NIH Diversity Program Consortium

Technical Report

BUILD Program Evaluation: Case Study Analysis

Prepared by: Michael W. Moses II, Ana Romero, Justin Gutzwa, Hector Ramos, Nicole Maccalla, Dawn Purnell, & Sylvia Hurtado, Coordination and Evaluation Center of the Diversity Program Consortium, UCLA

www.diversityprogramconsortium.org

DIVERSITY PROGRAM CONSORTIUM
Supported by the National Institutes of Health

Coordination & Evaluation Center | UCLA

Introduction

Funded by the National Institutes of Health (NIH), Building Infrastructure Leading to Diversity (BUILD) features a set of 10 linked awards granted to undergraduate institutions, each of which developed approaches intended to determine the most effective ways to engage students from diverse backgrounds in biomedical research. These awards also intend to prepare students to become potential future contributors to the NIH-funded research enterprise. The Coordination and Evaluation Center (CEC) at UCLA has been charged with conducting a longitudinal, multi-method evaluation of the BUILD programmatic interventions designed to diversify the biomedical workforce (Davidson, Maccalla, Afifi, et al., 2017). This technical report documents the methodology for the qualitative analysis employed by the CEC case study evaluation team. See our other white paper for details on the protocols and data collection process.

Our analysis draws upon methodologist Pat Bazeley's guidance about the importance of clear and documented audit trails as necessary evidence to convince readers about the strength and reliability of qualitative analytic procedures: "By making your perspective and your questions explicit, you can increase the consistency of your coding, and assist the reader to understand how and why you interpreted the data in a particular way" (2013, p. 148). Over the course of the evaluation, the case study team has therefore relied on ongoing strategies of dialogic engagement (Ravitch & Carl, 2016) to offer rigorous, trustworthy, and credible interpretations of each BUILD awardee. These strategies included individually and collaboratively written analytic memos, weekly virtual meetings (via the web-based platform Zoom), and peer-reviewed data workshop sessions. The following evaluation questions guided the CEC team's methodological approach to analysis and design of the protocols used throughout data collection and analysis. Our central questions are:

How are BUILD programs building capacity and infrastructure for Primary and Partner Institutions to advance diverse student success in biomedical research training?

Institutional Commitment to Diversity:
How are BUILD sites addressing diversity and inclusion in biomedical research training?
How are BUILD sites increasing the number of biomedical researchers from underrepresented groups?

Institutionalization and Sustainability:
How are BUILD sites actively building program sustainability?

Participant Experiences:
What is the student experience of BUILD activities? How is their experience related to program outcomes?
What is the faculty experience of BUILD activities? How is their experience related to program outcomes?

As part of the overall CEC multi-method approach, these research questions will also inform a more complete understanding of the findings that are emerging from the quantitative analyses that focus primarily on the "Hallmarks of Success" (see https://www.diversityprogramconsortium.org/pages/hallmarks_yr6-10). The Hallmarks provide a largely deductive analytic framework for analysis while the qualitative data allows a more inductive and discovery framework. In addition, a number of the institutional Hallmarks are best answered with the more holistic approach offered by case study methods.

METHODOLOGY

The CEC evaluation team employed a descriptive, embedded multiple-case study design (Yin, 2018) to gain "an in-depth, multi-faceted understanding of a complex issue [i.e., each BUILD site] in its real-life context" (Crowe et al., 2011, p. 1). Within each BUILD program (or individual case), we examined four embedded sub-cases including: BUILD students, BUILD faculty, BUILD key programmatic dimensions, and the intra- and inter-institutional level dimensions and partnerships of each BUILD site. This examination seeks to identify the conditions, features, and characteristics that allowed for challenges and successes within the BUILD programs. Cross-case analyses of the 10 BUILD sites commenced in fall 2019 to offer a compelling and robust depiction of the innovative intervention and sustainability strategies designed to promote engagement with underrepresented students in biomedical research training.

Data Collection

Data collection occurred in two phases, with five site visits completed in the first half of 2017 and the second five completed in the fall of 2018. Protocols and procedures were approved by the UCLA Human Subjects Review Board. In total, the CEC team collected hour-long interview data with over 500 participants (individual and focus group formats) and completed 40-60 hours of observations by the completion of data collection. Details about the development of the protocols and data collection are available in the (forthcoming) white paper on those topics.

ANALYTIC PROCEDURES

At the end of each site visit, the CEC case study team conducted a concluding debrief meeting with each site's PIs and integral program officials, offering key takeaways from interviews and observations that had been completed. The CEC case study team also wrote and shared with BUILD awardees institutional debrief reports soon after the completion of each visit. This provided the case study team a means to summarize first impressions and provide immediate constructive feedback to each site. These methods also provided a first level of member checking as each site was able to respond to and check the case study team's initial feedback for accuracy.

Creating the Codebook

Upon the completion of data collection, CEC evaluators created a codebook to systematically analyze data collected from each site visit. The codebook primarily served as a measure to examine how trends and concepts from existing literature on diversity in the biomedical sciences and the DPC "Hallmarks of Success" manifested within the data collected. Emergent codes or themes that were important to understanding BUILD implementation were generated inductively. The codebook has four categories of codes, each representing the distinct experiences of various stakeholders across all 10 BUILD programs. These categories include:

1. Program-level codes: These codes address BUILD programmatic structure and implementation along with the narrative experiences and perceptions of BUILD program from key stakeholders (e.g., Principal Investigators, program administrators, evaluators, faculty and students). For example, "intended recipients/audience" and "strategies to promote diversity in biomedical research" were codes designed to classify programmatic BUILD interventions.

2. Institutional-level codes: These codes addressed the culture, climate, and make-up of each BUILD institution along with intra and inter-institutional partnerships and the experiences of institutional leadership (e.g., chancellors, provosts, and deans). For example, "amplification of existing activities" and "institutional culture/climate" were codes designed to classify institutional components at each BUILD site.

3. Faculty-level codes: These codes addressed the experiences of faculty participants and their perceived impact of being involved in BUILD programming. For example, "faculty-to-faculty mentoring" and "training/faculty development" were codes designed to characterize faculty participant experiences within BUILD.

4. Student-level codes: These codes addressed the experiences, narratives and program impact of student participants (both undergraduate and graduate) involved in BUILD programming. For example, "perceived fit/sense of belonging" and "inspiration to pursue biomedical research" were codes designed to characterize student participant experiences within BUILD.

These four categories consisted of subsequent groupings of codes that captured descriptive characterizations of stakeholders' positive and negative experiences; institutional and programmatic successes, innovations, and challenges; sustainability, institutionalization, and amplification efforts; and the nature of intra and inter-institutional partnerships (see Appendix A list of codes).

Interrater Agreement and Reliability Process

The goal of the interrater agreement and reliability process was to assess the applicability of codes, finalize the design of the codebook and address discrepancies amongst team members' approaches to interpreting the codebook with the goal of accurately identifying codes and themes within the data.

The agreement process involved an initial review of the codebook using an example transcript from one student, faculty member, program official, and upper administrator from different BUILD sites. Transcripts were chosen from different sites to account for the variability of BUILD activities and developmental interventions across the 10 programs. The team coding the first five cases after their completion met weekly and multiple coders compared their codings and came to consensus on how cases should be coded. A different coding team worked on the last five sites and were onboarded by members of the site visit team who also worked on the first round of coding, who continued to be available to consult as needed. The four new researchers charged with coding round two of site visits began their initial review by reflecting on how the data from the second five cases fit within the existing codes. Work by those engaged with the first five sites as well as the second five expanded the first code book to include more finely targeted codes that were consistent with those from the first round (see interview protocols white paper for details). Selected data from round one of site visits were coded by the new coding team members to provide cross coding team validation.

During the initial review of the codebook, researchers asked themselves, how do participants' language and experiences fit within the codes identified? Does the codebook account for the breath of participant experiences and perceptions? Reflections were done in the form of analytic memos written by each coder as they reviewed the example transcripts; the case study team met weekly to discuss their reflections and initial coding. This reflective process culminated in the case study team's collective agreement upon a codebook of over 100 codes for use during data analysis.

In order to establish intercoder reliability to accurately, efficiently, and consistently code the entire data corpus, one researcher coded a different set of example transcripts representing one student, faculty member, program official, and upper administrator from the BUILD landscape. Of the four researchers, this researcher was the longest standing member and most informed about the CEC's evaluation. As such, their institutional knowledge of the CEC and DPC landscape was used to design an 18-question reliability test within the qualitative data analysis software Dedoose. The researcher assessed for comprehension of the four embedded sub-cases by assigning codes from each of the cases' respective groupings of codes (i.e., student, faculty, program and institutional levels) to excerpts of the data. This approach allowed for the test to assess coders' ability to accurately categorize data within the four primary categories of codes (i.e., student, faculty, institutional, and program). The test also assessed coders' knowledge of the subsequent level of codes characterizing nominal and descriptive categorizations across various stakeholder experiences (e.g., "negative experiences," "challenges").

Dedoose was employed to use this researcher's assignment of codes to data excerpts as the standard, or Kappa statistic, from which the remaining three researchers would establish reliability. Their intercoder reliability ratings of 1.0, .91, and .94 (or .96 across all ratings) were reached based upon a Pooled Kappa (de Vries, Elliot, Kanouse & Teleki, 2008). Given the extensive size of the codebook, the team deemed these ratings as sufficient levels of agreement to proceed with the coding of the remaining transcripts.

Overviewing the Coding Process

The CEC case study team embarked upon a two-phase coding process to ensure reliable sense-making and reporting of the data (Saldaña, 2013). Phase one occurred from January to March of 2019. Phase two began thereafter and ended in May of 2019. Each phase (outlined below) utilized distinct inductive and deductive approaches to ensure that our analytic conclusions spoke to both the lived experiences of case study participants and general body of knowledge about underrepresented student success and retention in STEM.

Phase one of coding. Phase one allowed coders an opportunity to intimately familiarize themselves with the unique qualities of each BUILD site. Additionally, this phase enabled coders to characterize meaningful patterns across the data, subsequently providing coders a general sense about "'What is going on here [at a particular BUILD site]?'" (Saldaña, 2013). During this phase, coders initially reviewed each site's data corpus by employing inductive approaches to coding [primarily in-vivo (Strauss, 1987) and descriptive (Miles & Huberman, 1994; Wolcott, 1994) strategies] while simultaneously documenting comments and questions in the margins of each transcript in Microsoft Word. Descriptive and in-vivo approaches to coding offered language to summarize chunks of the data by using "the terms used by [participants] themselves" (Strauss, 1987, p. 33). More importantly, these coding strategies were a means to prioritize and honor participants' voices and lived experiences (Bazeley, 2013).

Phase two of coding. Phase two featured a systematic, deductive application of the codebook for each BUILD site within Dedoose, a web-based qualitative data analysis software. This phase of coding specifically sought to establish reliable frequency counts of code usage across the data to identify salient themes for inclusion in the writing of institutional narratives. Coders conducted a chunk by chunk coding of the data, as opposed to line by line, to account for the context needed to accurately interpret quotes during the narrative reporting of each site. During this phase of coding, coders also revisited CEC and DPC literature such as each site's description in the DPC's supplement published in BMC Proceedings (Hurtado, 2017) and institutional debrief reports drafted upon the completion of each BUILD site-visit. Referring to this literature allowed for coders to triangulate and contextualize salient themes emerging from the analysis of each site's data corpus.

Institutional Narrative Report Writing

According to case study methodologist Robert Yin (2018), the writing of institutional narratives (or case reports) serve as foundational information to conduct cross-case analyses. These reports provide researchers deeply contextualized narratives to facilitate cross-case comparison and synthesize the programmatic features, challenges, and innovations learned from each individual case (Yin, 2018). To begin the writing of these reports, coders developed a PowerPoint presentation featuring the salient themes and relevant data excerpts that emerged from their coding of the data. Researchers specifically organized their presentations in the following topical arrangement: 1) student-level innovations, experiences, and challenges 2) faculty-level innovations, experiences, and challenges and 3) institutional and program-level innovations and challenges. With these presentations, coders guided their peers through a

data-informed tour of each site's challenges and successes in their efforts to diversify the biomedical pipeline. Coding team members provided constructive feedback to help each other think through their rhetorical presentation of the data and troubleshoot any analytic points of concern. More importantly, these presentations served as draft outlines from which coders began to write 15-20 page narrative reports for each site. These reports are part of the initial stage of analysis to capture context before cross-site comparisons are attempted in the next stage of analysis to answer the research questions (Yin, 2018).

We attempted to ensure that each report followed a standardized organization and structure. The complexity and nuance of each BUILD program necessitated contextualized considerations and deviation from a standardized report template in order to best capture the rich and unique data collected from each site. Despite these considerations, narrative reports reflect the four-categories of the CEC codebook by highlighting the institutional, program, faculty, and student-level dimensions of each BUILD site.

Multiple Cross Case Analytic Procedures

While analysis can occur within a single site using data at the four levels to study multiple realities or perceptions, the key focus and advantage of an embedded multiple case study design is that 1) selected embedded cases can be studied across institutional sites (e.g. examining only student participant experiences across all sites, or variations in faculty development approaches across all sites), 2) comparisons can be drawn across sites on an interrelated set of codes/themes to understand how they are shaped by context (e.g. forms of institutional commitment, identifying innovations specific to sites), or 3) a focused analysis can occur to identify similarities in approach, logic, or conditions that inform implementation and/or outcomes (e.g. pulling codes and quotes that inform a single Hallmark of Success across embedded cases and across sites). These analyses can become quite complex and because of the large number of codes, participants, and sites, it is important to have a systematic procedure to draw conclusions regarding similarities or differences. The primary technique for cross case comparison are to first focus on the guiding theory, hypotheses, and research questions, determine one's approach described above (select embedded case level, interrelated set of codes/themes, or focused analysis of single code/theme across embedded cases and/or sites), and then run data queries using the software. The codes and quotes are subsequently organized into matrices (columns and rows) that allow researchers to "see" the data, its frequency and qualitative differences across sites and/or embedded cases (Miles & Huberman, 1994). The matrix allows researchers to see how codes may be further related to each other as well as identify similar and unique approaches and perspectives across and within sites. The dialogue within the research team is a valuable step to further inform how and why they interpret the findings as writing begins to occur. As a method of ensuring trustworthiness, before conclusions are drawn, a systematic search for divergent cases or participants should be included in the analysis and discussed since this more often than not reflects the reality within institutions in the midst of innovation and change. Finally, because all participants have given their interviews under an understanding of confidentiality, any written results should ensure protection of informant identity by using broad description or roles. Member checking can also be employed to assure the informant's identity is protected before manuscripts are made public.

LIMITATIONS

As for any research and evaluation project, our methodological process was not without its own set of unique challenges and limitations. Our primary concern sought to find credible and trustworthy approaches to data analysis given a number of ebbs and flows throughout the research process. First, case study data collection was interrupted for a year to accommodate NIH site visits for purposes of grant renewal. While the same protocols were used, the campuses were in earlier or later phases of their program development, depending on when the case study visit occurred. We attempted to select campuses that were further along in organizing their program in the first round of site visits; however, what remained clear was that the first cohorts of BUILD participants had different levels of exposure to program activities than later cohorts of participants. Moreover, faculty development activities were still in formation, and thinking about institutionalization efforts was a discussion during the first round of site visits while more focused considerations took place during the second round. Those campuses visited later had evolved further and began to think of the next five years of funding, laying out their proposals, and changing initiatives. Rather than replications of a single experiment, the case studies capture institutions at different stages of evolution of their program along with contextually-based differences in serving populations and collaborations with nearby institutions. Methodologists Merriam and Tisdell (2016) state, "one of the assumptions underlying qualitative research is that reality is holistic, multidimensional, and ever-changing; it is not a single, fixed, objective phenomenon waiting to be discovered, observed, and measured as in quantitative research" (p. 242). This assertion certainly holds true for what was observed on the different sites, contrasting participants' views within each site, as well as thinking about comparisons across each site in participants' understanding of the program and their role.

Second, over the course of the evaluation, 15 researchers have been involved in various aspects of the design, collection, and analysis of the case studies. During the midpoint of data collection, changes in the team's makeup occurred as some team members went on to other career opportunities, for example. While three researchers spanned both sets of data collection, the coding teams had significant changes in membership. Given this limitation, some members of the current data analysis team frequently referred to internal literature, such as analytic memos, methodological reports, and other collected documents, to guarantee that their approach to analysis mirrored the evaluation's original analytic procedures. The current team also met with previous team members via phone and/or Zoom for clarity, background, and reasoning behind some of the team's earlier methodological decisions. Lastly, four senior members of the case study team, who have been with the evaluation since its inception, have been critical to the CEC's maintenance of institutional knowledge and analysis of case study data. Ravitch and Mittenfelner Carl (2016) describe dialogic engagement as "a requirement of rigorous, reflexive research and constitutes an approach to qualitative research that engenders and supports criticality": "Dialogic engagement processes allow you to co-create the conditions of collaboration by deliberately engaging thought partners, critical friends, and/or research participants to challenge your biases and interpretations" (p. 16). Given these limitations and challenges, we prioritized dialogic engagement as a means to ensure credibility and consistency throughout the ongoing nature of the data analysis process.

CONCLUSION

The case study method is best suited for understanding multiple contextual adaptations of an institution-wide capacity-building initiative such as the BUILD within the Diversity Program Consortium. More specifically, BUILD programming considers student training, faculty development, and institutional changes as areas of interventions to increase institutional capacity for enhancing diversity in biomedical research training (Hurtado, White-Lewis & Norris, 2017). Contextual differences are key in understanding not only how campuses have interpreted and implemented their charge but also what indications possibly explain variation in outcomes. The lessons learned in understanding institutions and their response to an NIH funding initiative to increase diversity in the biomedical workforce will be valuable for other institutions engaged in change across the four areas of students, faculty, program interventions, and institution-wide capacity-building.

Publications and Contact Information

Suggested Citation:

Acknowledgements

References

Bazeley, P. (2013). Qualitative data analysis: Practical strategies. Thousand Oaks, CA: Sage.

Davidson, P.L., Maccalla, N.M., Afifi, A.A., Guerrero, L., Nakazono, T.T., Zhong, S., & Wallace, S.P. "A participatory approach to evaluating a national training and institutional change initiative: the BUILD longitudinal evaluation." BMC Proceedings 2017, 11(Suppl 12):15. PMCID: PMC5773880 http://rdcu.be/AnJU

De Vries, H., Elliott, M. N., Kanouse, D. E., & Teleki, S. S. (2008). Using pooled kappa to summarize interrater agreement across many items. Field Methods, 20(3), 272-282.

Hurtado, S (Ed). (2017). The Diversity Program Consortium: Innovating education practice and evaluation along the biomedical research pathways. Biomedical Central Proceedings, 11(Suppl 12):28.

Hurtado, S., White-Lewis, D. & Norris, K. (2017) Advancing inclusive science and systemic change: The convergence of national goals and institutional aims in implementing and assessing biomedical science training, BMC Central Proceedings, 11(Suppl 12):17 DOI 10.1186/s12919-017-0086-5.

Merriam, S. B., & Tisdell, E. J. (2016). Qualitative research: A guide to design and implementation, 4th Edition. San Francisco: John Wiley & Sons.

Miles, M. B., & Huberman, M. (1994). Qualitative data analysis: An expanded sourcebook, 2nd Edition. Thousand Oaks, CA: Sage.

Ravitch, S.M., & Carl, N.M. (2016). Qualitative research: Bridging the conceptual, theoretical, and methodological. Los Angeles, CA: SAGE.

Saldaña, J. (2013). The coding manual for qualitative researchers. Thousand Oaks, CA: Sage.

Strauss, A. L. (1987). Qualitative analysis for social scientists. Cambridge: Cambridge University Press.

Wolcott, H. F. (1994). Transforming qualitative data: Description, analysis, and interpretation. Thousand Oaks, CA: Sage.

Yin, R. K. (2018). Case study research and applications: Design and methods, 6th Edition. Thousand Oakes: Sage.

Appendix A: CEC Coding Headings

Institutional Level
• Institutionalized Components
  • Curriculum
  • Research & Research Infrastructure
  • Tenure and Promotion
  • Faculty Diversity
  • Diversity Training
  • Student Advising
• Amplification of Existing Activities
• Institutional culture/climate
• Institutional Partnerships
  • Expectations
  • Incentives for participation in BUILD activities
  • Negative Experiences & Challenges
  • Positive Impact/Experiences
  • Sustainability & Recommendations
  • Inter-Institutional Partnerships
    • Pipeline Partnerships
    • Research Partnerships
    • Community Partnerships
  • Intra-Institutional Partnerships
• Sustainability
• Challenges
• Non-BUILD Diversity Initiatives

Faculty Level
• Mentoring
  • Barriers to mentoring
  • Structural supports
  • Undergraduate student mentorship
  • Faculty-to-Faculty Mentorship
• Positive impacts/experiences
  • Scholarly productivity
  • Interdisciplinary research
  • Pedagogical practices
  • Networking & Interdisciplinary Connections
  • Curriculum Development
  • Lab Supplies & Equipment
  • Travel Awards
• Negative experiences
• Challenges/Tensions
  • Incentives for participation in BUILD activities
• Institutional Culture/Climate

Faculty Level Cont.
• Training/Faculty Development
	• Research
	• Pedagogical
	• Mentoring
	• Diversity
	• Barriers to training participation
	• Satisfaction with training
	• Recommendations for future training
• Perceived Positive Student Experiences

Student Level
• Science identity
	• Barriers to science identity
• Perceived fit/Sense of Belonging
	• University
	• Research Community
	• Barriers to perceived fit
• Self-efficacy
	• Science/Researcher
	• Academic
	• Barriers to self-efficacy
• Graduate Student Challenges
• Graduate Student Positive impacts/experiences
	• Teaching
	• Research
	• Mentoring
	• Career Development
• Recruitment
• Persistence
• Mentorship
	• Culturally Relevant
	• Peer
	• Near-Peer
	• Research
	• Academic & Advising
	• Career
	• Challenges with mentorship
• Preparedness
	• Challenges to preparedness
	• Competitiveness

Student Level Cont.
• Positive impacts/experiences
     • Awards
     • Fellowships
     • Undergraduate Research Experience
     • Publications
     • Presentations
     • Conference Attendance
     • Skills
     • Professional Networks
     • Community Experience
     • Physical Space and Resources
     • Exposure to Real Science
     • Direction in Life
• Negative [program] experiences
• Inspiration to pursue biomedical research
     • Give Back/Make a Difference
• Challenges
     • Funding/Financial Concern
• Institutional Culture/Climate

Program
• Intended recipients/audience
• Feedback & Suggestions
• Commendations
• Strategies to promote diversity in biomedical research
     • Conceptual Frameworks/Approaches (One Health, Entrepreneurship Model, CRT, Stereotype Treat, Micro Aggressions, Micro Affirmations, Growth Mindset)
     • Entry Point into Program
     • URE
     • Mentor-Mentee Matching
     • Student Support (funding; housing; wrap around services; mentoring; advising)
     • Curriculum
     • Training (mentor, pedagogical, research, PD, diversity)
• Innovations to promote diversity in biomedical research
• NRMN partnership
• Challenges
     • Program Implementation
     • External Forces
     • Perceived Student Challenges
     • Challenges Meeting Diverse Needs

www.diversityprogramconsortium.org

DIVERSITY
PROGRAM
CONSORTIUM
Supported by the National Institutes of Health

Coordination &
Evaluation
Center | UCLA