Preparation for Expanding Equitable Pathways

for In-Demand STEM Careers at Hispanic Serving Institution*

(Evaluator Report on Year 1 Operation)

Jianjun Wang

February 20, 2023

* This project received $5 million funding from the U.S. Department of Education (Award number Project Grant P031C210093).

Contents

Abstract

California State University, Bakersfield (CSUB) was awarded a five-year grant, "An Equitable Pathway to In-Demand STEM Careers", that began in Fall 2021 to create and expand a streamlined pathway for Hispanic students to pursue their educational and career goals with servingness culture support. In the first year, the grant team has completed six tasks assured by the grant proposal: (1) Establishing a budget in alignment with federal and district protocols; (2) Pursuing recruitment of key personnel to lead K-12 and community college engagement in strengthening on-ramps into CSUB's STEM programs; (3) Securing campus space to maintain workflow for cross-functional teams; (4) Finalizing internal/external evaluation service agreements and data collection procedures; (5) Launching STEM Careers Council for local partnership building; and (6) Purchasing instructional equipment and supplies in STEM lab building. In addition, quantitative and qualitative data have been gathered to assess the effectiveness of a research program in Summer 2022 and guest speaker presentations in regular semesters. To track the outcome of student learning, an indicator of the *Rate of Progress* is configured to monitor the status of student achievement across milestones of the STEM program pipeline. In the *conclusion* section, evaluation findings are reviewed to support three recommendations for project improvement. The entire evaluation framework incorporates a *participatory, utilization-focused,* and *program theory-driven* mechanism that conforms to the *utility, feasibility, propriety,* and *accuracy* standards of program evaluation.

Preparation for Expanding Equitable Pathways

to In-Demand STEM Careers at Hispanic Serving Institution

On October 1, 2021, the *Hispanic-Serving Institutions - Science, Technology, Engineering, or Mathematics (HSI STEM) and Articulation Programs* (a.k.a., HSI STEM; Title III, Part F) of U.S. Department of Education awarded California State University, Bakersfield (CSUB) a five-year grant, "An Equitable Pathway to In-Demand STEM Careers", to increase access and degree completion rates in STEM education while closing opportunity gaps for Hispanic students in career outcomes. The support for workforce development is partnered between CSUB and its surrounding community colleges under an articulation agreement to strengthen STEM education and career preparation for Latino and other low-income students. As Dr. Lynnette Zelezny, the CSUB President, envisioned, "This $5 million grant will benefit our students with increased opportunities for STEM internships, research and graduate training, which will yield tremendous benefits for our regional workforce."[1]

To track the funding impact, the project team has delineated a plan to "prepare an annual report for the program's progress" that includes *instrument development, data analysis, result interpretation,* and *recommendation of improvement* (see p. 11 of the grant proposal). In fulfillment of these expectations, the first-year report is divided into five sections. Section 1 provides an overview of the evaluation framework according to a logic model of the grant proposal. Section 2 addresses the task completion in Year 1, as well as the challenges encountered during the grant setup. Section 3 aggregates qualitative data from guest speaker presentations and quantitative data from surveys of the summer research program to examine the effectiveness of project support. Section 4 covers the construction of an empirical index, *rate of*

[1] https://news.csub.edu/csub-receives-5-million-grant-to-increase-stem-opportunities

progress (ROP), for monitoring student progress across the program pipeline. Evaluation recommendations are adduced in the *conclusion* section to highlight strategies for program improvement next year.

Overview of the Evaluation Framework

The first evaluation report is delimited to the activities of project preparation and implementation up to January 1, 2023. To support project evaluation, Tom Angelo (1999), a former director of the National Assessment Forum, advised, "Though accountability matters, learning still matters most" (¶. 1). Others agreed that institutional learning would contribute to local capacity building for project improvement across multiple years (Miller & King, 2019). Based on the consideration of funding accountability and program improvement, an annual assessment of the ongoing progress has been conducted to fit the dual paradigm of a formative and summative evaluation of federal government funding.

To facilitate formative evaluation, a *participatory, utilization-focused,* and *program theory-driven* approach has been incorporated in the original proposal for collecting well-rounded evidence to assess the grant's effectiveness each year. From a perspective of summative evaluation, the annual progress is expected to accumulate credible results for seamless expansion of the equitable pathway across in-demand STEM careers. The broad impact further extends in both space and time dimensions – As a Hispanic Serving Institution (HSI), CSUB covers a broad service region with a land area as large as the state of New Jersey. In addition, the mechanism of capacity building in the first year introduces an institutional setup to support student career development throughout the period of grant funding.

Upon receiving the grant award, the project director has been working with the evaluator through a *participatory* approach to design survey instruments and ensure meaningful data

gathering that is closely aligned with the grant expectation. The participatory approach is advocated by the current literature – "Since it is important that the evaluation plan and program design are informed by each other, it is important to engage with the external evaluator in the early stages of program design and proposal planning".[2] Built on the professional cooperation, an agreement has been reached, as evidenced by an evaluation contract, on the logistics of implementing the evaluation design and its time commitment.

The pursuit of a well-rounded evaluation design aligns with the *utilization-focused* approach to address project needs. In particular, examples have been adduced by the evaluator to use R scripts for aggregating both structured and unstructured data from document collection, network description, and text/interview records for result reporting.[3] Creativity in data handling is especially needed in Year 1 when the IRB protocol has yet to be approved in the first semester shortly after the project funding. In coping with the delay in data availability, an innovative approach has been taken to first transcribe guest speaker videos into text files for the event documentation. R scripts are subsequently developed to extract credible evidence from the *qualitative data* (see Appendix 1), and thus, assess the alignment of *Career Talk* contents with the intended outcomes. The consideration of both qualitative and quantitative data was advocated by Donaldson, Christie, and Mark (2009) for enhancing the evidence's credibility. They asserted, "Stakeholders are provided with a wide range of choices for gathering credible evidence, which reinforces the idea that neither quantitative, qualitative, nor mixed method designs are necessarily superior or applicable in every applied research and evaluation context" (p. 244). Altogether, the *participatory* and *utilization-focused* mechanism fits the project context with flexible method choices for result triangulation.

[2] https://hsistemhub.org/portfolio-item/june-2020-newsletter/
[3] p. 108-111 of https://files.eric.ed.gov/fulltext/ED610410.pdf

Furthermore, the *utilization-focused* approach has addressed the need for information gathering pertaining to an *absolute priority* of the HSI-STEM program, i.e., "increase the number of Hispanic and other low-income students attaining degrees in STEM fields." Although the count of degree completers is a meaningful indicator upon grant completion, only a few students, particularly those with the Latino origin, can achieve a STEM degree within five years of grant funding. Consequently, the completer count cannot be accumulated for most students remaining in the program pipeline during the summative evaluation. To amend this void, a formative measure, "Rate of Progress" (ROP), is derived in the first year from the *utilization-focused* approach to track student advancement throughout the period of grant funding. In complement with the *degree completer count* for summative evaluation, the ROP indicator offers additional information "such that evaluation efforts measure both *ongoing development and implementation of the program* and *the success of the program*" (Demetrikopoulos, 2020, p. 1).

Table 1: Alignment between Project Goals and Year 1 Activities

Goals	Activities
1. Develop equitable on-ramps to STEM pathways and careers	-Launch STEM Careers Liaison Council -Hire STEM Outreach Liaison -Begin K-12 Outreach -Initiate Transfer Acceptance Guarantees -Support BC-CSUB Dual Admissions
2. Use a holistic approach to serving and foster a sense of belonging through culturally enhancing coaching and integrated support	-Implement Engagement Model -Offer STEM First Year Seminar -Start STEM Speaker Series
3. Strengthen educational and industry partnerships to scale experiential learning	-Support Undergraduate Research -STEM Careers Council -STEM Success Conference

The *theory-driven* approach further ensures conformation of the evaluation design to professional standards in program evaluation. In particular, "Sometimes terms like program logic, **logic models** and logical frameworks are used as equivalents to **program theory** (PT) or

intervention theory" (Leeuw & Donaldson, 2015, p. 469). By design, the logic model in Section A(3) of the grant proposal is employed in the *theory-driven* approach to track grant activities for Year 1 according to the project goals (Table 1).

As Donaldson, Christie, and Mark (2009) maintained, "This practical approach for gathering credible evidence is highly consistent with the profession's guiding principles, evaluation standards, and other mainstream approaches to practical program evaluation" (p. 244). More specifically, the current standards for program evaluation include four components, *utility, feasibility, propriety,* and *accuracy* (Yarbrough, Shulha, Hopson, & Caruthers, 2010). The *utility* consideration is addressed by the *utilization-focused* approach. The *feasibility* criterion is met by the *participatory* approach to engage key stakeholders in project design and evaluation planning. The *propriety* standard is upheld by IRB's approval of a data gathering protocol to ensure compliance of the project evaluation to federal, state, and local laws and regulations. The *accuracy* standards are employed to guide instrument development – The evaluator and the project director have collaborated on designing a *speaker presentation survey* that covered student benefits from the STEM career talk, including the *pace, clarity, interaction, attractiveness,* and *informative natures* of each presentation. The survey results are analyzed in Section III of this report to assess the effectiveness of program support. In summary, the *participatory, utilization-focused,* and *program theory-driven* approach offers a valid evaluation framework to guide the collection of credible evidence for result reporting.

Fulfillment of Year 1 Tasks

The grant team followed page 39 of the grant proposal to implement tasks and monitor milestones under the timeframe for Year 1. As a result, progresses are demonstrated in six aspects of the grant operation:

1. Establishing a budget in alignment with federal and district protocols;
2. Pursuing recruitment of key personnel to lead K-12 and community college engagement in strengthening on-ramps into CSUB's STEM programs;
3. Securing campus space to maintain workflow for cross-functional teams;
4. Finalizing internal/external evaluation service agreements and data collection procedures;
5. Launching STEM Careers Council for local partnership building; and
6. Purchasing instructional equipment and supplies in the STEM lab building.

Besides the grant setup, students are engaged in active learning processes throughout the year. In the Fall of 2021, four *Career Talks* were offered. Each of the first three guest speaker sessions was attended by 27 students, and one talk was given to 61 students. In Spring 2022, another *Career Talk* session was held for 73 students. Meanwhile, the summer research program was launched in three STEM disciplines, *Chemistry, Engineering,* and *Physics,* from July 11-August 4, 2022. The participants included 18 students from Bakersfield College (BC) and one student from CSUB. Evaluation data have been collected, and the findings will be reported in the next section.

The endeavor of lab-based inquiry is extended in a Year-Round STEM research program. Four CSUB students have been accepted based on a thorough review of their *activity description, mentorship plan,* and *transcript records.* The topics have been carefully chosen to support open-ended scientific inquiries:

- Tetrodotoxin Binding Protein Genes in Taricha granulosa, the Rough-Skinned Newt
- Enzymatic Sulfur Reduction in Fossil Fuels
- Identification of Key Mutations of Lysyl Oxidase
- Degradation of Ofloxacin in Varying Conditions: Kinetics and Mechanism

During Year 1, a growth pattern has been observed in all aspects of the learning process to expand the academic horizon of both BC and CSUB students. Besides the increase in participant counts from 27 to 73 across these *Career Talk* sessions in regular semesters, the demanding nature of summer research has led to a decision to expand the period from four to five weeks next year. The Year-Round program applicants also grew from one in Fall 2021 to three in Spring 2022.

In Fall 2022, the federal grant funding supported new enrollments of 70 students, surpassing the target of 50 new first-year students in the intake process. In addition, six peer mentors have been hired with a caseload of approximately 12 students per person since August 2022. A box folder has been made available to the evaluator to confirm the development of peer mentor training materials that have been proven effective in the past. New training materials will be gathered to support peer mentors in Summer 2023.

The ongoing progress is inseparable from addressing challenges in project personnel recruitment. In particular, a *STEM Outreach and Community College Liaison* has been hired to lead K-12 and community college engagement. Per the description on page 38 of the grant proposal, an Assistant to the Director (AD) is filled in the grant office of the School of Natural Sciences, Mathematics and Engineering (NSME) to help advertise, interview, hire, and train peer mentors each semester. In undertaking the tasks of the *Director for the Office of Institutional Effectiveness (OIE)*, a data researcher is hired recently to collaborate with staff of the Institutional Research, Planning, and Assessment (IRPA) for evaluation data gathering. Before an Academic Advisor (aka., Counselor) got on board in January 2023, the NSME student center assumed the role of establishing and tracking early alert protocols for student advising.

In contrast to the success of project management, an internship coordinator (IC) has yet to be hired, which delayed the completion of three tasks in the first year:

- Recruitment of interns for the Work-Integrated Learning (WIL) program
- Finalization of MOUs for internships/work-based learning employers
- Promote stipends for summer interns.

The situation hinges on both external labor shortage and internal budget configuration. In 2022, the market for qualified employees has been extremely tight because "Inflation is currently outpacing salary increases by a large margin" (Thoumyre, 2022, p. 1), which caused shrinkage of the applicant pool for the grant-supported position. The grant budget configuration has classified the IC position as a half-time job, making this service less competitive than a full-time employment. To pursue a solution, an arrangement has been made to add half-time funding from another grant and make the position fit for someone to earn dual half-time salaries.

In summary, a concerted effort has been made by the grant team to address all proposed tasks for the first year. Despite the unavailability of IC support, the project has successfully hired a STEM Outreach and Community College Liaison, an AD, a data researcher, and an Academic Advisor. As a result, student recruitment has exceeded the freshman enrollment target, institutional support has been established through the grant setup, peer mentors have been trained for student advising, *Career Talk* sessions have been offered for an increasing number of students in regular semesters, summer research opportunities have been extended to both BC and CSUB students, an IRB protocol has been approved for evaluation data collection, and participants have been identified from the Year-Round program applicants to sustain in-depth scientific inquiries with faculty mentor support. These accomplishments confirmed an effective delivery of the expected services in the first year.

Effectiveness of the Grant Implementation

While the previous section outlined the task completion to address the question on *how much* has been done in Year 1, the effectiveness of grant implementation is further examined in this section through qualitative and quantitative data analyses to address the question of *how well* the project performed in terms of assessment outcomes. After receiving IRB approval, the grant team has gathered evaluation data from two sources: (1) Pretest and posttest surveys from the 2022 Summer Research Program, and (2) The *Career Talk* Survey in Fall 2022. Accordingly, this section maintains dual foci on (1) assessing the effectiveness of summer STEM training, and (2) examining the impact of guest speaker presentations on STEM career preparation.

Benefits of the Summer STEM Inquiry

Nine students responded to questionnaires about the 2022 Summer Research training featuring faculty mentorship and collaboration with students. Seven responses were gathered before the training, and three were collected after the training. One student provided the data in both survey sessions. As a first-generation college student in physics and engineering, he set a degree objective to become an "Engineer with an emphasis in Aerospace", and is on track to achieve the degree in less than four years. At the sophomore stage, he described himself as:

> I come from your average low-income Latin home, where the environment teaches you that there is no future for us. I, on the other hand, started seeing the world differently. I noticed the errors in which my culture thinks. I always was a curious person and Engineering offers me to use that curiosity in a positive manner. I have some comp science background and very basic engineering background. I am almost done with an A.S. in engineering and I will be transferring to obtain my B.S in engineering from wherever I land.

Based on the reflection on self-preparation, he made a plan to find a job first, and consistently reported a *strong agreement* response to the following statements:

- I would like to collaborate with faculty in research in addition to my regular course work
- I intend to network with other students during the research process
- More research experiences are essential for pursuing my career objectives
- I am confident in research participation with the knowledge I gained from the past

After the summer training, he reconfirmed no change in his major and degree objective. The learning experience strengthened his commitment as he insisted, “I've come this far and won't turn back now.” He testified to the merit of summer learning by indicating a *strong agreement* with the following statements about the STEM research process:

- I understand my role and the roles of other in the research project
- I learned more scientific literature related to my inquiries
- I received encouragement and feedback on my participation in the team work
- I did real research work beyond following a textbook or lab manual
- I had the exposure to new research methodology, technique, or instrumentation
- I expanded my network with other classmates during the research process
- My research experiences were helpful in pursuing my career objective
- I am more enthusiastic about STEM knowledge inquiry in graduate school
- I would recommend this research program to other students.

Altogether, four respondents revealed their Latino identity in the survey before the summer training. They all *agreed* or *strongly agreed* with the statement, “I would like to collaborate with faculty in research in addition to my regular course work.” The same pattern applied to the entire group of respondents, not just those of Hispanic origin. Hence,

collaboration in faculty research, as exemplified by the summer research program, not only benefited the Latino student according to the data tracking, but also offered an effective mechanism to promote learning interest for all students, regardless of their ethnic identities.

Quality of Guest Speaker Presentation

Besides the improvement of STEM education, a unique feature of this project is to strengthen student readiness for in-demand jobs in the increasingly competitive market. Thus, *Career Talk* is an essential component of the grant support for expanding the career pathway. To justify the objective attainment, guest speaker sessions need to be evaluated to assess the outcomes of grant implementation. Based on the aforementioned student counts on page 9, one could have expected 251 responses in the evaluation data across five *Career Talk* sessions in Year 1. Nonetheless, the due process of IRB review took time, which postponed the data collection to Spring 2022. As a result, no survey was conducted in Fall 2021 for four *guest speaker* sessions. Without evaluation data from 178 respondents across these sessions, the missing data exceeded 70% (or 178/251) of the total possible data point, and no imputation method can be employed when the patterns of absence spread over an entire session (Wang & Johnson, 2019). Thus, video recording is employed in this section to examine the quality of the guest speaker presentation through text analytics.

In the past, qualitative research is a mainstream method for extracting in-depth messages from video analyses (Heath, Hindmarsh, & Luff, 2010). While the individual presentations illustrate authentic examples of career development for STEM students, the inductive approach from qualitative inquiries could be biased and inconclusive even under the most promising tool of grounded theories (Konecki, 2021). Fortunately, it is well-known that “Today’s natural language processing systems can analyze unlimited amounts of text-based data without fatigue

and in a consistent, unbiased manner."[4] The methodology advancement has overcome an insurmountable issue of replicability in information extraction (Sarkar, 2019). More recently, the NLP-based text synthesis has been spearheaded by an R package, *Quantitative Analysis of Text Data* (quanteda). According to Benoit et al. (2018),

> quanteda is an R package providing a comprehensive workflow and toolkit for natural language processing tasks ... Using C++ and multithreading extensively, quanteda is also considerably faster and more efficient than other R and Python packages in processing large textual data. (p. 774)

To date, the R package application has been widely adopted by large-scale assessment projects of the federal government (Caro & Biecek, 2017; Matta, Rutkowski, Rutkowski, & Liaw, 2018). Built on the quanteda platform, an innovative approach is taken in this section to handle the text analytics in three steps:

1. An online portal is adopted to transcribe the video content in text files;
2. Natural Language Processing (NLP) is applied to transform the unstructured text from *Career Talks* into normalized data suitable for analysis by machine learning algorithms;
3. R scripts are developed to extract the overall features of the career talk outcomes.

In addition, when survey data are available, statistical reporting is included to reconfirm the quality of the guest speaker presentation in Spring 2022.

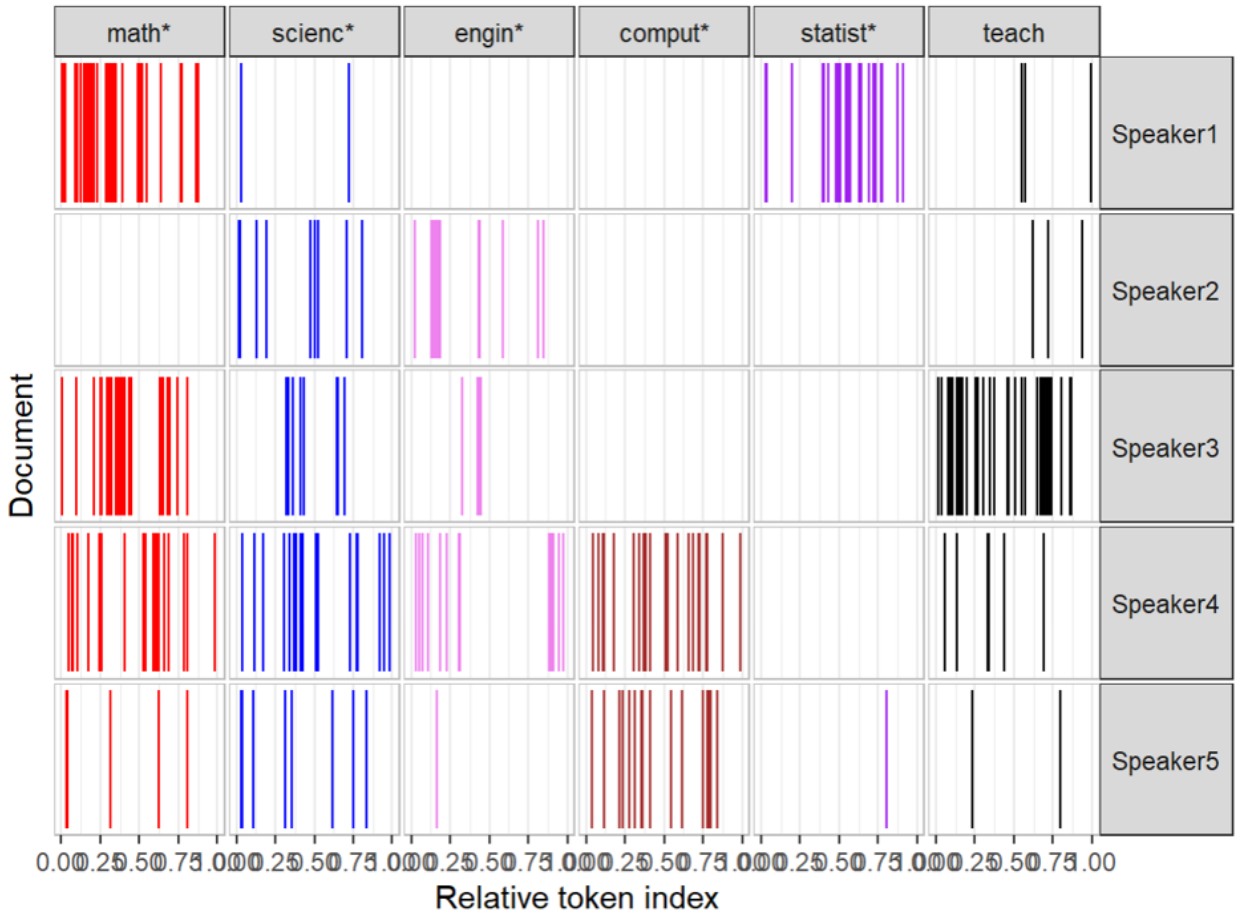
Findings from Text Analytics

After NLP's *text tokenization, stopping-word/punctuation cleaning*, and *dictionary stemming*, a Lexical Dispersion Plot has been drawn from the text data to compare frequently-mentioned words across the five guest presentations. In Figure 1, keywords stemming from

[4] https://www.linguamatics.com/what-text-mining-text-analytics-and-natural-language-processing

subject names are tracked to show emphases among the *Career Talk* sessions. Although all speakers mentioned the prefix "scienc*" or "teach", Speaker 1 clearly stressed career preparation in *math* and *statistics*. Likewise, Speaker 2 placed more focus on *science* and *engineering*. Figure 1 has displayed a pattern that STEM fields contain different domains, and thus, need multiple *Career Talk* sessions to cover. Altogether, five guest speakers in Year 1 have consistently illustrated the connection between STEM education and career development. The subject-specific impact may contribute to the expansion of student horizon on the pathway configuration across the fields of mathematics, science, engineering, computing, statistics, and teaching.

Figure 1: Dispersion of Subject Coverage in Career Talks



To capture all tokenized terms for each speaker, a word-cloud plot was generated in Figure 2. As shown by tokens of larger size, the plot indicates that these *Career Talks* centered on key components of "job", "work", "teach", "research", "phd", and "internship". While *teaching*, *research*, and *internship* are helpful stepping stones toward creating sustainable and equitable on-ramps for high-wage and in-demand employment, the *job* or *work* opportunities are built on STEM degree completion, such as a Ph.D., demonstrated by some of the guest speakers. Hence, Figure 2 shows the *Career Talk* alignment with Goal 1 of the grant funding in *Developing sustainable and equitable on-ramps to high-wage and in-demand STEM degree and career pathways* (see p. 11 of the grant proposal).

Figure 2: Word-Cloud Plot on Key Components of Career Talks

In terms of *fostering a sense of belonging* in Goal 2, Speaker 1 identified himself as a Latino professional during the video presentation to *improve servingness consideration to Hispanic students* (see p. 13 of the grant proposal). While three of the five speakers had a Hispanic origin, not all of them revealed their ethnic identity in the cross-culture dimension. Enrichment of the *Career Talk* with the cultural background seems to have added new ingredients to the guest speaker presentation. A keyness plot is created in Figure 3 to contrast the content coverage, and Speaker 1 mentioned *Ph.D.* more often as a landmark of educational accomplishment for students to *model* and *dream*. He also used other terms, such as Dr. and UC, to convey a high aspiration. In contrast, other *Career Talks* are more focused on practical considerations, including *internship*, *interest*, and *opportunity*, which are needed by everyone to expand the career pathway through STEM education.

Figure 3: Impact of Cultural Promotion Between Speaker1 and Other Presenters

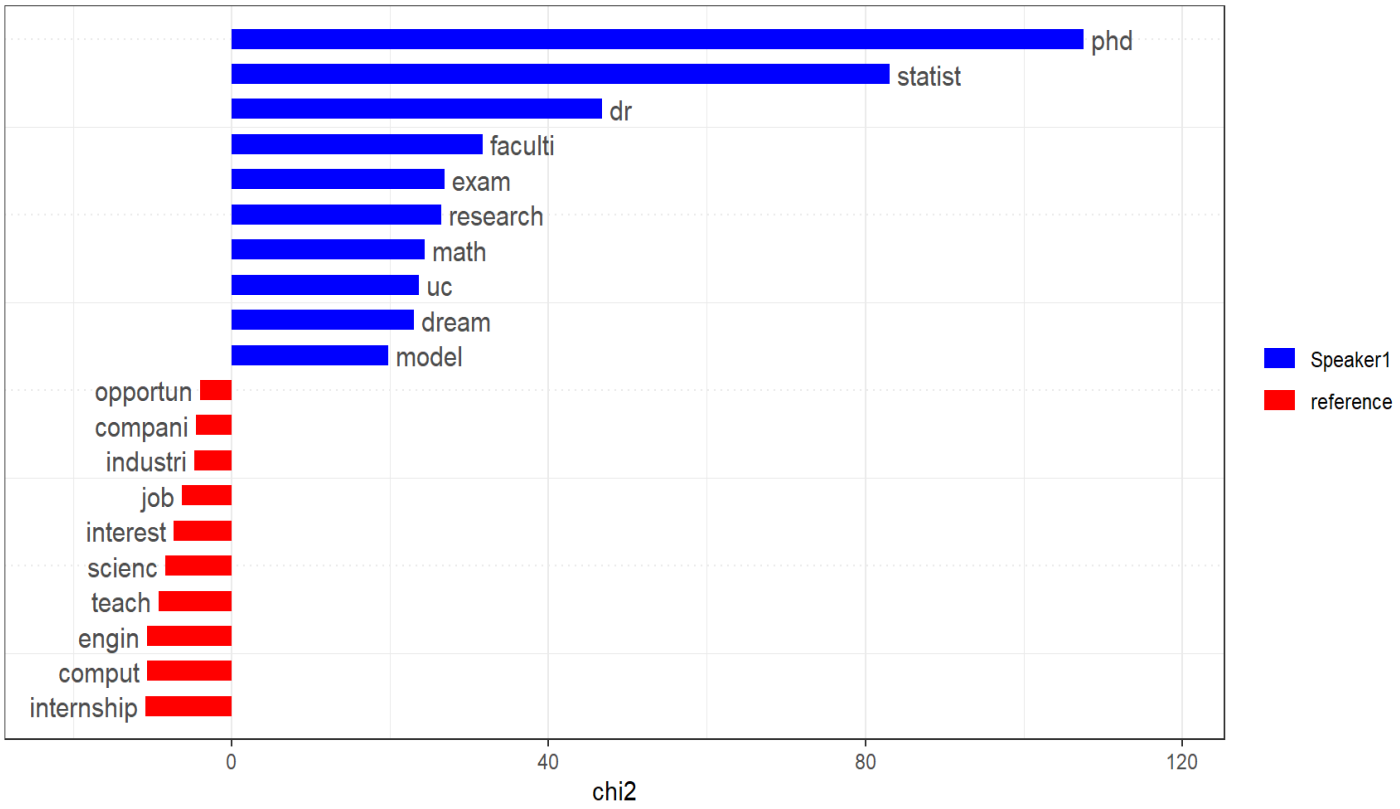


Figure 4: Indicators of Motivation Attribute in Guest Talks

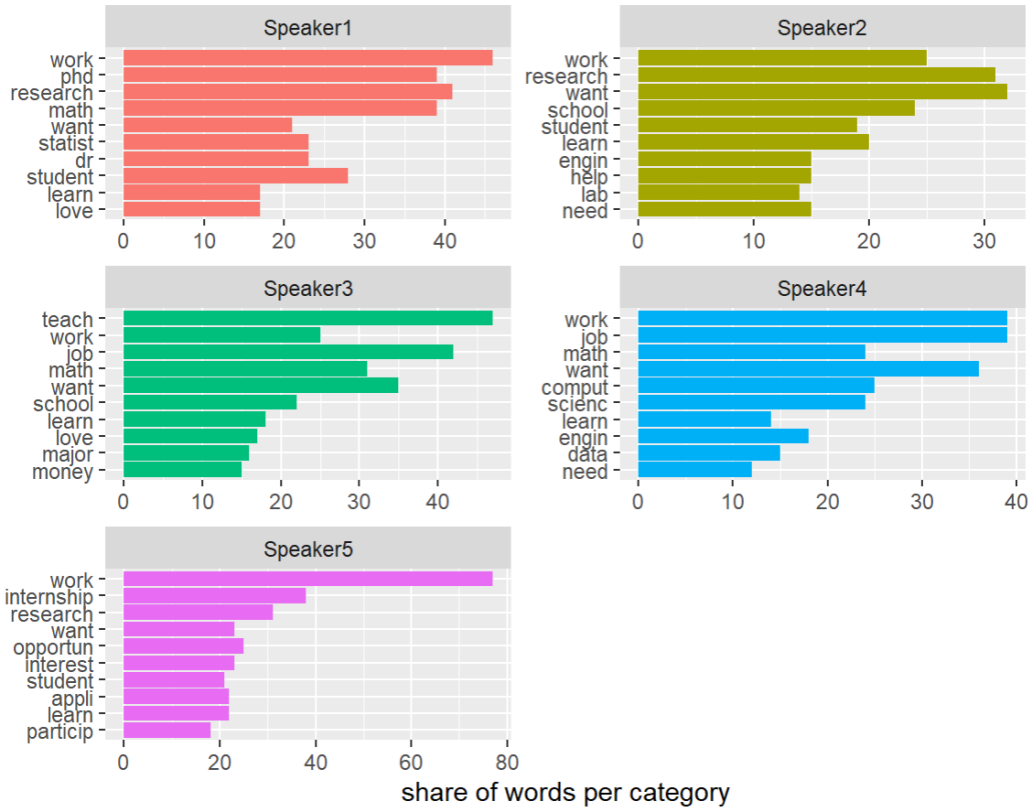
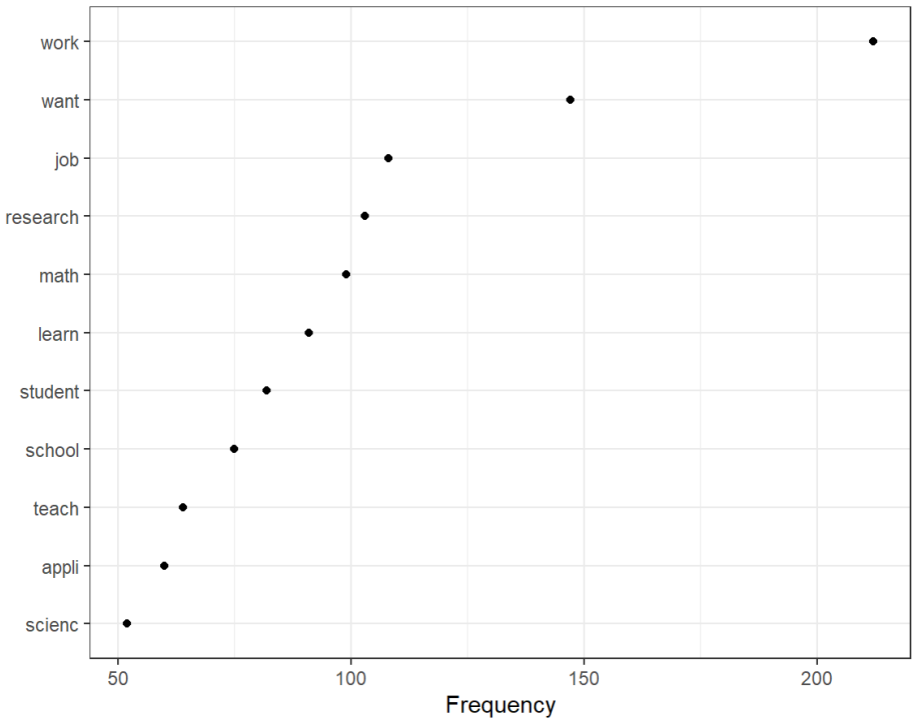
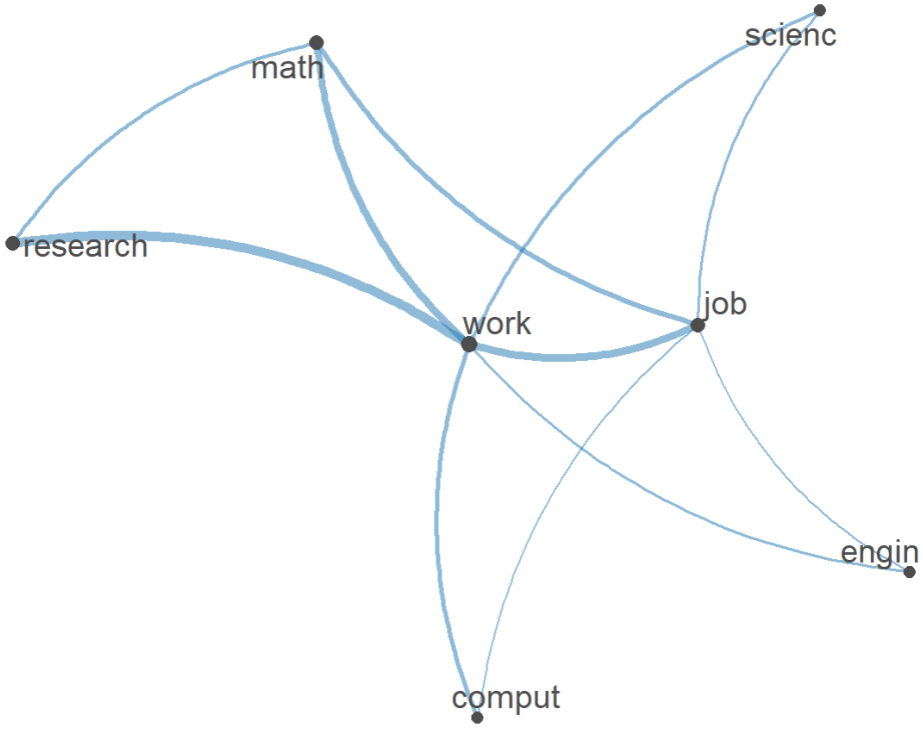


Figure 5: Top Impact Words across Career Talks



The common feature is reconfirmed by a plot of the top impact words for each presenter in Figure 4. After the text data aggregation, the speaker coverage of applied learning experiences is represented by tokenized terms of *teach*, *learn*, *research*, *appli*, *student*, and *school*. In addition, *work*, *job*, and *want* are the top-three impact words to stress the importance of workforce readiness in these *Career Talks* (Figure 5).

Figure 6: Token-Indicator Relations among Career Talks



In R computing, truncated terms are employed to reduce the matrix sparsity. A token-indicator plot is created in this report to extract the emphasis of *Career Talks* on expanding STEM pathways. In Figure 6, tokenized terms of *scienc[e]*, *engin[eering]*, and *comput[ing]* not only show mutual connections among the subject domains, but also illustrate the network support for *work* and *job* indicators at the center. Meanwhile, indicators of *math* and *research* are reciprocally connected, and the line thickness also shows their stronger support for the *work* and *job* foci of these *Career Talks*.

In summary, the text mining repeatedly suggests that *Career Talks* consistently meet their expectations of supporting the development of student pathways for in-demand STEM employment.

After the IRB protocol approval in Spring 2022, additional qualitative data were gathered from student comments in a questionnaire survey for the fifth guest speaker session. One student described the talk as "interesting and motivational." Another student thought that "seeing someone experience what I'm going through helps motivate me further." The presenter also "spoke clearly and at a great pace", according to the survey feedback. Like the unambiguous findings from text analytics (see Figures 1-6), positive comments from this survey added coherent evidence to suggest effective deliveries of *Career Talks*.

Quantitative Survey Results

Seventy-three students participated in a survey of the fifth guest speaker session. On a five-point scale (where 5 represents the best rating and 1 for the worst rating), the average student ratings are listed in Table 2.

Table 2: Average of the Rating Responses from Student Survey

Survey Item	Mean
Relevancy of the presentation to STEM career consideration	4.03
Informativeness of the presentation	4.53
Clarity of the presentation	4.74

All the results were above midpoint 3, suggesting that the presentation is informative and clear. The high rating on the *relevancy of the presentation to STEM career consideration* was backed by text mining results in Figures 1, 2, and 5 about the emphasis of Speaker 5 on key components of *work, internship, research, learning*, and *interest*.

In addition, 95.9% of the respondents believed that the pace of this presentation was *just right*. Hence, the *Career Talk* has effectively engaged students in the learning process. As a

result, 95.9% of the participants reported an increase in their interest in STEM career choices after the presentation, and 63% of the respondents *definitely would* recommend the speaker to others. During or after the presentation, 74% of the students believed that the speaker was responsive to their questions.

In retrospect, Year 1 involves ground-breaking explorations to cope with unexpected challenges in the project setup. While the issue in Internship Coordinator hiring is addressed by creating a dual half-time position in partnership with another grant, data collection delay is coped with in this section by triangulation of quantitative and qualitative data on the project effectiveness. Altogether, the quantitative data came from questionnaire surveys of the *Summer Research Program* and *Career Talk 5*; both occurred in 2022 after the IRB approval. Meanwhile, text analytics were applied to video recordings of all *Career Talk* sessions, including those in 2021, before the IRB protocol implementation. Built on the offering of guest speaker sessions in regular quarters and STEM research programs in the summer, the results aggregated in this section jointly support an assertion that the HSI STEM project has met the grant expectation by effectively enriching student learning opportunities throughout the year.

Configuration of the Rate of Progress

In describing the recent trend, Gurantz, Hurwitz, and Smith (2017) observed, "Hispanic students have graduated high school and entered college in growing numbers. Yet the rate of Hispanic college completion has remained persistently lower than that of whites and other ethnic groups in the United States" (p. 61). Advocated by *Excelencia in Education,* a national organization with a mission to *accelerate Latino student success in higher education*,[5] "measurement of student progress" is listed as a criterion for "Examples of Excelencia" in

[5] https://www.edexcelencia.org/

strengthening Latina/o student education. To support the grant set up in Year 1, the creation of a credible indicator is essential to revealing and tracking this issue for improvement.

From the perspective of formative evaluation, the low rate of degree completion is linked to slow progress in the program pipeline for Latino students. Thus, many of them, particularly those transferred from community colleges, might take longer to go through the *freshman*, *sophomore*, *junior*, and *senior* stages within the funding period. The data tracking also inevitably impacts the subsequent opportunity gap in career development after program completion. Therefore, monitoring student academic progress is needed to project the sustainability of grant contributions based on the trend data from the program pipeline.

In configuring the Rate of Progress (ROP), González and Ballysingh (2012) reviewed the literature and identified a common limitation as the "absence of a plan and process to track student cohorts served by the programs over time" (p. 283). Bahr (2009) concurred,

> Variables that address student enrollment patterns (e.g., persistence, enrollment inconsistency, completed credit hours, course credit load, course completion rate, procrastination) constitute a longstanding fixture of analytical strategies in educational research, particularly research that focuses on explaining variation in academic outcomes. However, nearly all measures of enrollment patterns are handicapped by untested assumptions about a more fundamental measure, namely students' rate of progress. (p. 691)

Fortunately, the PeopleSoft system at CSU has student classifications at the *freshman*, *sophomore*, *junior*, and *senior* levels. One feasible mechanism is to compute the median time that has been taken by STEM students in each major. Accordingly, ROP can be defined as

$$ROP = \frac{Median\ Year_{ps}}{Year_{ips}}$$

where i = individual, p = program (e.g., Math, Science, Engineering), s = stage (freshman, … senior).

When the ROP values are strictly greater than 1, it indicates faster student progress relative to the median group within a STEM major at the same stage; values strictly less than 1 imply slower progress, and values equal to 1 fit a student taking the median number of years to complete the particular stage of progress.

The index construction is deeply rooted in the research literature. For instance, Millett and Nettles (2009) testified that "We constructed our rate of progress measure by grouping individuals by their fields of study and reported stages of progress" (p. 68). Furthermore, they reconfirmed that the rate of progress measure was a ratio of dividing a field- and stage-specific median value by the time each individual reported being in the program at the time of data collection (see p. 68 of Millett & Nettles, 2009). Therefore, it is feasible to use the carefully-defined *ROP* indicator for monitoring student progress toward degree completion.

In summary, "Students' rate of progress is a fundamental concept in educational research, ... Only recently has the literature begun to hint at its import" (Bahr, 2009, p. 710). In the first year, the creation of this indicator not only meets the HSI STEM project needs in formative evaluation, but also supports the record articulation to project student attainment after the grant completion. González and Ballysingh (2012) stressed, "very few of the programs we investigated collected and used local data that shed light on the localized experiences and challenges of the Latina/o students at their institutions" (p. 284). Hence, the ROP tracking is particularly relevant to the grant support for local students of Hispanic origin.

Conclusion

According to González and Ballysingh (2012), effective programs share four common

characteristics, "(a) longitudinal, disaggregated cohort tracking, (b) utilization of formative evaluation data, (c) utilization of the scholarly literature, and (d) collecting and using local data to revise and enhance services to students" (p. 282). During the first year of the grant operation, preparations have been made by the HSI STEM project team to get a data researcher on board for implementing *longitudinal, disaggregated cohort* tracking. The utilization of formative data is demonstrated by the ROP indicator development that is rooted in the scholarly literature. The evaluation framework also incorporates the *participatory*, *utilization-focused*, and *program theory-driven* approach with solid literature support (e.g., Donaldson, 2007; Guijt, 2014; Patton, 2008). Efforts in collecting and using local data are exemplified by both quantitative and qualitative inquiries on the fulfillment of Year 1 tasks, as well as information extraction from text analytics and survey data reporting. Altogether, this project has met all four characteristics of an effective program.

Meanwhile, the evaluation data analyses have led to three recommendations for future improvement. In responding to the *Career Talk* survey in 2022, a student adduced the following considerations,

> Every survey tells us the stories of people who had it worse than others at some time and got up from there, but we should be told more stories about people who did it correctly their whole college career like (+3.8 GPA) so people can at least aspire or try to get there.

As a result, one recommendation can be derived from the student feedback to select guest speakers with a smooth journey during the STEM major preparation and school-to-work transition. While lessons from a setback experience are valuable for STEM career development, a carefully-chosen role model may inspire students to replicate the successful pathway with a

north star guidance.

Voorhees and Lee (2009) maintained that tracking a specific group of students is an effective way to identify achievement gaps and assess a program's impact on its participants over time. As a project funded by the Title III grant, a pertinent recommendation is to track the feedback from Latino students in future survey data collections. This recommendation could especially fit CSU with Graduate Initiative 2025 that pushes for STEM program completion within 4-6 years. Without the group tracking, no one can tell whether more program support is necessary for Latino students who might take an extra one or two years to graduate compared to their peers in the ethnic majority group. Avoiding this ignorance can help attain the servingness goal of this grant.

The third recommendation is to implement the ROP data collection. To date, the mechanism has yet to be incorporated into the existing data reporting from the institutional research sector of CSUB. However, another HSI-serving project funded by the Promoting Postbaccalaureate Opportunities for Hispanic Americans (*PPOHA*) program has already demonstrated the feasibility of ROP tracking for graduate students in different STEM departments (Wang, 2022). The result has been employed as an outcome measure to evaluate the initiatives of the Title Vb funding, such as the Faculty Fellows Program, Faculty Collaborative Research Program, and Student Travel Support Program (Jacobsen, 2022). Built on this foundation, it is recommended that the data researcher discusses with a Co-PI of PPOHA to create a system of ROP tracking that fits this HSI STEM project. Accompanied by a report of increasing Hispanic student enrollments in tertiary education (Gurantz, Hurwitz, & Smith, 2017), the ROP indicator can help tackle an acute and more profound question raised by González and Ballysingh (2012), "Why has the gap between White and Latina/o college degree attainment

widened over the past three decades instead of narrowed?” (p. 280).

References

Angelo, T. (1999, May). Doing assessment as if learning matters most. *American Association for Higher Education Bulletin*, pp. 1-2.

Bahr, P. (2009). Educational attainment as process: Using hierarchical discrete-time event history analysis to model rate of progress. *Research in Higher Education, 50,* 691-714.

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software, 3*(30), 774.

Caro, D., & Biecek, P. (2017). intsvy: An R package for analyzing international large-scale assessment data. *Journal of Statistical Software, 81*, 1-44

Demetrikopoulos, M. (2020). *Program evaluation development and standards*. Retrieved from https://hsistemhub.org/portfolio-item/june-2020-newsletter/

Donaldson, S. I. (2007). *Program theory-driven evaluation science: Strategies and applications*. New York: Routledge.

Donaldson, S., Christie, C., & Mark, M. (2009). *What Counts as Credible Evidence in Applied Research and Evaluation Practice?* London: Sage.

González, K. P., & Ballysingh, T. A. (2012). Increasing Latina/o college completion: Mistakes and opportunities. *Journal of Hispanic Higher Education, 11*(3), 279-290.

Guijt, I. (2014). *Participatory approaches, methodological briefs: Impact evaluation 5, UNICEF Office of Research, Florence*. Retrieved from:http://devinfolive.info/impact_evaluation/img/downloads/Participatory_Approaches_ENG.pdf

Gurantz, O., Hurwitz, M., & Smith, J. (2017). Boosting Hispanic college completion: Does high-school recruiting help more students graduate?. *Education Next, 17*(3), 60-68.

Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research*. Sage Publications.

Jacobsen, A. (2022). *Data driven graduate mentoring in the sciences: CSUB Title Vb STEM Graduate Excellence grant*. Bakersfield, CA: Author.

Konecki, K. T. (2021). Contemplative grounded theory: Possibilities and limitations. *Radical Interactionism and Critiques of Contemporary Culture*, *52*, 151-186.

Leeuw, F. L., & Donaldson, S. I. (2015). Theory in evaluation: Reducing confusion and encouraging debate. *Evaluation, 21*(4), 467-480.

Matta, T. H., Rutkowski, L., Rutkowski, D., & Liaw, Y. L. (2018). lsasim: An R package for simulating large-scale assessment data. *Large-scale Assessments in Education*, *6*, 15.

Miller, E. R., & King, T. (2019). *Promoting transformation of undergraduate STEM education: Workshop summary report.* Association of American Universities. Retrieved from https://files.eric.ed.gov/fulltext/ED605104.pdf

Millett, C., & Nettles, M. (2009). Rate of progress, degree completion, and time to degree. In R. G. Ehrenberg & C. V. Kuh (Eds.), *Doctoral education and the faculty of the future*. Ithaca, NY: Cornell University Press.

Patton, M.Q. (2008). *Utilization-focused evaluation (4th ed.).* Thousand Oaks, CA: Sage.

Sarkar, D. (2019). *Text analytics with Python: A practitioner's guide to natural language processing*. New York: Springer.

Thoumyre, E. (2022). *Tips on handling the impact of inflation on salary expectations.* Retrieved from https://accurservices.com/tips-on-handling-the-impact-of-inflation-on-salary-expectations/

Voorhees, R., & Lee, J. (2009). *Basics of longitudinal cohort analysis*. Indianapolis, IN: Achieving the Dream, Inc.

Wang, J. (2022). *PPOHA project evaluation: Strengthening STEM education of Latinx students for graduate degree completion (Evaluator Report on Year 2 Operation)*. Retrieved from https://eric.ed.gov/?q=jianjun+wang&id=ED618495

Wang, J., & Johnson, D. (2019). An examination of discrepancies in multiple imputation procedures between SAS and SPSS. *The American Statistician, 73*(1), 80-88.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The program evaluation standards* (3rd ed.). Thousand Oaks, CA: Sage & the Joint Committee on Standards for Educational Evaluation.

Appendix 1:

R Scripts for Information Extraction from Career Talks

```
install.packages ("quanteda", "readtext", "quanteda.textstats", "rlang", "ggplot2",
"quanteda.textplots")
library(readtext)
HSI <- readtext("D:/Lam/text/*", docvarsfrom = "filenames",
            docvarnames = "Guest_Speaker", encoding = "UTF-8")
library(quanteda)
d_corp1<-corpus(HSI)
HSI1<-tokens(d_corp1, what="word", remove_numbers=T, remove_punct=T,
remove_symbols=T, split_hyphens=T)
HSI1<-tokens_tolower(HSI1)
HSI1 <- tokens_select(HSI1, pattern = stopwords('en'), selection = 'remove')
HSI1 <- tokens_wordstem(HSI1)
d_corp1dfm<-dfm(HSI1)
library(quanteda.textplots)
library(ggplot2)
theme_set(theme_bw())
tplot <- textplot_xray(kwic(HSI1, pattern=c("math*", "scienc*", "engin*", "comput*", "statist*",
"teach")))
tplot + aes(color = keyword) + scale_color_manual (values = c("red", "blue", "violet", "brown",
"purple", "black")) + theme(legend.position = "none")
library("quanteda.textstats")
tstat1 <- textstat_frequency(d_corp1dfm)
ggplot(tstat1[1:11, ], aes(x = reorder(feature, frequency), y = frequency)) +
  geom_point() +
  coord_flip() +
  labs(x = NULL, y = "Frequency")
library(quanteda.textplots)
d_corp1_dfm<-dfm(HSI1)
d_corp1_dfm<-dfm_trim(d_corp1_dfm, min_termfreq = 3, verbose = F)
textplot_wordcloud(d_corp1_dfm, group = "Guest_Speaker", comparison=T, color = c("blue",
"brown", "orange", "purple"))
HSI1<-quanteda:::tokens_group(HSI1, groups = Guest_Speaker)
HSI2<-tokens_keep(HSI1, pattern=c("research", "learn", "work", "job", "degre*",  "school",
"want", "like", "love", "dr", "phd", "need", "help", "math*", "scienc*", "engin*", "comput*",
"statist*", "teach"))
HSI1dfm <- dfm(HSI2)
docvars(HSI1)
library(rlang)
fcmat_d1 <- fcm(HSI2)
dim(fcmat_d1)
feat <- names(topfeatures(fcmat_d1, 10))
fcmat_news_select <- fcm_select(fcmat_d1, pattern = feat)
```

```
dim(fcmat_news_select)
size <- log(colSums(dfm_select(HSI1dfm, feat)))
set.seed(144)
textplot_network(fcmat_d1, min_freq = 0.8, vertex_size = size / max(size) * 3)
textplot_network(fcmat_news_select, min_freq = 0.8, vertex_size = size / max(size) * 3)
d_corp1_dfm<-dfm(HSI1)
tstat_key <- textstat_keyness(d_corp1_dfm, target ="Speaker2")
textplot_keyness(tstat_key, color = c("blue", "red"), n = 10)
library(manifestoR)
feature_frequencies_categories <- d_corp1_dfm %>%  textstat_frequency(n = 10, group = docid)
library(dplyr)
feature_frequencies_categories %>%
  mutate(cmp_code = factor(group)) %>%
  ggplot(aes(x = reorder(feature, frequency) , y = frequency, fill = cmp_code)) +
  geom_col(show.legend = FALSE) +
  labs(x = NULL, y = "share of words per category") +
  facet_wrap(~cmp_code, ncol = 2, scales = "free") +
  coord_flip()
```