisticians usually struggled with only focusing on the interpretations and not on the analysis techniques, which was not of great interest to the audience:

Give me, you know, three to five bullet points, one sentence that somebody could gather this. And I think that was a little bit annoying for them. Because you put in potentially weeks or months of work into analysis into five bullet points. But in the end, that's what matters, that's what people are making decisions off. They're not making decision off of why you chose a GBM versus a random forest. They're making a decision off of ‘If you increase this do we get more money.’

Junior statisticians were probably used to present analysis techniques in academic settings, since courses tend to focus more on analysis techniques than on interpretations. Because David found the transition for junior statisticians challenging, he mainly observed discontinuity for the practice of communication.

Ethics. While communicating the results of a study, statisticians need to be transparent about any specific procedure that was undertaken during the design, data management, or analysis, to guarantee the ethics of the study. Florence provided an anecdote for which she had to remind her collaborator about the rules of ethics:

A particular collaborator [who] failed to mark down that we removed two patients in a publication that we're just about to send in. So even last week when I was on vacation and both my boss and I were reaching back out to that collaborator being like 'you missed this part and that's an important thing to add to the context of this study.'

During the interview, Florence kept referring to an ethics course she took for her master's degree because it "was probably one of the most valuable part of, of my education," making her aware of the rules but also giving her experience with the division of labor by taking different points of view while discussing case studies in class.

Boundaries in the Division of Labor for Communication. As participants transitioned to the workplace, they experienced boundaries in the division of labor by becoming the experts, having to explain statistics to non-statisticians, being in charge of presenting to stakeholders, or writing papers for publications.

Being the Expert. To communicate about statistics, participants had to position themselves as experts. Stella and Florence considered that aspect of the practice to be particularly challenging because as Stella stated, "when someone's coming to you, they expect you to be the expert and know everything." To become an expert in a domain,

Florence “spent time every week reading about different topics.” Indeed, at the workplace, statisticians had to learn fast to complete the projects and Florence found that it was “the weirdest thing to read about something and then walk into a team meeting the next day and be the expert on it. Having just learned about it the previous day.”

Explaining to Non-statisticians. As experts, participants needed to learn how to explain statistics to non-statisticians and one thing Janet learned from her master’s program was “the idea that you're going to have to... the majority of your time will be spent explaining your work to people who are not statisticians.” Indeed, Stella had to demonstrate her ability to communicate to non-statisticians during the interview process for her current job. She “had to give a presentation. They wanted to see how I could take technical language and present it in a way that everyone could understand.”

Communication to non-statisticians was valuable to employers and likewise, David recruited a person who “had a PhD in a non-quantitative field but was able to speak the language and knew what she was doing.”

Explanations were also needed to justify choices. For example, Gertrude was in charge of explaining the techniques to the stakeholders at her company who were skeptical about machine learning models. She had to reassure them and explain that the models were “not going to work very well at first because they need data to learn. But after a while they will get better.” Indeed, the stakeholders were afraid to lose money with these models and she had to convince them that they would eventually work better. She made sure to be “digging their concerns, translating them to, like, a data science thing, getting a data science solution, translating it back to English for them.” Gertrude also talked about justifying her choices of models within her programming. She said:

I wish I had learned a little bit more about explaining to other people, maybe teaching a little bit or um, documenting code in a way that other people can sort of pick up and understand like even a statistician, like what kind of model is trying to be run here.

To learn how to explain to non-statistician in academic settings, John referred to his experience as a teaching assistant:

I was a graduate TA when I was in grad school. Taught an introductory statistics course and I was really glad I did that because it really sharpened my communication skills, especially with the audience who doesn't have a strong statistics background.

John greatly valued this opportunity when presenting his work to an audience that did not have a background in statistics because he was able to “leverage difficult statistical concepts in a way that people can kind of follow and understand.”

To practice “translating statistics into English,” as a mentor in academic settings, Simone had students “at the board with others where people have to be able to question you, and you would have to be able to respond.” Indeed, as a mentor at the workplace, Simone shared that statisticians had to present both internally and externally, explaining “statistical methods to lay persons.” However, learning how to explain statistics to non-statisticians occurred progressively, “with increasing levels of difficulty.” As a statistician, she was in the managerial track, and at “every level you are talking to higher and higher management and just having to learning to do that.” From her perspective, developing communication practices “usually comes many, many years in the road” and for the first two years of her career:

When somebody would ask me how to find something, it was very hard for me to explain because I could just see it in my head, right, understanding in my head but how to explain it to somebody else. Yeah. I think that, um, you know, explaining math or statistics to a person, um, was not very strong for myself.

Because Simone experienced discontinuity for communication herself, she considered junior statisticians to have discontinuous experiences as well.

Presenting to Stakeholders. At the workplace, participants experienced boundaries in presenting to upper management because in academic settings they were presenting projects for a grade, rather than for decision making. As a statistician, David mostly communicated with upper management to provide insights for them to make decisions. In his presentations, he focused on the implications with “what the data is saying, this is the story it's telling us.” Similarly, Gertrude was “the point person for communication” at the workplace, “because the rest of the team doesn't really want to do it.” For example, she was designated by her coworkers to present at a conference with stakeholders:

We actually have a conference coming up with the people from like around the global offices of our company and from a few other companies and the team has decided like, ‘[Gertrude], you're the only one who was like even a little bit okay with speaking to people and probably about this in a way that anyone can understand.’

Indeed, Enid found that some statisticians on her team were nervous about presenting and she was also in charge of presenting at the workplace:

A lot of people in my, in my division, they're so nervous just even like getting in front of a table of five people from another division and talk. I'm like, but why? You know, what you're talking about? You're an expert, you're the subject matter expert, you know what you're talking about. And they're like, yeah, but I get so nervous.

Gertrude suggested that her coworkers were nervous because they were “introverts” and were “not super into answering things or answering clearly or concisely,” while Stella thought that it was because “a lot of researchers are so holed up in their offices that sometimes they forget how to interact with people.” The participants I interviewed may

have been naturally more comfortable presenting than the average statistician since I met them at conferences and they volunteered to talk about their experience.

Listening and Asking Questions. Janet combined her experience in customer service and some research online to develop soft skills such as “how to actively listen, things to look for in a conversation when it's going well, when it's not. How to redirect the conversation, things like that.” Indeed, part of her responsibilities at the workplace was “a lot of very careful like client relationship development and management, rather than just sort of sitting alone and working at my desk.” On the other hand, she did not identify opportunities in academic settings to learn “how to talk to people. I feel like that wasn't a skill set that was super common around my homework.”

Both mentors recommended to learn how to ask questions. Simone thought that “being able to stand in front of someone and um, you know, discuss the methods and the interpretation and have people ask questions and then being able to respond, understand what the question is that they're answering and being able to respond to it.” As a statistician, David learned from his mentor who “Asked some really in depth questions that really made me think about how I should be looking at data and analysis and really trying to get the best or what is most useful at the time kind of results.” Indeed, “I was definitely pushed on, you know, pushed pretty hard and really asked to question a lot of things. And even when there weren't any questions I had, I would have still asked the question, and I think really trying to think like that really kind of expanded the way I look at my work.” As a mentor, now he asked questions to his junior statisticians: “even if they had an analysis that looked really good. I would always ask him 10 different questions that really, even if I knew what the answer was going to be, I would make them

go in and really understand what was going on in the data. So I think that was something that I did a lot.”

Writing for Publications. In academic settings, most participants practiced how to write reports, however, at the workplace, reports were written to be shared with the community at the workplace or for publications. Thomas was required to write in academic settings and although he had “not an explicit course on communication, but it was generally something that we were often, when we were given assignments, we had to write up results in a way that was, I think it was called like ‘publication worthy.’ ” Similarly, Edwards had writing assignments in academic settings: “we had to write a paper summarizing the project. In a ASA type format. And then present that, then we submitted that for to be graded by our professor for the course.” At the workplace, Edwards did “authorship reviews to make sure that any sort of errors are fixed.”

In academic settings, Janet learned to communicate “in our code, make a lot of comments. Write out why we did something explain our choices.” Indeed, her professors emphasized that “you're going to have to do this all the time for people who have no statistical background. So get used to it.” Documentation practices are thoroughly described in next section.

Boundaries in the Tools for Communication. Participants referred to different tools to develop communication practices. They used data visualizations, breaking down statistical concepts into layman’s terms, and practice.

Data Visualization. Creating data visualizations was usually included in academic settings, as illustrated by Stella: “visualizations I, I kind of just learned as I went along in school.” As previously mentioned, the rules for data visualizations changed between

academic settings and the workplace and it also affected the tools participants were using to create data visualizations. For example, Enid had recently received training in Tableau because at her workplace they were “trying to think of different ways that we can do data visualization for next year” to prepare for her next big project.

Similarly, Florence took a summer course in data visualization that her supervisor recommended “which was actually really cool.” Indeed, she used data visualization “to communicate statistics to a non-statistics-based audience. Not to dig into the weeds necessarily, but give good overviews and a lot of pictures, a lot more pictures than I ever thought I would use. And very clear charts.” To foster collaboration at her company, Gertrude created interactive platforms to display visualizations and made them available to non-statisticians: “the dashboards I build are staying online so that people can actually use them, that's pretty important.”

John and Edwards referred to learning data visualization through programming with R. John was introduced to:

A lot of different packages for data visualization in R and [the professor] gave us like a project where we had to create like four really challenging data visualizations on our own. So that's a good way to learn how to create data visualizations, especially in R. I still use today in work.

Edwards mentioned that R specifically, had “more flexibility for techniques and visualizations and things like that.”

As an undergraduate student, David took communication courses with “a course on visual communication and a course on speech and, you know, I’ve taken a couple courses on that and it's something that's carried on till today.”

Layman’s Terms. Many participants mentioned using layman’s terms to communicate with non-statisticians. For example, Thomas formulated his explanations

“in layman's terms, but kind of making statistics and analyses and different approaches more accessible to kind of foster collaboration.” Similarly, Florence was required by her mentor to explain statistical concepts to domain experts without using statistical jargon:

I've got a mentor at [the workplace] and she in particular is really hardnosed about making sure that any layperson could read your table whenever you submit it to the rest of the group. Just for clarity sake.

Another boundary encountered at the workplace was shared by Janet and Enid in terms of choosing their words. They were afraid that using layman's terms would sound patronizing. Enid wished she had more experience communicating with non-statisticians to “be like, ‘Alright, you have no clue what I'm talking about. Like let me try and bring it down to a level that you understand’ without coming across condescending and arrogant.”

Practice. Communication is the only category of practices that participants uniformly identified as getting better at by practicing. Indeed, David said “I like to give presentations. I always, I'll often volunteer If I have to just, you know, the more you do it, the more practice you get, the more comfortable you feel.” The other mentor, Simone suggested to work on communication progressively and proposed training at the workplace with “internal courses that have it, provides for doing presentations, skills, um, that people can attend. A lot of it is just practice.” She gave the opportunity for her junior statisticians to rehearse with “people to review your work and give feedback because you have to understand how much detail you give, um, to the audience, is different from audience to audience and even culturally.” But once again, she emphasized “that all comes with, you know, practice.” Enid also thought that only practice can make you better:

Practice. So yeah, I think it really just boils down to practice. Like, I definitely, it's not something that's textbook taught, you know, I mean, you can talk about good techniques for, you know, talking to non-statisticians, but in reality like you just have to jump in there and do it. Like, you know, just getting that hands-on experience, um, talking to people that, you know, they don't do what you do every day.

Indeed, to get experience communicating with non-statisticians, Gertrude and Stella mentioned practicing with their friends and family. Gertrude used her family to practice her explanations:

I think if I can't explain something to my mother or my sister say, who are not statisticians, if they can't understand the basis of what's going on, then I don't really understand it very well. So just as a way for me to like, learn, I would always try to explain things to other people.

Similarly, Stella practiced with her husband for her presentations: “I look at the presentation and then anything that's on there that my husband described as alien language. I just take it off the slide.” She also practiced with her friends who “watched me present and they said, ‘Oh, I don't understand what you said here and there’ and I, so I knew to, to take that out.”

Overall, Thomas said to have enjoyed learning at the workplace “communication and technical writing have been like lessons I really enjoyed. It's not like a pain to learn about. Um, so I really, I'm very satisfied with my current position and the work I get to do now.”

Transition of Communication. Participants communicated to and with audiences that did not necessarily have a background in statistics which created boundaries in the community. The practice of communication involved activities such as explaining statistics to non-statisticians, presenting to stakeholders, and writing for publications. Participants identified some opportunities to learn in academic settings such

as writing assignments, consulting courses, or, more rarely, teaching experience.

However, they uniformly identified that communication practices mainly improve with practice at the workplace which explains why most participants experienced a detailing boundary crossing. By learning communication at the workplace, participants developed strategies to create data visualizations and jargon-free presentations, using layman's terms. Additional courses offered at conferences or training at the workplace supported junior statisticians in the transition while mentors provided opportunities at the workplace to learn progressively. Next, I describe the practice of documentation that fostered communication and collaboration.

Coordination of Documentation

Statisticians need to document their work so that it can be understood and shared with others. Documentation sustains collaboration and can help reproduce the work in case the statistician is out of the office or if there is an audit. Writing documentation included writing down the objectives and procedures involved in a project to extract, clean, or analyze data, for example. The tools for documentation took the form of comments written directly in the code and files created to keep track of the different steps of the project. Four junior statisticians and one mentor have mentioned this practice (see Table 22). Boundaries in the division of labor provoked a need to learn documentation at the workplace for Gertrude and Stella. The experiences of Gertrude and Stella differed from the experience of Janet in terms of boundaries in the rules, with the absence of rules leading to discontinuity. In particular, the experiences of participants were affected by the community at the workplace which created different rules. To support the learning of the

practice of documentation in academic settings, Simone introduced rules for writing reproducible code because that was expected of statisticians at the workplace.

Table 22

Important Elements for the Coordination of Documentation

Participants	Academic Settings	Workplace	Transition
Florence		Documentation was part of the daily routine	Discontinuous
Stella		Documentation not required but for audit	Discontinuous
Gertrude	Labeled files	Documentation not required but used for collaboration with GitHub	Discontinuous
Janet	Learned Jupyter Notebook in machine learning course	Documentation required for collaboration	Continuous
Simone (mentor)	Encourage students to write reproducible code	Documentation required for audit	Observed discontinuous

Boundaries in the Rules for Documentation. Rules for documenting work changed between academic settings and the workplace since statisticians were no longer working independently towards obtaining a degree for example, but they were contributing to a bigger picture, by collaborating with others or meeting audit requirements.

Collaboration. In order to exchange their work, members of Gertrude’s team used GitHub, which is a platform designed to collaborate on code with a version control system for repositories. However, she had not learned how to collaborate on her code in academic settings mainly because there was no rule for sharing her work. On the contrary, Janet was required to document her work in academic settings and she recognized how documentation fostered collaboration at the workplace via Jupyter

Notebook: “you could output your results right in line with the code like you just hand someone your lab notebook.... you could explain what your code is doing and how you got that code, that graph.”

As a mentor in academic settings, Simone made sure to teach her students “the highest level of rigor on programming” because from her industry perspective, code needs to be exchangeable between collaborators or shared with regulatory agencies for audit. She noticed that “when people learn programming, they don't typically have that rigor” that is why she established the rule of documentation in two courses she taught for a master’s program in biostatistics.

Audit. As Stella transitioned to the workplace, she had not developed the practice of documentation. It was not until six months on the job when she had to provide documentation for an audit that she realized it was missing. Indeed, there was no explicit rule set by the workplace for writing documentation even though she said that at her company:

We get audited, because we are a federal grant recipient and the auditors will come in and they'll pull our reports, they'll look at a particular objective ... and we need to reproduce those results for them. And we need to have documentation on how to do this. So that's why it's really important that you, that we, have to have something because if you don't, you're going to be fined.

Because the rules were set or omitted by the workplace, participants also experienced boundaries with the community.

Boundaries in the Community for Documentation. The boundaries introduced by the community at the workplace were caused by the rules, whether requiring documentation or not. The interactions between junior statisticians and collaborators also created boundaries.

Rules. Because the community did not set rules at the workplace, the experiences of Gertrude and Stella were discontinuous. Indeed, they learned that documentation could facilitate collaboration or the ability to reproduce their work after they encountered boundaries in the rules. On the other hand, Janet was helping promote documentation within the community at her workplace by mentoring others, teaching them how to use Jupyter Notebook. The Jupyter Notebook is an open-source web application with online repositories that supports workflows, code, data, visualizations, and interpretations, to facilitate collaboration.

Collaborators. Janet and Florence provided statistical support as consultants for projects with many different collaborators while Gertrude and Stella both worked on a limited number of projects within their company or organization. Because Janet collaborated with different principal investigators, she needed to write documentation to be able to share her work. This need for documentation was probably not as strong for Gertrude and Stella who did not collaborate with people outside of their team and therefore there was no rule for documentation. Thus, Gertrude and Stella were in charge of the division of labor, creating other boundaries in the transition between academic settings and the workplace.

Boundaries in the Division of Labor for Documentation. In academic settings, the professors were in charge of the division of labor as they were probably explaining and commenting on the code while teaching. For example, Gertrude said that her professors did not use comments in their code because “the code never really had to be shared with anybody” and therefore they did not require students to comment their code either. However, at the workplace, Gertrude was given code without explanations and

Stella was not given code at all.

Defining Procedures. Stella realized how important documentation was when a shift in the division of labor occurred as the database administrator at the workplace left unexpectedly. She became in charge of pulling the data from the database but there was no documentation on how to do so. The tools Stella used to document her work included writing step by step procedures in a document for each objective, reporting how to pull the data and how to conduct the analysis. She directly added comments to her code in R as well. Similarly, Florence organized the files and documented her code as part of a daily routine: “it means going back into the code and then making sure it's commented.”

Explaining Procedures. On the other hand, Gertrude encountered a boundary in the division of labor since she had to make sense of the code herself at the workplace. When she started her current position as a data scientist, she was asked to read the code on the company’s GitHub page and was expected to understand what the code was written for, even though none of the code was commented and there was no explanation given from the people who wrote the code. Because Gertrude was not familiar with the programming language used by her coworkers, she said: “it took me literally like a year and a half to understand what the code was trying to do.” In addition, Stella talked about documentation as a way to ensure activities would continue in case that she had to miss a day at the workplace:

What if somebody got really sick or something and you needed to have this done, you got, you have to have a backup person that can you know, run your analysis for you if you're not in the office and preferably the expert should be doing the analysis, but if not, at least I know now that the database administrator could run some of the [program] reporting for me if I'm not in the office.

Documentation can serve also as a tool to momentarily replace the role of statistician to

conduct routine analysis but some tools for writing documentation created boundaries between academic settings and the workplace.

Boundaries in the Tools for Documentation. Participants reported using different tools to document their work by creating files or using online repositories.

Files. Florence mentioned that organizing files and documenting code was part of her daily routine, although she did not specify how she learned documentation. In comparison, the only tool Gertrude was using in academic settings was to organize her code “piece of code one, piece of code two, piece of code two final, piece of ‘this one works.’ ” Gertrude decided to document and write notes about the code by picturing what would have been helpful to her when she started working. Other tools she used were examples of documentation she had been exposed to briefly during her first job to collaborate with others. Once she transitioned to her current workplace, Gertrude learned how to use online repositories.

Online Repositories. On the contrary, Janet learned tools in academic settings, dealt with consistent rules in academic settings and at the workplace, and she did not experience a significant shift in the division of labor at the workplace. Janet was required to document her code for one of her classes, and she was taught how to use tools such as the Jupyter Notebook. In academic settings, Janet hold some of the authority for the division of labor since she had to “explain the code and walk people through those kinds of things, that, I had to do a lot of that with [the] machine learning class and that was really helpful.” Indeed, Janet’s practice of documenting code developed in academic settings was of interest to her employers and she said that “one of the reasons I got hired

was because they wanted me to teach them how to do that and how to apply some of the concepts you use around your notebook.”

At her workplace, Simone provided samples of documentation to set the expectations. She said that new employees typically spend two months reading the documentation and “looking at good standardized coding” so that they “don’t have to start anything from scratch because we have so many things templated and standardized.” Indeed, Florence said that she read a lot of documentation for the first three months on the job.

Transition of Documentation. Participants identified two main rules that required documentation at the workplace but not in academic settings. First, documentation facilitated collaboration, with principal investigators or within a team of statisticians, using online repositories for example. Second, statisticians had to be able to reproduce their work in case of an audit, so they rigorously recorded procedures. Three out of the four participants who mentioned this practice experienced discontinuity while in the transition to the workplace. Indeed, they were not accountable for the division of labor nor developed tools in academic settings which actually facilitated a continuous experience. Except for one participant who experienced continuity, participants had to teach themselves how to document their work at the workplace and they did not mention the support of a mentor. However, one mentor observed that junior statisticians generally have not developed the practice of documentation when entering the workplace and offered ways to promote rigor through programming in academic settings.

Coordination of Time Management

With all practices previously mentioned, there is no wonder that participants had to balance their time between different tasks and activities on a typical day: designing studies, programming, cleaning data, performing analysis, participating in meetings, presenting, or writing documentation. To manage their time effectively, participants learned how to switch between different tasks throughout the day or week, estimating how much time should be allocated to a task, prioritizing some tasks over others, and finding work balance. The practice of time management was mostly characterized as being developed on the job, with 91% of the quotes referring to the workplace (see Table 14). Six junior statisticians and both mentors mentioned this practice (see Table 23). Participants recognized that time management was challenging at the workplace because of boundaries in the division of labor, although the role of the mentor facilitated the transition. Indeed, mentors recognized ways to support junior statisticians during the transition, with Simone focusing on how she supported future statisticians in academic settings while David shared his experience mentoring junior statisticians at the workplace.

Boundaries in the Rules for Time Management. Boundaries in the rules for the practice of time management between academic settings and the workplace mainly involved handling short deadlines and unexpected priorities.

Unexpected Priorities. Unlike academic settings where assignments were announced in advance, participants had to adjust their time when some tasks needed to be prioritized unexpectedly. For example, Enid had coworkers “schedule meetings for that day at like eight o'clock in the morning and it's like, surprise, we have a meeting at 11

Table 23*Important Elements for the Coordination of Time Management*

Participant	Academic Settings	Workplace	Transition
Florence		No mentor to set priorities	Discontinuous
John		Supervisor helped set priorities	Detailing
Edwards		Mentor provided time estimates	Detailing
Janet		Supervisor helped set priorities	Detailing
Thomas	Deadlines for assignments	Juggle between two workplaces	Detailing
Enid	Deadlines and assignment weights	Balance between tasks Supervisor helped set priorities	Detailing
Simone (mentor)	Encourage multitasking and include short deadlines		Observed discontinuous
David (mentor)		Encourage tradeoffs	Observed discontinuous

o'clock.” Participants also reported having to complete tasks on short notices.

Short Deadlines. According to Simone, being able to handle shorter deadlines was the main difference between academic settings and the workplace:

When we've had people who were statisticians and that, if you don't come out of school and go into industry and get used to that, statisticians who had been used to academia where it's like, ‘Oh yeah, I'm on a research project, whatever, it's a year project, it's a 2-year project.’ And they meet milestones along the way. Then, they, they can't handle that, um, you know, 2-hour deadline kind of thing.

Indeed, Thomas found that time management was “heightened” at the workplace compared to academic settings because of shorter deadlines. Enid had discussed the importance of deadlines and the importance of having a timeline for a project in her consulting class, however, she encountered boundaries in the rules because as opposed to having the entire semester to work on a project in academic settings, she only had a very limited time at the workplace to complete tasks. She added that there were different types of deadlines at the workplace, some are firm and some are flexible: “you have to learn,

like, you have to figure out which projects have air quote deadlines that are okay to kind of like push back in precedence of like other things more important.”

After observing junior statisticians struggling with time management at the workplace, Simone decided to offer her students an authentic experience in academic settings by “giving them short timelines to see if they are able to do that.” Indeed, she valued time management skills and, during interviews she conducted with future statisticians, she focused on:

Asking them how they would approach a situation in which you were given some work to do or projects to do or what have you, and a timeline, and it's a day or two before the timeline and you don't think you're going to get it completed. What do you think you should do? Right. So that is very important for us to see how people handle that or how they think about that in using the support of their management to set their priorities or notifying them on time when they know something's not going to work out.

Indeed, Simone mentioned that statisticians were not alone and should rely on the community, with supervisors and mentors, but also collaborators.

Boundaries in the Community for Time Management. Seeking the support of the community helped reviewing priorities, especially when working with many different collaborators but also introduced boundaries with conflicting expectations.

Collaborators. Along with boundaries in the rules, junior statisticians encountered boundaries with the community at the workplace. Florence, Janet, Edwards, and John worked in the field of medicine and pharmaceutical industry and worked with many principal investigators while Thomas and Enid tended to work on fewer projects within their team at government institutions. The experience of Thomas was unique because he worked at two different workplaces and had to juggle his time between the two positions. He held a fulltime position at a government institution, and he worked remotely as a

consultant for a Law school for approximately 15 hours a week. He did not mention time management as being specifically challenging, probably because he mainly focused on the workplace for the government institution during the interview. The differences between communities can be illustrated by the example of Enid, who had “meetings with other branches, other divisions, other non-statisticians,” in comparison to Janet, who worked with many principal investigators on entirely different projects. Similarly, Florence divided her time to serve 22 principal investigators between her and two other statisticians. Balancing work on numerous projects with different collaborators probably augmented the need to develop the practice of time management and widened the gap between the expectations of academic settings and the workplace.

Expectations. The transition was even more discontinuous for Florence because, in addition to working on many different projects and holding the division of labor, she was:

Managing the expectations in terms of what I can do. And learning how to say no, that was a big one. I still am not sure I know how to say no. I, a lot of times, just direct [principal investigators] to my boss and say you have to ask [my boss] if I could do that.

Indeed, Florence was challenged by her collaborators and had to learn how to say no and set priorities which in turn, created boundaries in the division of labor.

Boundaries in the Division of Labor for Time Management. At the workplace, junior statisticians became responsible for setting priorities, keeping a timeline, and balancing between different tasks.

Setting Priorities. Florence was in charge of the division of labor when she first transitioned to the workplace. She was dividing her time on projects herself, but she quickly felt overwhelmed. After about eight months at the workplace and following an

incident that had monetary repercussions on one of her projects, she experienced a shift in the division of labor with the intervention of a senior statistician who became her mentor. Together they discussed the responsibilities and expectations for a first-year statistician. In fact, Florence's supervisor originally intended to hire a statistician with about five years of experience, but there was no adjustment to the expectations of the position even though Florence had just graduated. The mentor helped bridge the gap between the expectations of the supervisor and the experience of Florence in managing her time. After her mentor offered some guidance, Florence was "limited to a certain number of projects with time allowed" and before she would start on a new project, she needed to get approval from her supervisor to manage priorities. The transition to the workplace would have been more continuous in terms of time management for Florence if she had not started by holding the division of labor.

Indeed, Edwards was not responsible for the division of labor and followed what his "mentor recommended. With some extra padding" during the first year at the workplace. He said that it was "almost impossible" to estimate how much time he would spend on a project "because you have to know what you're capable of," "know what you can accomplish in a certain period of time." After getting some experience, Edwards gradually undertook the division of labor to estimate how many hours a project would take, which was still challenging because he worked on about 100 to 150 projects a year which could each range from eight to 200 hours. Similarly, Janet did not hold the division of labor for allocating time to her different projects because her supervisor took the responsibility to "assign or like have everybody come into the meeting and just sort of figure out [what our] availability is to work with the assigned clients," consistently

making sure her team was not overwhelmed. In fact, Janet relied on her supervisor “to recognize our limitations as far as what we were able to contribute and not get us mired in a problem that can be way more than we bargained for. Which is really nice.”

Therefore, learning time management was considered as detailing when the division of labor was gradually passed down from the supervisor to the junior statisticians, as opposed to discontinuous when the division of labor was suddenly the statisticians’ own responsibility.

Keeping a Timeline. In academic settings, the division of labor was mainly held by professors who set the rules such as deadlines and weights for assignments. For example, Enid took a consulting course for which the professor emphasized the importance of deadlines and having a timeline to work on a project. She felt that she did not have to balance the workload in academic settings nearly as much as she did at the workplace because the timeline was already set by the professors. Even though some priorities were still defined by her supervisor, it became more of her own responsibility at the workplace.

Balancing Time Between Tasks. Because the rules have changed between academic settings and the workplace, tasks needed to be completed under short notice, prioritized over other tasks, or pursued simultaneously, causing more boundaries in the division of labor. Indeed, Enid was multitasking, balancing her time between programming and meetings:

You're in the mindset and you're like, SAS, SAS, SAS, code, run, code, run. Oh, it's meeting time. So then you have to kind of like detach yourself from the [coding], go to the meeting, get your brain into a different mindset, get finished with that, come back to your computer and you're like, all right, now I have to reattach myself to the coding.

Enid learned how to juggle between tasks, but “that's definitely something that can be challenging [...] getting back into the groove after you've had a disruption like that.”

Florence also shared that she did “so much multitasking,” had some projects to do urgently, and needed to update her priorities constantly.

As a mentor at the workplace, Simone recognized that handling several projects simultaneously was challenging for junior statisticians who are just transitioning from academic settings:

People are not, you know, they're used to have like tasks, very small oriented tasks, at the university or in academia, and here maybe the project goal is wide open and they have to learn how to budget their time and things like that. And I think that's something that is um, it's harder for them to do in the light of multiple projects.

To support students developing practices in time management in academic settings,

Simone offered opportunities to her students by assigning different tasks at once.

Furthermore, she gave her students “projects at different stages or different topics.”

On the other hand, David provided support to junior statisticians at the workplace by assigning progressively more and more projects: “the first project I gave all of them was kind of a simple project” and “as they kind of grew, I tried to get them a combination of projects that I thought would be interesting for them.”

Indeed, the role of the mentor at the workplace facilitated boundary crossing for John. He did not mention receiving guidance from his mentor explicitly but he said that his mentor was monitoring “that we are doing our work accurately and responsibly” and she would probably have let him know if he was working too much. John valued work-life balance and admitted “that is a kind of a difficult thing to navigate.” He was in fact

inspired by his mentor who “may not be the most technically strong but she's a very good person in terms of respecting the work life balance.”

Boundaries in the Tools for Time Management. Participants identified different tools and strategies to develop the practice of time management at the workplace. Tools such as setting a schedule, finding time estimates, or making tradeoffs, were used to facilitate time management at the workplace.

Schedule. Enid mentioned starting her day by checking her calendar for meetings that might have been added in the last minute. Similarly, John spent “30 minutes to an hour... organizing my day prioritizing what I need to accomplish that day” while Florence created a week-by-week schedule to manage her time and got into the habit of:

Sitting down with my boss at regular intervals and asking him what he thinks should be the priorities for that week. Knowing that of course someone's going to have a house on fire project, it has to be done right now, knowing that, that might happen. Setting up just kind of a structure of what needs to be worked on in this particular week. And once I know what's the most important, I can start to shift my time around.

Janet also organized her time over the week, dividing her time between independent work and meetings. For example, “Tuesdays, we typically have time at our desk working and then we have staff meeting” and “we have like different meetings with clients throughout the week.”

Within a year at the workplace, Florence “definitely [learned] time allocation and how to prioritize multiple, multiple projects cause I didn't know how to do that.” She tried to focus on one project a day or one project in a half day because her “mind doesn't switch very easily between different data sets.”

Time Estimates. To determine time allocation for each task, Florence and Edwards had the strategy to “do the math.” For Florence, the time allowed is calculated

to “have roughly the amount of time given to the amount of salary paid,” while Edwards had a rule of thumb to estimate his time:

So let's say I had a massive machine learning project with missing data and all this kind of stuff. So I have to like stack a bunch of methodologies here. You know how many of those could I handle simultaneously in a year. Maybe five? Maybe 10. So if I say 10 okay that's .10 FTE. We need to spend – I know I need to bill x amount of hours in a year to meet my quota. So 10% of that. And then that helps me find a budget. Or from a weekly standpoint of, I think I'd work on this, I have time to work on this project one day a week. Okay, what's 20% of a week. And how long is this project is set to last for? If it's six months, Okay, great. That's .1 FTE again. Things like that.

Calculating time estimates based on salary or annual working hours, contrasted the rules in academic settings, for example with Enid who considered that the weights attributed to assignments gave her a sense of what needed to be prioritized and she “might [have] put a little more importance on something that might be weighted more.”

Tradeoff. Constrained by time, some participants used tradeoffs to complete tasks on time. For example, John noticed that his coworkers made some concessions by deciding to “put time outside of work hours to really dedicate to their work and projects.” While working more hours was not ideal, Enid weighted different tools to complete a task and adjust her priorities. She shared an anecdote about one specific task:

I've got somebody that needs this report done by the end of the day. They need these estimates done by five o'clock into business. I don't get time to sit around and try and research how can I find the code in SAS.

To complete this task, she used a combination of Excel and SAS operations instead of writing a SAS macro because she needed to prioritize time over technical efficiency. Indeed, researching the best solution for writing a SAS macro would have taken too much time and completing the task was the priority. She eventually wrote a SAS macro

for that project when she had more time to dedicate to it. From his perspective, David talked about being flexible as making “tradeoffs”:

Sometimes I will sacrifice what might be the best analysis for something that can actually get done and have similar results. I think really understanding the difference between statistically significant and practically significant, you know, is really been something that I had to learn on the way.

Indeed, David believed that at the workplace “time is valuable” while in academic settings, students “were told certain time limits, ... were told to use something, some sort of technique or try to get some sort of conclusion” and they have time to research the best solution. As a mentor, David mentioned how he encouraged a junior statistician on his team making concessions to save some time:

He was very caught up in trying to do the best analysis possible and with deadlines you can't always do that and sometimes even the best analysis, you don't really get the best results. But if you get some results, it's better than no results, otherwise the people that are making decisions are making bad decisions based on no evidence as opposed to at least we can give them a little guidance. And I think really trying to teach that took a little bit of time and a lot of effort on my end to really just kind of push him and say ‘Okay, I know this isn't the greatest result and I know it's not the most meaningful, but it's something and they just need something, they need some guidance. It's better than no guidance.’

David was able to identify what his employee was struggling with and provided support to help him realize the different expectations between academic settings and the workplace. In academic settings, the rules usually require following directions and implementing the best technical solution whereas at the workplace the rules of deadlines and getting the task completed prevailed on the technical aspect. Finding a balance between these rules is what David called making “tradeoffs.”

Transition of Time Management. Because participants were responsible for completing many tasks at the workplace, they learned how to prioritize their time, weighing their time between different activities. Rules created boundaries in the

transition to the workplace with meeting short deadlines or handling projects that had to be prioritized unexpectedly. As the division of labor seemed to be an important element in the transition to the workplace for the practice of time management, all participants referred to the role of a mentor, except for Thomas, and particularly impacted the transition of Florence who was experiencing discontinuity on her own. Both mentors considered that junior statisticians typically experience discontinuity, however, most participants underwent a detailing transition with the support provided by their mentors.

Summary of Coordination

By reflecting on the transition from academic settings to the workplace, the analysis of the interviews of junior statisticians and their mentors revealed eight categories of statistical practices (see Table 24). Participants coordinated the elements between the two activity systems to reveal boundaries involved in the transition and their experiences were qualified as continuous, detailing, or discontinuous.

Table 24

Level of Continuity for Each Practice and Principal Boundaries

	Continuous	Detailing	Discontinuous	Boundaries
Programming	7	3	0	Rules
Analysis	6	4	0	Division
Design	3	5	2	Division
Time Management	0	5	3	Rules
Communicating	2	5	3	Community
Collaboration	3	3	4	Community
Data Management	4	2	4	Rules
Documentation	1	0	4	Rules

Note. The level of continuity of a practice is shown by how many participants had continuous, detailing, or discontinuous experiences for each practice.

Participants generally experienced continuity with *Programming* and *Analysis*, while the practices of *Collaboration*, *Data management*, and *Documentation* were mostly discontinuous across participants. Boundaries were often created in the rules, directing the division of labor, and set by the community.

Some rules had an impact across several practices. For example, *decision making* and *domain knowledge* created boundaries because statistics were used to inform decisions and contributed to significant domain research at the workplace while in academic settings, the assignments were often decontextualized and the difference between a correct and incorrect answer had limited impact, only affecting students' grades. Other boundaries in the rules encountered at the workplace included *time* and *big data*. Indeed, time constrained the practices of programming, data management, and analysis, weighing tradeoffs to complete tasks in a timely manner. Dealing with big datasets was also time consuming and required specific tools and techniques.

Recurring boundaries in the division of labor concerned the ability to choose the appropriate methodology, analysis technique, or program which junior statisticians often had no experience with as they entered the workplace because professors were holding the division of labor in academic settings. Another boundary was introduced as participants assumed different roles at the workplace, besides being statisticians, they also took over responsibilities related to the work of programmers and data managers.

Furthermore, working within interdisciplinary teams or with domain experts interposed boundaries in the community. Students and professors built common knowledge in academic settings while engaging in statistical activities while at the

workplace participants interacted with collaborators with no background in statistics for example.

To overcome boundaries during the transition, the community at the workplace sometimes offered training opportunities for professional development. But first and foremost, tools acquired in academic settings facilitated the transition for junior statisticians, indicating the importance of the education of future statisticians. Indeed, learning a variety of programming tools or analysis techniques in academic settings sustained the development of practices at the workplace. Next, the triangulation of the three data sources formulates recommendations for the transformation of statistical practices in academic settings.

Triangulation and Transformation of Statistical Practices

Besides ensuring the validity of the findings, the triangulation of the identification, reflection, and coordination of statistical practices through three data sources confronts different perspectives on statistical practices. The triangulation also reveals transformations suggested by members of the statistical community. I connect and contrast the importance of statistical practices identified through the sorting task, the reflection on developing practices in academic settings or at the workplace, with the degree of continuity for the transition of each practice. By recognizing elements from each system that facilitated boundary crossing, participants formulated recommendations to transform statistical practices in academic settings.

Triangulation of Categories of Statistical Practices

Findings from the three data sources generated different categories of statistical practices (see Figure 15). First, the list of 24 statistical practices that I created based on a

review of the literature and refined by my own experience and an overview of job offers, introduced five main categories of practices with *Design*, *Data*, *Analysis*, *Interpersonal skills*, and *Personal skills*. Participants identified the most important practices for the role of statisticians at the workplace with the sorting task. Second, members of the statistical community were asked to reflect on their experience with no predetermined categories and they unrestrictedly shared practices they learned in academic settings and practices they developed at the workplace. By analyzing their survey responses, I constructed seven categories of practices. Finally, I organized the findings of the interviews in eight categories, summarizing the coordination of practices.

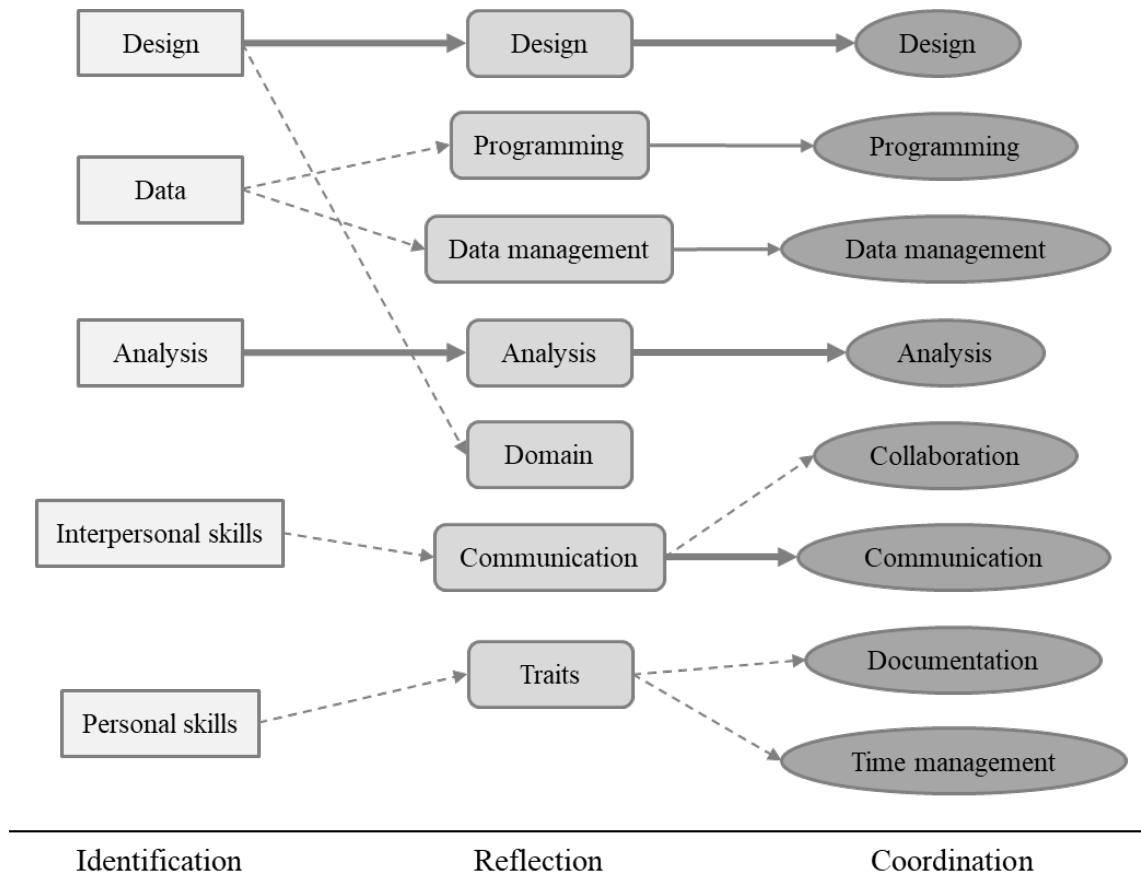


Figure 15. Connections between categories of practices across findings.

Design and analysis practices were present in the three categorizations but the connections between other categories of practices require further explanations. After the identification of the importance of interpersonal skills, members of the statistical community specifically emphasized on communication practices on the survey, establishing its own category. Furthermore, the insights shared by participants during the interview made the distinction between collaboration and communication practices. Collaboration practices were included in previous categorizations but not with enough importance or detail to separate from the categories of interpersonal skills or communication practices.

Categories of programming and data management practices both appeared in the findings for the reflection and coordination of statistical practices. These two categories were combined within the category labeled *Data* for the identification of important practices. Because participants focused on programming and on data management in their survey responses, these practices produced their own categories.

Similarly, practices involving domain knowledge which were first identified within design practices were sufficiently mentioned in the survey to justify a category. However, through the interviews, I found that domain knowledge operated as a rule across several practices and the category of domain knowledge was no longer supported.

Finally, some personal skills such as “Being skeptical /critical,” “Meeting deadlines,” or the additional practice of documenting work mentioned in the sorting task, appeared in the survey in the form of traits. The interviews gave a thorough description of traits related to the practices of documentation and time management, explaining the last two categories of practices arising from the analysis of the interviews.

The connections between categories of practices show the evolution of the identification of important statistical practices focusing on the perspectives of the members of the statistical community. Therefore, the eight categories of practices that emerged from the deepest level of introspection with the interviews of junior statisticians and their mentors constitute the main findings of this study. Next, I focus on the recommendations formulated by participants to better promote these eight practices in academic settings.

Triangulation of Statistical Practices

Following the triangulation of the categories of practices, I cross-examine the three data sources for each of the eight main categories, to summarize the identification, reflection, and coordination of practices. I conclude with transformations suggested by participants.

Triangulation of Design Practices

Practices associated with design appeared in the sorting task, with three statements contained in the list of existing practices and one that was added by members of the statistical community (see Table 25). Specific activities involved with designing studies were given by the analysis of the survey with choosing the appropriate question, measures, and techniques. Interviews with junior statisticians and their mentors further revealed the importance of the division of labor at the workplace.

Identification of Design. The members of the statistical community ranked designing studies in general relatively high with the sorting task but “translating a real problem into a statistical form” was the fourth most important practice, following practices of communication and analysis (see Figure 10). Survey responses and the

Table 25*Statements on the Sorting Task Related to Design Practices*

Practice	Category Sorting Task	Mean
Translating a real problem into a statistical form	Design	6.3
Designing studies / Aligning design with research goals	Design	5.2
Preparing sampling frames / Drawing samples	Data	3.5
Using power analysis for sample sizes	Added	5.7

interviews reinforced the importance of determining the appropriate question, especially while communicating to domain experts. Even though the preparation for sampling was listed for the sorting task, three participants specified that they had to calculate sample sizes using power analysis, which was also mentioned in the interviews.

Reflection and Coordination of Design. Comparing practices mentioned in the survey, more participants developed how to choose the appropriate method at the workplace. Indeed, junior statisticians reported experiencing boundaries in the division of labor specially to plan for the appropriate analysis, but also for formulating questions, defining measures, and writing protocols. Boundaries in the rules added to the challenge of the transition, qualifying most experiences as detailing. Rules changed between academic settings and the workplace, mainly with the implications based on the results of a study, facing real consequences. Mentors at the workplace particularly insisted on the importance to recognize these implications and offered opportunities to develop the practice since junior statisticians tended to focus more on the analytical plan in academic settings.

Transformation of Design. Mentors recommended to give students opportunities to design their own studies in academic settings, yet they should not only pursue their own interests but contribute to significant research. For example, in the interviews, junior statisticians recognized some opportunities by taking consulting courses and working with clients to design their research. In fact, one statistician mentioned on the survey:

It was a requirement of my program in graduate school to spend time volunteering in our statistical consulting lab, helping professors and other graduate students on campus with their research. I think a focus on getting students involved in situations like this more would be helpful.

This experience emphasized that students need to hold the division of labor in academic settings in order to have a continuous transition. Courses on experimental design, clinical trials, or sampling were found to facilitate the transition between academic settings and the workplace but sharing the division of labor with consulting or project opportunities were utterly valuable. In her interview, Florence mentioned a mockup proposal she did for a course on study design that she kept referring to at the workplace to help her write grants for instance. She also used case studies introduced in an ethics course to have examples of study designs. Indeed, an educator of statistics recommended on the survey to “introduce more ill-structured problems so that students are forced to grapple with ‘real’ situations that would occur in the workplace.”

Triangulation of Programming Practices

There was only one statistical practice in the sorting task that referred to programming and specific tools for programming were mentioned in the survey. It was only through interviews with junior statisticians and their mentors that the complexity of this practice was revealed, involving more activities than writing code.

Identification of Programming. In the sorting task, members of the statistical community ranked “Using statistical software package / Writing computer programs” in the first quartile of most important practices with a mean score of 5.8 (see Figure 10). Participants indicated that they learned programming in academic settings, being the second most common practice they cited in the survey. Programming was considered as a technical skill related to the category of Data at the beginning of my study, but after analyzing the survey, it became its own category because of how often participants cited it. Indeed, programming was the most cited category of practices in the interviews.

Reflection and Coordination of Programming. Both in the survey and interviews, participant specifically mentioned learning how to code with R and SAS or using statistical software to perform analysis in academic settings. They kept learning at the workplace, with more diverse tools such as R, SQL, or Python, mostly teaching themselves. The need for new tools emerged from dealing with rules at the workplace, with requirements for specific software and limitations in power, cost, and time. Survey respondents and interviewees also realized the importance of documenting their code at the workplace. However, that practice was even more emphasized in the interviews, forming its own category since it not only encompassed documenting code but also documenting other procedures. Junior statisticians and their mentors highlighted that they assumed different roles for programming: being statisticians to perform analysis, being data manager to maintain and extra data out of databases, but they also had to think like a programmer.

Transformation of Programming Practices. One of the principal recommendations from junior statisticians was to develop the logic of programming. To

instill this logic of programming, participants identified computer science courses as particularly beneficial. They also recommended to learn programming with different applications, in the context of data management, with data cleaning or with big data for example. Indeed, mentors found that junior statisticians mostly experienced a continuous transition for programming but were lacking writing program for other tasks than performing analysis. One statistician in the survey emphasized: “most [educational] programs teach theoretical statistics alongside code but continue beyond that to realize that data handling needs to be considered.” In her interview, Enid referred to a course in data management with SAS that was very valuable for her career. Before writing code though, she recommended to start exploring statistical software that are more user friendly with drag-and-drop menus:

An introductory to statistical programming, you know, be it in SAS or in R, I think, you know, instead of just diving headfirst and being like, all right guys, we're going to do some statistical programming with this software. Um, we're going to do, you know, mixed models with SAS, you know, but you have no experience with SAS whatsoever. Go.

Besides learning progressively to program, participants praised for a diversity in programming tools and give students the opportunity to learn many software in academic settings. Furthermore, John advised future statisticians to think about what activities they would like to engage on at the workplace and choose their educational path accordingly. He found differences in activities between statisticians who obtained a master's versus a doctoral degree:

If they're master's level statisticians, I would tell them to be careful when they're applying. Knowing that a lot of master level statisticians - I've spoken to some of my classmates - a lot of them also do a lot of statistical programming work and be aware that could be a possibility for you and if that excites you, do you like programming, and doing that, great. But if that's something that doesn't really excite you naturally be very aware and when you're interviewing for jobs and

different companies really kind of ask them and ask them about what a master level statistician does.

Triangulation of Data Management Practices

Data management practices were included in the sorting task, appeared in the survey, and were described in the interviews. There were three statements in the list of 24 practices for the sorting task that referred to data management, with data cleaning, data collection, and databases (see Table 25). Members of the statistical community mentioned having learned data management in academic settings and at the workplace with equal proportion (see Figure 10). However, junior statisticians and their mentors indicated that data management practices were mostly learned at the workplace.

Table 26

Statements on the Sorting Task Related to Data Management Practices

Practice	Category Sorting Task	Mean
Cleaning data / Managing missing data	Data	5.0
Collecting / Gathering data	Data	3.9
Creating / Maintaining databases	Data	3.2
Accessing and storing data	Added	6.0

Identification of Data Management. Data management practices such as cleaning data, handling missing data or extracting data were described in the interviews of junior statisticians and their mentors, after being identified on the survey by members of the statistical community. The practice of data cleaning had a mean score of 5 across all participants, indicating a relative importance. Indeed, in the survey, most practices of data management referred to data cleaning both in academic settings and at the workplace. Even though “Creating / Maintaining databases” was ranked as one of the

least important practices through the sorting task (see Figure 10), five participants added the practice of “Accessing and storing data” to the list, in the higher end of the distribution. In addition, junior statisticians shared in the interviews that they were responsible for pulling data out of the database at the workplace.

Reflection and Coordination of Data Management. According to the survey responses, the proportion of learning practices associated to data management was approximately the same in academic settings and at the workplace, respectively representing 7% and 8% of all practices. However, junior statisticians recognized mostly discontinuous or detailing transitions to the workplace, in terms of data management practices. They encountered boundaries in the rules at the workplace by having to deal with big data, databases, and messy data, which they had very little experience within academic settings because professors tended to provide clean datasets for instance. Therefore, they had not necessarily acquired the appropriate tools such as programming for the purpose of data cleaning, dealing with databases through cloud environments, or using strategies to handle missing data, and courses supporting some of these tools were found to be particularly beneficial.

Transformation of Data Management. Practices related to data management were the third most common recommendation on the survey. Members of the statistical community suggested to promote data cleaning in the education of statisticians, dealing with messy data or big data, and how to store data. For example, one statistician proposed to “focus on solving problems start to finish, including data cleaning.” Florence added in her interview that education should promote data cleaning across the curriculum, implemented in different courses. To offer such opportunities, Simone included real data

in her graduate courses for her students to experience data cleaning. However, one professor on the survey suggested to incorporate messy data “earlier in their education,” as John experienced in his introductory course in statistics with data cleaning assignments. Indeed, many participants on the survey mentioned using real data in academic settings and one graduate student wished to have had opportunities to be “dealing with messy or missing data.” In addition, to give experience with big data, one statistician recommended to develop and “teach classes on not only what to do with small amounts of data (fisher vs chi squared tests) but also large data.” Findings from the interviews gave more insights about tools junior statisticians learned to deal with big data such as SQL queries to extract data from cloud databases. Concerning databases, data storage and data quality, one graduate student wondered “how should the data be stored? How can you ensure and screen for data quality in the way that you collect and store the data?” Both mentors shared in their interviews that they trained junior statisticians on their database system as they transitioned to the workplace. Therefore, besides promoting data management practices in academic settings, mentors provided support at the workplace for junior statisticians.

Triangulation of Analysis Practices

The three data sources included practices related to analysis. Five statements of the sorting task represented statistical practices in the category of analysis with applying, researching, and developing new statistical methods and techniques as well as using advanced mathematics and interpreting data (see Table 26). At the beginning of my study, the practice of “Interpreting data” was considered as part of the category “Data” in the sorting task, however, through the survey and interviews, participants identified this

practice as part of the analysis category. In the survey, analysis was the most common practice acquired in academic settings while also mentioned at the workplace but with less emphasis (see Figure 11). In addition, it was the second most recommended practice to transform in academic settings. The category of analysis was also described in the interview, being the third most cited practice across the ten interviewees.

Table 27

Statements on the Sorting Task Related to Analysis Practices

Practice	Category Sorting Task	Mean
Applying statistical methods and techniques	Analysis	6.5
Interpreting data (Limitations of methods / Bias)	Data	6.3
Researching appropriate statistical methods and techniques	Analysis	5.5
Developing new statistical methods and techniques	Analysis	3.5
Using advanced mathematics	Analysis	3.2

Identification of Analysis. The practice of “Applying statistical methods and techniques” was ranked as the second most important practice according to the members of the statistical community on average (see Figure 10). Interpretations also received a high mean score in the sorting task, being the third most important practice. Indeed, in the interviews, mentors particularly emphasized that interpreting the results of an analysis was crucial for statisticians in order to provide insights and make informed decisions.

Reflection and Coordination of Analysis. Even though some participants mentioned learning interpretations in academic settings through the survey, only 6 out of 79 quotes about analysis actually referred to interpreting results. Indeed, the interviews of junior statisticians and their mentors showed that they tended to develop this practice at

the workplace. The most important boundaries for analysis concerned the division of labor, with professors setting techniques to use in academic settings, and statisticians having to choose the appropriate technique to apply to their datasets at the workplace. Statisticians were also constrained by the specific branches of statistics required by the community. Tools acquired in academic settings facilitated boundary crossing, making analysis one of the most continuous practices. Indeed, on the survey, members of the statistical community reported having learned advanced mathematics and probability theory, while at the workplace, they mentioned specific branches of statistics such as survival analysis, time series analysis, or topics in data science.

Transformation of Analysis. Through the analysis of the interviews, it was suggested to encourage students to hold the division of labor in academic settings by choosing the appropriate technique, for example. On the survey, members of the statistical community recommended to give opportunities to apply analysis techniques in real context, advising for “more emphasis on practical applications, from theory to real use cases.” In addition, practices such as checking assumptions, choosing the appropriate technique, or interpreting results, were suggested to be promoted in academic settings. One graduate student mentioned to “stress validation of results” while one professor advocated for “more emphasis on how to interpret/use results.” At the end of his interview, David recommended including long-term projects with emphasis on interpretations and decisions rather than techniques:

Maybe longer-term projects that, you know, where the emphasis really is on [interpretations] and less on the technical aspects.... Something you think about as deeply when you're actually in a real-world situation where what's actually important is the end. What does it actually mean?

Participants weighted the importance of theory on the survey, with as many recommendations to include more theory as to include less theory in the education of statisticians. Participants also suggested to offer courses in more specific fields such as survival analysis or time series. Indeed, Janet shared in her interview that:

I would have liked to have taken more on specific for clinical trial data, specific for high dimensional data, specific for time series experiments, survival analysis, things like that. I don't typically do, but it would be nice to have those in my wheelhouse, a little more.

Triangulation of Collaboration Practices

Practices associated with collaboration appeared twice in the sorting task, and a third statement was added to list of existing practices by members of the statistical community (see Table 27). The analysis of the interviews revealed specific activities involved in collaboration such as working with domain experts, understanding a problem, or explaining results, and in particular making the distinction with communication.

Table 28

Statements on the Sorting Task Related to Collaboration Practices

Practice	Category Sorting Task	Mean
Consulting / Working with a client to solve a problem	Interpersonal	5.6
Participating in teams / Collaborating	Interpersonal	5.6
Managing project / budget / people	Added by 3	5.0

Identification of Collaboration. Boundaries in the community identified in the interviews of junior statisticians and their mentors reflected the two statements that were associated with collaboration in the sorting task, which were ranked above average in the sorting task, revealing their importance. Some other aspects of collaboration were also

added to the sorting task within the category of “managing project / budget / people” with quotes referring to consulting with other departments in the company, developing organization-wide strategy, and explaining results. and specifically referred to working with domain experts. Other boundaries were mentioned by statisticians and their mentors, with rules framing relationships and attitudes.

Reflection and Coordination of Collaboration. Comparing quotes in the survey, learning collaboration seemed to occur at the workplace with consulting with clients, or domain experts, and participating in teams. Developing working relationships with domain experts were ruled by managing expectations and listening to their needs, diplomatically. Mentors at the workplace considered that collaboration was not the responsibility of junior statisticians and they had to develop this practice progressively. However, some junior statisticians identified consulting courses in academic settings to facilitate the transition to the workplace, especially if interactions with domain experts were involved.

Transformation of Collaboration. Members of the statistical community and more specifically junior statisticians and their mentors recommended to include opportunities to foster collaboration in academic settings. Indeed, practices associated with collaboration represented the highest number of recommendations in the survey. Overall, participants suggested consulting courses or laboratories, offering statistical support within the university or to local organizations for instance. One graduate student also mentioned peer tutoring:

New graduate students should have opportunities to peer tutor graduate students in other fields. This is extremely rewarding for both tutor and tutee. The tutor begins to learn about communicating statistical content on a variety of levels with

real people. As they begin to master these skills, they can then go on to peer statistical consulting.

Florence focused on the importance to work with interdisciplinary contexts and recommended to offer courses that gather students with different domain expertise:

I really think a communication class, like how to work in a team, would be incredibly helpful, especially a class that isn't just statistics students or math students, but instead has students from all over the college who have different backgrounds and have to work together and collaborate cause learning how to work in a team with people who don't speak the same technical language as you is really hard.

Next, I share recommendations that were focusing on one particular aspect of collaboration, which is communication.

Triangulation of Communication Practices

Communication practices were included in the three data sources. There were four statements related to communication in the sorting task, with written and oral communication, data visualization, and communication with non-statisticians (see Table 28), originally classified as interpersonal skills and associated with data. On the survey, communication was the most common practice developed at the workplace and received less focus in academic settings (see Figure 11). Communication was also the second most practice cited in the interviews across eight junior statisticians and their mentors.

Table 29

Statements on the Sorting Task Related to Communication Practices

Practice	Category Sorting Task	Mean
Communicating interpretations of statistics to non-statistical audiences	Interpersonal	6.7
Producing visual representations of data	Data	5.5
Communicating in writing / Writing reports	Interpersonal	5.5
Communicating orally / Making presentations	Interpersonal	5.4

Identification of Communication. Practices associated with communication created differences between the two main perspectives of the statistical community, and “Communicating interpretations of statistics to non-statistical audiences” received the highest mean score across all participants (see Figure 10). Indeed, interactions with non-statisticians were also mentioned in the survey and extensively referred to during the interviews. Other statements were also ranked higher than average in the sorting task.

Reflection and Coordination of Communication. According to the survey analysis, even though communication practices such as presenting and explaining statistics to non-statisticians were quoted in academic settings, they were heightened at the workplace, being the first category of practices developed at the workplace. Indeed, mentors at the workplace observed that junior statisticians usually experienced a discontinuous transition and provided support to cross boundaries. The main boundary affecting the transition concerned the audience, with communication with non-statisticians. In addition, interacting with domain experts or upper management for example also affected the rules with a focus on decision making and on the domain knowledge. Junior statisticians identified their transition as mostly detailing and developed strategies at the workplace to better communicate.

Transformation of Communication. Mentors at the workplace shared how they support the transition of junior statisticians by providing the opportunity to rehearse and emphasizing on the importance of interpretations for decision making rather than technicalities. By taking a course at a conference, Stella realized that professors who taught that course:

They actually have classes at their respective universities on this. And I just, I think it's very, um, it, it was important that you know, students learn how to

communicate, and they talked a lot about body language and things like that. And it would be good to have that. You know either in undergraduate courses or even, you know, even in graduate courses.

Indeed, David mentioned that his minor in communication from his undergraduate degree was very beneficial for presenting his work. Enid also suggested courses in public speaking. One professor shared on the survey to emphasize communication in “higher-level courses, I think it's more appropriate to expect students to write and present on their findings. Communication/ consulting is an important skill for students that they need many opportunities with to get comfortable.” While there were other quotes related to communication through the survey, participants mainly recommended to foster communication in academic settings but with very few suggestions about how to promote an authentic experience. The interviews of junior statisticians revealed some strategies such as using data visualization with simple graphics, avoiding statistical jargon and equations by using layman’s terms, and practicing. Tools outside of academic settings and the workplace were also found to support communication skills, by practicing with friends and family to learn how to explain interpretations to non-statisticians, or by having experience in customer service to learn how to listen.

Triangulation of Documentation Practices

Writing documentation about was added to the list of statistical practices in the sorting task. This practice appeared again in the survey under the categories of programming and communication at the workplace. However, junior statisticians and a mentor revealed boundaries at stakes at the workplace for this practice, complying with audit needs and fostering collaboration.

Table 30*Statements on the Sorting Task Related to Documentation*

Practice	Category Sorting Task	Mean
Participating in teams / Collaborating	Interpersonal	5.6
Working independently	Interpersonal	3.5
Documenting work	Added	4.3

Identification of Documentation. Members of the statistical community identified the practice of documentation in the sorting task and as being developed at the workplace in the survey. In Table 30, I included two practices that impacted writing documentation, comparing the importance of collaboration versus independent work at the workplace. Working independently was ranked as one of the least important practice as opposed to participating in teams in the sorting task (see Figure 10). Indeed, in the interviews, participants mentioned that documentation fostered collaboration and controlled for independent work with regulatory audit. On the survey, one statistician described the importance of documentation at the workplace:

Being very diligent with organization of datasets and code. While working on many projects concurrently, some of which span multiple years, it is necessary to have a system of labeling and commenting that allows work to be carried out efficiently, accurately, and thoroughly. This organization is also extremely helpful (and even necessary) for regulatory auditing purposes.

Reflection and Coordination of the Documentation. According to the survey responses, the practice of documentation was solely developed at the workplace while the interviews provided one instance of a continuous experience. Indeed, acquiring tools such as online repositories and assuming the division of labor in academic settings allowed for a continuous transition while most experiences were qualified as discontinuous.

Perspectives of participants provided recommendations to transform the practice in academic settings.

Transformation of Documentation. As a mentor in academic settings, Simone promoted the practice of documentation in her courses, in particular for programming and being able to collaborate on code. She said:

I teach them really, well we have to look at it from an industry perspective, um, looking at good standardized coding. It is a good practice so that, you know, it's exchangeable between persons.

Indeed, as a student of statistics, Janet was given the opportunity to learn how to document her code in a course about machine learning to be able to collaborate with non-statisticians. She practiced using Jupyter Notebook, and her professors taught students to:

In our code, make a lot of comments. Write out why we did something explain our choices, but I feel like their point of that was to say, like, 'you're going to have to do this all the time for people who have no statistical background. So get used to it.'

The lack of guidance and tools in academic settings created boundaries for Gertrude resulting in a discontinuous experience. To provide support for future statisticians, she mentioned sharing her experience with students at the university she graduated from, to inform them about the expectations of the workplace, including using GitHub:

I'm trying to do that for the masters' students at [my former university] now, going back and speaking on panels, saying 'it's all well and good but learn these things now to make your life much easier.' I'm hoping it's having an impact and it's helping people, but we'll see in a few years what they say about it, I guess. I don't know if I am doing the right thing. But also you know looking at code samples from working professionals, and after work in a business setting for example. Anything like, 'You should write your code like this. Because it's easier to collaborate with people.'

Therefore, the principal recommendation from participants for the practice of documentation was to establish best practices using sample codes written at the

workplace, informing students about the existence of this practice, and developing tools such as online repositories in academic settings.

Triangulation of Time Management Practices

Because statisticians engage in numerous activities at the workplace, participants have identified the importance of time management. Indeed, time was considered as a rule at the workplace, constraining the development of several practices. This practice was briefly touched upon by members of the statistical community in the sorting task and survey while it was mentioned extensively during the interviews of junior statisticians and their mentors. Indeed, participants identified boundaries and shared how to overcome them.

Identification of Time Management. In the list of 24 statistical practices, there was only one statement that was related to time management with “Meeting deadlines,” which was ranked in the lowest half of the practices. However, some participants added the practices of “project prioritization” and “tracking time spent on projects” on the sorting task, which fell into the category of managing projects. On the survey, three participants identified practices that were developed in academic settings such as “workflow skills,” “work organization,” and to “submit report on time.” In comparison, two practices were developed at the workplace with “workflow management” and how selecting an analysis might be constrained by time. Junior statisticians and their mentors identified this practice to be particularly challenging and highlighted the role of the mentor in the transition between academic settings and the workplace.

Reflection and Coordination of Time Management. Some elements from academic settings were identified to develop the practice of time management, such as

deadlines for assignments and their weights in the calculations for the final grade.

However, participants encountered boundaries in the rules at the workplace with short deadlines or dealing with last minute priorities. The role of the mentor was crucial in facilitating the boundary crossing to provide guidance and estimates on how much time to allocate to each task. In order to manage their time efficiently, participants planned their schedule for the day or for the week, found strategies to guess time estimates, and learned to make some tradeoffs on the choice of the analysis technique or the choice of programming tools.

Transformation of Time Management. No recommendation was suggested in the survey, however, Simone provided examples of how she supported time management practices in the graduate courses that she taught. For example, she assigned students to work on different projects simultaneously, giving them the opportunity to practice how to manage priorities. She also gave assignments on very short deadlines. Indeed, she explained that for her course:

[Students] need to be working on two product launches at one time. So they'll have maybe analysis that they have to do for this product and a different analysis they have to do for another project and they may be going on in parallel. So, um, they would have to be able to complete and work on two separate projects. Maybe it's totally two different domains and be able to support that. So they're supporting maybe two functional teams, cross-functional teams, a scientist and then another scientist for all the projects.

Summary

By engaging members of the statistical community, I not only identified practices but gave insight about their relative importance to one another. Members of the statistical community identified interpersonal skills such as communication to be crucial for the role of statistician at the workplace. The reflection and coordination of statistical practices

revealed ways to transform practices in academic settings from the perspective of the community and especially called for transformations in interpersonal skills. In the last chapter, I share my own perspective on these findings and discuss future research to be conducted.

V. DISCUSSION

A misalignment between statistical practices developed in academic settings and statistical practices performed at the workplace has been identified (Harraway & Barker, 2005), highlighting practices at the workplace that were not included in academic settings (Van der Berg, 2017). As a result, junior statisticians learn practices at the workplace and some topics for further training were proposed (Hijazi et al., 2019). However, what aspects of academic settings and the workplace sustain learning during the transition needed to be investigated. Thus, the main objectives of this study were to identify important statistical practices at the workplace, to reflect how these practices are coordinated between academic settings and the workplace, and to propose transformations of these practices in academic settings. Findings were presented in Chapter 4 and were based on a survey of 154 members of the statistical community and on the interviews of eight junior statisticians and two mentors. This chapter includes a discussion of the findings, how they address the research questions and align with current research. I also offer implications and provide recommendations for future research to support learning during the transition between academic settings and the workplace for statisticians.

Discussion of Findings

In the previous chapter, I presented two main perspectives from members of the statistical community on the importance of statistical practices at the workplace and described in detail emerging practices from the interviews of junior statisticians and their mentors. The eight categories of practices included design, programming, data management, analysis, collaboration, communication, documentation, and time

management. Next, I discuss how identifying important statistical practices and their transition to the workplace contribute to current research.

Important Statistical Practices

Previous research on statistical practices at the workplace focused on establishing the most common practices statisticians engage in, given a predetermined list of practices (Harraway & Barker, 2005; Hijazi et al., 2019). For these two studies, participants reported if they engaged in a practice at the workplace by answering yes-or-no type of questions. Most common practices identified by Harraway and Barker (2005) included analysis, reading, writing reports, and design practices. Findings from Hijazi et al. (2019) confirmed the prevalence of writing reports and performing analysis, followed by designing studies and survey. Even though these practices were frequently identified, there was no indication on what practices had the greatest influence on the role of statisticians. To address this gap in the literature, I engaged members of the statistical community in a sorting task of 24 practices. Participants revealed the importance of practices in communication, followed by practices in analysis, in design, and in programming. These results align, to some extent, with findings aforementioned, but add how statistical practices are regarded relatively to one another.

Besides addressing the relative importance of practices, the methodology used for my study allowed to consider different perspectives on the role of statisticians at the workplace by taking into account that not all members of the statistical community may share the same point of view. I summarized the different points of view with two main perspectives that emerged from the community, which contrast the importance of communication practices with analysis practices. The perspective representing most

participants considered communication and collaboration practices such as communicating with non-statisticians and working with clients as prevailing practices, as opposed to the second perspective which emphasized analysis practices such as choosing appropriate techniques or developing new ones. Participants contributing to each perspective did not differ in terms of their role, as statisticians, educators, or students of statistics, nor across disciplines.

To give statisticians the opportunity to identify practices at the workplace, Van der Berg (2017) included an open-ended question on the survey intended for interns in official statistics. Prevalent practices included communication along presenting and writing reports, programming as well as data collection, which were consistent with more systematic research studies (Harraway & Barker, 2005; Hijazi et al., 2019). Similarly, I encouraged members of the statistical community to share additional practices in the sorting task, but also through open-ended questions on a survey, and with a discussion during interviews. Since practices are constantly evolving and increasingly including topics in data science, there was a need to update the list of practices from the perspective of the practitioners.

More importantly, findings from my study revealed new categories of practices with data management, collaboration, documentation, and time management. Some aspects of data management were only emphasized with designing databases (Osman & Ismail, 2009) or data collection (Van der Berg, 2017) while there was absolutely no reference to practices of documentation or time management in the studies mentioned above. While communication practices were included in all studies, they focused primarily on presentation and writing, rather than communication practices that foster

collaboration. Collaboration only appeared in the list of practices suggested by Harraway and Barker (2005) with “understanding a consultant” (p. 47). The lack of collaboration practices may be due to the fact that most of these studies were conducted in the public sector, and more precisely in the domain of official statistics, which does not usually involve collaboration with clients or domain experts as shown by the experience of one of the participants in my study. To prevent from concentrating on certain profiles of statisticians, I recruited participants across various sectors and disciplines.

During the interviews, participants mentioned programming practices as taking an important portion of their time at the workplace and considered programming as a tool for performing many other practices. Participants also reported spending a consequent amount of time on data management practices with cleaning messy data, handling missing data, or extracting data from databases. They contrasted the importance of programming and data management practices with respect to time to practices in analysis. Indeed, participants expected to conduct more analyses than they actually did while previous research mostly focused on specific techniques for analysis. For example, Harraway and Barker (2005) identified a misalignment concerning 46 statistical techniques while statisticians surveyed by Osman and Ismail (2009) and Hijazi et al. (2019) felt mostly prepared for analysis practices at the workplace. Participants in my study also reported mostly continuous transitions for analysis practices compared to other practices. The findings suggest that the misalignment between practices developed in academic settings and practices required at the workplace lies somewhere else.

Transition of Statistical Practices

The fact that junior statisticians need to keep learning as they transition to the workplace was recognized by previous research and motivated my study. As just mentioned, Harraway and Barker (2005) identified a need to learn specific techniques of analysis while statisticians studied by Hijazi et al. (2019) demanded further training for practices in data management, survey design, and programming. Besides data management practices with creating databases and design practices with defining metrics, statisticians acknowledged not being prepared for communication practices with designing posters or leading discussions (Osman & Ismail, 2009). Van der Berg (2017) also identified practices in communication, among presenting and writing reports, that were not promoted in academic settings prior to the transition to the workplace. In agreement with these findings, the results of my study indicate that junior statisticians keep learning practices in the area of analysis, data management, design, programming, and communication, but also further develop practices in collaboration, documentation, and time management. Even more importantly, my study shed some light on how these practices are developed.

The main objective of my study was to investigate how statistical practices transition between academic settings and the workplace and identify elements that support learning for junior statisticians. Previously, there was no description for how statisticians develop appropriate practices, except for the indication that they read articles, attended conferences, took training courses, or were supported by mentors at the workplace (Hijazi et al., 2019; Van der Berg, 2017). The interviews of participants in my study not only revealed elements to support learning of a practice but also what aspects of

this practice were the most challenging. By considering the lens of boundary crossing, I was able to discern the tools, rules, division of labor and the community involved in learning practices during the transition between academic settings and the workplace. The adaptation of the framework described by Akkerman and Bakker (2011) helped me study the phenomenon of the transition with an innovative approach, allowing for an in-depth analysis of statistical practices.

Differences in the rules between academic settings and the workplace created boundaries as junior statisticians transitioned to the workplace. Concerning practices related to data management, the rules in academic settings often set that students are provided with clean whereas at the workplace statisticians are required to deal with various types of data, handling messy data, managing big data or coping with small sample sizes. More broadly, time constraints affected statistical practices and statisticians need to learn how to make tradeoffs between the best way to complete a task and getting the task done on time. Indeed, another rule at the workplace concerned the purpose of statistics which is mainly to inform decisions and participants suggested that some results should be prioritized over the best results that may concretely not yield much more information. The role of the mentor was found particularly important in managing these priorities and helping junior statisticians develop time management practices such as keeping a timeline and balancing priorities between tasks.

Likewise, boundaries in the division of labor challenged statistical practices in the transition to the workplace. As Hijazi et al. (2019) found that statisticians urged for training in data management, the reason that caused boundaries may not only concern rules but also the division of labor. In academic settings, educators are providing clean

datasets while at the workplace statisticians become responsible for cleaning data.

Assuming responsibilities such as choosing the appropriate tools for programming, for analysis, or for designing studies, also created boundaries. Courses in academic settings are usually built around one specific program, technique, or design methodology, not encouraging students to experience the division of labor as choosing the appropriate tools. Only two participants reported having options to choose the programming tools to complete their homework assignments in academic settings, which facilitated the transition to the workplace.

Because the community involved in academic settings and at the workplace generally differed in terms of the background knowledge, junior statisticians encountered boundaries as they interacted with domain experts who were often not experts in statistics. As I pointed out earlier, practices of collaboration may have been overlooked in previous studies because they involved statisticians who did not experience boundaries in the community. Indeed, the community mostly affected collaboration and communication practices with understanding the domain problem and explaining results to non-statisticians.

Boundaries in the rules, division of labor, and the community subsequently created needs to learn tools for developing statistical practices. Participants referred to acquiring a solid background in theory, being exposed to diverse techniques with some topics in data science as facilitating the transition to the workplace for practices in analysis, for instance. They mentioned computer science courses at the undergraduate level to develop the logic of programming or consulting courses at the end of the graduate program to experience collaboration. As statisticians surveyed by Osman and

Ismail (2019) recognized that their education prepared them well for designing graphs, participants in my study shared that data visualizations were tools they learned in academic settings to develop communication practices at the workplace.

By identifying important statistical practices and the elements involved in developing these practices, I can formulate implications in terms of promoting the tools, rules, division of labor, and the community, in academics settings and at the workplace. Next, I integrate recommendations from participants in my study with current efforts to transform statistical practices.

Discussion of Implications

The findings of this study have implications for statistical practices in academic settings, at the workplace, and concerning the statistical community in general.

Transformations suggested by participants add to current efforts that have been developed to promote authentic practices in academic settings. The experiences of participants also provide implications for elements to include at the workplace to support the transition for junior statisticians.

Implications in academic settings

Some efforts have been put into the education of statisticians, implementing authentic opportunities to experience statistical practices in academic settings. For example, Greenhouse and Seltman (2018) created a master's degree program called the Master's of Statistical Practice. The program includes courses on foundational concepts in analysis, with linear models or experimental design, and specific branches of statistics, such as time series analysis or topics in data science. Across courses, students emphasize on programming, data visualizations, and data management, and are encouraged to

develop written and oral communication practices. Indeed, participants in my study expressed that practices such as programming, data management, and communication should be implemented across the curriculum as well as being taught within their own course. For example, I just mentioned that computer science courses in undergraduate programs were beneficial to develop programming practices at the graduate level and subsequently at the workplace. Next, I focus on some statistical practices that are currently lacking in academic settings. I discuss implications for data management and documentation, collaboration, and communication practices in particular.

To this date, there are very few courses that promote data management practices as noted by Wiljes & Cimiano (2019). They only reported one other study about a research course in data management and noticed very few textbooks dealing with topics in data management. On Google scholar, only few studies refer to “teaching data management” which mainly concentrate on the fields of computer science or library science, with most of the research being published after 2010. To address this gap in the literature, Wiljes & Cimiano (2019) proposed an interdisciplinary course in data management that cover tools that were found to facilitate the transition to the workplace in my study. Indeed, topics discussed in this course included documentation, data storage along with cloud environments, version control repositories such as GitHub, and electronic lab notebooks. The course touched based on ethics in conjunction with sharing and publishing data. Students enroll in the course from many different disciplines and levels, from bachelor’s to doctorate degree programs, and recognized the importance of learning about these topics. Participants in my study reinforced the need for such courses in the specific field of statistics, emphasizing how to use programming tools to clean

messy data, handle missing data, or extract data from databases. They recommended to have courses to introduce these data management practices at the beginning of a program but also integrate data management progressively across all courses to promote the division of labor by letting students being responsible for cleaning data. As discussed in next section, research that focus on specific practices is needed to determine how the implementation of data management practices within specific courses and across the curriculum in academic settings affect the transition to the workplace.

Students enrolled in the data management course offered by Wiljes & Cimiano (2019) suggested that documentation practices require more attention in academic settings, which was explained by participants in my study who noticed that documentation practices are often neglected at the workplace, yet crucial for collaboration and audit purposes. The only example of documentation practices in academic settings that appeared in my study was incorporated in a machine learning course and was proven to be extremely valuable at the workplace. Other participants experienced discontinuity for documentation practices and suggested to promote tools such as online repositories for version control or electronic lab notebooks.

Moreover, the master's program offered by Greenhouse and Seltman (2018) concludes with a capstone consulting project in collaboration with a real-life client. Indeed, boundaries were introduced by the community for collaboration practices since working with real-life clients, who are domain experts, was not common in academic settings. Participants in my study gave examples of elements promoting collaboration practices in academic settings such as consulting courses or service, and their experiences showed that these opportunities were only authentic if they involved domain experts. To

coordinate how to learn and teach collaboration practices with domain experts, Vance and Smith (2019) proposed a framework along five components for collaboration: attitude, structure, content, communication, and relationship. They integrate these components in a series of courses at the undergraduate and graduate level and in combination with a consulting service across the institution, also involving external clients. Getting experience with the components of attitude and relationship in academic settings would have particularly facilitated the transition for participants in my study since they were challenged by the attitude to develop while working with domain experts, being diplomatic and understanding, yet not condescending.

Efforts initiated by Vance and Smith (2019) not only foster collaboration practices in academic settings but promote communication practices as well. However, Davidson, Dewey, and Fleming (2019) identified a gap in the literature for research on how to promote communication practices effectively within courses in academic settings. To remedy this gap, they proposed a two-semester course, first introducing how to interact with domain experts before getting experience with consultations. For the first semester, students engage in interviews of statisticians and in role-playing, using video demonstrations, case studies, and doing mockup presentations. To complement the tools suggested by Davidson et al. (2019), participants in my study formulated some recommendations to present to non-statistical audiences by using data visualizations with simple graphics, avoiding statistical jargon by using layman's terms, and practicing. Another approach was described by Gibbons and MacGillivray (2014) with an undergraduate mentored tutoring program which gives students the opportunity to tutor other students in statistics, reflecting how statisticians communicate with non-statisticians

at the workplace. For example, Gibbons and MacGillivray (2014) argued that students develop practices such as listening, asking questions, and establishing relationships with individuals who have different personalities and backgrounds, which were practices mentioned by participants in my study. In addition, participants John and Edwards shared that experiences in teaching were beneficial to develop the practice of explaining statistics to non-statisticians.

In the second part of the course proposed by Davidson et al. (2019), students collaborate within teams on five projects with experts with backgrounds in medicine. The interactions with domain experts and managing several projects simultaneously not only promoted communication practices, but also time management practices. One of the mentors interviewed in my study recognized that offering opportunities to work on different projects at the same time in academic settings helped prepare students for the reality of the workplace. The majority of junior statisticians in my study confirmed that they were challenged at the workplace by balancing their time between different tasks and projects, adjusting their priorities, and dealing with short deadlines.

Recently, a longitudinal study conducted by Tuononen, Parpala, and Lindblom-Ylänne (2019) followed graduates into the workplace and revealed that graduates were more successful if they had developed diverse practices in academic settings and were aware of their competences. Indeed, it is crucial to share explicitly with students what practices they develop in academic settings and how these practices are expected to transition to the workplace. For example, participants in my study were surprised by the amount of time they spend at the workplace on performing data management and programming compared to conducting analysis. To raise awareness, participants

suggested to organize a seminar before graduation presenting what are some reasonable expectations for junior statisticians at the workplace and how to find resources to keep learning at the workplace. Efforts to invite statisticians from the workplace to share their experience in academic settings were also mentioned with panel discussions and alumni reunions. The findings of my study provide examples of experiences at the workplace to be shared with current students, indicating important statistical practices to be developed in academic settings and also highlighting implications for the workplace.

Implications at the Workplace

Grosemans et al. (2017) recommended to include the perspectives of mentors involved in the transition to the workplace to identify a misalignment between the expectations and capabilities of junior statisticians. Indeed, the role of the mentor at the workplace was found to be crucial for managing the expectations and priorities during the transition. Through the analysis of the interviews, I observed that some workplaces offered a mentoring program by assigning a mentor to junior statisticians which supported learning for various practices. The importance of the role of mentors was particularly uncovered with the experience of one participant who was first isolated by being the only statistician present at the workplace. Most of the practices were qualified as discontinuous until a senior statistician who was working remotely decided to act as a mentor. In addition, the two participants who identified as mentors of junior statisticians outlined many practices for which they provided support at the workplace. As reported by Van der Berg (2017), mentors realize how essential their role is in the transition, especially if they had received mentorship training. As a result, the role of the mentor

needs to be promoted at the workplace by including mentoring programs and helping junior statisticians identify mentors.

The transition was also facilitated at the workplace by encouraging junior statisticians to seek professional development which was first observed by Hijazi et al. (2019). Indeed, the findings of my study across all practices showed that junior statisticians need to keep learning at the workplace and this need was particularly imperative for programming. In comparison, statisticians surveyed by Hijazi et al. (2019) also expressed needs for professional development in data management, design, and programming. During the interviews, participants mostly mentioned having to teach themselves new programming tools while they referred to examples of workshops, forums, or interest groups organized onsite at the workplace that sustained learning. These meetings focused on topics in programming, advanced statistical techniques, or were meant to share research and progress within teams. The importance of support by the workplace to go to conferences was also highlighted. Indeed, participants were able to go to conferences where I recruited them and they mentioned learning about communication, keeping up to date with the newest statistical techniques, or networking. Therefore, workplaces need to facilitate the transition for junior statisticians by offering opportunities for professional development whether it be onsite or offsite.

Findings from this study can be used by junior statisticians to guide them in the transition by making visible what aspects of practices should be developed as well as what elements foster learning. I just mentioned seeking the help of a mentor and opportunities for professional development for instance. In addition, practices such as documentation, communication, or data management were rarely included in academic

settings but became crucial at the workplace. Raising awareness about the need and meaning of these practices can help junior statisticians develop the appropriate tools. For example, learning how to use GitHub or Jupyter Notebook sustains documentation practices while learning SQL or how to extract data from cloud databases furthers data management practices. To render the description of practices needed at the workplace accessible, I share implications for the statistical community.

Implications for the Statistical Community

Denis et al. (2001) recommended to make the role of statisticians more visible across the community. In the interviews, many participants shared that they studied statistics by “happenstance” because they were not aware what they could do with a degree in statistics as they started their educational journey. The results of this study can be used to depict a picture of what statisticians do at the workplace to encourage future students to study statistics. Indeed, Cameron, Iosua, Parry, Richards, and Jaye (2017) suggested that the lack of visibility for the profession may contribute to the shortage of statisticians on the job market. The statistical community needs to promote the role of statisticians, with interventions early in education, by publishing in different journals, and organizing interdisciplinary conferences.

Participants shared making significant impacts through their work. For example, one participant mentioned contributing to a critical discovery in Alzheimer's genetics and the lives of college students were meaningfully improved by the hard work of another participant. The importance of statistics in making informed decisions should be brought to attention by the statistical community by sharing significant contributions and impacts resulting from statistical studies.

The role of statisticians also needs to be promoted at different stages of a statistical study, especially for defining the design. As an example, one participant encouraged principal investigators to consult with her before data collection to calculate sample sizes. Statisticians surveyed by Cameron et al. (2017) also emphasized how they should be involved in the design of a study and encourage non-statisticians to appreciate the power of statistics.

Recommendations for Future Research

Investigating the transition between academic settings and the workplace for statisticians is challenging. Indeed, academic settings are complex activity systems and vary greatly in the tools, rules, division of labor and community they represent, with variation in course offerings and requirements, the students and professors involved, or the mission statement of the institution, for example. Similarly, variation is introduced in the elements at the workplace by requiring different programming tools or branches of statistics, with teams and relationships with coworkers, supervisors, and clients. Participants from my study experienced various contexts in academic settings and at the workplace, with different degrees and domains of application as well as years of experience. Because participants identified common practices across their differences, the findings yield significant contributions. However, narrowing the profile of participants to share the same context in academic settings, graduating from the same program, or to work at the same workplace, may better illustrate the relationship between the levels of transition to the workplace and the background in academic settings across individuals. Further research focusing on specific educational programs or specific workplaces could uncover more similarities and differences between individuals.

On the other hand, findings indicated that the responsibilities of a statistician may differ depending on the highest level of education obtained. According to participants, master's level and doctoral level statisticians seemed to have different responsibilities while bachelor's level statisticians tended to engage in similar activities than master's level statisticians but had more difficulties finding a job. Because participants who took part in interviews exclusively had master's degrees, further research should explore these differences and similarities by including junior statisticians with various degree levels.

As discussed in the previous section, some implications of my study are to develop curriculum materials and course outlines to incorporate statistical practices across the curriculum in academic settings. The thorough descriptions of the eight statistical practices that constitute the findings of this study can be used to build boundary objects to ease boundary crossing. For example, one participant wrote a case study about the rules dictated by ethics while reporting results in a publication. Within the implications, there are recommendations for both undergraduate and graduate courses.

After implementing boundary objects in academic settings, research would be needed to follow graduates to the workplace to investigate how these interventions support students to gain an authentic experience in academic settings and how they impact the transition to the workplace. A longitudinal study involving a combination of reflective journals and a series of interviews could be considered. For example, Lutz (2017) recommended to use weekly journals to complement interviews and investigate how individuals learn in academic settings and at the workplace.

Even though practices are intertwined and not developed exclusively, I argue that future studies should explore the effect of elements in academic settings on the transition

to the workplace by focusing on one specific set of practices at once. As shown with the implications in academic settings, more research is needed for creating courses or curriculum materials that promote data management or communication practices for example. As a result, there is also a need to investigate how these courses will influence the transition of statistical practices to the workplace.

Although this study gives examples about the different levels of transition with continuous, detailing, and discontinuous experiences, more attention is needed to define these levels and how they relate to boundary crossing. Akkerman and Bakker (2011) denoted boundaries as ambiguous, but I found that they manifested through the different elements of an activity system. To analyze the challenges encountered by statisticians in the transition, I specifically considered boundaries in rules, division of labor, community, and tools. However, I established a framework for boundary crossing between academic settings and the workplace that did not integrate elements outside of these activity systems. Other systems or components should be added to include elements such as Janet's experience in customer service or the teaching experience of Edwards after he graduated, before he worked as a statistician. Adding to the framework would enable predispositions of the individuals, cultural backgrounds, and experiences not related to academic settings nor the workplace to be taken into account.

Lastly, the methodological approach taken for this study also failed to uncover practices referring to habits of mind such as critical thinking, skepticism, or curiosity, which I included in the sorting task but were only touched on in the analysis of the survey and not mentioned during the interviews. The questions I developed for the survey and the interview focused on practices in terms of what statisticians do rather than how

statisticians think. The study by Pfannkuch and Wild (2000) revealed that these habits of mind were essential for the role of statisticians, but further research is needed to understand how statisticians develop these practices. Indeed, there was an animated debate between members of the statistical community about the practice of being skeptical at some of my workshops that attested of the importance of these practices.

Conclusion

My study provided an in-depth investigation of the experiences of the transition between academic settings and the workplace from the perspective of statisticians. Even with recent efforts to promote authentic statistical practices in academic settings, employers and mentors at the workplace still consider that junior statisticians have not developed the appropriate practices. To gain a deeper understanding of how statistical practices transition between academic settings and the workplace, I first involved members of the statistical community in identifying important statistical practices at the workplace. Second, junior statisticians and their mentors described how they coordinated what they learned in academic settings and what they were expected to do at the workplace through semi-structured interviews. I analyzed the data through the lens of boundary crossing, modeling the transition between the two activity systems of academic settings and the workplace, and identified elements that facilitate the transition.

The results revealed different perspectives from the members of the community on the importance of statistical practices at the workplace, contrasting communication with analysis practices, but agreeing on having to keep learning. Junior statisticians referred to a variety of elements creating boundaries as they transition to the workplace as well as elements that helped them overcome these boundaries. Experiences were

considered more continuous if mentors accompanied statisticians in the transition at the workplace and if practices were incorporated across the curriculum in academic settings, including experiences of authentic rules and division of labor. Findings for this study both align with and widen our understanding of statistical practices, focusing on how statisticians develop practices at the workplace. Combining recent efforts and the recommendations from participants in my study, I suggest implications for transforming statistical practices. I concluded with calls for future research to examine the implementation of authentic statistical practices in academic settings to facilitate the transition for future statisticians.

APPENDIX SECTION

A. INFORMED CONSENT	275
B. LIST OF STATISTICAL PRACTICES FOR THE SORTING TASK.....	277
C. SORTING TASK HANDOUT	278
D. SURVEY QUESTIONS	279
E. INTERVIEW PROTOCOL FOR JUNIOR STATISTICIANS	281
F. INTERVIEW PROTOCOL FOR MENTORS	283

APPENDIX A: INFORMED CONSENT

Study Title: Transition between Education and Profession: Experiences of Statisticians.

This consent form will give you the information you will need to understand why this research study is being done and why you are being invited to participate. It will also describe what you will need to do to participate as well as any known risks, inconveniences or discomforts that you may have while participating. We encourage you to ask questions at any time. If you decide to participate, you will be asked to sign this form and it will be a record of your agreement to participate. You will be given a copy of this form to keep.

PURPOSE AND BACKGROUND

You are invited to participate in a research study to learn more about the transition to the workplace from the perspective of statisticians. The information gathered will be used to promote an authentic experience of statistical practices throughout statistics education. By attending this workshop session, you are being asked to participate because you are showing interest in statistical practices and education.

PROCEDURES

If you agree to be in this study, you will participate in the following :

First, we will engage in a task that will help you reflect on what statistical practices are crucial to your work. It will take approximately 20 minutes to complete the task. Your engagement in the task is part of the workshop session, you need to sign this consent form in order to give your permission to use your task handout for research purposes. If you prefer to not participate to this study, your individual task handout will not be used for research.

Second, a survey will give you the opportunity to share your experience of the transition between learning statistics and performing statistics at the workplace. It will take approximately 15 minutes to complete the survey. If you would prefer not to participate to this study, your survey responses will not be used for research.

If you would like to participate to this study outside of this session, you can share your contact email and subscribe to a mailing list to be invited to complete the task and/or the survey online. You can also share your contact email to participate in an online interview that would explore your experience in more detail.

RISKS/DISCOMFORTS

When reporting results for this study, no identifiable information will be included. I will make every effort to protect participants' confidentiality and identity. In the event that some of the interview questions make you uncomfortable or upset, you are always free to decline to answer or to stop your participation at any time.

BENEFITS/ALTERNATIVES

There will be no direct benefit to you from participating in this study. However, the

information that you provide will help understand the role of statistician and how we can ease the transition to the workplace.

EXTENT OF CONFIDENTIALITY

Reasonable efforts will be made to keep the personal information in your research record private and confidential. Any identifiable information obtained in connection with this study will remain confidential and will be disclosed only with your permission or as required by law. The members of the research team and the Texas State University Office of Research Compliance (ORC) may access the data. The ORC monitors research studies to protect the rights and welfare of research participants.

Your name will not be used in any written reports or publications which result from this research. Direct quotes may be used, but they will be attributed to a pseudonym. Data will be kept for three years (per federal regulations) after the study is completed and then destroyed.

PARTICIPATION IS VOLUNTARY

You do not have to be in this study if you do not want to. You may also refuse to answer any questions you do not want to answer. If you volunteer to be in this study, you may withdraw from it at any time without consequences of any kind or loss of benefits to which you are otherwise entitled.

QUESTIONS

If you have any questions or concerns about your participation in this study, you may contact the Principal Investigator, Layla Guyot, at 512-245-6298 or by email l_g244@txstate.edu, or her faculty advisor, Dr White, at 512-245-3429 or by email aw22@txstate.edu.

This project 2018000 was approved by the Texas State IRB on [date]. Pertinent questions or concerns about the research, research participants' rights, and/or research-related injuries to participants should be directed to the IRB chair, Dr. Denise Gobert 512-245-8351 – (dgobert@txstate.edu) or to Monica Gonzales, IRB Regulatory Manager 512-245-2334 - (meg201@txstate.edu).

DOCUMENTATION OF CONSENT

I have read this form and decided that I will participate in the project described above. Its general purposes, the particulars of involvement and possible risks have been explained to my satisfaction. I understand I can withdraw at any time.

Printed Name of Study Participant

Signature of Study Participant

Date

Signature of Person Obtaining Consent

Date

APPENDIX B: LIST OF STATISTICAL PRACTICES FOR THE SORTING TASK

Code	Practice
AL13	Interpreting data (Limitations of methods / Bias)
BH28	Being skeptical / critical
FD11	Researching appropriate statistical methods and techniques
FH22	Collecting / Gathering data
GI15	Participating in teams / Collaborating
GJ37	Communicating in writing / Writing reports
HD64	Considering ethical issues
HI57	Creating / Maintaining databases
HS21	Preparing sampling frames / Drawing samples
IO44	Using advanced mathematics (calculus, algebra, differential equations...)
IR84	Communicating interpretations of statistics to non-statistical audiences
LH16	Meeting deadlines
LL27	Translating a real problem into a statistical form
OV47	Developing new statistical methods and techniques
PM34	Cleaning data / Managing missing data
RT71	Designing studies / Aligning design with research goals
SK97	Being curious / willing to learn
TL97	Consulting / Working with a client to solve a problem
TX27	Applying statistical methods and techniques
UA40	Working independently
WA74	Producing visual representations of data
WD86	Using statistical software package / Writing computer programs
XC59	Communicating orally / Making presentations
YN31	Using knowledge of the context (e.g. finance, biology, ...)
B1	
B2	
B3	
B4	
B5	
B6	

APPENDIX C: SORTING TASK HANDOUT

Complete the following grid with the codes for the corresponding practices:

1	2	3	4	5	6	7	8	9

Please select all roles that apply to you and **circle what you consider to be your main role**:

- ☐ Undergraduate student (Major: _____)
- ☐ Graduate student (Degree in: _____)
- ☐ Lecturer (Department: _____)
- ☐ Professor (Department: _____)
- ☐ Researcher (Field: _____)
- ☐ Statistician (Field: _____)
- ☐ Other, please specify: _____.

APPENDIX D: SURVEY QUESTIONS

Educational Background

1. What is your educational background (type of degree(s), highest degree)?
2. What year did you graduate from your highest degree / are expected to graduate?
3. How many statistics courses have you taken? What was the last statistics course you took?
4. What statistical software package(s) were you **taught** to use? Please select all that apply:
 - ☐ Excel
 - ☐ Fathom
 - ☐ JMP
 - ☐ Minitab
 - ☐ R
 - ☐ SAS
 - ☐ SPSS
 - ☐ STATA
 - ☐ other, please specify: _____.
5. What are the most useful concepts or practices you have learned through your education?
6. Why are the practices you mentioned in the previous question important?

If you have experience performing statistics in a professional setting, please continue answering the survey. Otherwise, thank you for taking the time to complete the survey, please share your contact to participate in further research!

Professional Experience

7. How many years of experience do you have practicing statistics in a professional setting?
8. Describe your current position:

What is your principal role?

- ☐ statistician
- ☐ other, please specify: _____.

What field/sector do you work in?

What type of firm/organization are you working for?

9. Describe past positions practicing statistics in professional settings:
10. What statistical software package(s) do you **use** at the workplace? Please select all that apply:
- ☐ Excel
 - ☐ Fathom
 - ☐ JMP
 - ☐ Minitab
 - ☐ R
 - ☐ SAS
 - ☐ SPSS
 - ☐ STATA
 - ☐ other, please specify: _____.
11. What are the most useful concepts or practices you have learned through your professional experience?
12. Why are the practices you mentioned in the previous question important?
- Transition from school to the workplace
13. Give some examples of statistical practices required by your job that you wish you had learned more about throughout your education.
14. Reflect on how well your education has prepared you to perform the statistical practices required by your job.
15. Do you have suggestions on how to implement statistical practices in statistics education?
16. Is there anything else you would like to share?
17. If you would like to participate in future research, please share your contact details.

APPENDIX E: INTERVIEW PROTOCOL FOR JUNIOR STATISTICIANS

Describe your educational background

- In general: type of diploma(s), major/minor, type of institution
- In statistics: how many courses, what type of courses, format (online, small or large classes, focused on theory or practice, involved projects, types of assignments...), statistical software packages
- Why did you decide to pursue your degrees? What were your expectations?
- After graduating, what did you do?
 - what kind of positions were you looking for?
 - what were some challenges in finding a first job? (how many interviews, how long did it take, making compromises...)

How did you adjust to the workplace?

- Did you experience a continuity between your education and profession?
- Has anybody in particular supported you during your transition?

Describe your professional experience:

- List past and current positions with: job titles, responsibilities, how long, what type of statistical projects you work(ed) on
- Describe a typical day in your current position
- What are some important practices/tasks you are required to perform? (**refer to examples of practices**)

For each of these practices, give details about:

- an actual example with a **scenario**
- **resources** you use to perform this practice: software, tutorials, previous knowledge, support from coworkers, ...
- **who** is involved to perform this practice: client, managers, coworkers, ...
- the **constraints** involved: instructions, deadlines, missing data, ...
- how much of your work time you spend performing this practice
- **how you learned to perform this practice**: where/when/who
- Of the important practices mentioned before
 - which ones did you learn through your education? What are some other important/useful practices you learned through your education? Give examples/anecdotes.
 - which ones did you not learn through your education? What are some other practices that you developed at the workplace and were not taught through your education? Give examples/anecdotes. How did you learn to perform these practices?
- What do you wish you had been taught before entering the workplace? Why? Give examples/anecdotes.
- Of the practices that you learned through your education:
 - what is similar or different about performing these practices at school versus at the workplace?
 - in particular, is there any difference in the context? (resources, constraints, people you worked with, ...)

- Of the practices that you developed at the workplace and were not taught through your education:
 - what is similar or different about performing these practices at school versus at the workplace? In particular, is there any difference in the context? (resources, constraints, people you worked with, ...)
 - why do you think you were not taught these practices in academic settings? (resources, constraints, people you worked with, too specific...)

Future

- Do you believe it is possible to fully prepare a statistician before entering the workplace? why/why not?
- What can be done to facilitate the transition to the workplace for statisticians?
- Would you like to work with me and do a second interview to develop an activity/example/description of some of the important practices to be implemented in education?
- Is there anything you would like to add about your experience as a statistician? as a student of statistics?

APPENDIX F: INTERVIEW PROTOCOL FOR MENTORS

Describe your educational background

- In general: type of diploma(s), major/minor, type of institution
- In statistics: how many courses, what type of courses, format (online, small vs large, focused on theory/practice, involved projects, assignments...), statistical software packages
- Why did you decide to pursue your degrees?
- What were your expectations?

Describe your professional experience

- List past and current positions with: job titles, responsibilities, how long, how you are involved with statistics
- Describe your role towards statisticians (supervisor, teacher, ...)
- Describe a typical day working with (future) statisticians

Mentor from academic settings

- what kind of positions are you preparing your students for?
- what qualifications have they gained through their education?
- what do you do to support statisticians to adjust to the workplace?
- do you think there is a continuity between statisticians education and what they have to do on the job?

Mentor from the workplace

- what kind of positions are you looking to fill in?
- how difficult is it to recruit a statistician?
- what specific qualifications are you looking for? (education, experience, ...)
- how long does it take for a new statistician to be able to perform the job?
- do you think there is a continuity between statisticians education and what they have to do on the job?
- what do you do to support statisticians to adjust to the workplace?

Mentor from academic settings

- What practices are you preparing future statisticians to perform at the workplace? (refer to the right tail of the sorting task for examples of practices)
- For each of these practices, give details about:
 - an actual example with a scenario
 - resources used to perform this practice: software, theory
 - constraints involved: instructions, deadlines, missing data, ...
 - how this practice is taught: where/when/who

Mentor from the workplace

- What are some important practices/tasks statisticians at the workplace have to perform? (refer to the right tail of the sorting task for examples of practices)
- For each of these practices, give details about:
 - an actual example with a scenario

- resources used to perform this practice: software, tutorials, previous knowledge, support from coworkers, ...
- who is involved to perform this practice: statisticians, client, managers, coworkers, ...
- the constraints involved: instructions, deadlines, missing data, ...
- how this practice was learned: where/when/who
- Of the important practices mentioned before
 - which ones did statisticians learn through their education? What are some other important/useful practices they learned through their education? Give examples/anecdotes.
 - which ones did statisticians not learn through their education? What are some other practices that they developed at the workplace and were not taught? How did they learn to perform these practices? Give examples/anecdotes.
- What do you wish statisticians had been taught before entering the workplace? Why? Give examples/anecdotes.
- Of the practices that statisticians learned through their education:
 - what is similar or different about performing these practices at school versus at the workplace?
 - in particular, is there any difference in the context? (resources, constraints, people to work with, ...)
- Of the practices that statisticians developed at the workplace and were not taught through their education: what is similar or different about performing these practices at school versus at the workplace? In particular, is there any difference in the context? (resources, constraints, people to work with, ...)
- Why do you think statisticians were not taught these practices in academic settings? (resources, constraints, people to work with, too specific...)
- What practices should have been acquired by statisticians before entering the workplace? Why? Give examples/anecdotes.

Future

- Do you believe it is possible to fully prepare a statistician before entering the workplace? why/why not?
- What can be done to facilitate the transition to the workplace for statisticians?
- Would you like to work with me and do a second interview to develop an activity/example/description of some of the important practices to be implemented in education?
- Is there anything you would like to add about your experience as a mentor of statistician/as a statistician/as a manager/as a teacher of statistics?

REFERENCES

- Abrandt Dahlgren, M., Hult, H., Dahlgren, L. O., Hård af Segerstad, H., & Johansson, K. (2006). From senior students to novice worker: Learning trajectories in political science, psychology, and mechanical engineering. *Studies in Higher Education*, 31, 569-586. <http://dx.doi.org/10.1080/03075070600923400>
- Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research*, 81(2), 132-169. <https://doi.org/10.3102/0034654311404435>
- American Statistical Association. (2015, October 1). *More Students Earning Statistics Degrees, But Not Enough to Meet Surging Demand for Statisticians*. Retrieved from <http://thisisstatistics.org/more-students-earning-statistics-degrees-but-not-enough-to-meet-surging-demand-for-statisticians/>
- American Statistical Association. (2020, January 7). *2020: A New Decade of Opportunity for Statisticians*. Retrieved May 21, 2020, from <https://thisisstatistics.org/2020-best-jobs-a-new-decade-of-opportunity-for-statisticians/>
- Anderson, C. W., & Loynes, R. M. (1987). *The teaching of practical statistics*. John Wiley & Sons, Inc.
- Anderson-Cook, C. M., Hamada, M. S., Moore, L. M., & Wendelberger, J. R. (2017). Statistical mentoring at early training and career stages. *The American Statistician*, 71(1), 6-14. <https://doi.org/10.1080/00031305.2016.1200491>

- Bakker, A., Kent, P., Derry, J., Noss, R., & Hoyles, C. (2008). Statistical inference at work: Statistical process control as an example. *Statistics Education Research Journal*, 7(2), 130-145. Retrieved from [https://www.stat.auckland.ac.nz/~iase/serj/SERJ7\(2\).pdf#page=133](https://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2).pdf#page=133)
- Bakker, A., & Akkerman, S. F. (2014). A boundary-crossing approach to support students' integration of statistical and work-related knowledge. *Educational Studies in Mathematics*, 86(2), 223-237. <https://doi.org/10.1007/s10649-013-9517-z>
- Beach, K. (1999). Chapter 4: Consequential transitions: A sociocultural expedition beyond transfer in education. *Review of research in education*, 24(1), 101-139. <https://doi.org/10.3102%2F0091732X024001101>
- Brown, S. R. (1980). *Political subjectivity: Applications of Q methodology in political science*. New Haven, CT: Yale University Press.
- Brown, S. R. (1996). Q methodology and qualitative research. *Qualitative Health Research*, 6(4), 561-567. <https://doi.org/10.1177/104973239600600408>
- Bureau of Labor Statistics, U.S. Department of Labor (2020). *Occupational Outlook Handbook*. Retrieved from <https://www.bls.gov/ooh/math/mathematicians-and-statisticians.htm>
- Cameron, M. (2006). Training statisticians for a research organisation. *Proceedings of the International Association for Statistical Education Satellite Conference*, Durban, South Africa. Voorburg, The Netherlands: International Statistical Institute.

- Cameron, C., Iosua, E., Parry, M., Richards, R., & Jaye, C. (2017). More than Just Numbers: Challenges for Professional Statisticians. *Statistics Education Research Journal*, 16(2). Retrieved from [https://iase-web.org/documents/SERJ/SERJ16\(2\)_Cameron.pdf](https://iase-web.org/documents/SERJ/SERJ16(2)_Cameron.pdf)
- Creswell, J. W. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oaks, CA: Sage.
- Davidson, M. A., Dewey, C. M., & Fleming, A. E. (2019). Teaching Communication in a Statistical Collaboration Course: A Feasible, Project-Based, Multimodal Curriculum. *The American Statistician*, 73(1), 61-69. <https://doi.org/10.1080/00031305.2018.1448890>
- Denis, J., Dolson, D., Dufour, J. M., & Whitridge, P. J. (2001). *Preparing statisticians for a career at Statistics Canada*. Statistics Canada. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.197.6684&rep=rep1&type=pdf>
- Engeström, Y. (1987). *Learning by expanding. An activity-theoretical approach to developmental research*. Helsinki, Finland: Orienta-Konsultit. Retrieved from <http://lchc.ucsd.edu/mca/Paper/Engestrom/expanding/toc.htm>
- Filstad, C., & McManus, J. (2011). Transforming knowledge to knowing at work: The experiences of newcomers. *International Journal of Lifelong Education*, 30, 763-780. <https://doi.org/10.1080/02601370.2011.625573>
- Gibbons, K. S., & MacGillivray, H. (2014). Education for a workplace statistician. In *Topics from Australian Conferences on Teaching Statistics* (pp. 267-293). Springer, New York, NY. https://doi.org/10.1007/978-1-4939-0603-1_15

- Greenhouse, J. B., & Seltman, H. J. (2018). On teaching statistical practice: From novice to expert. *The American Statistician*, 72(2), 147-154. <https://doi.org/10.1080/00031305.2016.1270230>
- Grosemans, I., Coertjens, L., & Kyndt, E. (2017). Exploring learning and fit in the transition from higher education to the labour market: A systematic review. *Educational Research Review*, 21, 67-84. <http://dx.doi.org/10.1016/j.edurev.2017.03.001>
- Hahn, C. (2011). Linking academic knowledge and professional experience in using statistics: A design experiment for business school students. *Educational Studies in Mathematics*, 86(2), 239-251.
- Harraway, J. A., & Barker, R. J. (2005). Statistics in the workplace: a survey of use by recent graduates with higher degrees. *Statistics Education Research Journal*, 4(2), 43-58. Retrieved from [http://iase-web.org/documents/SERJ/SERJ4\(2\)_harraway_barker.pdf](http://iase-web.org/documents/SERJ/SERJ4(2)_harraway_barker.pdf)
- Hendricks, K. B. (2014). *Graduates' perspectives on transition from college graduation to the workplace: a qualitative approach* (Doctoral dissertation). Retrieved from https://uknowledge.uky.edu/hes_etds/16
- Hijazi, R., Saeed, R., & Alfaki, I. (2019). Role of statisticians in building the UAE knowledge economy. *Electronic Journal of Applied Statistical Analysis*, 12(1), 303-319. <http://dx.doi.org/10.1285/i20705948v12n1p303>
- Holmes, P. (1997). Matching education, assessment and employment needs in statistics: The MeaNs project. *Teaching Statistics*, 18, 1-3. <https://doi.org/10.1111/j.1467-9639.1997.tb00854.x>

- Hoyles, C., Noss, R., Kent, P., & Bakker, A. (2010). *Improving mathematics at work: The need for techno-mathematical literacies*. London: Routledge.
- Kenett, R., & Thyregod, P. (2006). Aspects of statistical consulting not taught by academia. *Statistica Neerlandica*, 60(3), 396-411. <https://doi.org/10.1111/j.1467-9574.2006.00327.x>
- Kent, P., Bakker, A., Hoyles, C., & Noss, R. (2005). Techno-mathematical literacies in the workplace. *Mathematics Statistics and Operational Research*, 5(1), 5-9.
Retrieved from https://www.researchgate.net/profile/Arthur_Bakker/publication/46662073_Techno-mathematical_Literacies_in_the_Workplace/links/02e7e5278d9c3ce0cb000000.pdf
- Kettenring, J. R., Koehler, K. J., & McKenzie Jr, J. D. (2015). Challenges and opportunities for statistics in the next 25 years. *The American Statistician*, 69(2), 86-90. <https://doi.org/10.1080/00031305.2015.1033987>
- Konkola, R., Tuomi-Gröhn, T., Lambert, P., & Ludvigsen, S. (2007). Promoting learning and transfer between school and workplace. *Journal of Education and Work*, 20(3), 211-228. <https://doi.org/10.1080/13639080701464483>
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.
- Love, K., Vance, E. A., Harrell Jr, F. E., Johnson, D. E., Kutner, M. H., Snee, R. D., & Zahn, D. (2017). Developing a career in the practice of statistics: The Mentor's perspective. *The American Statistician*, 71(1), 38-46. <https://doi.org/10.1080/00031305.2016.1255257>

- Loynes, R. M. (1987). *The training of statisticians round the world*. International Statistical Institute. <http://iase-web.org/Books.php?p=book2>
- Lutz, B. D. (2017). *Into the workplace: Exploring the learning experiences of recent engineering graduates during the school-to-work transition* (Doctoral dissertation). Retrieved from <https://vtechworks.lib.vt.edu/handle/10919/78276>
- McKeown, B., & Thomas, D. B. (2013). *Q methodology* (Vol. 66). Sage publications.
- Merriam, S. B. (1998). The design of qualitative research. *Qualitative research and case study applications in education*, 2, 1-25.
- Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. (2nd ed). San Francisco, CA: Jossey-Bass.
- Ministry of Business, Innovation and Employment in New Zealand (2020). *Occupational Outlook*. Retrieved from <http://occupationoutlook.mbie.govt.nz/social-and-community/statisticians/>
- Nilsson, S. (2010). On the meaning of higher education in professional practice: The case of physicians and engineers. *Journal of Education and Work*, 23, 255-274.
<http://dx.doi.org/10.1080/13639080.2010.486397>
- Osman, M. I., & Ismail, M. A. (2009). What do statisticians working in policy research need from statistical education? *Proceedings of the International Association for Statistical Education Satellite Conference*, Durban, South Africa. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://iase-web.org/documents/papers/sat2009/9_4.pdf

- Pfannkuch, M., & Wild, C. J. (2000). Statistical thinking and statistical practice: Themes gleaned from professional statisticians. *Statistical Science*, 15(2), 132-152.
<http://dx.doi.org/10.1214/ss/1009212754>
- Ritter, M. A., Starbuck, R. R., & Hogg, R. V. (2001). Advice from prospective employers on training BS Statisticians. *The American Statistician*, 55(1), 14-18.
<https://doi.org/10.1198/000313001300339888>
- Society for Human Resource Management (2016). *Jobs of the future: Data analysis skills*. Retrieved from <https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Documents/Data-Analysis-Skills.pdf>
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social studies of science*, 19(3), 387-420.
<https://doi.org/10.1177/030631289019003001>
- Suchman, L. (1994). Working relations of technology production and use. *Computer Supported Cooperative Work*, 2, 21-39. <https://doi.org/10.1007/BF00749282>
- Tuomi-Grohn, T., & Engestrom, Y. (Eds.). (2003). *Between school and work: New perspectives on transfer and boundary-crossing*. Amsterdam, Netherlands: Pergamon.
- Tuononen, T., Parpala, A., & Lindblom-Ylänne, S. (2019). Graduates' evaluations of usefulness of university education, and early career success—a longitudinal study of the transition to working life. *Assessment & Evaluation in Higher Education*, 44(4), 581-595.

- Van der Berg, G., H. (2017). *A framework to integrate the formal learning with the informal workplace learning of statisticians in a developmental state* (Doctoral dissertation). Retrieved from <http://uir.unisa.ac.za/handle/10500/22638>
- Vance, E. A., & Smith, H. S. (2019). The ASCCR Frame for Learning Essential Collaboration Skills. *Journal of Statistics Education*, 1-10. <https://doi.org/10.1080/10691898.2019.1687370>
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wiljes, C., & Cimiano, P. (2019). Teaching research data management for students. *Data Science Journal*, 18(1). <http://doi.org/10.5334/dsj-2019-038>
- Zabala, A. (2014). qmethod: A Package to Explore Human Perspectives Using Q Methodology. *The R Journal*, 6 (2), 163-173. Retrieved from <http://journal.r-project.org/archive/2014-2/zabala.pdf>