

1 TITLE: Reading Outside the Lines: A Systematic Approach for Detecting Bias in Scientific
2 Communications

3 AUTHORS

4 *Melisa Osborne, Research Assistant Professor, Graduate Program in Bioinformatics, Boston University,
5 Boston, MA

6 TJ McKenna, Clinical Assistant Professor, Wheelock College of Education and Human Development,
7 Boston University, Boston, MA

8 Felicity Crawford, Clinical Associate Professor, Wheelock College of Education and Human
9 Development, Boston University, Boston, MA

10 Theresa Rueger, Lecturer, School of Natural and Environmental Sciences, Newcastle University, UK.

11 Barkha Shah, Laboratory Supervisor, CAS Department of Biology, Boston University, Boston, MA

12 Mae Rose Gott, Bioinformatics Analyst, Graduate Program in Bioinformatics, Boston University, Boston,
13 MA

14 Adam Labadoff, Assistant Professor, Graduate Program in Bioinformatics & Department of Neurology,
15 Boston University Chobanian & Avedisian School of Medicine, Boston, MA

16 *Corresponding author (melosbor@bu.edu)

17

18 ABSTRACT

19 Consciousness of the social impact of science and the potential biases of its authors is critical to
20 understanding, interpreting, and using scientific findings responsibly. This is especially true for sciences
21 concerned with human health and behavior, where societal and unconscious biases may reinforce existing
22 inequities and discriminatory practices. Considering backgrounds and biases, we may notice bias
23 influencing scientists' methodological choices and conclusions, even when a work is otherwise
24 scientifically sound. To this end, we created the pedagogical tool Finding inEquity in Literature and
25 eXperimentation (FELIX). FELIX is a tool that systematizes the detection of bias and subjectivity in
26 scientific communications by using a three-phase progression of (i) Annotation, (ii) Analysis and (iii)
27 Synthesis, where students form a unified argument about the text with a focus on its relationship to social
28 or ethical context. Results from a mixed methods approach indicated the efficacy of our approach in
29 supporting student learning related to reading comprehension, critical thinking skills, and in
30 understanding the social and ethical implications of the research they were reading. We put forward
31 FELIX as a universal method for training students in the reading of scientific communications and as a
32 tool for addressing systemic inequities in science and science education.

33 INTRODUCTION

34 Since the establishment of the field of biology, scientific work has both been impacted by the
35 society in which it is practiced and has in turn had an impact on policy and society. The example of
36 taxonomic hierarchies of human groups ideated in the late 1700's by Linnaeus, Blumenbach, and others
37 was a reflection of societal views on race and in turn a means for continued justification of colonialism
38 and slavery around the globe (Saini, 2019; Muller-Wille, 2014). The science-and-society feedback loop
39 regarding race thus having been established - assumptions about biological differences between people
40 from different so-called racial groups have been maintained throughout the history of science and have
41 been hard to remove, even in the light of their pseudoscientific nature. The history of biology and
42 genetics is littered with examples of how devaluation of groups lower in the social hierarchy has led to
43 unethical scientific practices and personal harm. This is most apparent in considering harms generated
44 during the Eugenic era in America and Europe in the 20th century, including the Tuskegee incident
45 (1930's -1970s) and cruel scientific experimentation carried out by Nazi Germany (Washington, 2008;
46 Gould, 1996; Okrent, 2020). Changes to scientific ethical standards post World War II directly addressed
47 the horrors of eugenic experimentation but have not eradicated ideas of race based biological difference in
48 societies (i.e. the USA and Western Europe) that remain racialized (Saini, 2019). Modern examples of
49 gynecological abuses and the use of race-based data correction are two contemporary instances where
50 Black and Indigenous People of Color have been harmed by these lingering ideas (Montgomery, 2016;
51 Aguilera, 2022; Vyas et al., 2020). The question then becomes how to educate current students of science
52 to examine scientific findings for bias and how to train future scientists, medical professionals, and
53 educators to avoid the incorporation of racially biased ideas into science and medical research
54 methodologies, data interpretation, and clinical practice in professional settings.

55 The history of biological racism makes an awareness of the social impact and implications of a
56 scientific work and the potential biases of its authors critical to appropriately understanding, interpreting,
57 and using scientific findings responsibly (Nature, 2022). For science and medicine concerned with human
58 health and behavior, societal and unconscious biases may reinforce or worsen existing inequities and
59 discriminatory practices based on lingering myths about biological distinctions between different groups
60 of people, e.g. races (Graves and Goodman, 2022; Dasgupta, 2020; Manali, 2018; Donovan et. al, 2019;

61 Hind, 2023; Green et al., 2020). We posit that when scientists are considered first as people who possess
62 their own backgrounds and biases, we may notice bias influencing their methodological choices and
63 scientific conclusions, even when a work is technically sound. We created a teaching tool Finding
64 inEquity in LIterature and eXperimentation (FELIX) to systematize the detection of bias and subjectivity
65 in scientific communications.

66 **Theoretical Framework: Constructivist/Active Learning**

67 The educational model used by our team was based on the Constructivist/Active Learning
68 Theoretical Framework (Brandon & All, 2010). A constructivist approach uses instructional design where
69 learning is framed as the development of personally meaningful understandings of content, developed
70 through interactions with tools and others in a social context. Simply put, constructivism states that
71 people construct their own understanding and knowledge of the world through experiencing things and
72 reflecting on those experiences (Bereiter, 1994).

73 Honebein (1996) summarized seven pedagogical goals of constructivist learning environments as:

- 74 1. To provide experience with the knowledge construction process (students determine how they
75 will learn).
- 76 2. To provide experience in and appreciation for multiple perspectives (evaluation of alternative
77 solutions).
- 78 3. To embed learning in realistic contexts (authentic tasks).
- 79 4. To encourage ownership and a voice in the learning process (student centered learning).
- 80 5. To embed learning in social experience (collaboration).
- 81 6. To encourage the use of multiple modes of representation, (video, audio text, etc.)
- 82 7. To encourage awareness of the knowledge construction process (reflection, metacognition).

83 These pedagogical goals highlight a shift away from prior conceptions of teaching and learning where the
84 teacher (i.e., expert) passively presents course materials to students who are empty vessels waiting to be
85 filled. Rather, a constructivist approach acknowledges that learners are often confronting current
86 understandings in light of new situations, leading to either a confirmation of what they know or the need
87 to change or assimilate this new information. The modification of knowledge arises from the active
88 process of applying current understandings, noting new elements in novel learning experiences, and
89 adjusting based on the consistency of prior and emerging knowledge (Phillips, 1995).

90 **Theoretical Framework: Critical Reading**

91 The importance of interpreting written texts in terms of authorship and societal influence has long
92 been at the heart of critical pedagogy (Friere, 1985) and specifically critical reading pedagogies (Wolf and
93 Barzillai, 2009). These frameworks are well established within education (Molden, 2007) and the social
94 sciences (Van, Li and Wan, 2022; Jewett, 2007; Jensen and Scharff, 2019). The process of gathering
95 information to place a text in its broader context to better understand it (sometimes called “deep”, “slow”,
96 or “close” reading) requires time and effort. While several strategies have been proposed in the context of
97 other disciplines (Jensen and Scharff, 2019), a formal or rigorous approach to performing this type of
98 reading in the biological sciences is not widely practiced. In the sciences, critical reading is often
99 presented in terms of scientific or information literacy with regard to public consumption of scientific
100 texts (Priest, 2013). Students of biology and medicine learn methods for the technical reading and
101 analysis of primary scientific literature with an emphasis on understanding the scientific context of an
102 article, the methods used for the study, and the results of the study. However, it is rare to place these

103 technical readings into a larger societal or historical context. The scientific method itself provides a
104 framework that can be employed to systematize a deep reading of a text.

105 In scientific articles, important pieces of existing evidence are annotated with citations that
106 provide the basis for the validity of a claim or assumption. However, while the same approach may be
107 employed in fields and contexts outside scientific publications, this convention of explicitly citing prior
108 evidence within a text is not as commonplace. It may be unclear in a text that a claim requiring
109 substantiation has even been made, which statements are opinions vs facts, etc, thus making it difficult to
110 know which statements to investigate. An intuition and sensibility for detecting aspects of a text that are
111 biased or unsubstantiated may be developed with practice. FELIX combines these ideas of developing a
112 systematic approach to critical analysis of text and using a standardized set of “measurements” into a tool
113 for identifying and understanding a text with the goal of identifying potentially hidden biases. FELIX
114 uses a three-phase progression: (i) Annotation, (ii) Analysis and (iii) Synthesis. Completion of each
115 phase builds on the step before and guides the student from a detailed understanding of the text to a more
116 conceptual understanding (Figure 1). In the annotation phase, students use a short vocabulary list of
117 terms related to bias in scientific writing to highlight passages in the text. During the analysis, students
118 answer critical analysis questions about the text and its authors to place the work in a social and ethical
119 context. Lastly, via the Synthesis, students bring together elements from the annotation and analysis
120 phases to form a unified argument about the text with a focus on its relationship to social or ethical
121 context. We view this approach as key to placing modern research that deals with human health, genetics,
122 and genomics, in both the context of current society and in relation to the history of biological racism and
123 the subsequent harms that have been caused in the name of science.

124 **Research Questions**

125 Here, we present the bias-detecting reading tool FELIX and results from its use in the context
126 of an upper-level undergraduate biology course on institutional racism in health and science. Our
127 instrument is intended to encourage students to deeply engage with course reading and was developed to
128 achieve the following learning goals (i) develop student critical analysis skills when reading texts, (ii)
129 help students build an intuition for identifying bias and opinions in texts, (iii) measure how student skills
130 from (i) and (ii) change over the course of a semester, and (iv) create a dataset of annotations for a corpus
131 of documents that capture the specific instances of bias-related aspects of the component texts. Our
132 specific research questions revolve around evaluating item (iii), the measurement of changes in student
133 skills over the course of semester. To specifically address this learning goal, we designed our approach
134 with following research questions in mind:

135 Research Question 1: How do the perceptions and critical assessments of students change throughout the
136 course of a semester using FELIX? We will address this question using quantitative analysis of the in-text
137 annotations produced by four cohorts.

138 Research Questions 2: How do students enrolled in the Institutional Racism in Health and Science course
139 at Boston University perceive the usefulness of FELIX? We will address this question using qualitative
140 analysis of anonymous surveys from one cohort.

141

142 **METHODS**

143

144 **Positionality**

145 In relation to the proposed work, we are driven to dismantle and diversify the manifest, monolithic culture
146 of whiteness in science. Although our white team members are beneficiaries of that culture, we recognize
147 the toxic and counterproductive consequences of the exclusionary atmosphere it sustains. We are biased
148 (we think positively) toward viewing systems through an antiracist lens, which may influence the design
149 and interpretation of this work. As a team, we span a range of identities that has enabled us to reveal and
150 mitigate many of our individual biases and blind spots. Each of us also sits at our own intersection of our
151 various identities, providing intersectional perspectives on the issues we tackle. Over our years of work
152 together, our team has built trust and cohesion that has been essential to supporting each other and
153 fortifying this highly transdisciplinary, conceptually and emotionally challenging work on the legacy and
154 reality of racism in our society. However, for all of our collective strengths and diversity, we recognize
155 that many experiences and perspectives are not represented on our team. As faculty and staff at a major
156 university in an affluent US city, we occupy a position of privilege which limits our ability to see the
157 world from many perspectives. We further acknowledge that there are many perspectives which we have
158 never encountered or imagined.

159 **Methodology**

160 The researchers applied GTM (Grounded Theory Methods), an iterative qualitative approach designed to
161 generate theory from data collected through interviews, observations, focused discussions, and document
162 analysis. This choice was due to the complex and intertwined nature of the use of FELIX in a college
163 course, the novel nature of our instrument, and the need to build theory based on empirical data collected
164 during the semesters of use. From a constructivist approach, meaning making is a complex and iterative
165 process and GTM aims to systematically collect and analyze data using an interpretive lens when the
166 learning does not clearly include variables that can be statistically linked (Corbetta, 2003).

167 Researchers developed initial codes by parsing the data into segments, identifying key words, and
168 concepts. Categories based on similarities and differences were developed from which themes and
169 relationships were identified and then associated with categories, causal conditions, and consequences.
170 Every new iteration of the course provided opportunities for theoretical sampling: researchers selected
171 additional data to refine the emerging explanatory model (Charmaz & Bryant, 2019). Guiding their
172 interpretation was the recognition that their own background and biases influenced their analyses. To
173 counter potential pitfalls, the researchers constantly reflected on their own interpretations and biases,
174 while raising questions and seeking alternative explanations. Their explicit goal was to center multiple
175 perspectives, give voice to marginalized voices, challenge dominant narratives, and avoid imposing a
176 single interpretation of the data (Charmaz & Bryant; Corbin & Strauss, 2008).

177 **Data Collection**

178 *Student Cohort and Survey Collection*

179 Students enrolled in the course BI/BF510 Institutional Racism in Health and Science were surveyed
180 anonymously as part of the university course evaluation process. The study cohort consisted of students
181 of junior and senior undergraduate standing as well as graduate MS or PhD candidate status from five
182 semesters of the course for qualitative data (Total N=78; F21 (N=18); S22 (N=16); F22 (N=12); S23
183 (N=22); F23 (N=10) and one semester of the course for quantitative data (F23, N=13). The student pool

184 is enrolled in degree granting programs in Biology and Bioinformatics. Students were given 5 minutes at
185 the beginning of class on the first and final days of the course. Surveys was generated using Qualtrics
186 with QR link generated via open web tools (qr.io); thereby adding accessibility via mobile devices.
187 Students used the projected link or QR code to complete the assessment. Pre and post survey questions
188 included perceptions of science and racial bias. The post survey also asked about student experiences with
189 FELIX. The assessment used a simple 1-5 scale indicating level of agreement (disagree completely – 1;
190 disagree somewhat – 2; not sure/don't know -3; agree somewhat – 4 and agree completely – 5). A
191 condition of the assessment is that instructors do not assess the exit results until after grades are submitted
192 for the semester.

193 *Application of FELIX to Course Readings: The Hashtagulary and Annotation Analysis.*

194 FELIX is composed of three phases: 1) annotation, 2) analysis, and 3) synthesis. During the annotation
195 phase, students use the Perusall web application (Clarke, 2019) to perform and record hashtag annotations
196 (e.g. #opinion) of assigned articles throughout the semester. Students may use any hashtag they deem
197 appropriate, but are instructed to first consider a standardized set of hashtags (the “hashtagulary”,
198 portmanteau of hashtag and vocabulary) that we developed to label elements that might be related to bias.
199 By annotating a text with a controlled vocabulary of hashtags, we can create a consistent dataset that
200 enables meaningful quantitative analysis. The purpose of the hashtagulary is to enable meaningful textual
201 analysis of annotations made by many people of the same text, thus enabling algorithmic characterization
202 of the presence and specific passages that might contain bias. The hashtagulary is listed in Table 1.

203 Perusall allows export of all annotation data that includes date, article, annotator, annotation comment,
204 and annotation position in the text. This data is then downloaded, cleaned, and analyzed to generate a
205 processed dataset of various hashtag counts and locations. The cleaning and analysis are performed in
206 Python and the current analysis is performed using Jupyter notebooks.

207 *Data Sharing*

208 Jupyter notebooks were used for hashtag analysis and can be found at the following locations:
209 https://colab.research.google.com/drive/1aUl-mRkmqKl6ItIA9a3UyeqmuQ1_5kZ3?usp=sharing

210 *Quantitative Assessment of FELIX – Survey Analysis*

211 Quantitative surveys were used to investigate the effect of the course on the effectiveness of FELIX as a
212 teaching tool. Surveys were generated using Qualtrics with QR link generated via open web tools (qr.io);
213 thereby adding accessibility via mobile devices. Students used the projected link or QR code to complete
214 the assessment. Students were given 5 minutes at the beginning of class on the first day of the course for
215 the entry survey and were sent the same survey after final grades were submitted via email. Pre and post
216 survey questions included perceptions of science and racial bias. The post survey asked about student
217 experiences with FELIX. The assessment used a simple 1-5 scale indicating level of agreement (disagree
218 completely – 1; disagree somewhat – 2; not sure/don't know -3; agree somewhat – 4 and agree
219 completely – 5). A condition of the assessment was that instructors do not assess the exit results until
220 after grades are submitted for the semester.

221 *Qualitative Analysis of Student Feedback – Student Questionnaire*

222 Qualitative feedback from the students was collected from students each semester through an anonymous
223 exit survey completed when the students handed in their final projects during finals week. A condition
224 of soliciting feedback was that instructors do not assess the exit results until after grades are submitted for
225 the semester. We posed the following question to the cohort regarding the use of FELIX in the course -

226 “Did you find the instrument to be effective at understanding bias in a text?” Students were asked to
227 further elaborate on why or why not they found the instrument to be helpful. Students were also asked to
228 specify which phase of FELIX they found to be the most useful to their learning. We compared terms in
229 the evaluations that indicated both positive student experiences and conversely, student confusion.
230 Criterion for the different themes were as follows - positive themes included language that indicated a
231 deeper understanding of the material; improved reading comprehension; new perspectives, increased
232 critical thinking, and intention to use skills from the instrument beyond this classroom; negative themes –
233 included expression of confusion and dislike of any aspects of the process; neutral themes – were the
234 absence of positive or negative themes.

235 **RESULTS**

236 *Hashtag and Annotation Analysis.*

237 The data from the annotation phase of this study was used to construct heatmaps showing key
238 moments in the text. A heatmap is a data matrix where coloring is used to illustrate an overview of the
239 numeric differences. Our heatmaps were created in Python to note which hashtag students used to
240 annotate the text (on the vertical axis) and the location of that annotation in the overall journal article (on
241 the horizontal axis). Figure 2 shows student annotations for the article Race Crossing in Jamaica [13]. The
242 top left (A) provides overall counts of the annotations used, the top right shows (B) shows how a
243 correlation matrix of the hashtags used, and the bottom heatmap (C) shows specific locations in the text
244 where annotations were used. These visualizations highlight the effectiveness of our implementation of
245 FELIX in this course as they provide a window into what students choose to focus on and how they are
246 interpreting that passage of text. A heatmap was created for each article, allowing our teaching team to
247 assess how students were grasping the course content and inform the focus of our in-class discussions.

248 We also examined the annotation for evidence of course-wide trends in student learning and
249 development. After cleaning, students used 541 unique hashtags (14 are in the hashtagulary) across
250 13,231 annotations made by 82 students. 68% of the annotations made utilized the hashtagulary hashtags.
251 We compared the annotation count for hashtagulary or student created hashtags and found some hashtags
252 were more frequently annotated than others, where some student created hashtags were used with greater
253 frequency than the standard set (Figure 3A). When hashtags patterns were compared with the
254 chronological course as measured by number of days into the semester (to normalize across all four
255 semesters) students consistently used hashtagulary annotations throughout but used increasingly more of
256 their own hashtags as the semester progressed (Figure 3B). We interpret this to mean students gained
257 more comfort and skill using FELIX annotations with practice. Finally, we performed a sentiment
258 analysis of annotations by manually annotating each hashtag with either a -1, 0, or 1 based on whether the
259 hashtag expressed negative, neutral, or positive sentiment, respectively. For example, hashtags #injustice,
260 #opinion, and #substantiated were labeled as -1, 0, and 1 respectively. We see a trend toward positive
261 sentiment across the semester, which we interpret as due to the course readings being published closer to
262 the present as the semester proceeds.

263 *Student Perspectives and Analysis of FELIX*

264 Quantitative feedback from the students in IRHS indicated an overall positive viewpoint on
265 FELIX as a reading method (Table 2), with students agreeing statements positing the usefulness of each
266 phase of FELIX (Annotation; Analysis; Synthesis). The effectiveness of FELIX as providing a new
267 perspective to understanding the intersection of race, genetics, and biology was also indicated in the
268 results (Figure 4). Student perspectives shifted over the course of the semester with less agreement at the

269 end of the semester with statements linking race with genetics and biology (orange dots, Figure 3).
270 Furthermore, student perspectives about the objectivity of science and scientists indicated a greater
271 awareness of the bias present in both the system and the people in the system (orange dots, Figure 3).

272 We collected additional qualitative feedback from the students in IRHS indicating the
273 effectiveness of FELIX in improving student learning and in providing a new perspective to training in
274 the biological sciences (Table 2). We posed the following question to the cohort after submission of the
275 final project at the end of the semester - “Did you find the instrument to be effective at understanding bias
276 in a text?” Most students from across all four cohorts answered “yes” to the effectiveness of the
277 instrument (87%, 68/78 responses), with 12% of the cohort responding neutrally (9/78) and 1%
278 responding negatively (1/78).

279 Students were asked to further elaborate on why or why not they found the instrument to be
280 helpful. For example, one student responded:

281 The instrument was incredibly effective at understanding bias in a text. I felt that using the three
282 steps of the instrument made you engage with the readings in a way that just reading it would not.
283 I particularly thought that the annotations were the most impactful in making me understand bias
284 in the readings. (S8; 12/13/21)

285 This student highlights how FELIX allowed them to engage with the readings in a way that allowed for a
286 deeper understanding of the course materials. Another student shared how the steps of FELIX helped
287 them in this course and how they plan to apply these in their future:

288 I absolutely will be using the techniques I learned while reading papers in the future; it made me
289 look at the papers more deeply than just absorbing the scientific information presented. I felt the
290 annotation and analysis were very helpful and the analysis questions are something I can use in
291 my everyday life while reading papers to think about who is included/excluded. (S5; 12/18/2022)

292 We compared terms in the evaluations that indicated both positive student experiences and conversely,
293 student confusion. Positive themes we saw indicated a deeper understanding of the material (N=20);
294 improved reading comprehension (N=6); new perspectives to learning the material (N=6), increased
295 critical thinking (N=6) and intention to use skills from the instrument beyond this classroom (N=4) (Table
296 2). Because these labels were not overlapping, this indicated that 42/68 evaluations had language
297 specifically denoting new skills attained by the students from using this tool.

298 Students were also asked to specify which phase of FELIX they found to be the most useful to
299 their learning. Students specifically cited more utility for annotation (N=32) vs Analysis (N=30) vs
300 Synthesis (N=20). These data are more reflective of themes that developed in the analysis. First, students
301 indicated a positive appreciation of the application of the phasewise process of FELIX (N=7), in which
302 the annotation and analysis phases were linked together. Secondly, those students indicating appreciation
303 of analysis also indicated appreciation for digging deeper into the backgrounds of the authors of scientific
304 papers (N=3).

305 DISCUSSION

306 This study contributes a unique pedagogical approach to reading scientific articles in the context
307 of upper level, undergraduate biology courses - one in which the history of biological racism is brought to
308 bear in looking for lingering racial bias in genetics, genomics, and human health literature. Our approach
309 is autodidactic in nature, with student learning that is generated from the engagement of the students with

310 the reading process via the FELIX framework. Additionally, knowledge is generated about the texts at
311 hand via the annotation process that is carried out during the annotation step of the instrument and
312 generates a dataset of student hashtags for each reading throughout the semester (Figure 3).

313 Our quantitative analyses of student perceptions from one cohort (Fall 2023) suggests the
314 effectiveness of FELIX in changing student viewpoints on the objectivity of scientists and the link
315 between biology and racism over the course of one semester. However, there are several limitations to
316 the initial analyses. Since the quantitative survey was a pilot ($N = 13$ pre / 8 post), the sample size for the
317 data is small and requires larger numbers to corroborate the pilot findings. Secondly, the pilot questions
318 require revision to include control questions regarding student perception of learning. Currently, all the
319 survey questions are stated positively and in a specific order. We would seek to include questions for
320 which a negative or neutral response would be expected and to randomize the order in which questions
321 are posed. The authors recognize the irony of trying to quantify student perceptions of their critical (and
322 subjective) analyses of texts. It is inherently difficult to introduce a tool for critical thinking and to widen
323 perspectives and then to assess that tool in an objective way. We are excited to refine our methods and
324 expand our data collection in current and future cohorts to face these challenges.

325 Overall, qualitative student feedback also indicated FELIX provided a systematic approach that
326 students found to be a useful tool for their learning. Students reported that it helped them to improve their
327 reading comprehension, critical thinking skills, and in understanding the social and ethical implications of
328 the research they were reading. Nonetheless, there are areas of our methodology that can be improved.
329 Students indicated a preference for the earlier steps of the instrument - annotation and analysis. Feedback
330 suggested that the current implementation of the synthesis phase could be improved to tie in the previous
331 two phases more strongly and to engage students more effectively. When it came to the annotation
332 process, students also indicated some confusion with the hashtags and feeling restricted by the
333 vocabulary. Future iterations will work to make the hashtag vocabulary a more dynamic process with
334 student input. Given the positive qualitative feedback collected over all five of the course offerings
335 ($N=78$), we are currently working to develop shorter workshop versions of FELIX that can be offered on
336 a condensed time frame compared to a full semester course.

337 Lastly, our assessment surveys - both qualitative and quantitative - have focused on perceived
338 learning on the part of the students. While this may be a good way to indicate student appreciation for the
339 material being learned; a more concrete assessment method would be ideal for determining if critical
340 reading skills are indeed increasing from using FELIX. Future studies in which students do a pre and post
341 annotation of articles is in the process of being designed and implemented to better quantify changes in
342 use of hashtags and indication of biases before and after becoming familiar with use of our instrument.

343 The result of our pilot leads us to a lingering research question: does learning about the history of
344 harms perpetrated by science and medicine due to racial bias change how scientists and medical
345 professionals think? Does it generate more empathetic and aware practitioners who will carry out their
346 work and research in ways that ultimately cause less harm? How can one quantify life-long effects from
347 encouraging students to think about the social and ethical implications of their scientific research?
348 Regardless of how we answer these questions, history shows us that bias can influence the design,
349 conduct, and interpretation of scientific research. This history is a reason why learning how to become
350 critical readers of scientific literature should be a key component of university teaching and learning.
351 Future work will include building on this foundation to compare learning outcomes to cohorts with and
352 without training using FELIX and tracking student use of FELIX in contexts outside of our course. Our
353 piloting of this method and the preliminary indications of its efficacy lead us to put forward this approach
354 as a universal method for training students in the reading of scientific literature and communications.

355 **ACKNOWLEDGMENTS**

356 Thank you to the Lab of Daniel Segre at Boston University for discussion and helpful feedback on FELIX
357 and the course, Institutional Racism in Health and Science, as well as for reading and discussing drafts of
358 this manuscript. Thank you to all the IRHS510 students, who inspire and motivate us in our work every
359 day.

360 **REFERENCES**

- 361 Aguilera, J. (2022). *Senate Committee Finds Medical Abuse of Detained Women at Georgia ICE*
362 *Facility.* " *Time Magazine.* Published Nov 16, 2022. Accessed Jan 25, 2024.
363 <https://time.com/6234031/medical-abuse-georgia-women-detained/>
- 364 Bereiter, C. (1994). Constructivism, socioculturalism, and Popper's World 3. *Educational Researcher*, 23
365 (7), 21-23.
- 366 Brandon, A. F., & All, A. C. (2010). Constructivism theory analysis and application to curricula. *Nursing*
367 *education perspectives*, 31(2), 89-92.
- 368 Cell Editorial Team (2020). Science Has a Racism Problem. *Cell*, 181(7), 1443–1444.
369 <https://doi.org/10.1016/j.cell.2020.06.009>
- 370 Charmaz, K., & Bryant, A. (2019). Qualitative data analysis (2nd ed.). Sage Publications.
371
- 372 Clarke, A. (2019). Perusall: Social learning platform for reading and annotating (perusall LLC,
373 perusall.com), *Journal of Political Science Education*, DOI: 10.1080/15512169.2019.1649151.
374
- 375 Corbetta, P. (2003), *Social Research: Theory, Methods and Techniques*, Sage, Thousand Oaks, CA.
- 376 Corbin, J. & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing*
377 *grounded theory* (3rd ed.). Thousand Oaks, CA: Sage
- 378 Dasgupta S. (2020). An Anti-Racism Toolkit for the Genetics Educator. *Genet Med.* 2020
379 Nov;22(11):1911-1912. doi: 10.1038/s41436-020-00937-6. Epub 2020 Aug 13. PMID: 32788666.
- 380 Davenport, C. B. (1928). Race Crossing in Jamaica. *The Scientific Monthly* 27 (3): 225–38.
- 381 Donovan, et, al. (2019). Humane genetics curricula: Toward a more humane genetics education:
382 Learning about the social and quantitative complexities of human genetic variation research could reduce
383 racial bias in adolescent and adult populations. *Science Education*. 103 (3) 529-560.
384 <https://doi.org/10.1002/sce.21506>
- 385 Freire, P. (1985). Rethinking critical pedagogy: The politics of education, culture, power, and liberation.
386 Granby: Bergin and Garvey.
- 387 Gould S. J. (1996). *The mismeasure of man* (Rev. and expanded). Norton.
- 388 Graves J.L. Jr, Kearney M., Barabino G, et al. (2022). Inequality in science and the case for a new
389 agenda. *Proc Natl Acad Sci U S A.* 2022 Mar 8;119(10):e2117831119. doi: 10.1073/pnas.2117831119.
390 PMID: 35210356; PMCID: PMC8915968.

- 391 Graves J. L. & Goodman A. H. (2022). *Racism not race:answers to frequently asked questions*. Columbia
392 University Press.
- 393 Green et al. (2022). Deconstructing Racism, Hierarchy, and Power in Medical Education: Guiding
394 Principles on Inclusive Curriculum Design. Academic Medicine 97(6):p 804-811, June 2022. | DOI:
395 10.1097/ACM.00000000000004531
- 396 Hind A. Al-Abadleh (2023). A critical look at the practice and culture of science with calls to
397 action, Communications Chemistry, **6**, 1.
- 398 Honebein, P.C. (1996). Seven goals for the design of constructivist learning environments.in
399 Constructivist Learning Environments: Case Studies in Instructional Design. Brent G. Wilson (Ed.).
400 Englewood Cliffs: Educational Technology Publications: 11-24.
- 401 How Nature contributed to science's discriminatory legacy. (2022). Nature, 609(7929), 875–876.
402 <https://doi.org/10.1038/d41586-022-03035-6>
- 403 Jensen, M., & Scharff, L. (2019). Improving Critical Reading with E-Texts: A Controlled Study in a
404 Collegiate Philosophy Course. The Journal of Scholarship of Teaching and Learning, 19(3), 49-.
405 <https://doi.org/10.14434/josotl.v19i2.23978>
- 406 Jewett. (2007). Reading Knee-Deep. Reading Psychology., 28(2), 149–162. DOI:
407 [10.1080/02702710601186365](https://doi.org/10.1080/02702710601186365)
- 408 Manali J. Sheth. (2018). Grappling with racism as foundational practice of science teaching. First
409 published: 22 May 2018 <https://doi.org/10.1002/sce.21450>
- 410 Mills, J., Bonner, A. and Francis, K. (2006), The development of constructivist grounded theory.
411 International Journal of Qualitative Methods, Vol. 5 No. 1, pp. 25-35, doi:
412 [10.1177/160940690600500103](https://doi.org/10.1177/160940690600500103).
- 413 Molden, K. (2007). Critical Literacy, the Right Answer for the Reading Classroom: Strategies to Move
414 beyond Comprehension for Reading Improvement. Reading Improvement, 44(1), 50-.
- 415 Montgomery, D. Sterilized against their will in a Los Angeles hospital: Latinas tell the story in a new
416 film. *The Washington Post*. Published Jan 10, 2016. Accessed online Jan 25, 2024.
417 <https://www.washingtonpost.com/news/arts-and-entertainment/wp/2016/01/10/sterilized-against-their-will-in-a-los-angeles-hospital-latinas-tell-the-story-in-a-new-film/>
- 419 Müller-Wille S. (2014). Race and History: Comments from an Epistemological Point of View. Sci
420 Technol Human Values. 2014 Jul 1;39(4):597-606. doi: 10.1177/0162243913517759. PMID: 25684833;
421 PMCID: PMC4326670.
- 422 Okrent, D. (2020). The guarded gate : bigotry eugenics and the law that kept two generations of jews
423 italians and other european immigrants out of america. Scribner.
- 424 Phillips, D.C. (1995). The good, the bad, and the ugly: The many faces of constructivism. Educational
425 Researcher, 24 (7), 5-12.
- 426 Priest, S. (2013). Critical Science Literacy: What Citizens and Journalists Need to Know to Make Sense
427 of Science. Bulletin of Science, Technology & Society, 33(5–6), 138–145.
428 <https://doi.org/10.1177/0270467614529707>

- 429 Saini A. (2019). *Superior : the return of race science*. Beacon Press.
- 430 Strauss, A. and Corbin, J. (1998), *Basics of Qualitative Research: Techniques and Procedures for*
431 *Developing Grounded Theory*, Sage, Thousand Oaks, CA.
- 432 Van, L. H., Li, C. S., & Wan, R. (2022). Critical reading in higher education: A systematic review.
433 *Thinking Skills and Creativity*, 44, 101028-. <https://doi.org/10.1016/j.tsc.2022.101028>
- 434 Vyas, D. A., Eisenstein, L. G., & Jones, D. S. (2020). Hidden in Plain Sight - Reconsidering the Use of
435 Race Correction in Clinical Algorithms. *The New England Journal of Medicine*. 383(9), 874–882.
436 <https://doi.org/10.1056/NEJMms2004740>
- 437 Washington H. A. (2008). *Medical apartheid : the dark history of medical experimentation on black*
438 *americans from colonial times to the present*. (First Anchor books (Broadway Books). Anchor Books.
- 439 Wolf, M., & Barzillai, M. (2009). The importance of deep reading. In *Educational Leadership* (Vol. 66,
440 Issue 6, pp. 32-). Association for Supervision and Curriculum Development.
- 441

442 **Captions**

443

444 **Figure 1:** FELIX, a teaching tool for Finding inEquity in LIterature and eXperimentation. The
445 conceptual framework and practical implementation of FELIX via the three phases – annotation, analysis,
446 and synthesis.

447 **Figure 2:** Annotation analysis of Davenport “Race Crossing in Jamaica” 1928 from four semesters. A)
448 Total count of hashtag annotations. B) Correlation of pairwise hashtags, i.e. hashtags frequently
449 annotated to the same passage have high correlation. C) Positional analysis showing passages in text
450 annotated with corresponding hashtag. Left to right is position in text, darker colors indicate more
451 annotations of a hashtag to that location. Text passages (1,2,3) with notable annotations (#generalization,
452 #opinion, #speculation, respectively) are quoted from Davenport as indicated.

453 **Figure 3:** Annotation Trends. A) Number of annotations vs number of students using each hashtag. B)
454 Number of unique hashtags used in annotations by all students vs chronological day in the semester. C)
455 Mean sentiment for all hashtags in all students for each reading across each semester.

456 **Figure 4:** Students' perceptions of key course concepts at the beginning of the semester (Pre; blue dots;
457 N=15) and at the end of the semester (Post; Orange dots; N=8).

458

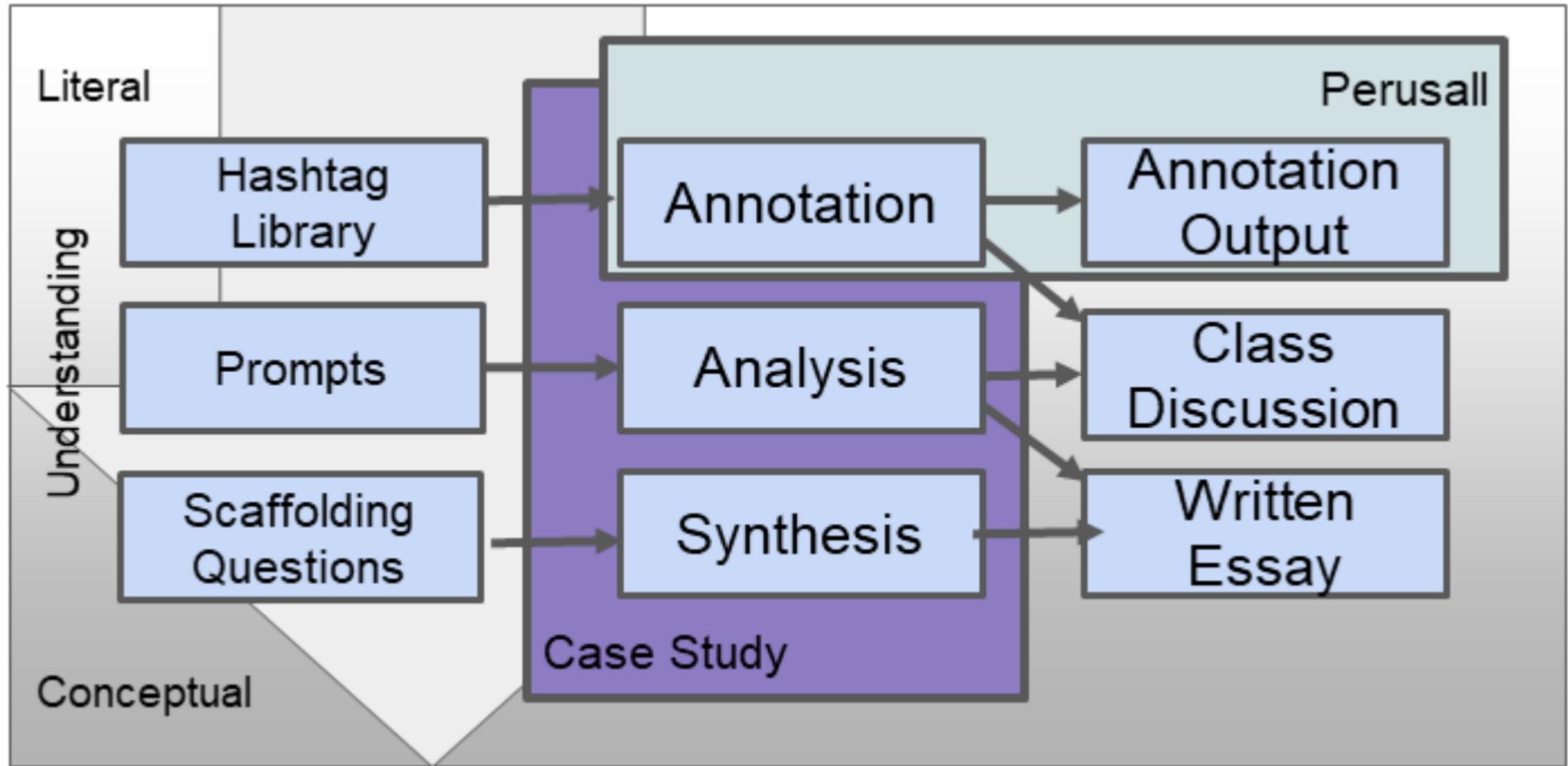
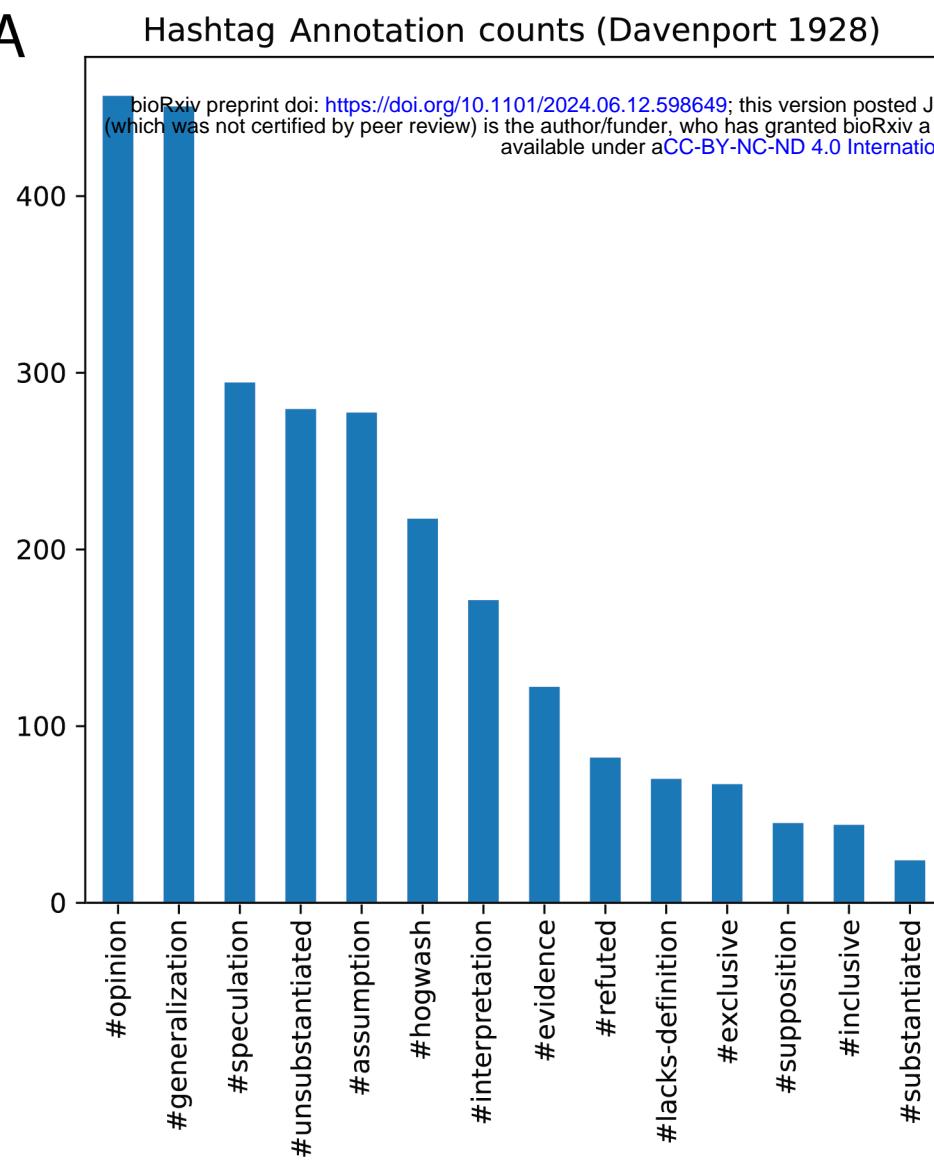


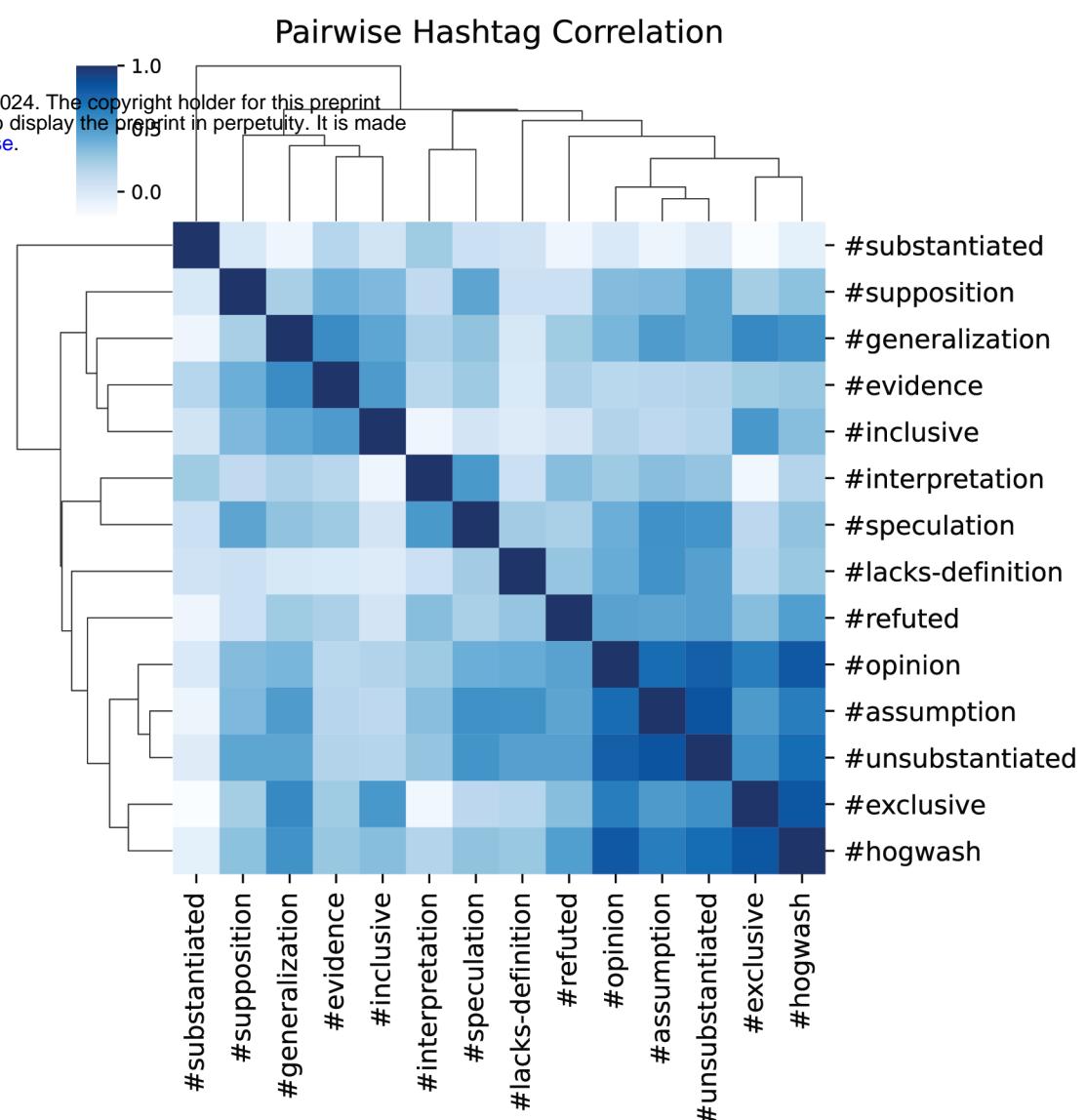
Table 1: The Hashtagulary - a hashtag vocabulary defined as set of hashtags used by students to annotate texts. Students may use other hashtags as they see fit but are instructed to consider these definitions first.

Hashtag:	Defined As:
assumption	accepted as true or correct without explicit justification or evidence
evidence	a fact or result that provides support for a claim
exclusive	idea or claim explicitly excludes a group of people
generalization	a general statement or concept obtained by inference from specific cases
hogwash	nonsense (tongue in cheek)
inclusive	idea or claim explicitly includes a group of people
interpretation	a belief or opinion the author holds to follow from evidence or claim
lacks-definition	a key term used without explicit definition, or assumes reader has a particular definition
opinion	a view or judgment about something, not necessarily based on fact or knowledge
refuted	other existing evidence directly contradicts the idea or claim
speculation	the forming of a theory or conjecture without explicit or comprehensive evidence
substantiated	adequate evidence is provided for idea or claim
supposition	an uncertain belief
unsubstantiated	no supporting evidence provided for idea or claim

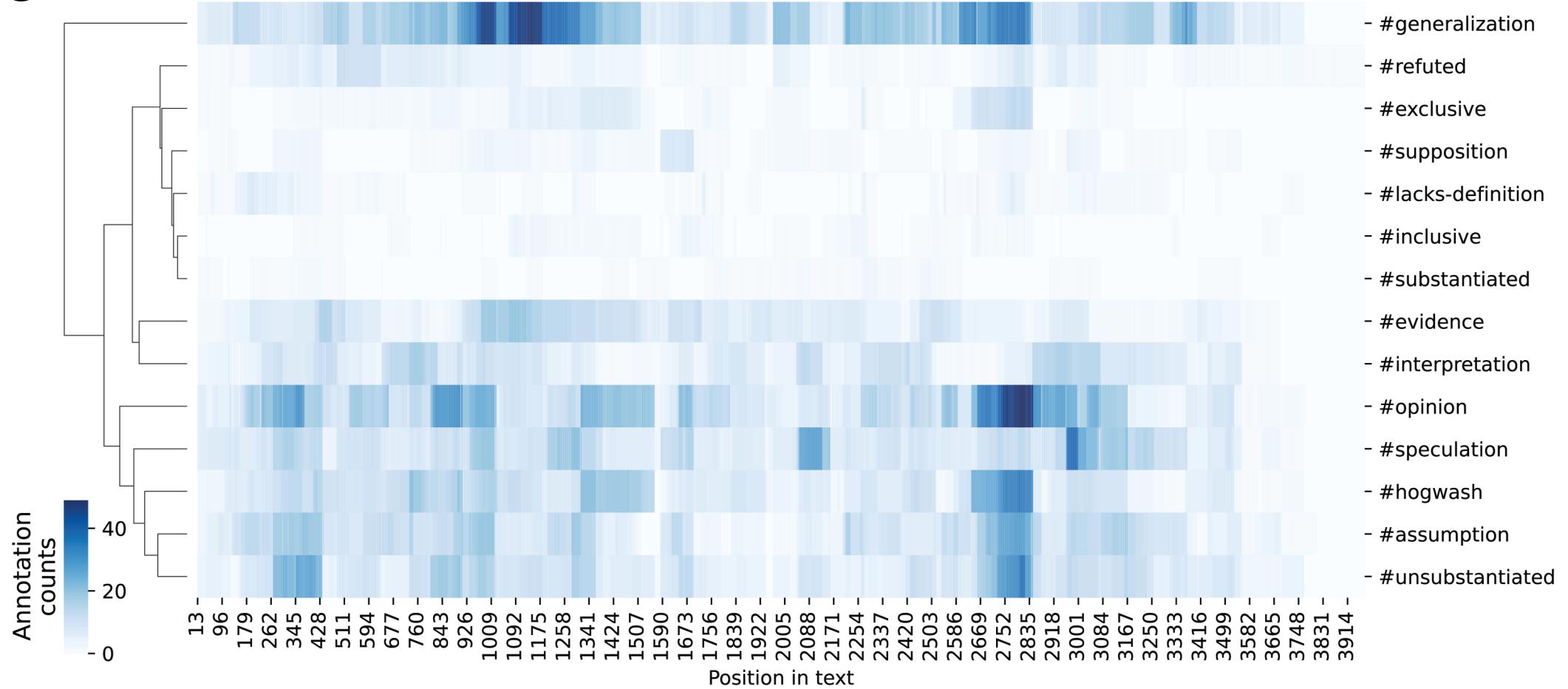
A



B

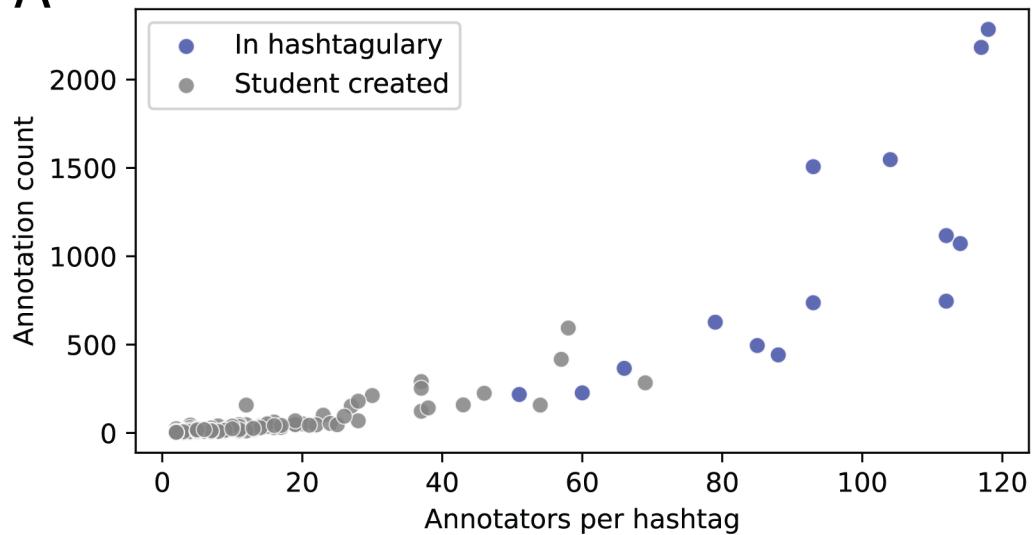


C



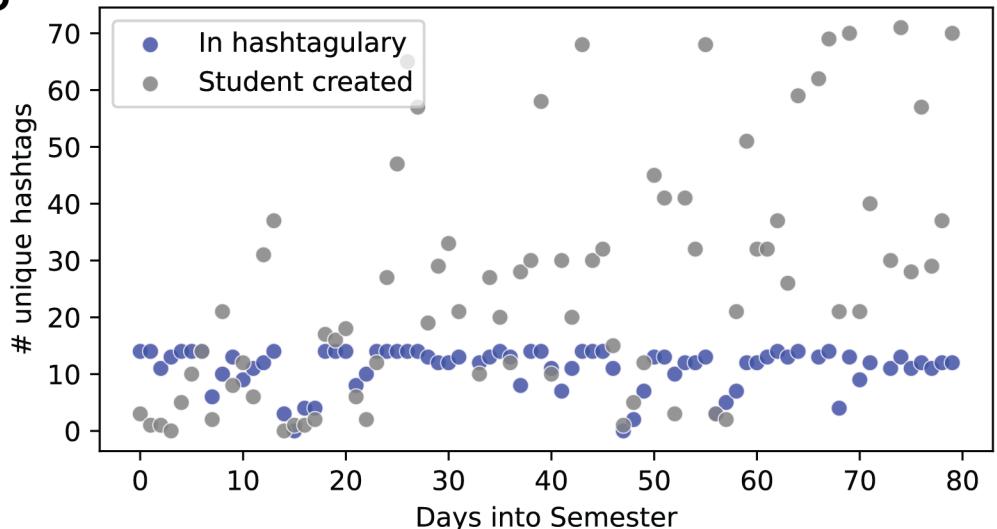
A

of annotators per hashtag



B

Hashtagulary vs student created annotations



C

Annotation Sentiment Analysis of All Readings

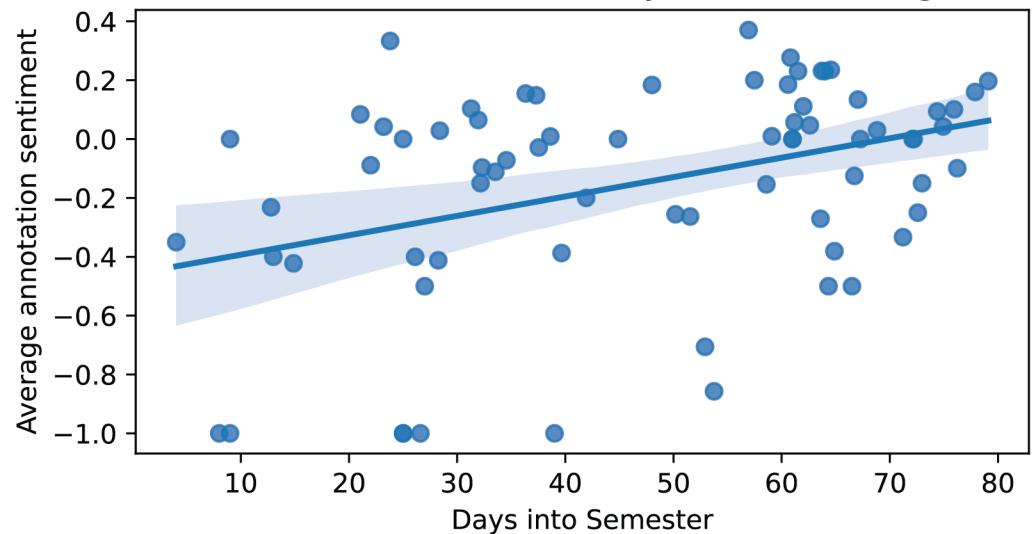


Table 2: Results from Quantitative Evaluation of FELIX Post Semester (F23; N=8). The assessment used a simple 1-5 scale indicating level of agreement (disagree completely – 1; disagree somewhat – 2; not sure/don't know -3; agree somewhat – 4 and agree completely – 5). Higher scoring indicates MORE agreement. (N=8).

Statement	Mean (5=agree completely)	Standard error
I found completing annotations to be useful to understanding the assigned text.	3.9	0.41
I found completing the analysis questions in groups during class discussion time to be useful to understanding the assigned text.	4.7	0.15
I found completing the analysis questions in the case studies to be useful to understanding the assigned text.	4.7	0.15
I found the synthesis step (case study) useful to understanding the assigned text.	4.3	0.34
I found the reading approach (FELIX) to be useful/helpful.	4.6	0.24
Using the reading approach (FELIX) in class put science papers into larger context.	4.7	0.23
The authorship of a paper can influence its conclusions.	5.0	0.0

Pre/Post Survey Responses

