

Student Perceptions on Reproducible Research in Introductory Statistics Courses

Nicholas Bussberg

Assistant Professor of Statistics

Department of Mathematics & Statistics

Elon University

nbussberg@elon.edu

Background and Definitions

Working definitions as I defined them in my class:

- **Reproducibility:** ability to obtain an identical result using the same procedures, data, and code
- **Replication:** ability to obtain similar results using a different dataset and/or approach

There's broad agreement that reproducibility is important, but comparatively little on how to teach students how to implement it.

- Baumer et al. (2014) – R Markdown as a reproducible tool
- Project TIER (www.projecttier.org) – framework for teaching reproducibility

Motivation for My Research

If reproducibility is important, then students should learn and practice it early

- This develops good habits and a foundation on which they can build
- This implies that we need a transferable, accessible way to introduce reproducibility to students

Research Study Information

Conducted an IRB-approved study in Fall 2021 and Spring 2022

- Administered a survey at the beginning and at the end of the semester

Beginning of semester survey (during 1st week of classes)

- Provided definition of reproducibility
- Asked questions on what students' thoughts were on reproducibility

End of semester survey (during last 2 weeks, including finals)

- Asked some of the same questions to see how perceptions shifted
- Also had questions related to how learning reproducibility concepts helped in the class and on their projects

Course Details

I taught 4 sections of STS 2120 Statistics in Application, 2 in the fall (60 students total) and 2 in the spring (60 students total)

- Introductory-level course with focus on statistical tests
- Students learned and used SAS via SAS Studio
- Students completed regular homework assignments and 2 larger projects
- Course had traditional grading structure in the fall and “ungrading” structure in the spring
- Taught at Elon University, NC - a mid-sized private university

How Reproducibility Was Taught

Had a short discussion (20 min) on reproducibility and replication

- This was done just after we had our introduction to SAS, during the first two weeks of the course

Did an in-class, group activity (10 min) to think more about the concepts

- Peanut Butter and Jelly activity

Assessments

- Two (out of 8) homework assignments were checked and graded for reproducibility
- Both projects were checked and graded for reproducibility

Basic Stats from Fall 2021

There were 60 students in both sections of my course

- 26 agreed to participate in the study and completed at least the first survey (43% response rate)
- 14 completed both surveys (54% of students who opted in)

Class standing

First-Year	11
Sophomore	11
Junior	2
Senior	2

Comfort w/ Math/Stats

Very Uncomfortable	1
Uncomfortable	3
Neutral	6
Comfortable	9
Very Comfortable	7

Anxiety w/ Math/Stats

Very Uncomfortable	3
Uncomfortable	6
Neutral	6
Comfortable	8
Very Comfortable	3

Basic Stats from Fall 2021

Organized w/ Digital Files

Strongly Agree	2
Agree	16
Neither	3
Disagree	5
Strongly Disagree	0

Detail-Oriented

Strongly Agree	6
Agree	15
Neither	3
Disagree	2
Strongly Disagree	0

Percent who had heard about reproducibility-related concepts before the class: 54%

Basic Stats from Spring 2021

There were 58 students in both sections of my course

- 31 agreed to participate in the study and completed at least the first survey (53% response rate)
- 9 completed both surveys (29% of students who opted in)

Class standing

First-Year	18
Sophomore	10
Junior	1
Senior	1

Comfort w/ Math/Stats

Very Uncomfortable	0
Uncomfortable	1
Neutral	13
Comfortable	13
Very Comfortable	4

Anxiety w/ Math/Stats

Very Uncomfortable	2
Uncomfortable	10
Neutral	6
Comfortable	6
Very Comfortable	7

Basic Stats from Fall 2021

Organized w/ Digital Files

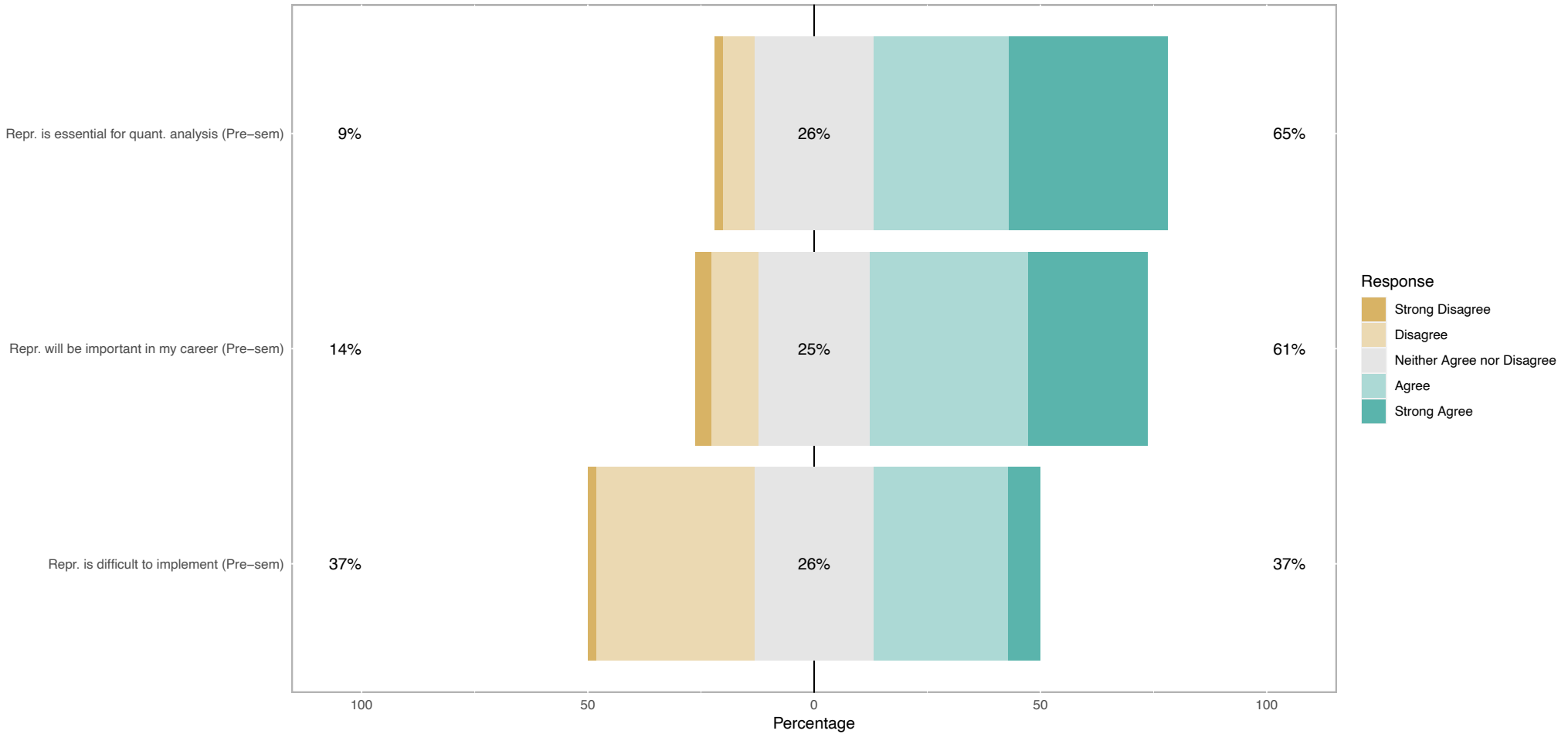
Strongly Agree	10
Agree	19
Neither	1
Disagree	1
Strongly Disagree	0

Detail-Oriented

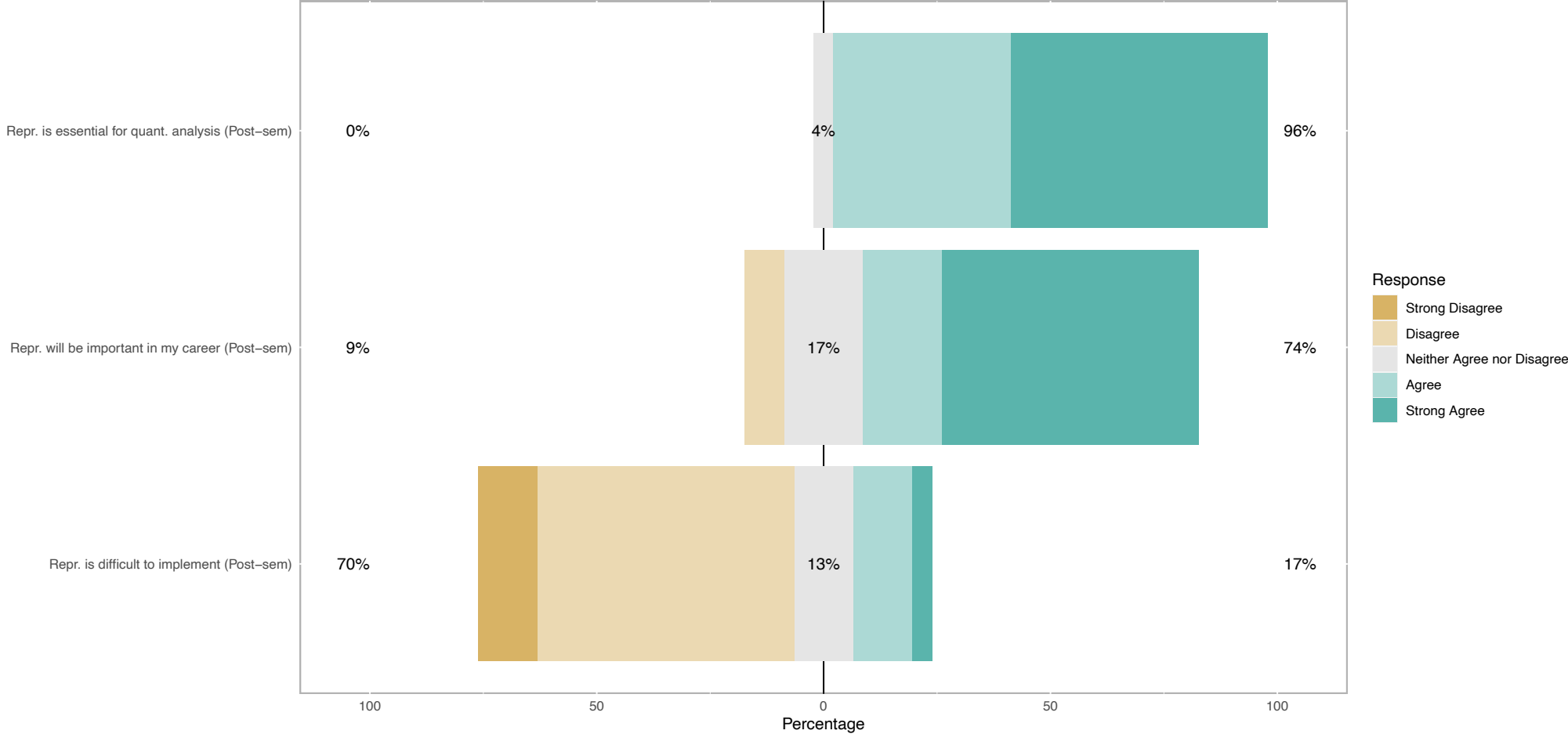
Strongly Agree	12
Agree	16
Neither	2
Disagree	1
Strongly Disagree	0

Percent who had heard about reproducibility-related concepts before the class: 52%

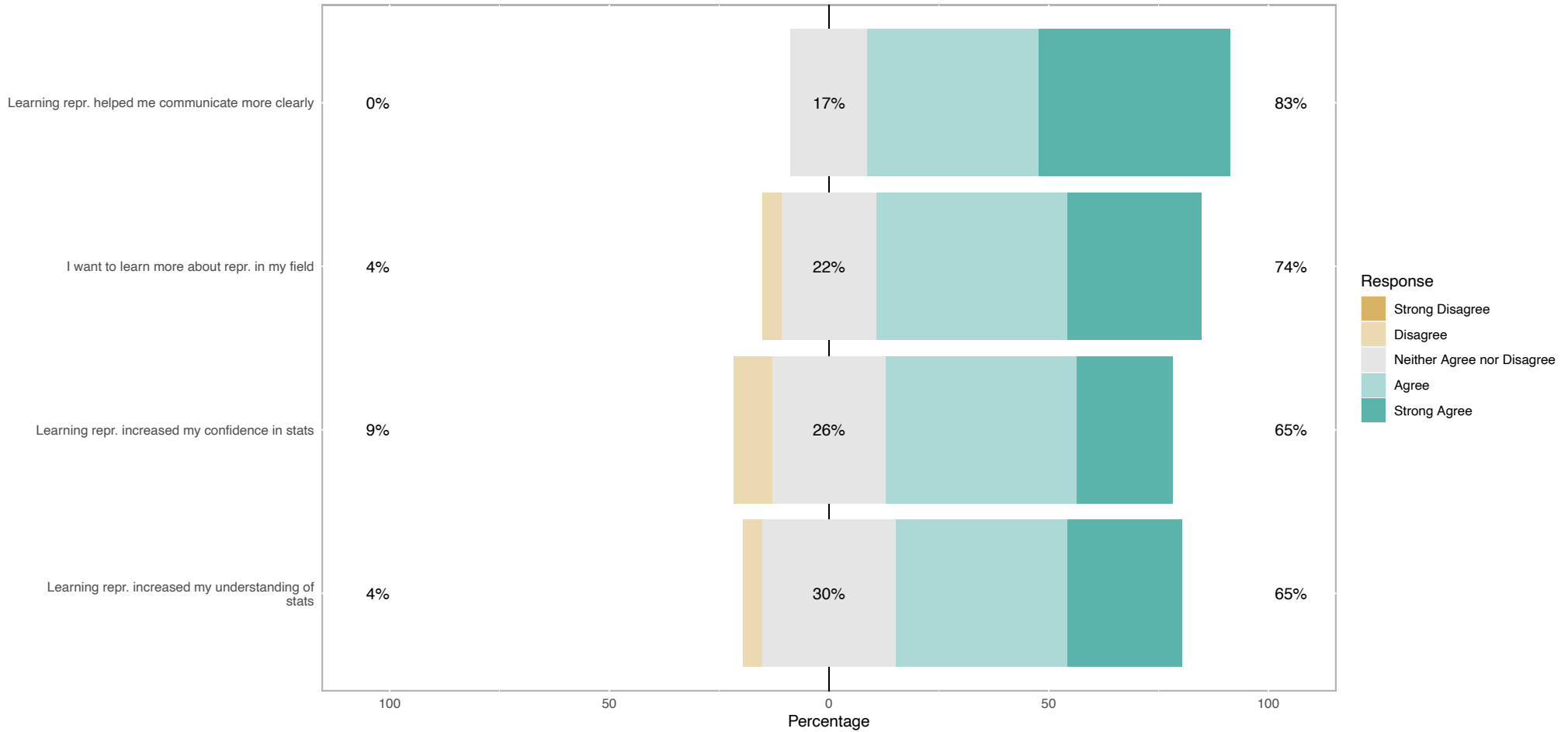
Beginning of Semester Reponses – Combined



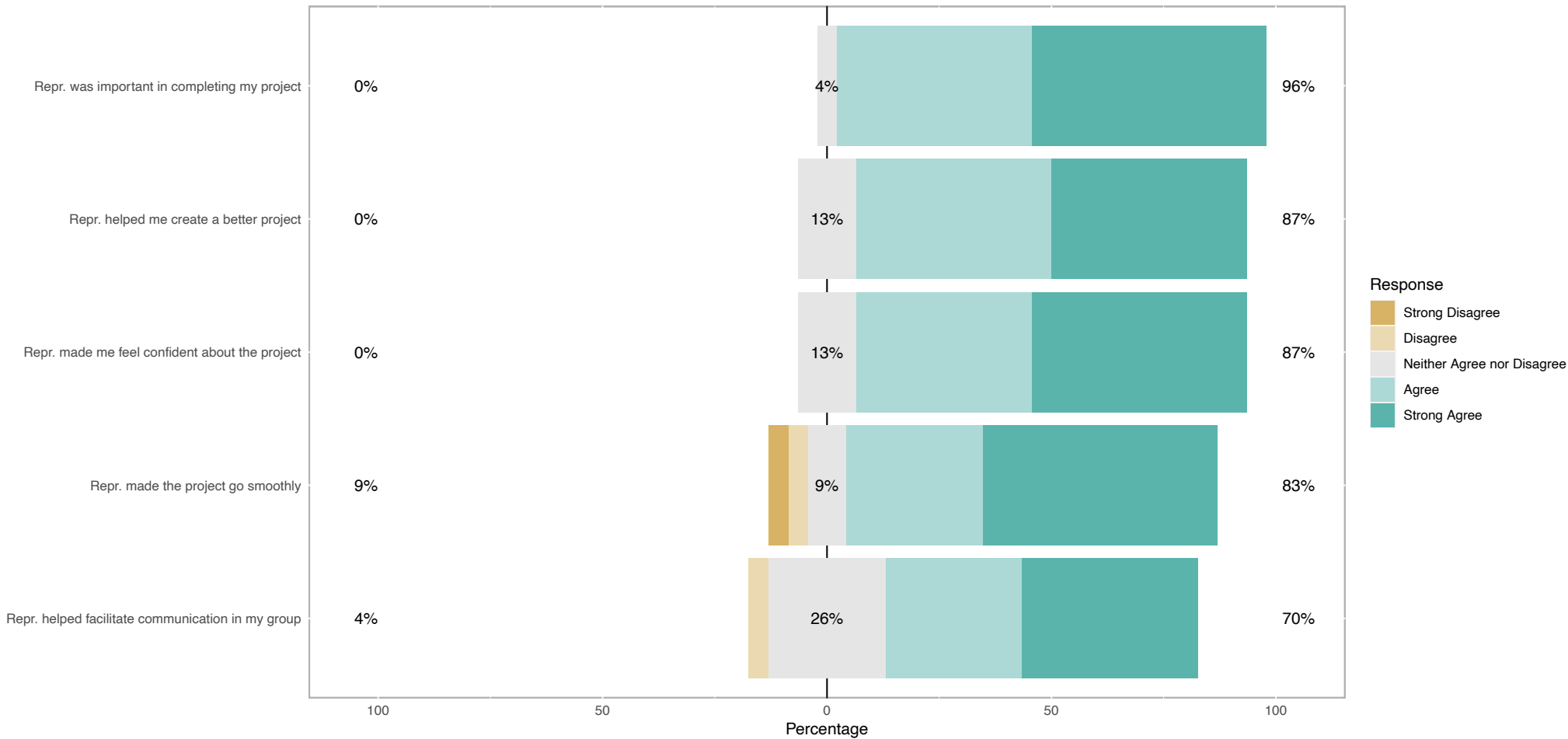
End of Semester Responses – Combined



Responses Related to Learning – Combined



Responses Related to the Project – Combined



Selected Student Free Responses

Some thoughts from students:

For Project 2, “If one person couldn’t run the code, there was most likely an error that had to be solved before proper results were achieved. Having the group run the code smoothly made for more confidence when documenting the results...later”

“I think all science and math courses should have this because I feel like you need to know how to it regardless.”

“...I think that the concepts are instructive for students who are relatively new to the field to remind them that statistics and related...fields are always a part of ongoing...processes, and that the more work you put into a...project, the better chance it has to stand up to scrutiny”

Thoughts on How Much Extra Effort Was Needed

There were two areas where I spent additional time to incorporate reproducibility concepts

1. Developing and teaching the concepts (just one class period)
2. Grading/checking for reproducibility on assessments

For grading/checking:

- Students copied their code into their homework and projects
- I copied that code into a (sometimes prepped) SAS script and ran the code, checked for errors (if any), and ensured the results matched what they had
- Example of time spent: for Project 2 in Spring 2022, I spent 75 minutes checking 25 projects (about 3 minutes per project)

Closing Thoughts

I kept the reproducibility concepts a minimal amount, attempting to assess whether even a small exposure to reproducibility would affect students.

1. Students seems to gain an understanding of the importance of reproducibility, though about half were already familiar with the concept
2. Students felt like reproducibility improved their overall learning and their ability to succeed in Project 2
3. The extra time that I spent preparing and grading was not overwhelming, and it helped me catch some errors

References

Baumer, B., M. Çetinkaya-Rundel, A. Bray, L. Loi, and N.J. Horton. 2014. R Markdown: Integrating a Reproducible Analysis Tool into Introductory Statistics. arXiv:1402.1894.

Project TIER (Teaching Integrity in Empirical Research).
<https://www.projecttier.org>

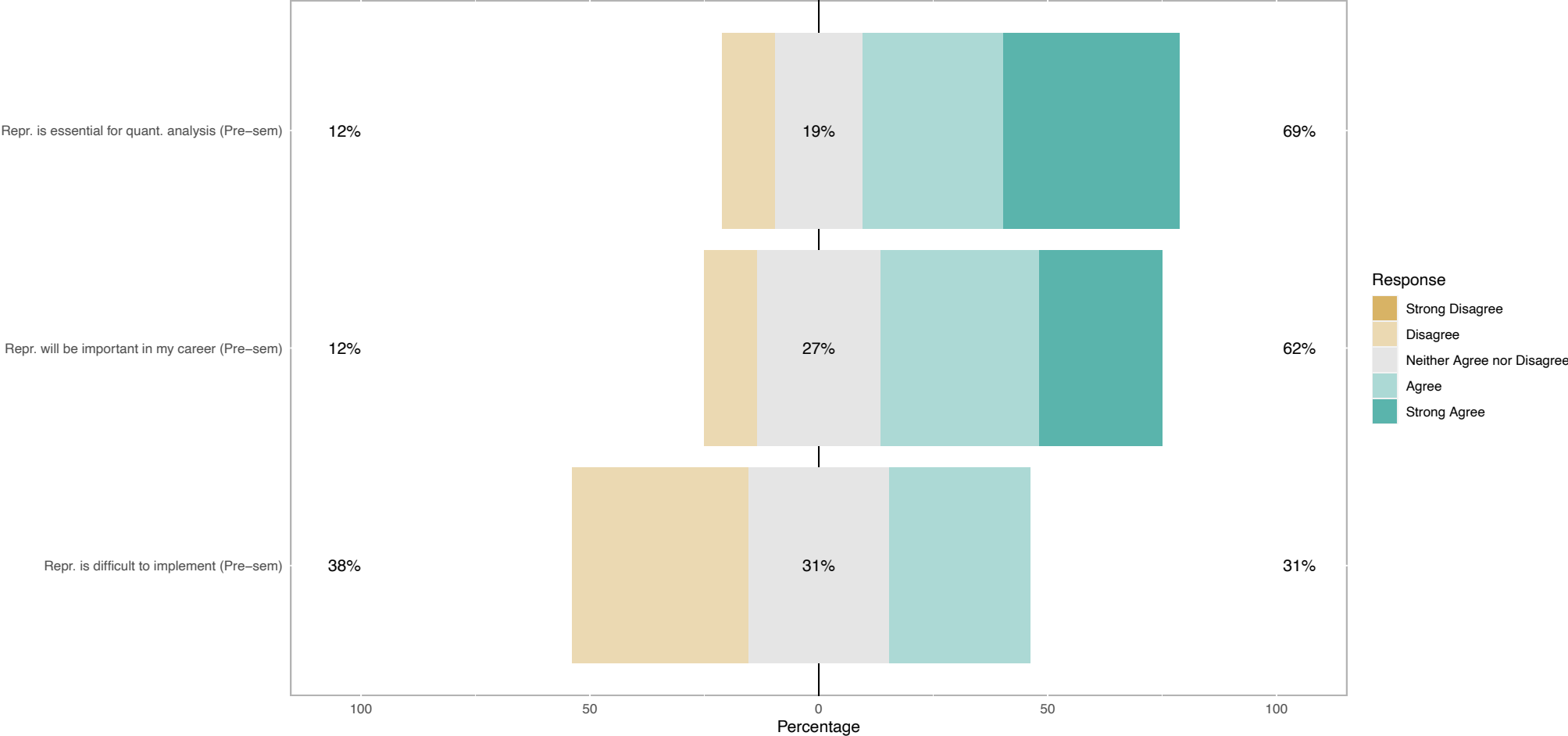
Acknowledgements

Dr. Scott Long (Indiana University) for teaching me to appreciate reproducibility and introducing the concept to me

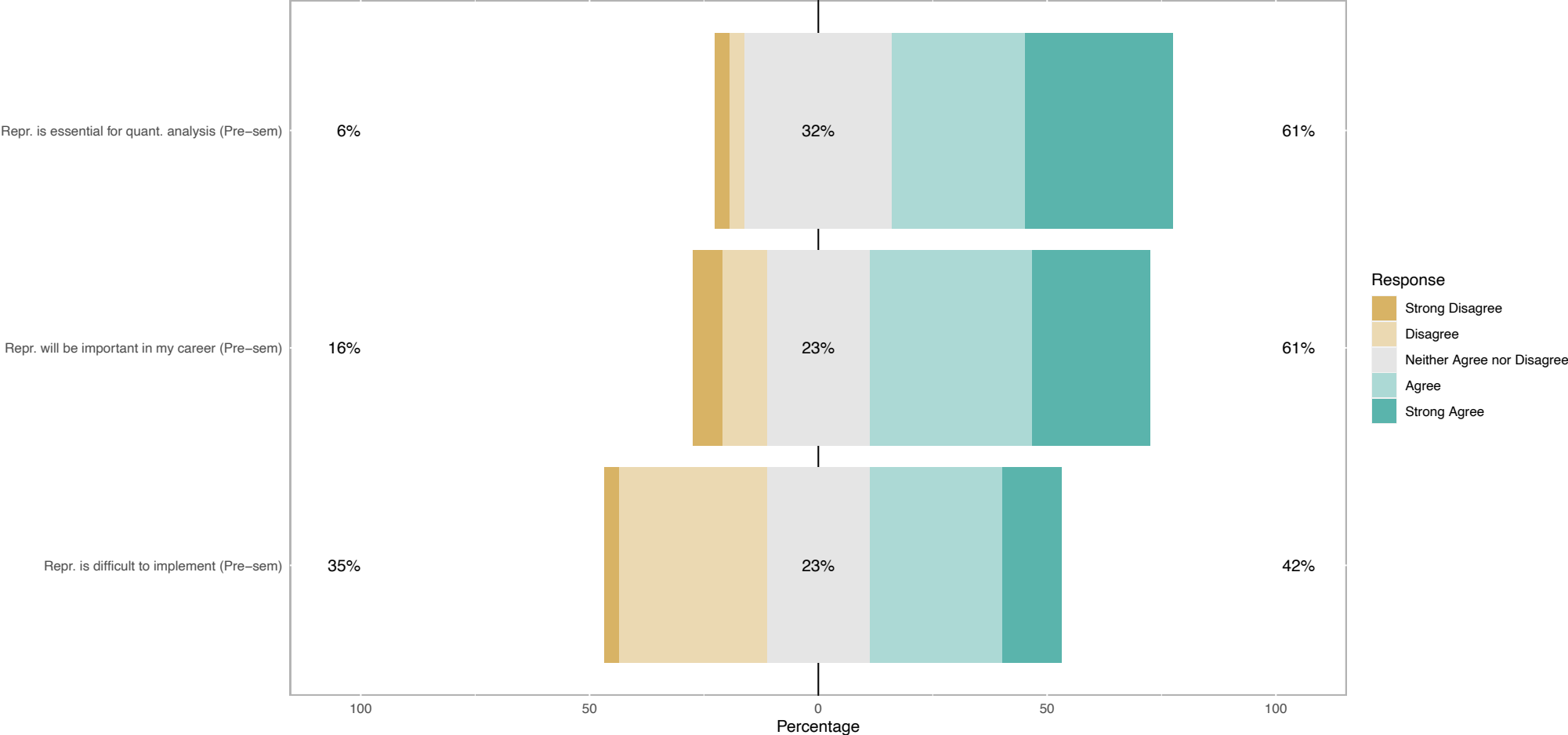
Elon University Dept. of Mathematics and Statistics for being supportive and encouraging in this endeavor

Elon University IRB committee

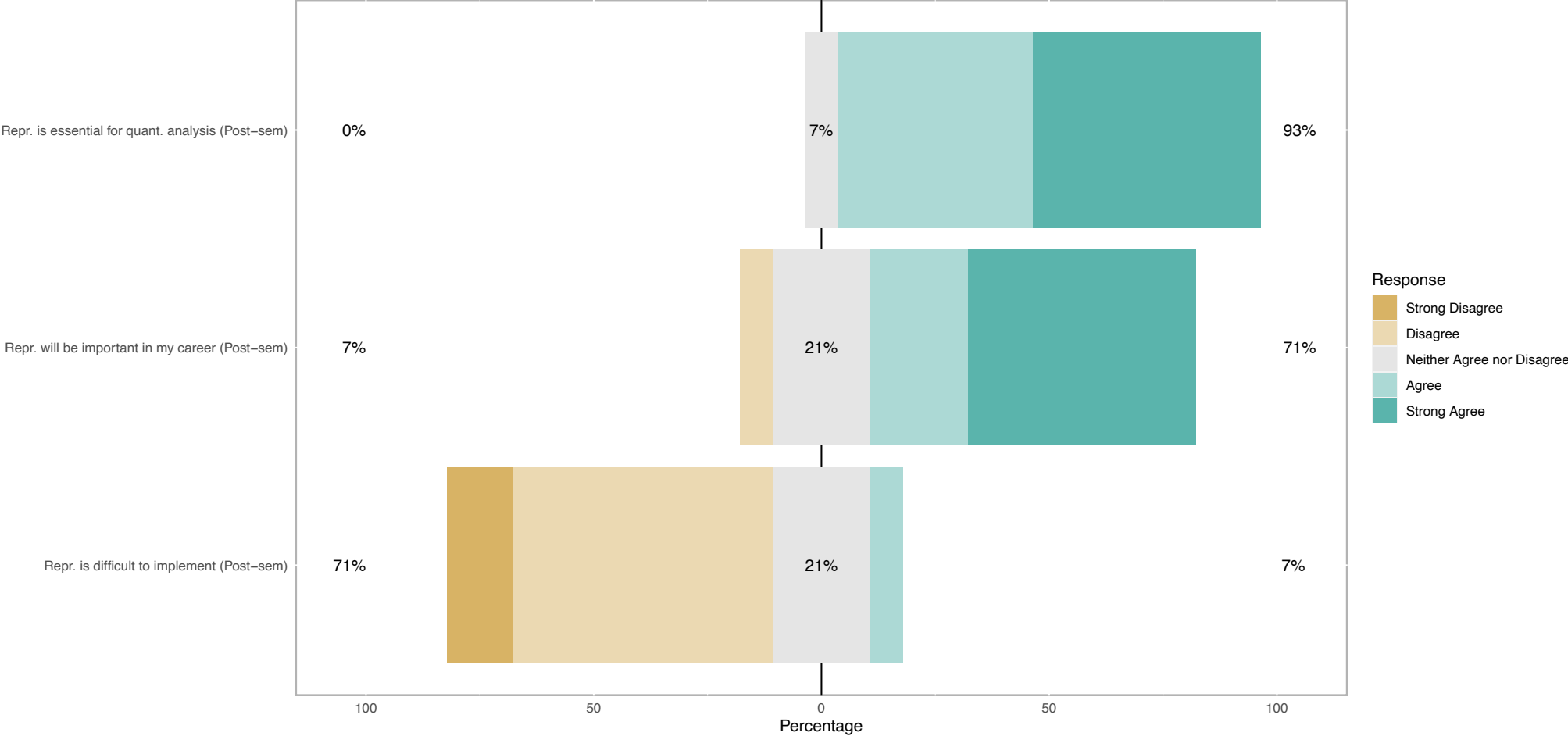
Beginning of Semester Reponses – FA21



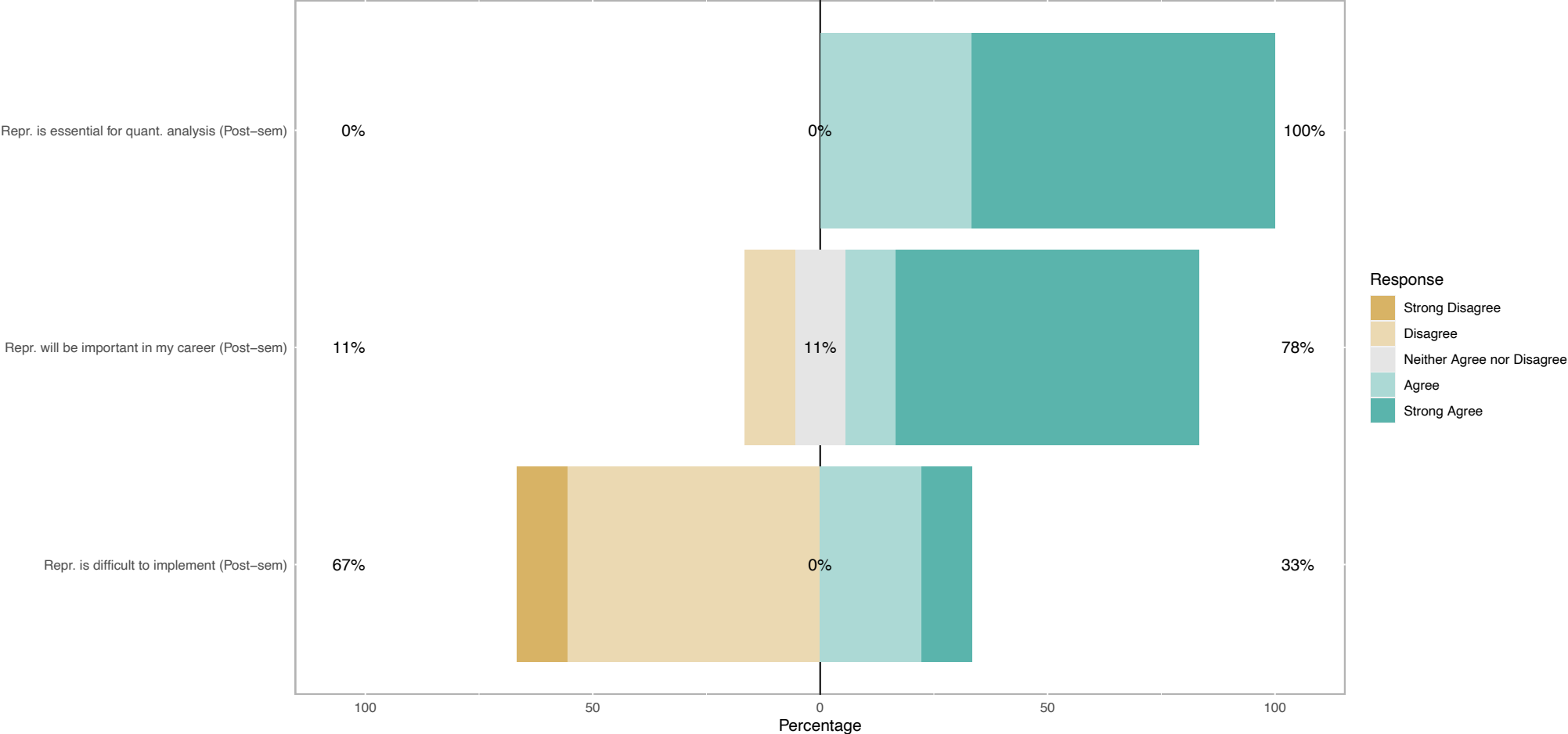
Beginning of Semester Reponses – SP22



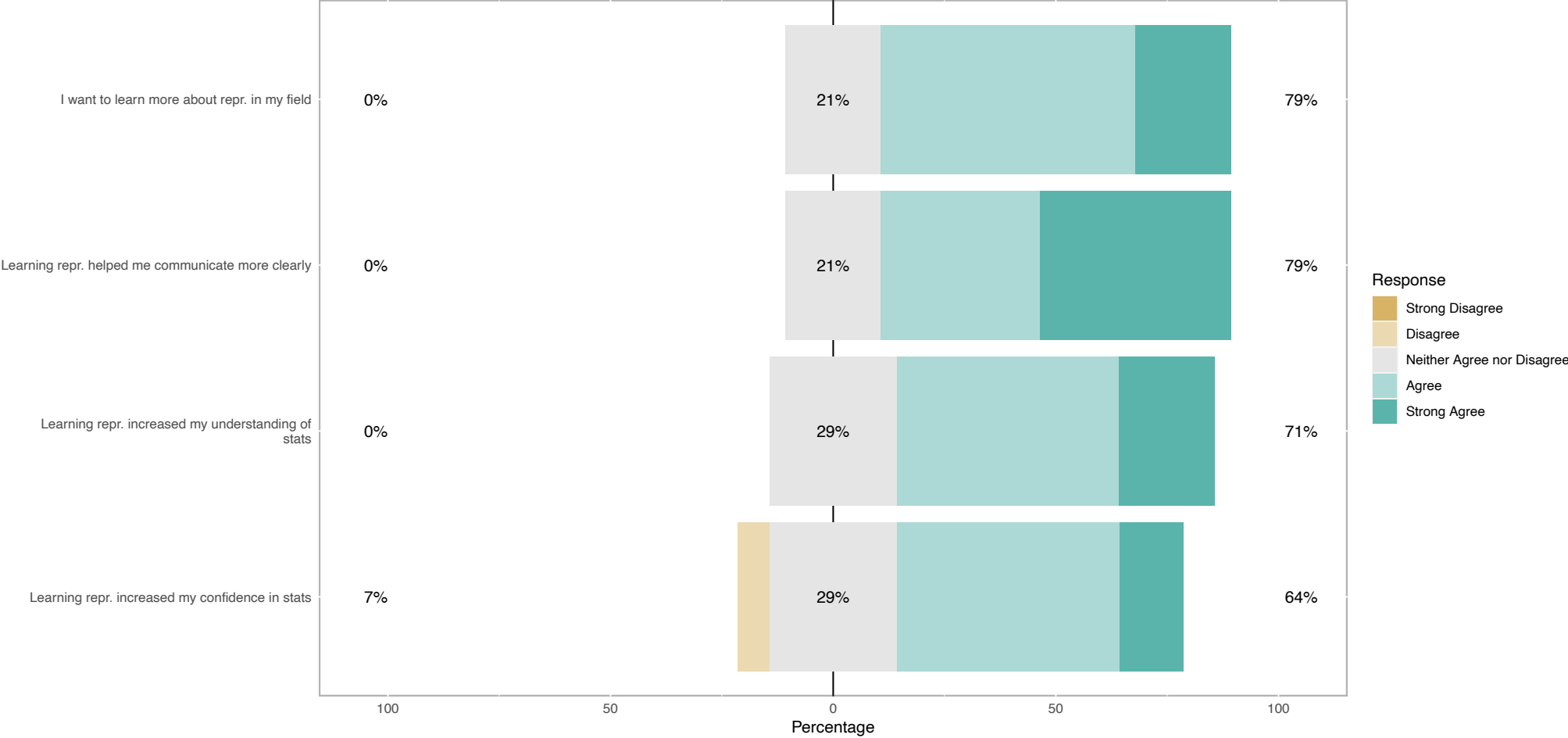
End of Semester Responses – FA21



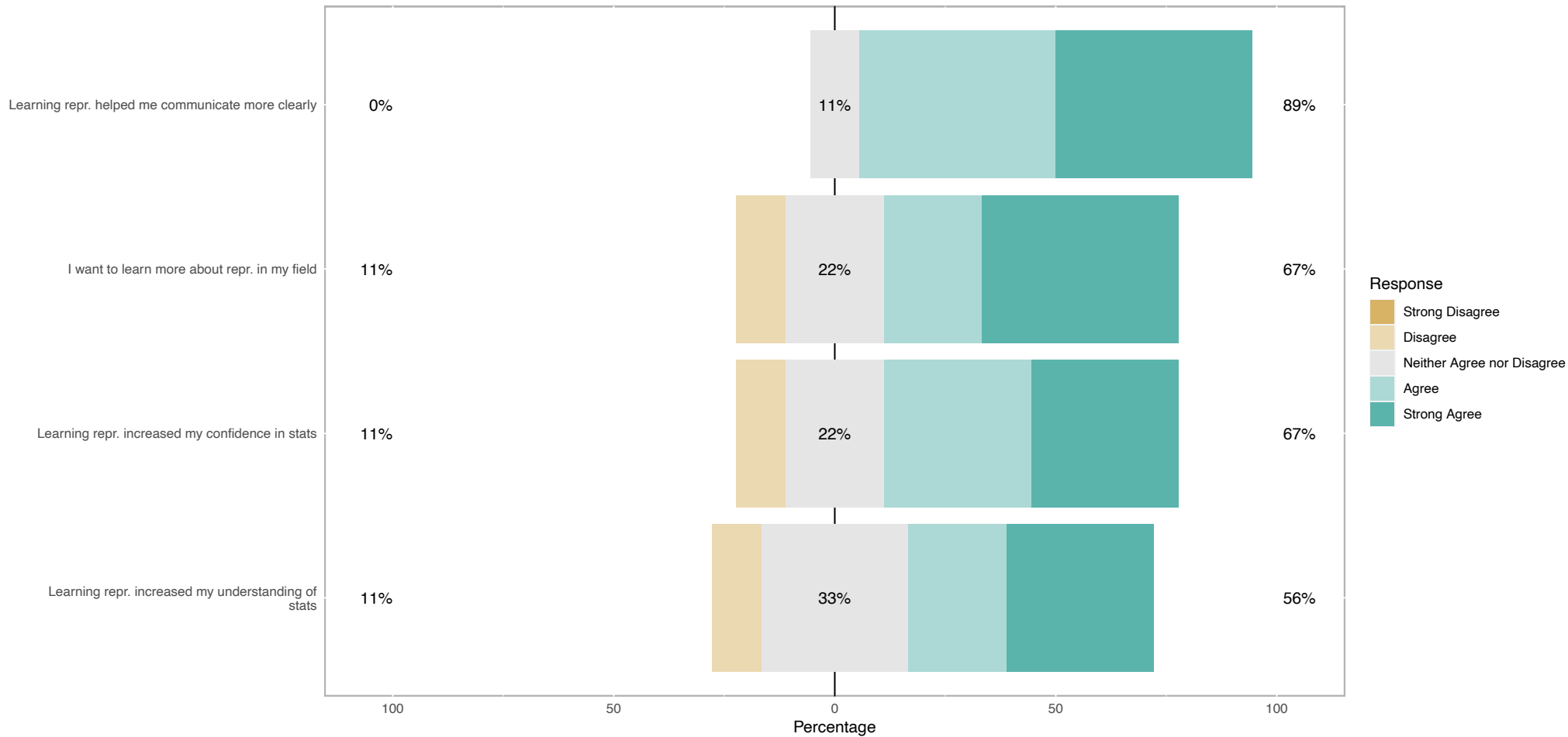
End of Semester Responses – SP22



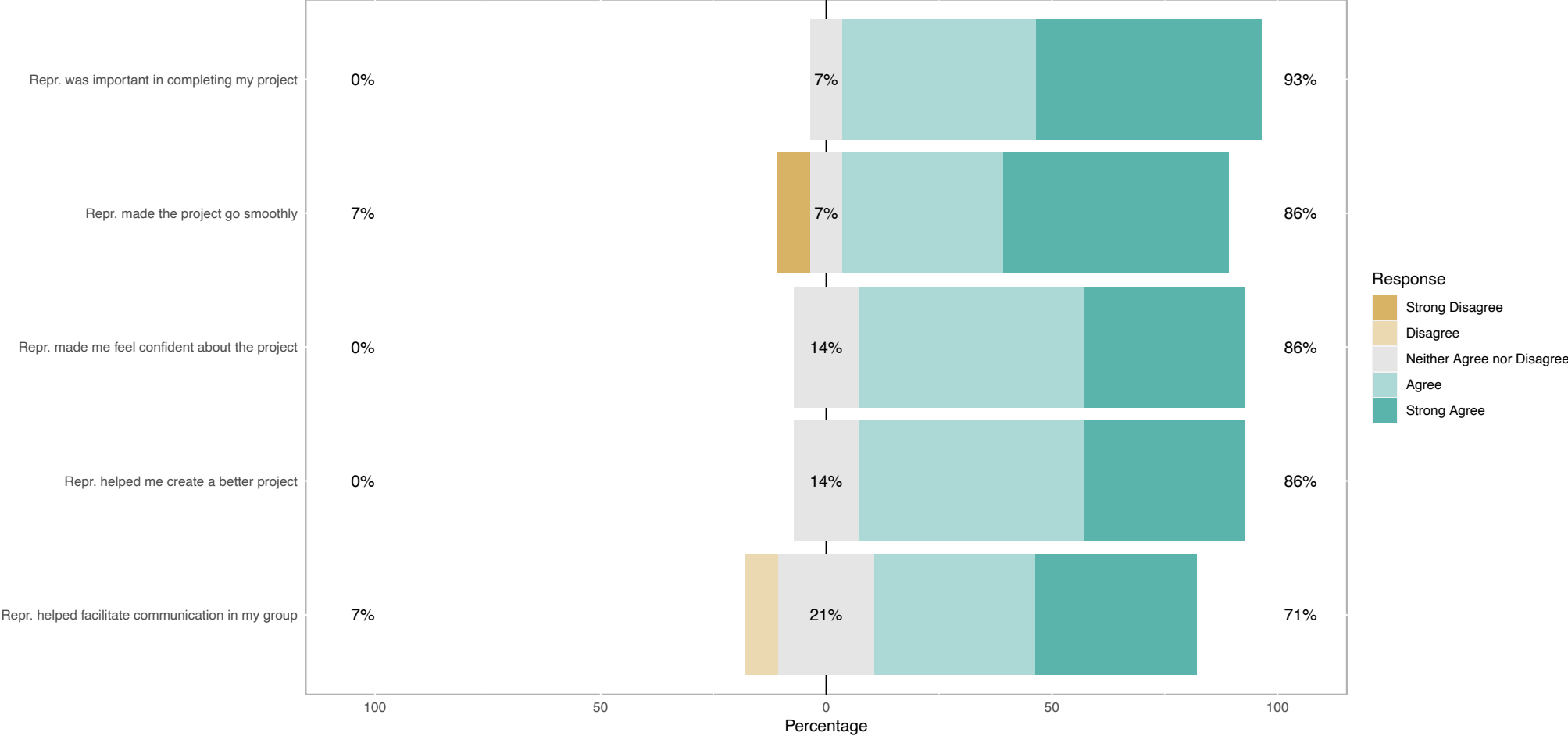
Responses Related to Learning – FA21



Responses Related to Learning – SP22



Responses Related to the Project – FA21



Responses Related to the Project – SP22

