Examining the Language of Equity and Inclusion in STEM Education Policy

A tale of two data science education reports

Carrie Diaz Eaton[1] and Anelise Hanson Shrout[1]
with Jude Hidgon-Topaz[2]

[1]Digital and Computational Studies, Bates College

[2]QSIDE Institute

IAS site, Blackwell-Tapia Conference 2021

Bates | DIGITAL & COMPUTATIONAL STUDIES

Outline

Bates | DIGITAL & COMPUTATIONAL STUDIES

Outline

Bates | DIGITAL & COMPUTATIONAL STUDIES

Current Context

"Data Scientist – The sexiest job of the 21st century.[1]*"*

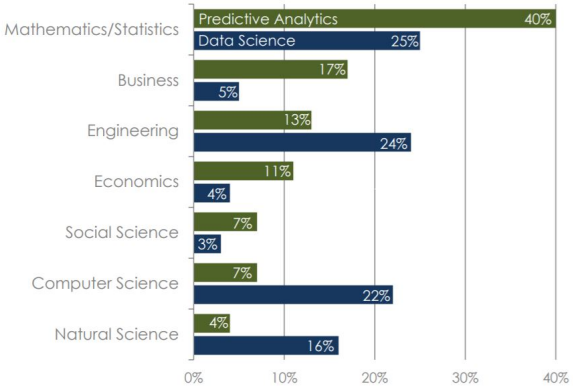


Figure: Source: Burtch Works Study [2]

- An emergent field with robust outlook, sector growth, and pay
- Disciplinarily diverse
- Opportunity to engage a new generation of scientists from whom STEM opportunities have been systemically withheld

Bates | DIGITAL & COMPUTATIONAL STUDIES

Current Context

"Data Scientist – The most sexist new job of the 21st century."

- Data science as a field is and will continue to be a driving influence decision-making and policy
- In the U.S., we are already failing at making sure that folx in these key positions represent who is in the U.S.
- Even the information about gender assumes a gender binary.

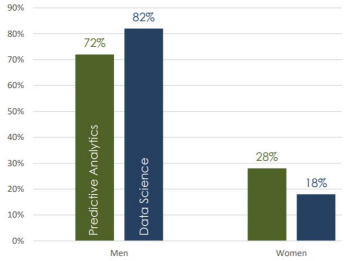


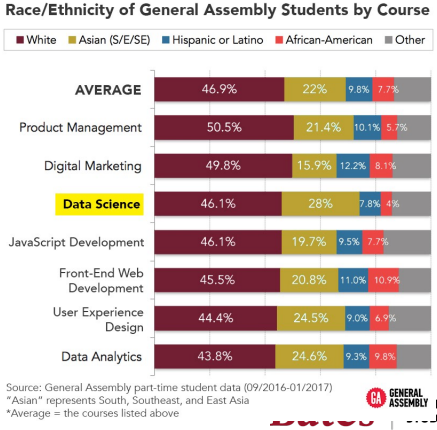
Figure: From The Burtch Works Study [3]

[3] Burtch, *The Burtch Works Study: Salaries of Data Scientists and Predictive Analytics Professionals* (Aug. 2020).

Current Context

"Data Scientist – The job most illustrative of systemic racism in STEM education in the 21st century."

- Data science education is not reaching our PEER students[4,5,6]
- This manifests in the ways our society is molded by decisions and technology is constructed - *e.g.* see work by Benjamin, Boulamwini, Noble, O'Neil and STS scholars



[6] Asai, "Race matters" (2020), p. 4.

Policy and STEM Education

- Professional societies need to take strong leadership to set the agenda in STEM education policy with bold attention, language, and actions that address the exclusion of mathematicians and scientists due to social constructions such as gender identity and race
- What are the mechanisms we use to do this work? Emerging statements, convening of conversations, and policy reports.
- In particular policy reports can serve as a text archive of institutional commitment.
- Informs funding allocation and program building

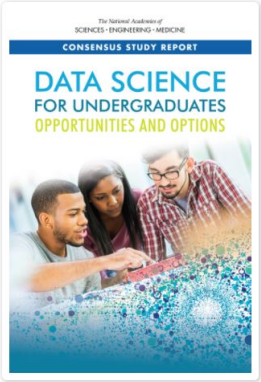
Here we analyze two reports in Data Science education which emphasize undergraduate data science education and workforce development.

Bates | DIGITAL & COMPUTATIONAL STUDIES

NASEM reports

National Academies Report on Undergraduate Data Science

- Committee is convened which takes into consideration expertise, conflict of interest and a "balance of perspectives"[7]
- Two reports commissioned
- Released an interim report in February 2018 with initial findings[8]
- Research process
- Released a final report in August 2018 [9]



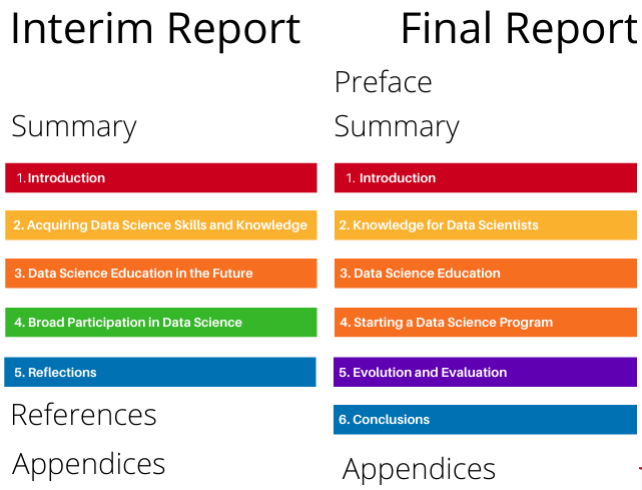
[8] https://www.nationalacademies.org/about/our-study-process

[9] National Academies of Sciences, Medicine, et al., *Envisioning the data science discipline: the undergraduate perspective: interim report* (2018).

[9] National Academies of Sciences, Medicine, et al., *Data Science for Undergraduates: Opportunities and Options* (2019).

What do you notice, what do you wonder?

- 68 pages, 35 of which were numbered report pages
- 11 Findings



- 138 pages, 91 content pages not including Appendices
- 13 Findings, 11 Recommendations

Bates | DIGITAL & COMPUTATIONAL STUDIES

Final NAS report

"The National Academies report would have been much different had this group been authoring it"

- Interim report included an entire chapter on broadening participation
- No such chapter exists in the final report
- What happened?
  - Is this loss real?
  - Or was it infused throughout?

Bates | DIGITAL & COMPUTATIONAL STUDIES

Outline

Bates | DIGITAL & COMPUTATIONAL STUDIES

Positionality

- Who are we?
- What epistemologies do we bring to this process?
- Why is this question meaningful?

Bates | DIGITAL & COMPUTATIONAL STUDIES

The Challenges and Benefits of Working with Text

- The meaning of text changes over time.
- Are computational methods the best way to understand text?

but...

- Text might be one of several powerful indicators of our values

Identifying Keywords for Language Analysis

- Compare words related to diversity, equity, and inclusion across two related documents - which words
- To identify a set of words, both internal and external to the NAS report, we use a contemporary report - EDSIN
- Identified seed words and topic modeling - Latent Dirichlet Allocation (LDA) - to derive additional keywords
- Close reading of all NASEM report works

Identifying Keywords: Seed keywords

"Seed" keywords:
"diverse", "diversity", "underrepresented", "underrepresentation", "access", "accessibility", "inclusive", "inclusion", "equity", "race," "racial."

Bates | DIGITAL & COMPUTATIONAL STUDIES

Identifying Keywords: Topic Model

Applying LDA to the EDSIN report

Assumptions:

- Textual documents are comprised of multiple and sometimes-overlapping topics
- These topics manifest as collections of words that likely to occur in proximity to one another[10]

[12] David M. Blei, "Probabilistic Topic Models," Communications of the ACM 55, no. 4 (2012): 77–84.

Identifying Keywords: Close Reading

Final Keyword List:

"ability", "african", "access", "accessibility", "accessible", "allies", "ally", "background", "backgrounds", "barriers", "communities", "community", "cultural", "cultures", "intercultural", "dei", "diverse", "diversity", "equitable", "equity", "ethic", "ethical", "ethics", "ethnic", "ethnicity", "hispanic", "identified", "identify", "identity", "impact", "inclusion", "inclusive", "indigenous", "justice", "perspectives", "pipeline", "poc", "racial", "race", "recruitment", "retention", "underrepresent" and "underrepresented."

Bates | DIGITAL & COMPUTATIONAL STUDIES

A Note on Presence and Absence

Processing Text

- Divide the report into sentences
- Remove "stopwords"
- Standardize case
- Strip punctuation
- Stem words

For example: access, accessibility, accessible, accessed all stem to "access"

Bates | DIGITAL & COMPUTATIONAL STUDIES

The NASEM report: Interim to Final

Table: Words experiencing the largest increase from interim report to final

Word	Interim count	Final count	Frequency change
data	578	1211	0.0047
February	0	73	0.0035
accessed	22	115	0.0034

Table: Words experiencing the largest decrease from interim report to final

Word	Interim count	Final count	Frequency change
can	108	127	-0.0038
diverse	42	11	-0.0033
questions	44	19	-0.0031

Bates DIGITAL & COMPUTATIONAL STUDIES

Theorizing Distribution of Words

Poisson Processes A Poisson process can be theorized to describe the number of occurrences of a certain word across documents of fixed length [11, 12]

[12] Ogura, Amano, and Kondo, "Gamma-Poisson distribution model for text categorization" (2013), p. 13.

[12] Inouye et al., "A review of multivariate distributions for count data derived from the Poisson distribution" (2017).

Simulated Keyword Distribution

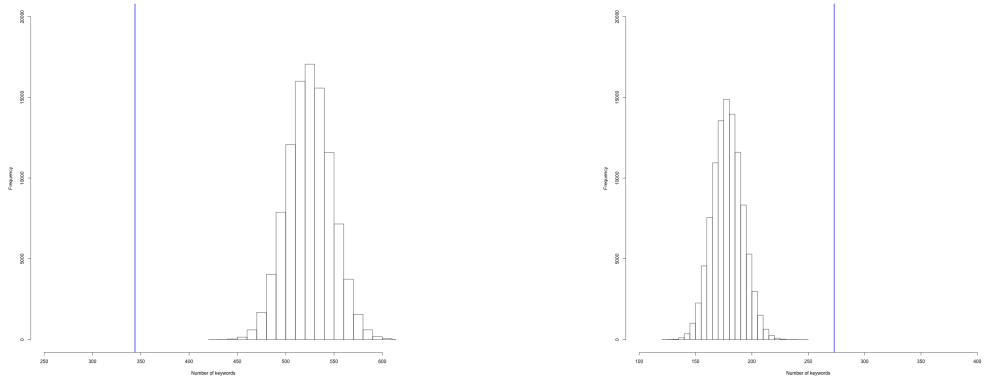
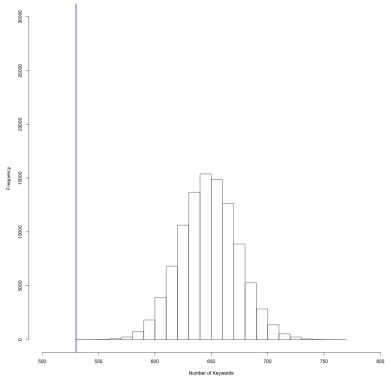


Figure: Predicted frequency distribution of keywords in the interim report (left) and final report (right), given the relative frequency of keywords in the other report. Actual frequency of keywords in the report in blue.

Bates DIGITAL & COMPUTATIONAL STUDIES

Simulated Stemmed Keyword Distribution



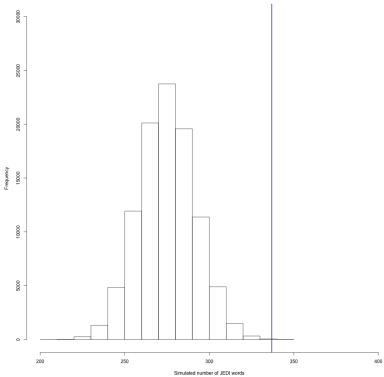


Figure: Predicted frequency distribution of stemmed keywords in the interim report (left) and final report (right), given the relative frequency of stemmed keywords in the other report. Actual frequency of stemmed keywords in the report in blue.

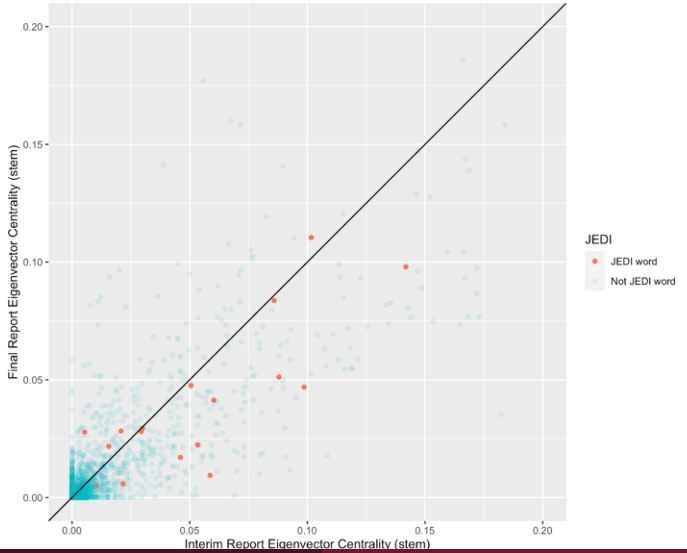
Bates | DIGITAL & COMPUTATIONAL STUDIES

Networks of words across all reports

Figure: Networks of word co-occurance in the interim report (left) and final report (right). Justice, equity, diversity and inclusion keywords in blue.

Bates | DIGITAL & COMPUTATIONAL STUDIES

Who are We Centering? Centrality of words across all reports



- Red dots - JEDI keywords, cyan dots - all others
- At the origin - low centrality
- On the line - roughly equal distribution
- Below the line - less central in the final report

Bates | DIGITAL & COMPUTATIONAL STUDIES

Findings and Recommendations

Questions: Aren't these two separate reports? But isn't an NASEM report all about the Findings and Recommendations, not the text?

- Out of the eleven Findings in the interim report, more than half (six), had clear analogs in the final report.
- Of the 6 which reappear, five had an associated Recommendation
  - **Interim Finding 4.2** Partnerships between 2- and 4-year institutions provide a valuable opportunity to develop innovative curricula, reach more diverse student populations, and expand the reach of data science education.
  - **Final Recommendation 3.1** Four-year and two-year institutions should establish a forum for dialogue across institutions on all aspects of data science education, training, and workforce development.

Findings and Recommendations

Question: What did not merit a Recommendation?

- What about the 6th Finding which appeared in both reports, but had no Recommendation?
  - **Interim Finding 4.1** Data science has the potential to draw in a diverse set of students and build in broad participation from the onset, rather than trying to broaden participation later. However, strategies are needed to recruit and retain these students.
  - **Final Finding 4.2** Data science would particularly benefit from broad participation by underrepresented minorities because of the many applications to problems of interest to diverse populations.

Bates DIGITAL & COMPUTATIONAL STUDIES

Findings and Recommendations

Question: What was left out?

- Five findings did not reappear in the final report.
  - Findings 2.1, 2.2, and 3.4 spoke to interdisciplinarity and using real data to solve complex problems.
  - Finding 3.1 connected these practices to enhancing data science curricula.
  - Finding 3.2 which suggested shared structures for multidisciplinary collaboration on data science.

Bates | DIGITAL & COMPUTATIONAL STUDIES

Outline

Bates DIGITAL & COMPUTATIONAL STUDIES

Construction and Revision of Policy

- How these reports were assembled is important - who constructed this report, the ideas for this report, how was it vetted
- Backed by research/good practice on this (listening, centering, compensating people of color)

Revision of Policy

- The revision processes can strip away our commitment to equity and inclusion due to biases that favor white centering to maintain control
- White centering can look/sound like:
  - Referencing majority vote/input as the only valid
  - Changing the message to appease a hypothetically conservative audience ("too controversial," "what would our donors think," "meet people where they are")
  - Counter-argument "Who are we centering here?"
- The removal of a chapter removes the message
  - Spreading it throughout does not work
  - This "well-meaning" act becomes a form of suppression [13]
  - Counter-argument - do both, build a foundation and literacy, then capitalize on it throughout

[13] Bowers et al., *The rhetoric of agitation and control* (2009).

Avoid Dichotomies

Why does our community continue to buy into the idea that you can embrace and encourage diversity OR research as an exclusive OR? The Blackwell-Tapia conference is a great counterexample.

Leadership

"The National Academies report would have been much different had this group been authoring it"

In conclusion - these reports tell us that we professional organizations and other leaders need to

1. STEP UP
2. Center people of color in the process of creating STEM educational policy,
3. Don't revise the equity and justice out of your message
4. Devote resources (and chapters) specifically to equity and justice AND infuse it throughout

Bates | DIGITAL & COMPUTATIONAL STUDIES

Acknowledgments

Thank You!

- Thank you to the Blackwell-Tapia 2021 organizers for the invitation
- EDSIN INCLUDES NSF OIA #1812997
- Feedback from students in DCS 375
- QSIDE support for research & collaboration to revitalize this project, in particular research intern Ethan Siau & Chad Hidgon-Topaz

Bates | DIGITAL & COMPUTATIONAL STUDIES

Questions?