

Social Media Mining to Understand the Impact of Co-operative Education on Mental Health

Mohammad S. Parsa and Lukasz Golab
University of Waterloo
{mohammad.parsa,lgolab}@uwaterloo.ca

ABSTRACT

Co-operative education is a form of work-integrated learning that includes both classroom study terms and paid work experience. Research on co-operative education focuses on its benefits to students, employers, and academic institutions. In contrast, we study the impact of co-operative education on students' mental well-being. To do so, we mine social media content on the Reddit platform, which includes, among many other topics, discussion forums for major U.S. and Canadian colleges. Specifically, we perform topic modelling of discussions related to mental health and co-operative education. We find that students report feelings of self-doubt resulting from a competitive co-op job market, especially when placed in entry-level jobs that are not related to their academic programs, and anxiety due to job interviews, especially when they coincide with exams and other academic deadlines.

1. INTRODUCTION

Co-operative education (co-op) programs combine academic content with paid work experience. For example, students may alternate between classroom study terms and work-terms. Co-operative education programs, both at the undergraduate and graduate levels, have become popular as they offer practical work experience for students and a talent pipeline for employers [3, 26].

Prior work has examined the effect of co-operative education on students and employers. From a student point of view, studies have illustrated the impact of co-op on skill and career growth (see, e.g., [22, 13]). From an employer point of view, there has been work on understanding employers' expectations (see, e.g., [4, 16, 19]). On the other hand, there is less work on the effect of co-op on students' mental well-being, aside from small-scale studies of specific issues such as failing to obtain co-op employment (details in Section 2). This is, however, a pressing issue as recent work reports a rise in mental health problems among college students [1].

Mohammad S. Parsa and Lukasz Golab "Social Media Mining to Understand the Impact of Co-operative Education on Mental Health" In: *Proceedings of The 13th International Conference on Educational Data Mining (EDM 2020)*, Anna N. Rafferty, Jacob Whitehill, Violetta Cavalli-Sforza, and Cristobal Romero (eds.) 2020, pp. 653 - 657

To fill this gap, we analyze social media to discover what students say about the impact of co-operative education on their well-being. Specifically, we perform topic modelling of U.S. and Canadian university discussion forums on the Reddit social media platform (reddit.com), followed by a detailed inspection of topics related to co-op.

In contrast to prior work based on surveys of small groups of students from a single institution, our study is based on a large dataset containing student-generated social media content from over 50 institutions, and is not limited to specific issues or students in specific circumstances. Furthermore, it has been recognized that the anonymity of social media makes it suitable for discussing sensitive issues. However, while there has been prior work on using social media such as Reddit and Twitter to understand mental health issues [7, 5, 18, 15, 14, 21, 8], including issues experienced specifically by students [1], these studies have not reported any issues related to co-operative education.

Our main findings are as follows. First, we find indications of self-doubt resulting from competition, specifically by students unable to secure highly-paid and popular co-op positions, and by students placed in entry-level jobs that are unrelated to their academic programs. Second, interviews for co-op positions appear to be causing anxiety: students fear being unprepared or unqualified, especially when interviews coincide with exams. These findings suggest actionable insights for academic institutions, including managing students' expectations and ensuring that co-op interviews do not conflict with academic deadlines.

2. RELATED WORK

In the context of social media mining, the closest work to ours is that of Bagroy et al. [1], which proposed a mental well-being index for college campuses. The index was computed by measuring the fraction of a given college's Reddit discussions that were related to mental health issues, as determined by a classifier. In a related study, Saha et al. [24] computed the fraction of these discussions that was classified as hate speech, and identified expressions of stress linked to exposure to hate speech. However, these studies did not report any issues related to co-op.

Next, we review related survey-driven studies. Drysdale and McBeath [12] surveyed 1970 students about psychological attributes such as hope, procrastination, self-efficacy, and study skills. They found that co-op students had lower anx-

ity, a better attitude, better use of study aids, and better time management. Drewery et al. [10] surveyed 1989 co-op students and found that students who see a strong connection between the work term and their academic program are more likely to feel satisfied and perform well. Rowe [23] surveyed 29 researchers about neglected negative aspects of co-op. Some of the reported issues were related to mental health, e.g., depression of students unable to find co-op jobs or placed in jobs that are unrelated to their academic programs, and disconnect from campus life caused by alternating work and study terms. Cormier and Drewery [6] surveyed 82 students and found that those who did not find co-op employment reported negative feelings. On a similar note, Drewery et al. [11] tested two interventions, on 74 participants, to improve unemployed co-op students' well-being: a writing exercise and information about coping with stress. The first was found to be effective, but not the second. Finally, Deziel et al. [9] surveyed 312 students about their mental health and found that it is related to academic and demographic factors such as program, year of study, and gender. However, the effect of co-op was not considered.

3. DATA AND METHODS

3.1 Data Collection and Pre-Processing

Reddit is an online social media platform divided into over 100,000 user-created discussion communities referred to as *subreddits*. A subreddit contains a number of *posts* that initiate discussions, and a post is followed by (zero or more) *comments*. Subreddit names begin with “r/” and are indicative of the content. For example, r/Fitness contains discussions of fitness and physical exercise, r/StarWars is a forum for fans of Star Wars movies, etc. As of 2019, there are over 400 million users on Reddit. Each user has a Reddit ID, but is not required to reveal any personal information.

Previous work [1] has identified the subreddits corresponding to top U.S. colleges according to U.S. News¹. We also use these subreddits in our analysis, listed in Table 1. Additionally, we collected the subreddits corresponding to top Canadian universities according to McLean’s Magazine², listed in Table 2. We downloaded these subreddits (posts and comments) from a publicly accessible database on Google Big Query³, spanning from September 2015 to September 2019. The sizes of each studied subreddit are shown in Table 1 and 2, in the “before” columns; the numbers in the “after” columns refer to content relevant to mental health and co-operative education, as determined by our filtering methods described in Section 3.2.

Next, we perform standard text pre-processing. Following previous work on Reddit data mining [17], we remove posts and comments with fewer than 256 or more than 4096 characters: short ones are unlikely to be meaningful (and may instead correspond to URLs), while long ones may mention more than one topic. We also remove stopwords and we lemmatize the remaining words (i.e., we group together all the *inflected* forms of a word) using the Python NLTK parser.

¹<https://www.usnews.com/best-colleges/rankings/national-universities>

²<https://www.macleans.ca/education/university-rankings-2020-canadas-top-comprehensive-schools/>

³<https://cloud.google.com/bigquery>

Table 1: U.S. academic subreddits: number of posts and comments before and after processing.

Subreddits	Posts		Comments	
	before	after	before	after
r/UIUC	5893	423	27258	1864
r/rutgers	4062	263	12858	913
r/UMD	2861	198	10748	916
r/UCSD	2638	163	10991	771
r/Purdue	2408	183	10540	883
r/berkeley	2254	170	15946	970
r/UTAustin	2134	133	8744	507
r/utdallas	1974	146	6476	524
r/gatech	18623	305	14605	1386
r/Cornell	1718	75	5865	453
r/udub	1571	97	7422	456
r/uofm	1550	99	6254	546
r/SBU	1450	90	4367	246
r/rit	1363	96	7080	684
r/UWMadison	1322	93	5040	343
r/RPI	1207	91	7119	571
r/SJSU	1187	51	3955	306
r/nyu	1146	82	2975	169
r/PennStateUniversity	1134	85	5255	369
r/NCSU	1110	58	3679	293
r/msu	1074	52	4191	316
r/UGA	1026	69	3511	278
r/USC	931	51	2851	197
r/UVA	616	49	2273	116
r/uichicago	532	37	1178	93
r/UNCCharlotte	512	32	1635	94
r/stanford	510	42	1416	112
r/UPenn	491	35	1276	68
r/columbia	411	30	1428	55
r/cmu	333	25	1320	118
r/Baruch	324	19	796	57
r/IndianaUniversity	320	25	1311	79
r/mit	316	25	1487	121
r/UMBC	286	11	822	59
r/Harvard	241	18	1081	73
r/BrownU	219	14	603	28
r/byu	198	20	1404	75
r/duke	187	10	502	19
r/UNC	184	9	416	21
r/washu	179	15	622	30
r/Vanderbilt	156	9	334	16
r/bostoncollege	96	3	315	9
r/Caltech	77	11	232	20
Total	66824	3512	208181	15224

3.2 Content Filtering

Academic subreddits discuss a variety of topics related to the corresponding college, such as admissions, academics and campus events. Thus, the next step is to *filter* the data and identify discussions that are relevant to our analysis, namely those which 1) are related to mental health, and 2) are related to co-operative education.

First, we apply a *classifier* that predicts whether a post or a comment is likely to be related to mental health. We use the logistic regression classifier from Bagroy et al. [1], which was originally used to compute the percentage of discussions on academic subreddits that are related to mental health. This classifier was trained by considering all posts on the subreddit r/mentalhealth to be mental-health-related and all posts on control subreddits (among them r/food, r/technology, and the FAQ forum r/AskReddit) to be unrelated.

Next, we only retain posts and comments that appear re-

Table 2: Canadian academic subreddits: number of posts and comments before and after processing.

Subreddits	Posts		Comments	
	before	after	before	after
r/uwaterloo	8912	1836	43382	5215
r/UofT	7895	588	32929	2490
r/UBC	3577	406	20504	1485
r/uAlberta	2766	146	8968	576
r/yorku	2612	182	9877	531
r/mcgill	2603	171	10517	635
r/Concordia	1643	129	4042	286
r/uwo	1599	112	6167	401
r/ryerson	1383	82	4018	374
r/CarletonU	1320	134	4988	473
r/McMaster	928	63	2354	218
r/queensuniversity	763	35	2452	163
r/uvic	665	41	2400	181
r/wlu	458	38	1159	114
r/uoguelph	399	29	960	99
r/Dalhousie	293	24	667	39
r/humanitoba	165	23	354	18
r/brocku	103	11	220	22
r/memorialuniversity	86	5	116	15
r/usask	74	4	157	11
r/uottawa	35	5	27	2
r/UdeM	22	2	101	4
r/University_Of_Regina	21	3	30	3
r/lakeheadu	20	0	47	4
r/uleth	17	0	18	3
r/laurentian	14	0	21	0
r/AcadiaU	14	1	44	1
Total	38387	4070	156519	13363

lated to co-op, and we do this by only keeping those which contain at least one of the following co-op related terms: “coop”, “interview”, “resume”, “workterm”, and “intern”. Note that we lemmatized the words during pre-processing, so “interview” also captures similar words such as “interviewer” or “interviewing”.

3.3 Topic Modelling

We then apply topic modelling to the posts and comments that passed the above mental health and co-op filters. First, we vectorize the comments and posts in a standard way. For each post or comment, the i th entry of its vector corresponds to the Term Frequency - Inverse Document Frequency (TF-IDF) of the i th word. We compute the TF-IDF score of a given word for a given post or comment as follows: we divide the number of times the word appears in the given post or comment (TF) by the fraction of total posts and comments that contain at least one occurrence of this word (DF). TF-IDF is frequently used when vectorizing text as it takes into account both the uniqueness of a word in the entire dataset and the importance of the word to the specific document (in our case, the specific post or comment).

Next, we run the Non-negative Matrix Factorization (NMF) topic modelling algorithm [27] on the vectorized posts and comments. NMF clusters the data into topics and produces a list of representative terms called *topic descriptors* for each topic. Each such term has a “representativeness” score, and we select the top-10 highest-scoring terms for each topic. Additionally, for each topic, we report the top-10 most frequent word n -grams (for n up to three, i.e., sequences of up to three consecutive words) within the posts or comments

belonging to the given topic.

NMF requires the number of topics as input. To select an appropriate number of topics, we ran NMF to produce between 2 and 100 topics, and computed the *coherence* [20] of each output (higher is better). We obtained the highest coherence for ten topics.

Another issue with NMF is that despite our text pre-processing, some topic descriptors were uninformative. Following prior work on topic modelling [17, 25, 2], we repeatedly remove uninformative terms from the posts and comments and re-run NMF until the topic descriptors no longer contain any uninformative terms. After two iterations, all the top-10 descriptors became informative.

Finally, we extract issues affecting students from the NMF topic descriptors, the frequent n -grams, and a manual inspection of a 5% sample of posts and comments assigned to each topic.

4. RESULTS

Table 3 shows the topic modeling results for posts and comments related to both mental health and co-op, including topic descriptors, a sample of frequent n -grams, and the percentage of content assigned to each topics. After inspecting these results, plus a sample of posts and comments assigned to each of the ten topics, we manually group the topics into issues, as shown in Table 4 (where we also point out which topics from Table 3 describe which issue).

Topics 1, 2, 4, 9 and 10 cover over 60 percent of the content and appear related to competition, specifically the competitive nature of the co-op job market. Upon manual inspection of a sample of posts and comments, we found that students express self-doubt and feelings of inadequacy when unable to secure a desirable co-op job, especially when one’s classmates and friends are able to obtain such jobs. There were also some discussions about choosing a good co-op program that enables interesting and highly-paid co-op job opportunities, concerns over not having enough experience to qualify for these desirable jobs, and the stress of maintaining a high GPA to qualify for or remain in such programs. Notably, many of the posts and comments related to competition referred to technology and software roles, as well as large technology employers such as Facebook and Google. This is likely due to the fact that co-operative programs are mainly in science and engineering.

Next, topics 3 and 5 are about questions students ask about co-op programs. This includes general questions related to admissions, and specific questions such as how to write a work report.

Topic 6 describes issues with interviews. Many posts and comments belonging to this topic referred to interviews for co-op jobs being stressful, especially because they often coincide with exams and other academic deadlines, and because interview processes for software positions may include lengthy programming tests. Students also reported feelings of uncertainty about how to prepare for interviews, how to acquire required skills, and what to expect. Additionally, some students reported anxiety after an interview while

Table 3: Topic modeling output for co-op related posts and comments

Topic descriptors	Frequent n-grams	%
1 work, time, people, like, make, want, hard, day, need, know	'work hard', 'people work', 'school work'	29.9
2 job, apply, degree, graduate, student, want, like, people, look, pay	'apply job', 'job market'	14.1
3 project, code, like, use, course, make, time, start, personal, create	'work project', 'start project', 'personal project'	12.2
4 experience, internship, year, co-op, school, summer, gpa, company, graduate, program	'work experience', 'grad school', 'work hard'	9.5
5 class, easy, semester, final, hard, pretty, time, exam, course, lecture	'class work', 'final project', 'work time'	7.8
6 resume, interview, look, company, ask, skill, apply, recruiter, employer, page	'work experience', 'career fair', 'cover letter'	7
7 lab, research, professor, prof, student, grad, undergrad, paper, ask, email	'research project', 'work lab', 'grad student'	5.6
8 group, member, people, meet, person, presentation, individual, make, facebook, fb	'group project', 'work group', 'class group'	5.3
9 team, game, member, join, play, club, engineer, player, people, design	'project team', 'work project', 'team work'	5
10 letter, cover, apply, write, application, position, make, generic, company, tailor	'cover letter', 'resume cover', 'resume cover letter'	3.6

Table 4: Issues extracted from co-op related posts and comments

Topics	Issue	Description	%
1,2,4,9,10	Competition	E.g. not qualifying for a desired co-op job	62.1
3,5	Questions	About co-op programs (e.g., seeking clarification when instructions are not clear enough)	20
6	Interviews	E.g., not knowing what to expect or how to prepare	7
7	Research opportunities	Not directly related to co-op	5.6
8	Group projects	Not directly related to co-op	5.3

waiting to find out if they have been hired.

Finally, topics 7 and 8 are not directly related to the effect of co-op on students' well-being; they instead refer mainly to research opportunities and participation in group projects during internships.

5. DISCUSSION AND CONCLUSIONS

By performing topic modelling on subreddits corresponding to U.S. and Canadian universities, we obtained the following insights into the impact of co-operative education on students' well-being.

1. **Competition** for internships, especially in the software and information technology fields, is a frequently discussed negative aspect of co-operative education. Prior work has observed that co-op unemployment can lead to mental well-being issues [6, 11]. However, our results further indicate that not securing a desirable, high-paying, challenging, and relevant employment can be a source of stress, self-doubt, and disappointment. This is especially true if one's friends and classmates are able to secure desirable jobs that are directly related their programs of study.
2. **Co-op interviews** are a source of stress for several reasons. First, students fear being unprepared or un-

qualified, especially when competing for sought-after jobs. Second, interviews often coincide with midterm examinations and other academic deadlines, meaning that students may have to choose between preparing for interviews (including preparing for programming tests) and coursework. Previous work has argued that co-operative education research should consider work-related variables in addition to education-related ones; these work-related variables include skills, job satisfaction, performance assessments, and selection interviews [23]. Our findings on co-op interviews align with this suggestion, providing data-driven evidence of another source of anxiety for co-op students.

3. As reported in previous work [23], we also found some reports of **loneliness during workterms**. Additionally, **moving and finding a place to live** during a workterm can be a source of stress.

Our findings suggest actionable insights for academic institutions and students. First, it is important to manage co-op students' expectations. For example, universities may want to offer workshops that explain the competitive nature of the co-op process and help students find jobs they qualify for. Junior students, specifically, should keep in mind that they may not immediately qualify for the sought-after positions secured by their senior colleagues. Additionally, these workshops should provide advice on interview preparation, coping with frequent moving, and finding short-term living arrangements during internships. Second, co-op interviews should not be scheduled during peak academic times. Having more time to prepare, especially for software interviews with programming tests, may reduce anxiety.

One limitation of this study is that it only reflects the opinions of students who are active on Reddit. Nevertheless, our findings can be used as a starting point for additional focused research. In future work, we plan to survey students to confirm our findings about the competitive nature of the co-op job market. Additionally, we will analyze course discussion forums to further investigate the impact of co-op interviews on class schedules and academic deadlines.

6. REFERENCES

- [1] S. Bagroy, P. Kumaraguru, and M. De Choudhury. A social media based index of mental well-being in college campuses. In *Proc. CHI Conference on Human factors in Computing Systems*, pages 1634–1646, 2017.
- [2] A. Blanchard. Understanding and customizing stopword lists for enhanced patent mapping. *World Patent Information*, 29(4):308–316, 2007.
- [3] L. A. Braunstein and W. A. Stull. Employer benefits of, and attitudes toward postsecondary cooperative education. *Journal of Cooperative Education*, 36(1):7, 2001.
- [4] S. Chopra and L. Golab. Job description mining to understand work-integrated learning. In *Proc. Int. Conf. on Educational Data Mining*, pages 32–43, 2018.
- [5] G. Coppersmith, M. Dredze, and C. Harman. Quantifying mental health signals in twitter. In *Proc. Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality*, pages 51–60, 2014.
- [6] L. Cormier and D. Drewery. Examining the effect of co-op non-employment and rejection sensitivity on subjective well-being. *Asia-Pacific Journal of Cooperative Education*, 18(3):213–224, 2017.
- [7] M. De Choudhury and S. De. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *Proc. Int. AAAI Conf. on Weblogs and Social Media*, 2014.
- [8] M. De Choudhury, S. S. Sharma, T. Logar, W. Eekhout, and R. C. Nielsen. Gender and cross-cultural differences in social media disclosures of mental illness. In *Proc. ACM Conf. on Computer Supported Cooperative Work and Social Computing*, pages 353–369, 2017.
- [9] M. Deziel, D. Olawo, L. Truchon, and L. Golab. Analyzing the mental health of engineering students using classification and regression. In *Proc. Int. Conf. on Educational Data Mining*, pages 228–231, 2013.
- [10] D. Drewery, T. J. Pretti, and S. Barclay. Examining the effects of perceived relevance and work-related subjective well-being on individual performance for co-op students. *Asia-Pacific Journal of Cooperative Education*, 17(2):119–134, 2016.
- [11] D. W. Drewery, L. A. Cormier, T. J. Prettti, and D. Church. Improving unmatched co-op students’ emotional wellbeing: Test of two brief interventions. *International Journal of Work-Integrated Learning*, 20(1):43–53, 2019.
- [12] M. T. Drysdale and M. McBeath. Exploring hope, self-efficacy, procrastination, and study skills between cooperative and non-cooperative education students. *Asia-Pacific Journal of Cooperative Education*, 15(1):69–79, 2014.
- [13] J. Gault, J. Redington, and T. Schlager. Undergraduate business internships and career success: are they related? *Journal of Marketing Education*, 22(1):45–53, 2000.
- [14] G. Gkotsis, A. Oellrich, T. Hubbard, R. Dobson, M. Liakata, S. Velupillai, and R. Dutta. The language of mental health problems in social media. In *Proc. Workshop on Computational Linguistics and Clinical Psychology*, pages 63–73, 2016.
- [15] S. C. Guntuku, D. B. Yaden, M. L. Kern, L. H. Ungar, and J. C. Eichstaedt. Detecting depression and mental illness on social media: an integrative review. *Current Opinion in Behavioral Sciences*, 18:43–49, 2017.
- [16] D. Hodges and N. Burchell. Business graduate competencies: Employers’ views on importance and performance. *International Journal of Work-Integrated Learning*, 4(2):16, 2003.
- [17] A. Khan and L. Golab. Reddit mining to understand gendered movements. In *Proc. EDBT Workshop on Data Analytics Solutions for Real-Life Applications*, 2020.
- [18] C. McClellan, M. M. Ali, R. Mutter, L. Kroutil, and J. Landwehr. Using social media to monitor mental health discussions- evidence from twitter. *Journal of the American Medical Informatics Association*, 24(3):496–502, 2017.
- [19] C. Nevison, L. Cormier, J. Pretti, and D. Drewery. The influence of values on supervisors’ satisfaction with co-op student employees. *International Journal of Work-Integrated Learning*, 19(1):1–11, 2018.
- [20] D. O’callaghan, D. Greene, J. Carthy, and P. Cunningham. An analysis of the coherence of descriptors in topic modeling. *Expert Systems with Applications*, 42(13):5645–5657, 2015.
- [21] U. Pavalanathan and M. De Choudhury. Identity management and mental health discourse in social media. In *Proc. Int. Conf. on World Wide Web*, pages 315–321, 2015.
- [22] E. Ralph, K. Walker, and R. Wimmer. Practicum-education experiences: Post-interns’ views. *International Journal of Engineering Education*, 25:122–130, 01 2009.
- [23] P. M. Rowe. Researchers’ reflections on what is missing from work-integrated learning research. *Asia-Pacific Journal of Cooperative Education*, 16(2):101–107, 2015.
- [24] K. Saha, E. Chandrasekharan, and M. De Choudhury. Prevalence and psychological effects of hateful speech in online college communities. In *Proc. ACM Conference on Web Science*, pages 255–264, 2019.
- [25] A. Toulis and L. Golab. Social media mining to understand public mental health. In *VLDB Workshop on Data Management and Analytics for Medicine and Healthcare*, pages 55–70, 2017.
- [26] G. Van Gyn, J. Cutt, M. Loken, and F. Ricks. Investigating the educational benefits of cooperative education: A longitudinal study. *Science*, 97(180):277, 1997.
- [27] W. Xu, X. Liu, and Y. Gong. Document clustering based on non-negative matrix factorization. In *Proc. Int. ACM SIGIR Conf. on Research and Development in Informaion Retrieval*, pages 267–273, 2003.