Summer 2021

Applying the Comprehensive Evaluation of Return-On-Talent-Investment Model (CERTi) In Higher Education

Mazen Aziz

APPLYING THE COMPREHENSIVE EVALUATION OF RETURN-ON-TALENT-INVESTMENT MODEL (CERTI) IN HIGHER EDUCATION

by

Mazen Aziz

Bachelor of Science
University of South Carolina, 1998

Master of Human Resource Development
Clemson University, 2015

Submitted in Partial Fulfillment of the Requirements

For the Degree of Doctor of Philosophy in

Educational Administration

College of Education

University of South Carolina

2021

Accepted by:

Henry Tran, Major Professor

Spencer Platt, Committee Member

Christina Yao, Committee Member

Brian Chen, Committee Member

Michael Seaman, Committee Member

Tracey L. Weldon, Interim Vice Provost and Dean of the Graduate School

Dedication

I dedicate this dissertation first and foremost to my family. Without your love, support, and encouragement, along the way, this would not have come to fruition. Each of you contributed to my success and perseverance. I also dedicate this dissertation to my late grandparents, who instilled in us the value of education and fostered a work ethic that facilitated the successful completion of this dissertation.

ACKNOWLEDGEMENTS

I want to acknowledge several people for contributing to my success as a doctoral student. First, I acknowledge my dissertation chair, Dr. Henry Tran, whom I consider a friend at this stage. His support and mentorship were invaluable to my development as a researcher and my growth as an individual. I am eternally grateful for his guidance, patience, and insight in shaping me for the next stage in my professional life. Second, I acknowledge my dissertation committee, Dr. Platt, Dr. Yao, Dr. Chen, and Dr. Seaman, for their instrumental contribution to completing my dissertation. I appreciate their time, effort, feedback, and guidance along the way. Third I acknowledge Dr. Beets, Dr. Weaver, and my Arnold Childhood Obesity Initiative (ACOI) work-family. Thank you for supporting my decision to pursue doctoral studies, making resources available, and giving your invaluable feedback and expertise when needed. Fourth, I acknowledge my Ph.D. cohort (Alex Jordan, LeCinda Jennings, Reena Patel, and Rushondra James) for making the journey enjoyable. I especially acknowledge Dr. Jordan, who became a close friend for his support, encouragement, and collaboration. Lastly, I acknowledge Bill Barlow, a friend that was instrumental to me pursuing doctoral studies in the first place. His advice, encouragement, and support were instrumental to me completing my doctoral studies.

Abstract

Economic turmoil has forced higher education institutions (HEI) to reduce expenditures in many areas, including human resource operations such as talent development. Before considering these cuts, HEI should conduct robust assessments of their efficacy, including whether they generate more revenue than they cost to administer. These assessments were critical contextually as state divestment in higher education and mounting public pressure against tuition hikes forced HEI to heavily rely on external funding sources, which became essential in these economically uncertain times. This three-article dissertation critically examined existing evaluation methodologies of a type of talent development initiative, faculty research development (FRD) programs, that seek to enhance faculty grant acquisition skills. Building on the scholarship, this work proposes a new comprehensive talent-centric evaluation model known as The Comprehensive Evaluation of Return-on-Talent-Investment Model (CERTi) (Aziz & Tran, in press). An exhaustive review of existing measurement and evaluation methodologies of FRD program efficacy in the literature precedes the novel CERTi Model presentation. The model combines multiple evaluation frameworks from varying scientific disciplines into a comprehensive approach to evaluation that advances theory on talent development at HEI. CERTI's holistic Macro-Micro assessment approach employs an overarching (Macro-level) adult-learner faculty-centric theoretical framework for this research while simultaneously incorporating (Micro-Level) qualitative, quantitative, and economic evaluations to assess FRD efforts at HEI jointly. This

dissertation presents a case study of an FRD program for grant acquisition to demonstrate the utility of the model and its application for practice and scholarship. The dissertation utilizes a sequential explanatory observational study design. The first article of the dissertation examined the program's effectiveness (i.e., quantitative assessment), the second examined its' implementation (i.e., qualitative assessment), and the third its' return-on-investment (i.e., economic assessment). As HEI face an organizational environment characterized by state divestment, accountability demands, and requests for financial returns-on-investment, the CERTI approach is critical for efficacious assessments of talent development efforts at HEI.

TABLE OF CONTENTS

List of Tables

LIST OF FIGURES

Chapter One: Introduction

Organization of the Dissertation

This dissertation relies on a single study interdependent or recursive format. This format allows for data utilization from a similar sample population across multiple studies or articles while allowing each article to explore unique variables and applying differing methodological approaches from varying data sources (i.e., quantitative, qualitative, and economic). The articles build on the findings of each other and show a clear link between the identified variables. This introduction aims to provide an overview of the structure of this three-article dissertation. Chapter 1 of the dissertation provides details on the dissertation's organization and serves as a comprehensive introduction for all three articles to orient the reader towards understanding the proposed holistic CERTi approach (Aziz & Tran, in press). This approach is a new comprehensive evaluation methodology that combines a talent-centric macro-level assessment approach with Micro-Level qualitative, quantitative, and economic evaluations to appraise FRD efforts' efficacy at HEI jointly. It includes a statement of the problem, purpose, significance, research questions, and study design. Chapter 2 of the dissertation provides a detailed literature review of the theoretical framework and three evaluation methodologies (i.e., quantitative, qualitative, and economic) utilized for the comprehensive evaluation. Chapter 3 of the dissertation provides methodological context for the rest of the dissertation (i.e., FRD program overview, participant and site selection strategies, researcher reflexivity, and study trustworthiness and validity strategies). Each chapter

following this comprehensive introduction is an independent article with its specific methodology, findings, and discussion sections. Article 1 will be chapter 4, article 2 will be chapter 5, and article 3 will be chapter 6. Lastly, a comprehensive conclusion for all articles that provides a general discussion of the research process's progression and identifies the application to practice and future research serves as the seventh and final chapter of the dissertation.

Problem Statement

State appropriations for HEI have steadily declined (Webber, 2017). These institutions progressively raised student tuitions (Mitchell et al., 2016) and increased the ratios of higher tuition-paying non-resident students (Jaquette & Curs, 2015) to compensate for state divestment. However, public and political pressure against such revenue-generating measures caused these HEI to seek alternate revenue sources such as federal grants (Haycock et al., 2010; Laderman & Carlson, 2017; Rizzo & Ehrenberg, 2004). As a culture of growing reliance on grant funding emerges at public research universities, research and tenure-track faculty, once held to research publishing and instruction performance standards, became increasingly held to a grant acquisition one (Musambira et al., 2012). This shift from a publish or perish to a grant or perish measure of performance manifests as the ability to obtain external funding became a core criterion for hiring and evaluating faculty (Musambira et al., 2012). Competition between universities for limited federal grant funds and reduced funding for federal agencies (i.e., a proportional decline in federal research funding relative to other sources such as private industry funding)(AAAS, 2019) created a need for FRD. As research productivity

becomes a standard measure of performance for faculty, FRD emerged at HEI as a field concerned with developing faculty research skills.

HEI invests in FRD programs on the premise of a positive ROI. However, examining current literature on the effectiveness, implementation practices, and return-on-investment of these programs highlights limitations. First, most studies examining these programs' effectiveness lack randomization or control measures for confounding factors, rendering the findings suspect about the program's actual effect (Fox et al., 2016; Pirracchio et al., 2012). Second, failing to account for the moderating influence of implementation fidelity (O'Donnell, 2008) per program guidelines can skew results (Durlak & DuPre, 2008). Third, a mere measure of program effectiveness (e.g., grant dollars acquired) that neglects to compare the total cost of provision of the program in a formal cost-benefit analysis cannot produce the necessary information to determine if the program was financially worth university investment (Levin et al., 2017). Consequently, a rigorous and comprehensive examination of each of these areas is crucial.

Purpose

The purpose of this dissertation is to introduce the CERTi Model and demonstrate its applicability in comprehensively assessing the efficacy of talent development efforts (e.g., FRD programs) in an HEI context. This model provides an innovative, comprehensive, and interdependent approach that combines quantitative, qualitative, and economic methodologies to advance talent management theory in a higher education setting. It presents a case study of a faculty FRD program for grant acquisition to illustrate the applicability of the evaluative framework for practice and scholarship. As HEI face an organizational environment characterized by declining public financial

support and an atmosphere wrought by accountability demands and increased requests for financial returns-on-investment, the CERTi approach is ever more critical for evaluating talent development programs' efficacy in advancing scholarship. Three main articles comprise the dissertation. The first assesses the program's effectiveness. The second investigates its implementation fidelity, and the third determines its ROI to answer the following research questions.

Research Questions

1. *Are faculty recipients of FRD programs more likely to increase their grant acquisition?*
2. *What can be learned from implementing FRD programs to improve their delivery and maximize their potential effectiveness?*
3. *What is the cost of implementing an FRD program, and do the financial benefits derived from the program worth the investment cost?*

Study Design

This study relies on a sequential explanatory observational study design. This design consists of two distinct stages: quantitative and qualitative (Creswell et al., 2003). Quantitative data will be collected and analyzed first, followed by collecting and analyzing qualitative data, explaining or elaborating on the first stage's quantitative results. First, quantitative data analysis (i.e., program effectiveness) will explain whether faculty recipients of FRD programs are more likely to increase their grant acquisition (i.e., the effect or lack thereof). Second, analysis of the qualitative data will help explain what can be learned from implementing these programs to improve their delivery and maximize their potential effectiveness (i.e., understand the results of the quantitative

analysis to recommend improvement). Third, the same qualitative data and their analysis will facilitate tabulating the total cost associated with the FRD program’s provision. Program total cost is subsequently compared to the program's total benefit (i.e., total grants acquired by all participants) to ascertain program return-on-investment (i.e., cost out the effect) (Rossman & Wilson, 1985; Tashakkori et al., 1998). This design illustrated in Figure 1.1 is lengthy and resource-intensive but valuable, especially when unexpected results may arise from the quantitative data (Morse, 1991).

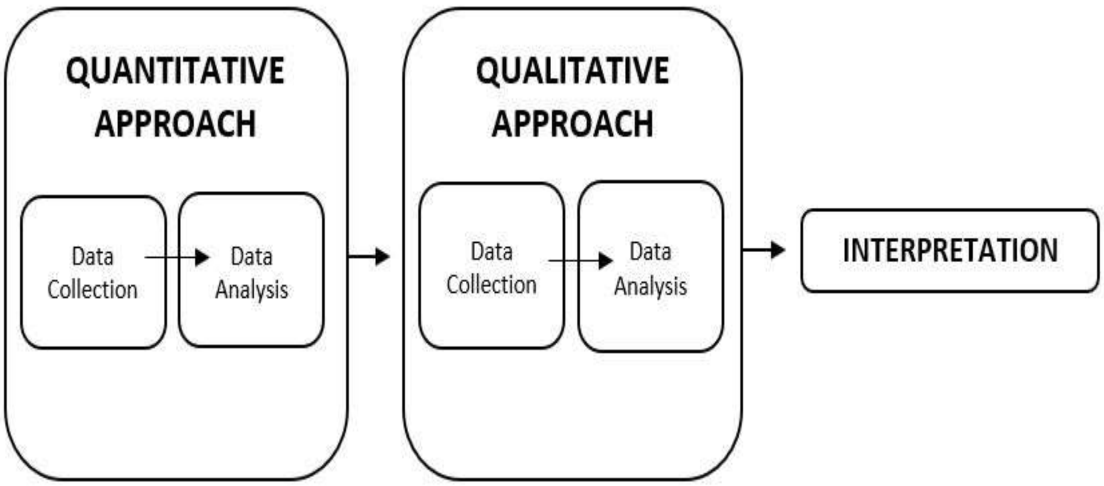

Figure 1.1 Study design

Chapter Two: Literature Review

HEI Financial Sustainability

A New Funding Norm

Although state appropriations for higher education institutions equated to $213 billion in the fiscal year 2019 (NASBO, 2019), state divestment in these institutions is evident from the percentage decline in state general fund expenditures towards higher education, dropping from 14.9% in 1990 to 10.1% in 2018 (Sigritz et al., 2018). Today, per-student spending remains $1,000 below pre-recession (i.e., 2008) levels and nearly $2,000 below 2001 levels (Laderman & Carlson, 2017), representing a 30% decrease over the past 30 years (Webber, 2017).

Traditionally, the bulk of public university funding has been state-appropriated, representing the third-largest priority for states after elementary education and Medicaid (Sigritz et al., 2018). However, this share has been steadily declining in recent decades (i.e., 1990-2018). Discretionary higher education funding has taken a backseat to the burgeoning cost of maintaining mandatory programs such as Medicaid (Hebel & Blumenstyk, 2014). In 1995, higher education spending as a percent of state general fund expenditures was 12.9%, while Medicaid was 14.4%, but by 2018 the share of higher education spending had shrunk to 9.7% while spending on Medicaid had expanded to 20.2% (Sigritz et al., 2018). Facing this new financial norm, characterized by state divestment, many universities employed cost-cutting and alternate revenue-generating strategies as they vied for financial sustainability. Universities engaged in diverse

expense reduction activities such as reducing programs, faculty/staff layoffs, and campus closures (Johnson et al., 2011). The two primary financial sustainability strategies employed by universities have been raising student tuition fees and increasing the enrollment of out-of-state students who pay higher tuition rates than their in-state counterparts (Curs & Jaquette, 2017; Hoover & Keller, 2011; Webber, 2017)

Tuition Hikes Financial Sustainability Strategy

Research shows that virtually all states have shifted tuition costs to students over the last 25 years (Mitchell et al., 2016). The average published in-state tuition and fees at public four-year institutions increased 3.2% per year beyond inflation between 2008 and 2017. Comparably, the increase is 4.0% between 1988 and 1998 and 4.4% between 1998 and 2008. In monetary terms, the 3.2% average annual growth rate in tuition and fees equated to $270 in 2017 dollars compared to $160 and $250 for the years spanning 1988 to 1998 and 1998 to 2008, respectively (Ma et al., 2017). The continuous rise in higher education costs has pushed many students and their families to near financial turmoil balancing between the importance of a college education to one’s success in society and the financial burden associated with it. While families have been forced deeper into debt, many continue to pay the increasing tuition rates as the benefits of a college education remain invaluable to both students and society. These benefits include upward mobility and personal development for the former and providing an educated workforce and economic stimulation for the latter (Association of Public & Land-Grant Universities, 2016; Christensen & Eyring, 2011; Tandberg, 2008). On average, over time, college graduates find it easier to secure jobs and earn more than their non-college-educated peers, creating an ever-widening income gap (Greenstone & Looney, 2012).

The ever-widening income gap between college and non-college graduates has increased from $4,000 in the 1980s to $12,000 currently, exacerbating the disparity and making attaining a college degree vital despite the cost (Greenstone & Looney, 2012). Moreover, the educational cost transference from states to students and their families could not have come at a worse time. As the cost of education has been on a steady incline, family income has been suffering stagnation for years. Mitchell et al. (2016) state that although tuition and income both grew at the same rate, by the end of the 1980s, tuition increased at a much faster rate. Parallel to this tuition increase, the only income level keeping up with tuition increases has been that of families representing the top 1% of the population. This shift especially walloped non-white families, who have historically had lower median income levels than their white counterparts. (Baum et al., 2014).

Historically, higher education institutions enjoyed a favorable rating among the American public. Driven by the perception that a college degree is crucial to societal success, people perceived universities as purveyors of the American dream. In communicating the public mood towards higher education institutions Callan and Immerwahr (2008) state:

> Colleges have lived a charmed life. According to the public-opinion studies that we have conducted over the past 15 years, this continued while many fields — athletics, accounting, politics — have lost the public's trust, but higher education continues to receive praise for its accomplishments, while criticisms usually fail to stick. (p. 1)

However, this perception seems to be eroding as noted, “We have also seen erosion in the public's appreciation of the altruistic mission of higher education.” They continue, “In our recent focus groups, we were surprised by how many people spoke of higher education as "a growing business" with "money coming in from everywhere. The

honeymoon may be slowly coming to an end” (Callan & Immerwahr, 2008, p. 2). The public's concern is that continued tuition increases would put a college education out of many students' reach.

The continuous rise in higher education costs has pushed students and their family's ability to pay for college to a near-breaking point, exacerbating student debt. Schoen (2015) highlights this by stating, “To make up the gap, millions of students and families every year are forced deeper into debt to make up the difference—around $100 billion a year is borrowed through a cottage industry of private and publicly-funded loan programs” (p. 2). This situation, in turn, exacerbates the public’s negative view of higher education institutions. Public dissatisfaction with higher college tuition is not a new phenomenon, yet student enrollment in colleges is steady despite the exponential increase in tuition, the financial burden on families, and burgeoning student debt (Laderman & Carlson, 2017). Rizzo and Ehrenberg (2004) explain that public universities' raising tuitions is not a popular move because it is perceived negatively, despite state divestment forcing them to do so. This predicament has pushed universities to seek alternate avenues of revenue, such as increasing the rates of out-of-state student ratios to take advantage of these students' higher tuition.

Out-of-State Tuition Financial Sustainability Strategy

Higher education institutions' second financial sustainability strategy to compensate for reduced state funding has been to increase out-of-state student ratios (Curs & Jaquette, 2017; Hoover & Keller, 2011). The benefit of such a strategy for universities is that it is generally unregulated by state governments, allowing them to continuously increase tuition rates to compensate for state divestment and pressure

against in-state tuition increases (Jaquette, 2017). State divestment, financial strain, and budgetary constraints contributed to a surge of nonresident students at higher education institutions (Jaquette, 2017) who pay a much higher tuition rate on average, $10,560 compared to $27,020 (Ma et al., 2020). However, even though this strategy benefited higher education institutions, just like the tuition increase strategy, it came with a cost.

The ramifications of this strategy are twofold. First, when examining out-of-state student enrollment, Haycock et al. (2010) and Curs and Jaquette (2017) warn of a crowding-out of talented in-state low-income students attending public flagship universities, as these institutions increasingly favor out-of-state students and the higher tuitions they pay. Jaquette (2017) mentions that "One primary consequence of raising tuition is a reduction in the number of high-achieving, moderate-and low-income students because financial aid has not kept up with tuition hikes" (p. 11). This situation, combined with state policymakers divesting in state institutions and setting a price ceiling on in-state tuition rates (e.g., 14 states as of 2018), has created a financial incentive for universities to seek out-of-state students (Kim & Ko, 2015; Pingel, 2018). The increase in out-of-state students who tend to be White or Asian and less likely to be Black or Latino, and more affluent than their in-state counterparts' is associated with a 2.7% point decrease in the share of Pell grant recipients. They add, "Thus, the aggressive shift towards out-of-state enrollment by many public research universities is associated with socioeconomic uniformity which, in turn, is associated with negative student development outcomes for moderate- and low-income students" (Jaquette, 2017, p. 18).

Second, Haycock et al. (2010) report increasing pressure on public universities to cease increasing out-of-state student ratios from state legislators, policy think tanks, and

national media outlets fearing insufficient access and opportunities for resident students, especially from within low-income non-white communities. Universities have argued that out-of-state tuition enables them to finance in-state enrollment in light of state divestment. However, Curs and Jaquette (2017) report that although out-of-state enrollment did not affect in-state enrollment at less-prestigious public flagship universities, it negatively affected in-state enrollment at prestigious public flagship universities (i.e., crowded out in-state students at those institutions). Specifically, they state that "out-of-state enrollment at prestigious public flagship research universities grew by 80 students per year on average from 2013-13 to 2014-15. "Our model suggests that these 80 additional out-of-state students crowded out 46 in-state students annually" (Curs & Jaquette, 2017, p. 661). Because of a fear of this trend, in recent years, some state systems have gone as far as enacting legislation limiting out-of-state student enrollment, causing universities to seek other revenue generation avenues.

Federal Grants Financial Sustainability Strategy

Federal Grants Addiction

One potential alternate financial sustainability strategy for revenue-deprived universities is for institutions to compete with their peers to obtain external grant funding to alleviate budgetary pressure. Because funds have become increasingly limited (Hourihan & Parkes, 2016; Howard & Laird, 2013; NIH, 2020), this competitive environment continues to exist in the present day. It has become the new normal for a growing number of these institutions (M. B. Chun, 2010). As a result, there is an increased demand for these funds from university administration, which trickles down to faculty increasingly pressured to secure them (Gallup & Svare, 2016). Although

universities seek all forms of external funding (i.e., private, foundation, local, state, and federal) in their efforts to compensate for state divestment, a particular focus on attaining federal grants is at the forefront of these efforts. This focus has manifested because federal grants provide indirect cost revenue to universities for the support of general administration and facilities cost incurred for research, as opposed to private foundation grants that seldom pay for such expenses (Ammons & Salterio, 1999; Canizares, 2008; Ledford, 2014; Noll & Rogerson, 1997; Sale & Sale, 2010; Zuiches & Vallely, 1987). Additionally, beyond their financial benefits, earning a competitive and nationally recognized grant, like those funded by the NIH, brings the institution prestige (i.e., national recognition, rankings, and publications). Prestige is vital because it is an institutional effectiveness metric for many research universities (Ali et al., 2010; Devine, 2009).

Growing dependence on federal grants has manifested itself at institutions of higher education. Rabovsky and Ellis (2014) state, “Traditionally, student tuition and subsidies from state governments have been the primary sources of funding for public universities in the United States, and as a result, the vast majority of the literature concerning higher education finance focuses on these two streams of revenue” (Rabovsky & Ellis, 2014, p. 741). However, they continue by explaining the growing significance of federal grant funding by adding, “In recent years, however, as costs at institutions of higher learning have increased faster than state governments can match, other revenue sources, such as funds related to research grants and contracts, have taken on an increased significance” (p. 741).

Cantwell and Mathies (2012) highlight an exponential increase in academic research expenditures that are supported by the federal government, “In the US, as in other countries, academic research expenditures have increased markedly over the past several decades” (p. 311). They also note that the bulk of “R&D support in the US comes from the federal government” (p. 312) and that the federal government sponsors 60% of all public university research funding, while the remaining 40% comes from state and local governments. Gallop and Svare (2016) highlight this developing dependence on these funds by public research universities and the subsequent pervasiveness of their revenue generation mindset, “because of a growing reliance on federal grants, awards from private foundations, alumni donations, tuition, student fees, tax revenues and proceeds from athletics, a focus on revenue generation has come to pervade the daily operations of higher education institutions” (p. 1). At the forefront of securing federal grants are university faculty increasingly pressured by administrators (e.g., provosts, research administrators, college-level deans, deans of research, and department chairs) to secure these funds.

A Pressured Faculty

Administrators at higher education institutions increasingly demand that faculty increase their research productivity, with one productivity metric being grant acquisition. Research addressing this trend cuts across all disciplines as faculty, who were not traditionally relied on to secure such funds, are now pressured to do so. In describing the pervasiveness of this trend, Gallop and Svare (2016) reported, “Decisions at many universities about hiring, term renewal, promotion, tenure, post-tenure review, merit

salary increases, and performance-based salary adjustments have come to emphasize individual differences in the ability to attract external funding" (p. 1).

They continue by pointing out that although "published research does far more to enhance the reputation of a department and a university than grant money, administrators continue to cling to grant dollars as the single most important criteria for decision making" (p. 1). Musambira et al. (2012) discussed growing pressure by administrators on communication program faculty to secure such grants and attribute the origin of this to the economic downturn and subsequent decline in state funding. Peck (2008) speaks of a climate in which grant obtainment is fast becoming a principal goal at public research universities, and attaining such funds is touted and celebrated to a greater degree than research and publishing. Speaking about mounting pressure on social work faculty, Thyer (2011) notes growing demands by university administrators for federal grant funding as revenue resources shrink. Finally, Cronan (2012) mentioned a "hyper-competitive grant climate" (p. 2) where increased pressure by administrators had pressed many faculty to seek federal grants. They also highlight how many new and junior faculty struggle to attain such funds in this hyper-competitive environment. This hyper-competitive environment they add can be best described as a grant or perish one for these faculty.

Grant or Perish

Increasingly, research grant attainment has become a measure of prestige as mentioned by Ali et al. (2010), "Research grants play a pivotal role in the development and dissemination of new knowledge; securing competitive research grants also enhances a faculty members' individual reputation, which in turn contributes to a positive perception of the academic institutions employing them" (p. 164). This pivotal role is

evident in that “Success among administrators nowadays is often measured in terms of their ability to garner additional financial support for the university” (Gallop & Svare, 2016, p. 1). They add, “Predictably, this mindset has meant that faculty members are increasingly evaluated based on their ability to secure externally funded research grants” (p. 1). Additionally, “Advertisements for faculty openings often feature the ability to obtain external funding as one of the principal criteria upon which prospective candidates will be screened, evaluated, and eventually hired” (p. 1).

According to Anderson and Slade (2016), the adage of “Publish or Perish” has shifted to a “Grant or Perish” one, for acquiring federal grants became the norm for faculty of all rank at public research universities. They further stated, “Our findings indicate that faculty time spent pursuing grants increases in response to pressure from administrative superiors” (p. 99). Increased pressure to pursue federal funding has made those funds more competitive as faculty compete with one another to secure them. This pressure is especially true for those seeking funding from the NIH, the predominant funder for public health schools at research universities, the context for this dissertation work. Furthermore, federal agencies are not immune to the economic conditions that caused a decline in state funding, and in recent years these funds have been on a steady decline, resulting in more competition between public universities for these scarce grant dollars, further pressuring faculty seeking these funds.

NIH Funding Woes

Research revenue is vital to both universities and their respective communities because it carries crucial social, economic, and societal benefits. Between 2007, the year preceding the great recession, and 2019, public research universities have seen a 26

percent decline in state investment (Association of Public & Land-Grant Universities, 2016). Complicating this is a corresponding decrease in federal funding for research despite a slight rebound in 2016, making the current funding model insufficient for sustaining financial viability for higher education institutions (American Academy of Arts and Sciences, 2016).

The NIH is one of the federal government's primary funding agencies to support research in many areas. Recent cuts in NIH budgets and decreases in state-allocated funds, coupled with expected increases in funding demand, are detrimental to public research universities and their respective faculty. M. B. J. Chun (2010) highlighted this by saying, “Due to the current economic crisis, research administrators at public universities are grappling with declining state funding and are faced with identifying other potential sources of revenue to support operations” (p. 77).

Despite comprising the most significant source for research funding for higher education institutions, rising to a high of 73% in the late 1960s, federal research funding has declined to around 55% presently (AAAS, 2019). Competition between universities for these limited federal grant funds, coupled with reduced funding for federal agencies, caused demand to exceed supply significantly as federal agencies witnessed a doubling of grant applications (Forero & Moore, 2016). This decline is especially true for the NIH, which has seen funding declines over the years from $40 billion in 2004 to $25 billion in 2018 in constant dollars (Hourihan & Parkes, 2016), despite a doubling in grant applications from 31,937 in 1997 with a 33% funding success rate to 67,496 in 2018 with a 22% funding rate (NIH, 2020).

Despite the decline in state support, the higher education industry has seen an increasing amount of overall spending to satisfy a growing diversity of needs, with that figure climbing to over $584 billion in the academic year 2016-17 from $296 billion in the academic year 2010-11 (Kena et al., 2016). The intense external grant funding competition has inspired some universities to strategically direct a proportion of their funds as an investment in FRD programs to give their faculty a competitive advantage in securing these funds.

Research Productivity

The decline in federal funding and consequent increased competition is symptomatic of both a push by the federal government for greater efficiency and productivity in the use of federal grant funds and university efforts to maximize faculty potential in securing grant funding. Forero and Moore (2016) explained this shift in funding agency priorities, “As research funding becomes less abundant and more competitive, it is more important than ever to focus on efficiency and productivity.” Explaining the reasoning behind this shift, they add, “This is because funding agencies want to see their limited resources have a bigger impact, and researchers need increased productivity to compete for highly-prized research grants” (p. 1).

Gallop and Svare (2016) add that “A disturbing corollary to this trend is that, when it comes to program development and resource allocations, decisions that affect academic matters are increasingly being made on the basis of a parallel, trickle-down, ‘what’s the return on the investment?’ mentality” (p. 1). They explain that as it relates to faculty, university administrators increasingly measure success financially rather than scholarly. They state, “And more and more frequently, those returns are being measured

in dollars and cents rather than good teaching, scholarly achievements, national prominence and academic excellence" (p. 1). Concluding, they explain that grant attainment impacts the university's bottom line by bringing in operating expense funding (i.e., indirect cost), "Importantly, among grants obtained by faculty members, those that bring in overhead are given the highest priority" (p. 1). These funds represent the expenses incurred for conducting the research, such as building, equipment depreciation, general administration, and maintenance expenses. These funds benefit the university, colleges, departments, and the principal investigator of the grant.

As a culture of growing reliance on grant funding emerges at public research universities, research and tenure-track faculty once held to a research publishing standard of performance find themselves increasingly accountable to a grant acquisition one (Reiser et al., 2015). The shift from "publish or perish" to "grant or perish" as a measure of performance is evident as the ability to obtain external funding becomes a core criterion for hiring and evaluating faculty (Musambira et al., 2012), begging a talent development approach that facilitates securing them.

Talent Management - Research Development

Talent management is an "integrative system that includes the selection, development, and retention of employees to meet an organization's end goals." (Rothwell et al., 2018, p. 818). It is a methodic, organized, strategic process of attracting talented employees, facilitating them to grow their optimal competencies while having organizational strategic objectives in mind. Talent development is a critical element of overall talent management. Many definitions exist for talent development. Caplan (2013) defined it as "a process that delivers capabilities that the organization needs, identifies

future capabilities, delivers team capabilities, creates innovation, inspires, seeks out people's ideas, [and] actively communicates" (p. 17). Hedayati Mehdiabadi and Li (2016) defined it as "a comprehensive system that consists of a set of values, activities, and processes with the aim of improving all willing and capable individuals for the mutual benefit of individuals, host organizations, and society as a whole" (p. 287). While Garavan et al. (2012) communicated that:

> Talent development focuses on the planning, selection, and implementation of development strategies for the entire talent pool to ensure that the organisation has both the current and future supply of talent to meet strategic objectives and that development activities are aligned with organizational talent management processes. (p. 6)

Faculty are increasingly becoming vital to public research universities' financial competitiveness. Because of their potential to generate revenue, institutions are beginning to invest in developing faculty as they navigate an ever-increasingly competitive federal grant environment (Forero & Moore, 2016). Boucher et al. (2006) confirmed this by adding, "The most important resource that any institution of higher education has is its faculty members. As such, faculty development must be considered an essential element in nurturing and supporting this invaluable resource" (p. 1). Consequently, recent years have witnessed an increase in faculty development programs spawning a whole field of RD that strategically approaches research faculty development to respond to the competitive federal funding environment (Evans, 2011b; NORDP, 2019). Along with their colleges and departments, public research universities have embarked on a path of developing their faculty's grant acquisition skills as they compete for federal dollars.

Faculty RD might also address equity concerns as the lack of grant acquisition experience, and expertise explicitly disadvantages new, junior, non-white, and women faculty in a declining and ever more competitive grant acquisition field (Freel et al.,

2017b). Research by Ginther et al. (2011); Waisbren et al. (2008) have shown that these groups submit fewer grant applications and ask for lesser funding amounts and years of funding than their white male counterparts. Beech et al. (2013); Ginther et al. (2011) found that it currently takes, on average, 4-5 years for new faculty to secure federal funding as compared to one year in 1980. In the same year, 16% of NIH grant awardees were 36 years of age or younger than only 3% presently (Alberts et al., 2014). Grant funding data from the NIH spanning from 2000 to 2006 showed that funding rates for African American faculty were 10% lower than their white peers (Ginther et al., 2011). Moreover, research has also detected a gender disparity in grant awards as funding for female faculty was lower than that of their male counterparts (41% versus 45%). Women faculty also received fewer years of funding, and on average, $27,000 less funding than their male counterparts (Waisbren et al., 2008).

As external grant acquisition competition increases between universities, RD, a field encompassing strategic and proactive approaches to attracting extramural funding, has emerged at research universities across the nation (NORDP, 2019). To support their RD efforts, universities began investing in FRD programs. One model increasing in popularity across research universities are cohort-based, peer-led FRD programs. These programs' design aims to leverage the expertise and experience of senior faculty (i.e., Mentors) with a successful track record of grant acquisition, to mentor new and junior faculty (i.e., Mentees) as they seek and apply for external grant funding (Van der Weijden et al., 2015). Duke University, The University of California at San Francisco, Dartmouth College, New York University, and Northwestern University are among the many institutions using this model. This innovative approach of strategic investment

towards RD efforts to secure institutional funding rests on the premise of a successful return-on-investment in the form of acquired research grant dollars. The next section delves into previous research that attempted to evaluate such research development efforts at HEI and highlights their inadequacies.

Evaluation of Research Development

FRD Program Effectiveness Evaluations

Paul et al. (2002) surveyed 189 junior and senior faculty from 20 nursing colleges to understand how mentorship affected research productivity (i.e., publications and grant acquisition). The survey gathered information on junior faculty roles and functions, research productivity, productivity factors (i.e., weekly hours spent on scholarly and research-specific activities such as grant submissions and acquisition), and institutional factors (e.g., demographic information, Dean/Chair support, grant writing seminars). Descriptive statistics analyzed participant demographic data and response rates, while ANOVAs and T-tests compared faculty's research productivity with and without a mentor in light of institutional support factors. Spearman-rho correlation analysis determined any significant relationships between institutional support factors and productivity scores. The results suggest that mentoring plays a crucial role in increasing junior faculty’s research productivity and shows a low-to-a-moderate positive correlation between this productivity and certain institutional factors. Although the findings support the positive effect of mentoring on research productivity, one cannot be sure if the mentorship truly affected the outcomes in the absence of randomization, a control group, and statistical methods to control selection bias and confounding influence.

Steiner et al. (2002) surveyed 215 graduates of the NRSA's Health Research Extension Act, which allocated federal funds to train primary care fellows to increase their research productivity to assess the impact of fellowship training on research productivity. The survey targeted those who participated in this program from 1988-1997 and achieved a response rate of 68% (n=146). Over 2/3 of the participants were full-time faculty, and 44% reported having funding as a principal investigator (PI) before the training. Participants reported whether they had a mentor during the training and acquired grant funding from any federal or non-federal source as a PI. A 2-group comparison using Mann-Whitney U or X2 tests to compare research productivity showed that participants who had influential mentors (n=64) were more productive in acquiring grants. Specifically, 43.8% of participants with mentors earned more than one grant post-fellowship compared to 23.4% of those who did not have a mentor. Among the limitations of this study are low sample size, a response rate limiting generalizability, self-reported data, and the fact that many of the grants received were small-in-scope project grants, compared to the much more desired large-scale research grants. Additionally, although this study identified mentorship as a predictor of research productivity, it failed to isolate the actual effect of mentoring on research productivity outcomes by employing causal estimation statistical methodologies in the absence of randomization.

Gardiner et al. (2007) evaluated a mentoring program for junior female academics in which they found that mentees were more likely to acquire grants as a result of mentorship. The study utilized a longitudinal design that evaluated the program at baseline (1998), at the end of the pilot phase (9 months in 1999), and several years later

(2004) to assess behavioral changes in mentees relative to not receiving mentorship. The study created a control group from a list of 46 female faculty that did not participate in the FRD program, which had similar academic standing (i.e., level B lecturers, expected to teach and conduct research) and length of employment to the mentees. Surveys collected grant acquisition data from the participants, and descriptive statistics illustrated the differences between the intervention and control groups. On average, mentees accumulated $41,896 in external grants over the six years compared to the control group, which averaged $14,647.

A close examination of participant selection in Gardiner et al. (2007) study reveals selection bias. Head faculty and senior staff identified a pool of potential mentees. Final study participants were selected based on Junior and less experienced status. "Faculty with a more established research career and those who stated they had no interest in research were less likely to be approached" (Gardiner et al., 2007, p. 431), denoting selection bias. Study authors created a synthetic control group from nonparticipating faculty based on the covariates of rank and experience. What they failed to do, though, is to account for other covariates (i.e., race, gender, tenure, grants, and publishing) that also influence the outcome per the literature (Ali et al., 2010; Conn et al., 2005; Ginther et al., 2011; Waisbren et al., 2008). Not accounting for those covariates renders causal estimation of treatment effect suspect.

Santucci et al. (2008) tracked three postdoctoral fellows and one junior faculty over 24 months to assess how mentorship influenced their research productivity (i.e., publishing, grant applications, presentations, and job talks) through self-reports. The participants reported involvement in 91 research projects that included publications

(61%), grant applications (24%), presentations (13%), and other (e.g., job talks; 2%). None of the participants reported any grant acquisition. The study lacked randomization of participants, a control group and neglected to utilize statistical methods to control for confounding and estimating causal effects. The study's main focus was on topics discussed, the structure of meetings, and the type of research productivity (i.e., publishing, grant applications, presentations, and job talks), which the study reported descriptively.

Evaluating what they call "Institutional investment in mentored research training for junior faculty," Libby et al. (2016, p. 1666) sought to address a gap in research development literature regarding grantsmanships' influence on junior faculty persistence (i.e., retention and progression in rank utilizing data from the Clinical Faculty Scholars Program (CFSP) at the University of Colorado Anschutz Medical Campus, an intensive interdisciplinary mentoring program spanning the years 2000-2011. The researchers matched a cohort of 25 junior faculty to a comparison group comprised of 125 faculty using time in rank and pre-period grant dollars awarded. A quasi-experimental difference-in-difference design was employed to identify the FRD program's effects on grant outcomes (i.e., grant counts and dollars). Although this study's results highlight an increase in annual counts and dollars of grants awarded for both groups, the mean yearly dollars increase for the observed cohort was significantly higher than the matched group (Δ$83,427 vs. $27,343, $p < .01$). The authors only controlled for two covariates, time-in-rank and pre-period grant dollars awarded with the previous study. They admit, "We had no access to individual or organizational variables such as gender, race/ethnicity, research training, years since training, department, start-up package, in-kind research resources

(e.g., laboratory or administrative access), or protected time for research.”. (Libby et al., 2016, p. 1674). This omission of these variables is essential to consider, given that each of them was predictive of grant acquisition.

Freel et al. (2017a) evaluated two faculty FRD programs at the Duke Universities’ School of Medicine in which senior faculty successful track record of grant acquisition mentored cohorts of their junior peers. The four-month programs aimed to increase junior faculty's grant success rates by developing complete grant applications during the four-month program. Each program was a 20-hour curriculum, including lectures, mentoring workshops, consultations, and the structured review of grant applications. Enrollment was limited to 24 junior faculty per cohort, and participant selection emanated from enrollment packets that included project abstract, specific aims, and a memorandum of understanding signed by the applicant, their department chair, and their primary research mentor. Final participant selection stemmed from the time of submission and completeness of applicant enrollment materials. Program designers deferred non-selected applicants to later cycles of the FRD program. Program participant surveys (n=197) with a 36% response rate garnered participant perceptions on their increased competency (e.g., designing and communicating research plans) before and after the mentorship and to assess their level of satisfaction with the program. At the same time, university-sponsored awards data records determined grant success rates. Descriptive statistics highlighted the program’s results. As with previously mentioned studies, this one did not employ any statistical method to control selection bias or account for extraneous factors that could confound results in the absence of randomization.

The CERTi Comprehensive Evaluation Approach

As the world grapples with the financial implications of the COVID-19 global pandemic, HEI, already under fiscal strain, is sure to reduce funding for talent development efforts and potentially eliminate FRD programs. Comprehensively assessing the efficacy and ROI of such programs is ever more crucial. Although past research evaluated FRD programs in terms of their effectiveness, implementation practices, and ROI independently, no model suggested addressing all three concurrently and simultaneously to assess these programs' worth comprehensively. The CERTi Model provides an innovative, comprehensive, and interdependent approach that combines quantitative, qualitative, and economic methodologies to advance adult Talent development theory in a higher education setting. Future work should empirically examine the viability of the model in the field setting and expand the model to include a talent-centered focus (Tran, 2020), which emphasizes the needs of employees (e.g., support, growth, satisfaction, engagement) and assesses the degree to which FRD programs meet those needs.

The proposed Comprehensive Evaluation of Return-on-Investment Model (CERTi) provides a holistic assessment approach by merging several evaluation frameworks from varying scientific disciplines into a comprehensive evaluation methodology that advances theory on adult talent development at HEI. CERTI's all-inclusive Macro-Micro assessment approach utilizes an overarching (i.e., Macro-level) adult-learner faculty-centric theoretical framework (Aziz & Tran, in press). Then, concurrently integrating (Micro-Level) qualitative, quantitative, and economic evaluations to assess FRD efforts at HEI jointly. The Micro-level approach employs the

Kirkpatrick (1994) four-stage Human Resource Development evaluation framework in assessing FRD program effectiveness (i.e., Quantitative Assessment). It then supplements that framework with Evans (2011b) RD conceptual model to explain what can be learned from the implementation of FRD programs (i.e., Qualitative Assessment) to improve their delivery and maximize their potential effectiveness. Finally, the model is extended by the principles of economic evaluations to methodically account for total program costs associated with FRD programs' provision compared to its total benefits/effectiveness to determine program ROI (i.e., Economic Assessment) (Levin et al., 2017). The following sections delve into the macro-level theoretical framework and micro-level evaluation approaches that comprise the CERTi Model. Each section systematically builds on the other, culminating in the interdependent holistic approach of the model.

The Adult Learning Model for Faculty Development

The theoretical framework underlying this study is the Adult Learning Model for Faculty Development. Lawler and King (2000) introduced this adult learner-centric model for higher education faculty as a four-stage process: Pre-planning, Planning, Delivery, and Follow-up. The model incorporates multidisciplinary adult learner-centric approaches from various scientific disciplines (i.e., adult learning, program development, professional development) to evaluate adult talent development effectiveness. This framework would “broaden and inform the perspectives of those responsible for faculty development in post-secondary institutions” (Lawler & King, 2000, p. 2). The model is grounded in adult learning principles such as “developing a climate of respect, utilizing collaborative modes of inquiry, building on participant experiences, learning for action, and cultivating a participative environment” (Lawler & King, 2000, p. 2). These

principles make this model a suitable theoretical framework for holistically evaluating FRD programs. Figure 2.1 illustrates the theoretical model and its four steps.

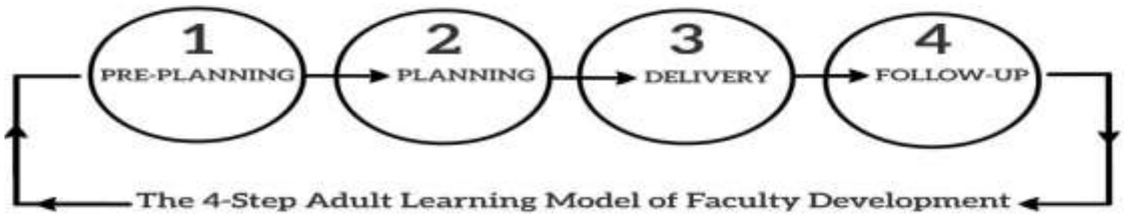

Figure 2.1 Adult learning model of faculty development

The model addresses three concerns relating to faculty development: first, the realization that faculty are "content experts" and sometimes "leaders" in their field. A faculty developer who is conscious and respectful of this fact stands to build a foundation of respect, facilitates them taking ownership of their development, and motivates them to progress through the development process. Secondly, faculty developers need to realize that adult learning and development do not happen in a vacuum. Research is rife with the influence of organizational climate (i.e., social, political, and financial context) on program and participant outcomes. An astute faculty developer must be mindful of these environmental factors that could facilitate or hinder learning and development. Lastly, most of those tasked with administering faculty development programs are not experts in training and development themselves. They are often faculty or administrators who have demonstrated a skill set or expertise that the leadership would like to see in their peers, hence tasking them to develop and administer development programs. These individuals should be conscious that faculty are usually suspicious of their capability to train them (Lawler & King, 2000).

Pre-planning—Program Design

The first step of The Adult Learning Model of Faculty Development entails understanding the purpose that underlies the development process, its relation to the institutional mission, and what resources must support the development effort. This stage focuses on faculty needs for learning and change versus what the institution deems those needs to be (Tran, 2020). Faculty characteristics, motivations to learn, change stimuli, and nuances of how faculty operate within their professional roles should undergird the pre-planning process. This perspective allows the faculty development program designer to address them better. The focus of the pre-planning stage is two-fold. The first entails paying close attention to organizational goals, needs, and the organization's climate. The second is learner-centric in that it assumes that the training's structure and goals of desired outcomes should originate with the faculty. Faculty development planners can accomplish this by asking a series of questions, performing a set of tasks per the model, keeping faculty needs in mind, and considering their experiences (Tran & Smith, 2020).

Faculty developers should ask the following questions in anticipation of the planning stage: "What is the purpose of faculty development? What is the purpose of this specific faculty development initiative? How is faculty development tied to the mission of the institution? What resources are available to support a faculty development initiative at this time?" (Lawler & King, 2000, p. 4). These questions are equivalent to a reflective process that facilitates scanning the organizational, social, and political context to plan the development program. Tasks associated with this stage are "Identifying the role of the faculty developer, assessing needs, evaluating resources, and establishing goals" (Lawler & King, 2000, p. 4).

Planning—Program Structure

This stage ought to be faculty-centric and is guided by the following questions, “What exactly is to happen? Who will be involved? How will it all be organized?” (Lawler & King, 2000, p. 4). Asking these questions allows for the building of a positive environment versus one characterized by fixing a deficiency. Tasks associated with this stage are selecting a topic, identifying a presenter, preparing for delivery, preparing for support and transfer of learning, scheduling the event, and beginning the evaluation. Although the pre-planning process is concerned with a faculty development program's overall direction, the planning stage’s concern is program structure. This stage “involves structured preparation for what specifically will happen during the program” (Lawler & King, 2000, p. 3) and is where developers will decide what will happen, who will be involved, and the organization of the entire program.

This stage should not operate under the assumption that a deficiency needs to be fixed, as is the case with many developmental programs; instead, it suggests a positive approach that values faculty members' input in the planning process. This learner-centric approach should consider faculty “needs, interests, experiences, and capabilities” (Lawler & King, 2000, p. 3). Of utmost importance in this stage is the choice of development topic, identification, and selection of the presenters (i.e., trainers), preparation for delivering the training, a preparation that facilitates learning, and structuring events that enable participation, and including evaluation in the process. Developers can utilize many formats in structuring faculty development programs such as institutionally supported self-teaching, cohort mentoring, peer-led mentoring, collaborative course design, workshops, online training, and quality assurance evaluation programs (Herman, 2012).

Among those, cohort and peer-led mentoring are most effective (Bryant, 2005; Mandzuk et al., 2005); programs that do not adopt these methods do not reach their full potential.

A cohort is a group of individuals persisting together in a program from its' beginning to its' end. In doing so, these individuals develop a sense of community as they grow throughout the learning process by navigating the learning environment, experiencing the same stimulus material, and managing the challenges of their work environment (Ashworth & Goodland, 1990; Mather & Hanley, 1999; Sapon-Shevin & Chandler-Olcott, 2001). Greenlee and Karanxha (2010) add that "A cohort group is acknowledged as being a distinct, interdependent group, markedly different from non-cohort groups" they add that a cohort "as separate learners proceed course by course with random groupings of other students. Cohort structures are a collegial support system to improve the teaching and learning process" (p. 358).

The benefits of cohort-based structures are far-reaching. Within the cohort structure research literature, Norris and Barnett (1994), Reynolds and Hebert (1998), Barnett et al. (2000) have linked cohort-based learning to positive attitudes towards learning, fostering critical thinking, and facilitating knowledge transfer. Most notably, among the benefits of these structures are that they yield better overall program outcomes. Twale and Kochan (2000) found that cohort members exhibited stronger feelings of belonging, confidence, and motivation toward group tasks than their non-cohort counterparts. Yerkes (1995) investigated the impact of cohort structures on engagement and cooperation and found that both increased due to the cohort-based learning environment. She also notes better socialization by cohort members in the

learning environment. Increased cooperation between cohort members was also confirmed by Barnett et al. (2000) in a later study.

Later research confirms the benefits of cohort-based learning environments. Lawrence (2002), Mandzuk et al. (2005), Maher (2005), Harris (2006), and Miller (2007) have all indicated the ability of cohorts to enhance members' learning experiences, foster meaningful professional and personal connections, increase learning motivation, reduce turnover in the face of challenges, and to facilitate program completion. Cohort models promote logistical support among groups of learners, give them a focused direction, provide a clear timeline, persistence in the face of challenges, a sense of community, increased social capital, competence, and higher confidence levels (Browne-Ferrigno & Maughan, 2014).

In combination with cohort design, university leadership has relied on internal talent by leveraging the knowledge, skill, and knowledge of senior faculty with a demonstrable record of securing extramural grant funding to mentor their less experienced peers. The peer mentorship approach is pervasive as its' benefits are apparent. Peer mentorship increases research productivity (Steiner et al., 2004), grant acquisition (Tsen et al., 2012), and overall performance among faculty mentees (Van der Weijden et al., 2015). Many definitions of mentorship exist, yet consensus remains elusive. Berk et al. (2005), examining the literature on mentorship, uncovered more than 20 definitions and, in describing their findings, state, "These definitions are extremely diverse, plus there is no professional consensus on any "acceptable" definition" (p. 66). Kram (1988) defines mentorship as a relationship in which a senior, experienced individual commits to providing developmental help and guidance to a less experienced

one. Files et al. (2008) mention, “Traditionally, mentoring has been thought of in terms of the dyadic model, in which an experienced mentor is paired with a less experienced mentee” Files et al. (2008, p. 1009). Waddell et al. (2016) said that mentorship is “an important strategy to help socialize new faculty to their roles and the expectations of the academic environment. It also helps them learn new skills that will position them to be successful in their academic career” (Waddell et al. (2016, p. 60).

Perhaps the most thorough analysis of the definition of mentorship has been detailed by Jacobi (1991). In examining her work, Berk et al. (2005) gave us a glimpse into what is agreed upon in the literature regarding mentorship by stating the following:

> Jacobi distilled five elements in the mentoring relationship on which there is general agreement. A mentoring relationship (1) focuses on achievement or acquisition of knowledge; (2) consists of three components: emotional and psychological support, direct assistance with career and professional development, and role modeling; (3) is reciprocal, where both mentor and mentee (aka protege) derive emotional or tangible benefits; (4) is personal in nature, involving direct interaction; and (5) emphasizes the mentor’s greater experience, influence, and achievement within a particular organization. (p. 66)

Mentorship in the higher education field is both prevalent and impactful. Lunsford et al. (2013) spoke of the impact of mentorship by stating, “First, mentoring activities are prevalent in college, and university work settings and costs to faculty have been understudied” (p. 127). For example, “in the USA alone, there are more than 1.7 million faculty members in higher education” (McFarland et al., 2018, p. 184). Keyser et al. (2008) state that research mentorship has been “increasingly recognized as an essential catalyst for providing researchers with the skills needed to advance successfully in their careers—and for fostering the highest environments within which researchers work, levels of research integrity and professional practice” (p. 217).

Mentees themselves acknowledge the benefits and impact of mentorship as Savundranayagam (2014) reported mentee sentiment regarding the issue, “Peer mentorship was an equally important component of the institute that contributed to my increased sense of self-efficacy regarding the grant writing process” (p. 273). Mentorship in the research field is beginning to manifest itself as demands for external federal grant funding increase. Administrators are demanding that faculty increase their research productivity and utilize RD techniques to develop them (Smith, 2016). They often rely on senior faculty's expertise with a proven track record of grant acquisition to mentor their peers into productive researchers capable of obtaining large-scale federal grants (Brutkiewicz, 2012; Freel et al., 2017b). The prevalence of FRD programs is evident as public research universities, colleges, and departments invest in peer faculty FRD programs to enhance research productivity (Raymond & Kannan, 2014; Van der Weijden et al., 2015).

Delivery—Program Implementation

The delivery stage is rooted in the assumption that successful program implementation facilitates future successful programs. Research has linked poorly implemented programs to a lack of program effectiveness (Pettigrew et al., 2015). Failing to account for the moderating influence of implementation practices potentially diminishes the program’s impact, resulting in a less than comprehensive evaluation of program outcomes. “Only by understanding and measuring whether an intervention has been implemented with fidelity can researchers and practitioners gain a better understanding of how and why an intervention works, and the extent to which outcomes can be improved” (Carroll et al., 2007, p. 1).

Implementation broadly refers to the “process by which interventions are put into action” (Graczyk et al., 2003, p. 306). Implementation research has yielded five components that influence implementation outcomes. These components, collectively referred to as “Fidelity of Implementation” (FOI), are Adherence, Dosage, Quality, Participant Responsiveness, and Program Differentiation, and are critical measures of implementation quality (IQ) (O’Donnell, 2008).

Adherence refers to how program components are delivered as prescribed by the model (O’Donnell, 2008). Program content, methods, and activities are regarded as indicators of adherence and typically reported as the proportion of program components delivered compared to the number prescribed. Dose or exposure refers to the amount of the program given compared to the amount specified by the designers (i.e., number of training sessions, attendance per training, and frequency/duration of each training) (Bradley et al., 2016). Quality refers to how a program is delivered. Quality is contingent on many aspects such as the deliverer’s level of preparedness, enthusiasm in providing the material, the use of pertinent examples, proper interaction with participants, mutual respect, and confidence in answering questions relevant to the topic. Poor delivery of program components (i.e., low quality) of an excellent program may result in poor participant outcomes (i.e., program effectiveness) (Carroll et al., 2007). Participant Responsiveness refers to the degree to which participants are stimulated by the program, how well it retains their interest, and how well they react to and engage in the program (Durlak & DuPre, 2008). The appeal, perceptions of relevance, enthusiasm, and engagement are all aspects of participant responsiveness that can moderate program adherence. Participant responsiveness has been compared to “Reaction Evaluation” in the

evaluation literature (Carroll et al., 2007). Program Differentiation, according to Dusenbury et al. (2003), is defined as “identifying unique features of different components or programs so that these components or programs can be reliably differentiated from one another” (p. 224). Differentiation of program critical components (e.g., training protocols and manuals) is vital to producing positive outcomes and is also known as component analysis. Differentiation is also critical when comparing interventions because if the critical components of two interventions differ, then comparing those two interventions' outcomes is confounded by this difference (Moncher & Prinz, 1991).

Lack of implementation fidelity has been linked to poor intervention outcomes by a considerable amount of research (Durlak & DuPre, 2008; Dusenbury et al., 2003; Lipsey, 2009; Mihalic, 2004). Fidelity is vital to the “validity of any intervention study and is closely related to the statistical power of outcome analyses. . . Failure to establish fidelity can severely limit the conclusions that can be drawn from any outcome evaluation” (Dumas et al., 2001, p. 39). Furthermore, a growing body of work suggests measuring fidelity to ensure internal validity, circumvent compromising external validity, and maximize statistical power (Chen & Rossi, 1983; Cook, 1998; Dumas et al., 2001; Maynard et al., 2013). Lastly, research (Lakin & Shannon, 2015) suggests that evaluating implementation fidelity leads to better future program design, leading to higher program fidelity.

Fidelity of implementation is contingent on the first two stages (i.e., pre-planning and planning) successful administration (i.e., implemented). The leading question developers should ask in this stage is, “Are we building on this preparation?” This

question precedes additional ones that include: "How do we effectively promote the program? How are adult learning principles implemented? How do we monitor the program?" Tasks associated with this stage are "utilizing all needs assessment, faculty input, and environmental scanning information, promoting the program, implementing adult learning principles, and monitoring the program" (Lawler & King, 2000, pp. 4-5). Lawler and King posited that successful programs bring forth more successful ones. They state that faculty will be the ultimate judge of the development program, "if the faculty find that the program meets their needs, is tied to their reward system, has meaning for their work and is delivered professionally and appropriately; they are more likely to be positive towards faculty development and change" (Lawler & King, 2000, p. 3).

Follow-up—Program Fidelity of Implementation

Undergirding the final stage is that many faculty development program developers stop at the "Delivery" stage and fail to follow up on their development efforts. This stage is rooted in continuous improvement and, as such, challenges the developer by asking the following questions: What is the evaluation plan? How will ongoing support be provided for what was learned? Moreover, as faculty developers, what can be gain from reflecting on our role in this endeavor? The principle underlying this stage is empowering the faculty to apply the newly acquired knowledge on the job. Knowledge application empowers faculty to change, which is foundational to development efforts. Lawler and King (2000) state:

> Learning does not end at the close of a seminar or workshop. Interest in the faculty's continued learning promotes a positive climate and promotes ownership and interest in future initiatives. Finally, we come to our role as a developer. Just as we ask the participants to reflect on the event and learning they attended, we too need to reflect on the entire process and the outcomes of

our planning. Such reflective practice will enable the developer to offer ever-improving faculty development programs. (p. 5)

Understanding an FRD program's structure is essential to assess its effectiveness. Comparing how the program was structured (i.e., intended) to how it was delivered (i.e., implemented) advances understanding of the gap between theory and practice. Additionally, "Understanding this gap provides all of the stakeholders of an intervention with the opportunity to use that knowledge not only to judge the merit of the program but also to improve the program in future iterations" (Stolovitch & Keeps, 2006, p. 1176).

Stolovitch and Keeps (2006) discussed a comprehensive six-step process for assessing program implementation fidelity. This process involves comparing a program's design with its implementation using logic models. A logic model is "a flowchart that summarizes key elements of a program: resources and other inputs, program activities, and the intermediate outcomes and end outcomes (i.e., short-term and longer-term) that the program hopes to achieve" (McLaughlin & Jordan, 2004). They holistically describe and illustrate how and why the desired change happens within a particular context. They map out the "missing middle" between what a program does (i.e., its' activities) and how these lead to desired goals (i.e., its' impact). They can also explain why the desired change did not occur (i.e., detail, why a program's full potential did not come into fruition). As depicted in Figure 2.2, logic models are flowcharts that summarize the program's key elements, such as 'Inputs,' which are the resources needed to operate the program (i.e., human, financial, organizational, or material). 'Activities' are the allocation of inputs or the events, while 'Outputs' are the activities' direct/immediate results. 'Outcomes' are the short-term, intermediate, and longer-term results of the program evidenced by specific changes in participant skills, knowledge, behavior, or performance,

and 'Impact,' which is the ultimate change to the organization resulting from the program. (McLaughlin & Jordan, 2004).

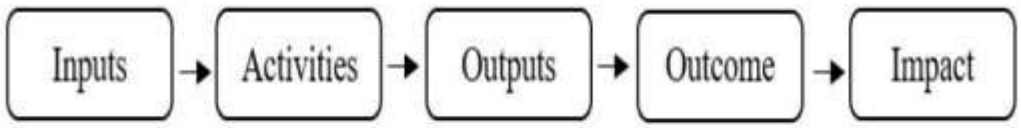

Figure 2.2 Generic logic model

Read from left to right, a logic model describes the program as it should work; inputs feed into activities that yield individual outputs resulting in specific outcomes and producing the desired impact. Read from right to left, the model describes the theory behind the program; creating an individual impact necessitates accomplishing particular outcomes resulting from specific outputs, emanating from critical activities, and requiring unique inputs. Understanding the theory of change underlying an FRD program is essential because it explains the linkages between activities and outcomes and how and why the desired change is expected to happen, based on past research or experiences. A logic model is essentially a graphic representation of change theory that illustrates the linkages among resources, activities, outputs, audiences, and short, intermediate, and long-term outcomes.

Stolovitch and Keeps' approach to comparing a program's design with its implementation progresses in five steps (p. 1177):

1. Develop a logic model representing the program-as-intended
2. Develop measures of key program indicators
3. Develop a logic model representing the program-as-implemented
4. Compare program-as-intended to program-as-implemented logic models

5. Improve the model

(1) *Program-as-intended logic model*. This step involves the documentation of inputs, activities, outputs, outcomes, and impact. The inputs are performance variables that the program aims to improve and are program-specific, emanating from the actual program's context.

(2) *Program Indicators*. This step entails the development of quantifiable indicators for the program-as-intended logic model. These indicators are derived for each of the logic model’s components (i.e., inputs, activities, outputs, outcomes, and impact) to compare to the model-as-implemented logic model in an attempt to ascertain program implementation fidelity.

(3) *Program-as-implemented logic model*. Once the program-as-intended logic model is developed, and key program indicators identified, a detailed collection and analysis of data will enable evaluators to develop a program-as-implemented logic model. This model represents the program as it happened, which may or may not resemble the program-as-intended logic model.

(4) *Program-as-intended to program-as-implemented Comparison*. This step allows an evaluator to compare the program as its designers intended it to be to how it was implemented in actuality to uncover incongruities. Stolovitch and Keeps (2006) state that the findings resulting from such a comparison may lead an evaluator to one of the following conclusions:

1. The program was implemented as intended and was successful: good planning, good implementation, positive result.
2. The program was implemented as intended and was not successful: poor planning, good implementation, negative result.
3. The program was not implemented as intended and was not successful: good planning, poor implementation, negative result.
4. The program plan was not clear, program implementation was poor, and the program was unsuccessful: poor planning, poor implementation, and negative results.

(5) *Program Improvement Model*. The evaluator utilizes this final step when comparing the program-as-intended and program-as-implemented logic models' results in an unsuccessful program. Using comparison data as feedback, the evaluator can develop a new and improved program that addresses the first model's shortcomings.

As mentioned, implementation practices moderate program or intervention outcomes, ensuring implementation fidelity is fundamental to producing positive results. Stolovitch and Keeps' methodic framework facilitates the assessment of implementation fidelity components (i.e., Adherence, Dosage, Quality, Participant Responsiveness, and Program Differentiation). Surveying and interviewing stakeholders and participants will provide insight into how they responded to the program and its quality. Finally, the process of developing and comparing program-as-intended and program-as-implemented logic models allows the evaluator to scrutinize the adherence and program differentiation aspects of program implementation fidelity.

Conclusion

The adult learning model provides an overarching (i.e., Macro) approach to guide an evaluation of adult learning efforts within a higher education context. It considers what happens before the learning (i.e., the development effort) and appraises the program following the development. It also provides questions and tasks that would guide the adult learning process. Lawler and King (2000) stressed that faculty development does not end when a particular development program ends; faculty development is an iterative process that aims for effective programs that produce a change in both thinking and behavior. They encourage faculty empowerment and creating a collaborative environment that facilitates continuous learning and knowledge application on the job after program delivery.

To ensure continuous improvement of development programs, they encourage talent development program designers to incorporate a rigorous multi-faceted evaluation to gauge program effectiveness and improve program delivery in the future. They suggest that the developers ask themselves the following questions: "What is the evaluation plan? How will ongoing support be provided for what was learned? What can we, as faculty developers, gain from reflecting on our role in this endeavor?" (Lawler & King, 2000, p. 3). Although they strongly encourage the use of evaluation methods that take into consideration faculty feedback about program delivery, they also stress assessing knowledge attainment and its' application once back on the job by stating:

> We encourage developers to use more than one method of evaluation to get an overall picture of not only the feelings of the faculty regarding the event but exactly what they have learned and how they can transfer that learning to their work. Analyzing the data and what it means provides not only feedback on the program but begins the needs assessment process for the continuation of development activities within the institution. (p. 5)

Based on this recommendation, the CERTi model embeds a three-pronged approach (i.e., quantitative, qualitative, and economic) within the Adult Learning Model for Faculty Development to comprehensively assess talent development efforts (e.g., FRD programs) (Aziz & Tran, in press). This multi-faceted approach is also the rationale behind this study's use of a sequential explanatory design. Quantitative data are given priority and will be used to determine program effectiveness. Qualitative data will subsequently aid in explaining the program's effectiveness or lack thereof. Finally, economic data will provide a measure of program return-on-investment. This approach relies on evaluation models from HRD, RD, and economic evaluations to achieve this, as discussed in the next section.

Human Resource Development Evaluation Model

Definition & Relevance

HRD is a field synonymous with evaluations of training, education, and development of employees. McLean and McLean (2001) defined HRD as the processes and activities that can develop "adults' work-based knowledge, expertise productivity, and satisfaction, whether for personal or group/team gain or the benefit of an organization, community, nation or ultimately the whole of humanity" (p. 324). The field of HRD is all-encompassing and focuses on the micro (i.e., training) and the macro (i.e., development) of employee learning to improve performance, facilitate behavior change, and achieve goals that align with a stated organizational mission.

HRD research has a far-reaching influence on all aspects of talent management. It includes human performance technology, action learning, needs assessments, career development, organizational commitment, organizational development, performance

appraisals, culture and diversity, talent development, training effectiveness, workplace learning, knowledge management, and leadership development (McLean & McLean, 2001). HRD aims to continuously develop an organization's human resources for better performance, provide development opportunities through needs-based training programs purposefully designed to maximize effectiveness, measure outcomes, evaluate interventions, and examine implementation. This field's rigor has resulted in HEI leadership employing many HRD evaluation models and frameworks designed to assess their program's effectiveness. The field produced many such evaluation models, none more cited than the Kirkpatrick model (Tamkin et al., 2002).

The Kirkpatrick Evaluation Model

Almost every mention of evaluation in HRD literature begins with the work of Donald Kirkpatrick. Kirkpatrick developed what is arguably the most well-known and used evaluation model in the field. Over two years (1959; 1960), Kirkpatrick published what came to be commonly referred to as the "four steps to evaluation" and is now known as the Kirkpatrick four-level framework. The four steps or levels are:

- Level 1 (Reaction): An assessment of participants' perceptions of the program's training favorability, engagement, and relevance to their jobs.
- Level 2 (Learning): An assessment of changes in knowledge, skills, attitudes, confidence, and participants' commitment based on their training participation.
- Level 3 (Behavior): An assessment of changes in job behavior resulting from the training program to gauge learning applicability.
- Level 4 (Results): An assessment of the targeted outcomes of the training program.

Kirkpatrick’s model is famously depicted as a triangle with the base being Level 1 (i.e., Learning) and the top being Level 4 (i.e., Results), as illustrated below in Figure 2.3.

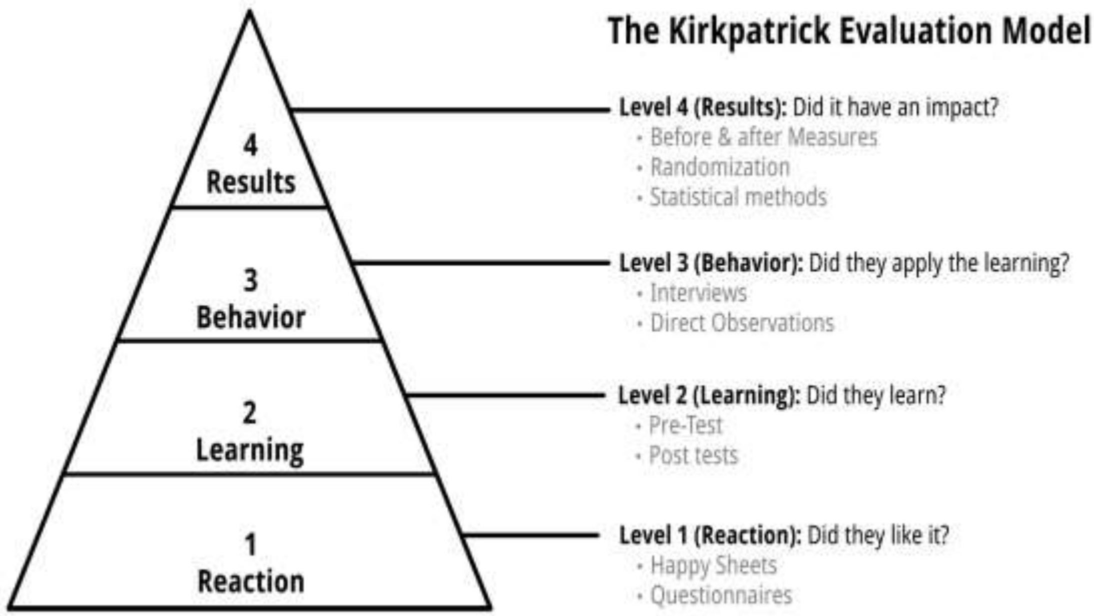

Figure 2.3 The Kirkpatrick four-level HRD evaluation model

Level 1—Reaction

According to D. L. Kirkpatrick (1998), the first level of his evaluation model (i.e., Reaction) assesses program favorability, engagement, and relevance to the job. This level constitutes a baseline measure of whether participants liked and perceived the program as valuable. HEI leaders can gauge participant satisfaction in this level using “Happy sheets” (i.e., surveys), feedback forms, verbal reactions, and post-training interviews. D. L. Kirkpatrick (1998) states that the aim that underlies the assessment of participant reactions to development efforts is to ensure motivation and immersion in the learning process. Kirkpatrick provided a guideline for implementing an evaluation of participant reactions that HEI leaders can utilize. First, determine the purpose (i.e., what do you want to find out). Design a form that will quantify reactions, attaining a rapid response rate of

100% if possible. Then develop an acceptable standard of measure, such as surveys or interviews. Reio et al. (2017) add that this is important because “positive reactions to a training program may encourage employees to attend future programs.” They add that in contrast, “negative comments about the program may discourage learners from attending and or completing the program. Both the positive and negative comments can serve to modify the program and to ensure organizational support for the training program” (p. 36).

Level 2—Learning

D. Kirkpatrick (1998) describes his second level as a measure of how much information participants retain during the training compared to learning objectives. HEI leaders should understand learning because it entails how the program affects participants in altering their approaches, enhancing their knowledge, and developing their skills. In essence, learning is “the degree to which participants acquire the intended knowledge, skills, attitude, confidence, and commitment based on their participation in the training” (Partners, 2018). Reio et al. (2017) state that “Kirkpatrick’s Level 2 is content evaluation, the examination of what employees learned as a result of participating in the training program” (p. 36). Kirkpatrick defined learning “as the extent to which participants change attitudes, improve knowledge, and or increase skill as a result of attending the program” (Kirkpatrick & Kirkpatrick, 2006, p. 22). Kirkpatrick expounded that, in essence, this level constitutes a shift in attitudes, increased knowledge, and promoted skills resulting from participating in the program and is usually assessed through pre-training and post-training knowledge tests. (D. L. Kirkpatrick, 1998).

Level 3—Behavior

Kirkpatrick's third level relates to the degree to which a learner applies the newly acquired knowledge on the job. D. Kirkpatrick (1998) points out that even though learning occurs, that does not always translate into action within real-world settings (i.e., on the job). Kirkpatrick adds that Level 1 (i.e., Reaction) and 2 (i.e., Learning) evaluation data can inform behavior level outcomes. He encourages the use of control groups, allowing for ample time to gauge behavioral change before evaluation and using surveys and interviews with trainees and trainers to garner behavior modification data. This level is critical for HEI leaders to comprehend because it measures the employee's actual job performance by gauging how the employee applies the newly acquired knowledge and skills when back on the job (Kirkpatrick, 1960). Kirkpatrick stresses that this level is critical. If employees do not apply the knowledge in Level 3 (i.e., behavior), the training effort cannot impact Level 4 (i.e., organizational results), which is of utmost importance for HEI leaders. Due to the complexity and time-consuming nature of evaluating behavior, Kirkpatrick and Kirkpatrick (2006) stated, "I believe that Level 3 is the forgotten level. Lots of time, energy, and expense are put into Levels 1 and 2 by training professionals because these are the levels that they have the most control over", while "Executives are interested in Level 4, and that is as it should be. That leaves Level 3 out there on its own with no one really owning it" (p. 83). Hence, HEI leaders should pay close attention to this level as part of their talent management efforts.

Level 4—Results

Kirkpatrick's fourth level seeks to link targeted outcomes and changes in performance to the development program itself. It aims to measure the degree to which

desired outcomes resulted from changes attributable to the application of knowledge acquired during the developmental program by participants once back on the job. This linkage proves to be challenging in the absence of randomization, which is why Kirkpatrick precisely suggested using before-and-after measurements, control groups, or statistical analyses to estimate program effects. This level is by far the most difficult to evaluate according to Reio et al. (2017) synopsis of the literature, “Level 4 is the most important and also the most challenging level to assess (D. Kirkpatrick, 1998; Kirkpatrick, 1960; Phillips, 1996; Werner & DeSimone, 2011).

Typically, at Level 4, organizations seek business results for their training efforts. At this level, organizations attempt to measure actual organizational changes due to training and place a monetary and or numerical value on those changes such as increased sales, reduced accidents, lowered turnover, decreased costs, or increased production (Kirkpatrick & Kirkpatrick, 2006). Alternatively, in HEI, this would manifest as grant revenue generated due to RD talent development efforts. According to McNamara et al. (2010), the results level contains an organization’s participants' ability to learn, alter, and improve performance according to talent development or training program-specific objectives. This level is the most challenging level to evaluate adequately, which is why the CERTi model provides a framework that HEI leaders can employ to causally link the results of their talent management effort to their talent development programs in the absence of randomization.

Limitations of the Kirkpatrick Model

Aside from their widespread use, HEI leaders should be cognizant of some limitations of level-based evaluation models such as Kirkpatrick’s model. The majority of

level-based evaluations only provide summative judgments of overall program effectiveness. Notwithstanding Kirkpatrick’s contribution to the field of assessment, the literature has criticized his model’s limitations. Guerci et al. (2010) criticized Kirkpatrick’s model for its excessive simplicity and for disregarding stakeholder needs, while Kaufman and Keller (1994) and Stokking (1996) lambasted its’ disregard for societal impact. Finally, Watkins et al. (1998) called into question its narrow focus. Nevertheless, Kirkpatrick's Model's simplicity and pragmatism have made it the most widely used model by practitioners in the evaluation field and the most cited in the evaluation literature.

Conclusion

This research aims to provide a comprehensive evaluation methodology of talent development on both the macro and micro levels. Kirkpatrick’s evaluation framework can provide organizational leaders with programmatic insight regarding such programs’ effectiveness by telling them ‘What’ happened (i.e., quantitative measures of participant reactions, knowledge attainment/transfer, and organizational impact). However, it fails to broaden their insight theoretically about the ‘why.’ Why did some participants have a favorable reaction to their counterparts' training vis-a-vis? Why did some participants attain the knowledge while others did not? Why were some able to apply what they learned when back on the job while others failed to do so?

Organizational leaders need to understand what underlies these attitudinal, intellectual, and behavioral changes to design talent development programs better. Additionally, Kirkpatrick’s model does not provide an economic assessment of program financial worth by neglecting to account for costs associated with the provision of such

programs and subsequently comparing that cost to the total benefit derived from them to ascertain whether they were worth organizational investment. This facet is specifically essential for HEI leaders in light of the new financial norm. HEI faces an organizational environment characterized by continued state divestment, limited federal grant funding, increased competition for those limited funds, and governmental demands for efficient public funding use. Therefore, supplementing Kirkpatrick's evaluation model with qualitative and economic evaluations provides a more comprehensive approach for HEI leaders to navigate this challenging organizational climate.

Research Development Evaluation Model

Research Development

The National Organization of Research Development Professionals (NORDP), a national peer network of RD professionals aiming to enhance multi-, inter-, and transdisciplinary research, defined research development as a set of "strategic, proactive, catalytic, and capacity-building activities designed to facilitate individual faculty members, teams of researchers, and central research administrations in attracting extramural research funding, creating relationships, and developing and implementing strategies that increase institutional competitiveness" (NORDP, 2019, p. np). Although it encompasses publishing, the main aim of RD is research productivity (i.e., grant acquisition); As (Evans, 2011b) explains, "Research development may be equated to capacity building for the purpose of increased output in the form of publications and successful funding bids, with the ultimate aim of increasing income" (p. 19) and is recommended for HEI by Mason and Learned (2006) to overcome financial challenges through the acquisition of external funding. RD, if conducted properly, ought to result in

attitudinal (i.e., mindset, attitudes, and perceptions), intellectual (i.e., knowledge, understanding, and competence), and behavioral (i.e., performance) changes culminating in the desired result (i.e., grant acquisition), (Evans, 2011b).

Evans RD Evaluation Model

Evans (2011b) summarized RD as the process “whereby people’s capacity and willingness to carry out the research components of their work or studies may be considered to be enhanced, with a degree of permanence that exceeds transitoriness” (Evans, 2011b, p. 21). She qualifies her descriptions by adding that researcher development “is not only about making researchers better at researching, but also about transforming into researchers, people representing other constituencies or who do not currently identify themselves principally as researchers” (Evans, 2011b, p. 20). A key term that she uses is ‘capacity.’ This term encompasses all human capital terminology such as skills, knowledge, attitudes, understanding, competence, and procedures and includes external factors such as resources, academic freedoms, and professional status.

Evans (2011b) categorized RD into Attitudinal, Intellectual, and Behavioral components. Figure 2.4 illustrates Evan’s emerging conceptual framework with its three main developmental components, and subsequent developmental changes potentially occur within each resulting from research development. This conceptual model catering to RD considers university faculty research development specifically, stating, “the academy judges by the theory and scholarship emerging from a particular field and discipline” (Evans, 2011b, p. 17). Since this field is emergent, a conceptual framework to guide research development evaluation efforts is necessary.

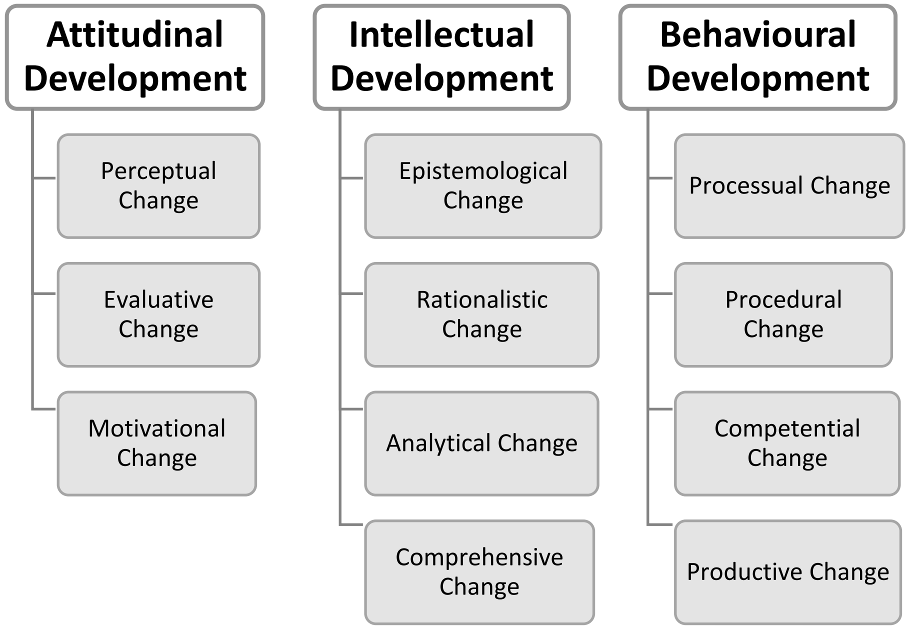

Figure 2.4 Research development conceptual model

Evans uses the term "modification" when describing research development and states that this modification should be "understood as ameliorative modification-change for the better, which constitutes what may be considered the enhancement of researcher capacity. It also means that the specific modificatory activity referred to in the three subsidiary definitions must be specifically research-capacity enhancement focused" (Evans, 2011b, p. 22). As aforementioned, Evans defined capacity as all-inclusive (i.e., skills, knowledge, attitudes, understanding, competence, procedures, external factors such as resources, academic freedoms, and professional status), encompassing professional and personal development. This model is appropriate for explicating the 'why' aspect of researcher development as she states, "My model can elucidate the complex multidimensionality of researcher development, and the relationship between the different dimensions and increased research activity and output" (Evans, 2012, p. 429). She explains in more detail:

> Used as an analytical framework, my conceptual model allows us to home in on individuals’ reported development experiences and identify the specific components and dimensions that evidently constituted the experience. More broadly, such analyses will allow us to identify patterns, trends, and atypicality and to identify causal links. It will allow us to identify which specific components and dimensions of researcher development occur most frequently, under what circumstances, and with what results. It will allow us to identify which specific components and dimensions evidently occur in the most effective researcher development experiences or opportunities, and which are evidently missing from the least effective ones. (p. 432)

Evans emphasized that the RD components' arrangement does not denote a hierarchical ranking, the organization in its form is only due to space limitations (Evans, 2012).

Attitudinal Development

Attitudinal development encompasses the process whereby researchers’ attitudes are amelioratively modified (i.e., enhanced for the better). Behavioral change is most effective when underpinned by attitudinal change (Evans, 2011b). This form of development represents a genuine commitment to change that transcends imposition, is not driven by compliance or lack of conviction but is driven by a genuine commitment to the development process. Attitudinal development’s subcomponents are perceptual, evaluative, and motivational change.

Perceptual Change refers to change concerning people’s perceptions, viewpoints, beliefs, and mindsets. It is concerned with “views about whether, for example, research should have relevance and usefulness and impact upon policy and practice; whether it should be “applied” or “pure;” whether it may—and should—be done by inexperienced and untrained amateurs/practitioners” (Evans, 2011b, p. 23). Perceptual change “relates, too, to perceptions of research as a component of one’s work, or a constituent of one’s professional identity; as such, it incorporates self-perception” (Evans, 2011b, p. 30). Perceptual change can manifest through external (e.g., the guidance of a mentor) and

internal (e.g., realizing one's naivete) factors. This change could lead to self-perception undergoing ameliorative change and also implicitly encompasses, among other things, people's beliefs, ideologies, self-conception, self-efficacy, and self-esteem. Evaluative Change is associated with a researcher's research-related values (i.e., what they consider important); "it means changes to people's research-related values, including the minutiae of what they consider important: that is, what matters to them about research and researching" (Evans, 2012, p. 428). Motivational Change revolves around a researcher's morale and job satisfaction relating to researching within their respective work environment. Evans defines it as a condition, or creation of a condition, "that encompasses all of those factors that determine the degree of inclination towards engagement in an activity" (Evans, 2000, p. 179). She stresses that this does not necessitate activity occurrence but merely the extent of inclination towards it.

Intellectual Development

Intellectual development encompasses the process whereby people's (i.e., researchers) "knowledge, understanding or reflective or comprehensive capacity or competence is modified" (Evans, 2011b, p. 22). Intellectual development includes four sub-components: epistemological, rationalistic, analytical, and comprehensive dimensions of change. Ameliorative modification in these areas can result in the necessary intellectual development to increase research productivity (i.e., grant acquisition) (Evans, 2011b).

Epistemological Change is a change in research-related knowledge structures. This shift means "change to the bases of what people know or understand about research and researching and to their research-related knowledge structures, as well as the

theoretical and conceptual frameworks within which they locate and undertake their research activity" (Evans, 2011b, p. 23). Epistemic beliefs affect how individuals deal with the essential requirements of a modern-day knowledge-based work environment, such as acquiring and evaluating knowledge. This change is imperative due to repeated findings stressing the beneficial effects of advanced epistemic beliefs. A prime example being the belief in weighing and evaluating knowledge claims (Kerwer & Rosman, 2018; Kienhues et al., 2016; Phan, 2008; Strømsø & Kammerer, 2016).

Rationalistic Change—relates to the nature of reasoning that researchers apply to their practice. It is equivalent to their thought process pre-and post-development. Bourke (1962) defined rationalism as a theory or methodology in which the criterion of truth is not sensory but intellectual and deductive. Put merely, rationalism gives high regard to the capacity of consciously making sense of things, establishing and verifying facts, applying logic, and adapting or justifying practices, institutions, and beliefs based on new or existing information (i.e., Reason) (Kompridis, 2000). 'Reasoning' is also associated with thinking, cognition, and intellect and includes logical, deductive, inductive, and abductive reasoning (Hintikka, 1973).

Analytical Change - "refers to the degree or nature of the analyticism applied to research-related activity" Evans (2011b, p. 23). What is meant by analyticism is the ability to conduct logical analyses and breaking down research into accomplishable parts. For example, developing and writing the various aspects of a grant proposal includes breaking up the grant proposal submission into parts per the funding agencies' requirements. Additionally, completing each required task, navigating the submission process within their institution, and abiding by the funding agency submission guidelines

are examples of analyticism. Comprehensive Change - Encompasses heightening or increasing research-related knowledge and understanding. It entails grasping new concepts that were not previously tenable. In doing so, comprehensive change alters perceptions and increases comprehension.

Behavioral Development

Behavioral development is the "process by which people's behaviour or performance are modified" (Evans, 2011b, p. 22). This component encompasses the physical act of researching all its forms and stages instead of the attitudinal and intellectual components, which only entail mental activity and includes four sub-components or dimensions of change: processual, procedural, competential, and productive. Behavioral development is most effective when underpinned by attitudinal and intellectual development, signifying genuine commitment on the developee rather than resulting from compliance, pressures, or imposition by the developer, reflecting a lack of conviction. In addressing the leadership of research development programs, Evans states:

> Research administrators need to understand that the effectiveness of initiatives aimed at increasing research productivity and output, at raising the quality of output by enhancing people's research skills, and at building and strengthening research cultures by making research a more prevalent feature of people's work (all of which are behavioural components of researcher development) will be dependent upon the extent to which, correspondingly, values are modified, perceptions are shifted (or widened), knowledge bases and structures are re-aligned, understanding is deepened, analyticism is increased, rationality is enhanced, and the motivation to participate and co-operate is heightened (all of which are attitudinal and intellectual components of researcher development). (p. 30)

A holistic approach undergirded by genuine concern for the developees' attitudinal and intellectual development should be central to HEI's research development efforts. However, HEI leaders' main concern is a behavioral modification that yields

desired results post-development (e.g., grant funding), which will only occur if attitudinal and intellectual development leads to ameliorative behavior modification. For this to happen, enhancement in the four behavioral subcomponents or foci of change (i.e., processual, procedural, competential, and productive) must occur. Processual change – relates to a researcher's processes in conducting their research practice (i.e., the various elements related to undertaking the research). Procedural change - revolves around a researchers' capacity to manage institutional procedures and can be referred to as "playing the game" (Evans, 2011b, p. 22). Competential change entails applying enhanced research-related skills and competencies due to research development on the job (e.g., writing, analytical, or presentation skills). This change entails the enhancement of skills (e.g., writing) "such as the development or refinement of writing, analytical or presentation skills" (Evans, 2012, p. 428). Productive change - deals with a research output resulting from the research development effort (i.e., grant application submissions). This change represents the result of the developmental process in which an individual's attitudinal and intellectual development manifests itself in the physical form as tangible evidence of the development program's effectiveness.

Research Productivity

Evans (2011b) postulated that favorable modification in these componential developmental areas would yield greater research productivity by modifying each component, which she states is understood as change for the better (i.e., enhanced research capacity). Evan's conceptual framework's developmental components, along with their 'foci of change,' provide a detailed accounting of why change might occur

during a developmental process. These developmental components are congruent with Kirkpatrick’s evaluation levels.

Replacing Kirkpatrick’s first and second levels (i.e., Reaction and Learning), which are quantitative assessments (i.e., surveys and pre-post test), with Evan’s first and second qualitatively assessed (i.e., participant interviews) developmental components (i.e., Attitudinal and Intellectual), provides for more insight into attitudinal and intellectual development between participants. This replacement allows HEI leadership to understand better why development occurs or not, versus just providing them with a summative assessment of what took place during the development process. The third level (i.e., Behavior) is where overlap occurs between Kirkpatrick’s and Evan’s models providing a combined quantitative and qualitative assessment. For example, HEI leaders can assess how many participants in an FRD program submitted a grant proposal to a funding agency. Parallel to this, they can interview program participants to delve into their behavioral development and assess why some submitted and acquire grants versus their peers that did not. The fourth level (i.e., results) is research productivity (i.e., grant acquisition) and is quantitative. HEI leaders can utilize propensity score matching statistical techniques (e.g., inverse probability treatment weighting, IPTW) to control selection bias and confounding variables to best estimate a causal link between the program and its outcome in the absence of randomization, as is recommended by Kirkpatrick. Figure 2.5 illustrates this combined qualitative/quantitative evaluation approach by combining Kirkpatrick’s and Evan’s evaluation models.

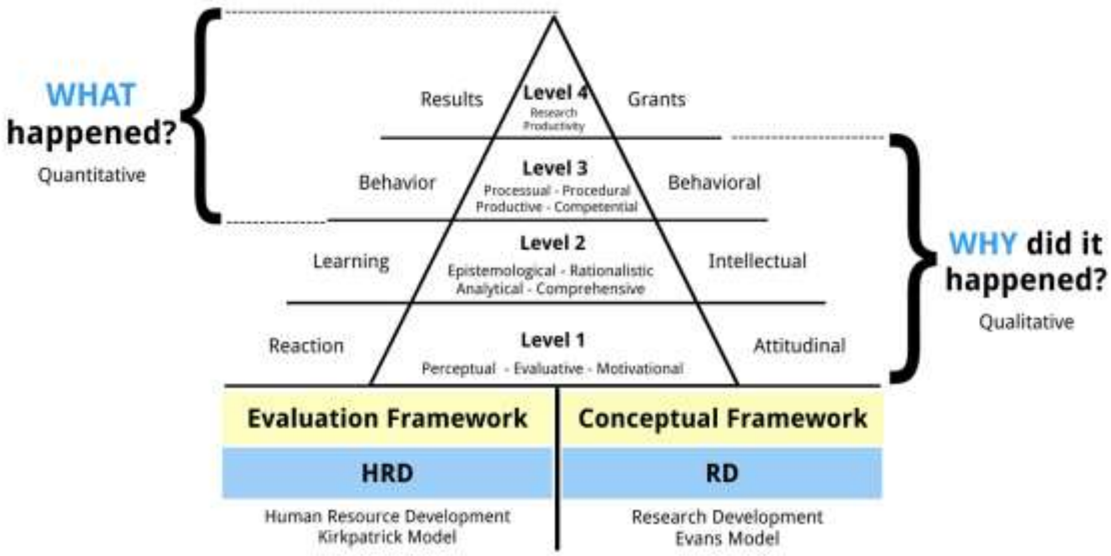

Figure 2.5 Kirkpatrick & Evans Combined frameworks

Conclusion

Research development is an intensive process that drains institutional resources such as faculty and administrator time and effort, hoping for a positive result that ultimately benefits the institution in extramural grant revenue. The two-pronged Kirkpatrick and Evans evaluation approach provides organizational leaders with a detailed account of what and why things transpired during a talent development program's implementation. However, it does not rise to the necessary level of a comprehensive evaluation based on external pressures such as the new financial norm facing higher education institutions, characterized by state divestment, mounting financial pressures, and demands of efficient use of public funds. Only by complementing the combined Kirkpatrick/Evans evaluation approach with a sound economic evaluation can organizational leaders attain an accurate measure of their talent development programs' return-on-investment. This three-pronged approach provides a comprehensive

methodological evaluation framework to assess talent development efforts' worth holistically.

Economic Evaluation Model

Definition

Economic evaluations combine knowledge from both economics and evaluation research to provide for better decision-making. Robbins and Robbins (1935) defined economics as "the science which studies human behavior as a relationship between ends and scarce means which have alternatives" (p. 16), while Levin (1975) stated that the purpose of evaluation research is to "obtain information that might be used to choose among alternative policies or programs" (p. 89). Drummond et al. (2015) defined economic evaluations as a comparative analysis between alternative courses of action in terms of both their costs (i.e., resource use) and consequences (i.e., outcomes).

The purpose of economic evaluations is to "inform decisions, so the key inputs to any economic evaluation are evidence about the effects of alternative courses of action" (p. 1). Levin et al. (2017) add that the purpose of economic evaluations is to help decision-makers make "better decisions" (p. 25). They state that fundamentally economics is the "study of the allocation of scarce resources, with an emphasis on the term "scarce" (p. 3). They emphasize that mere investigation of intervention effectiveness does not constitute an economic evaluation. Instead, economic evaluations incorporate the concept of scarcity by weighing the direct costs of interventions and the opportunity cost incurred in obtaining the effects of an intervention.

Cost Estimation

Levin and Belfield (2015) defined opportunity cost as “the value of what is sacrificed by using a specific resource in one way rather than in its best alternative use” (p. 403). They stress that although the assumption among decision-makers and evaluators is that cost information is readily available from budgets and business personnel, these methods are unreliable as a source for cost estimation because they fail to systematically account for all costs associated with the provision of programs and interventions. Furthermore, they neglect to account for opportunity cost of the value of what is sacrificed using a specific resource in one way rather than in its best alternative use. In contrast, the ingredients method of cost estimation is based on the economic principle of opportunity cost and provides more accurate cost estimations. This method begins with the assumption that all the ingredients (i.e., components) associated with programs or interventions have cost implications. Operating under this assumption, the ingredients method documents all resources utilized in the program or intervention, regardless of whether each resource has a budgetary cost or not, to fully capture the program's actual cost. The next step involves costing out each of those ingredients (i.e., matching each ingredient with its respective costs), which should not be confused with finance, as explained by (Shand et al., 2018):

> Costs differ from finance in the sense that finance deals with the way the costs are paid for and who pays for them. For example, consider an education program that relies on volunteer time for its implementation. Volunteer time is a resource that is necessary for the implementation and is necessary to achieve the impact the program generates. This resource will not appear in any budget or financial analysis, as it is a resource borne by the volunteer. However, if one were to replicate the program elsewhere, where there was no availability of volunteers, one would need to hire workers to replace volunteer time. Therefore, restricting costs only to those accounted for in the budget would

> understate the program's overall costs because they do not include costs borne by other sources. (p. 14)

The method concludes with the calculation of total program or intervention costs and, in doing so, provides evaluators with a proper accounting of the cost of each alternative (i.e., program or intervention) to conduct their economic evaluation of choice.

Economic Evaluation Approaches

There are multiple types of economic evaluation approaches. For example, *Cost analysis (CA)* - calculates all resource costs associated with implementing an intervention, including personnel, facilities and utilities, travel, materials, and supplies. These costs are essential for determining who incurs the costs, such as the program itself, participants in the intervention, or external community resources. Programmatic cost analyses include both financial costs that appear in a budget and economic costs that are in-kind services. According to Shand et al. (2018), conducting a cost analysis that relies on the ingredients method as part of a broader evaluation framework can provide program evaluators with the answer to the question of what it takes to implement a program to achieve observed results and sets the stage for comparative analyses (e.g., Cost-Effectiveness Analysis) (p. 10).

Cost-feasibility Analysis (CFA) - A cost-feasibility analysis measures a program's practicality given a specified budget or when significant investment is at stake. This type of analysis allows for eliminating non-feasible options before evaluating outcomes (Levin et al., 2017). This type of economic evaluation helps determine if a program is technically feasible to undertake within an estimated cost and if it will be profitable. Evaluators often employ cost-feasibility analyses when large sums of money are at stake.

Cost-Effectiveness Analysis (CEA) - Cost-effectiveness analysis, according to Levin and McEwan (2000), refers to "the consideration of decision alternatives in which both their costs and consequences are taken into account in a systematic way" (p. 381). Put merely, CEA is a decision-making tool that aims to establish which alternative is the most efficient. This economic evaluation compares the relative costs of the outcomes (effects) of two or more courses of action.

Cost-Utility Analysis (CUA) - Levin et al. (2017) define cost-utility analysis as "the evaluation of alternatives according to a comparison of their costs and their utility (a term that is often interpreted as value or satisfaction to an individual or group)" (p. 16). In essence, CUA weights outcomes based on the decision maker's preference. This type of economic analysis allows one to compare two different interventions or programs whose benefits may differ.

Benefit-Cost Analysis (BCA) - Boardman et al. (2017) characterize BCA as an analytical tool that compares alternatives based on the differences between their costs and a monetized measure of their effect. Essentially, this type of analysis monetizes program benefits and compares them to its' cost to determine an intervention's return-on-investment. Benefit-Cost analyses evaluate all potential costs, including opportunity cost (i.e., the forgone missed opportunity resulting from a choice or decision), which allows program evaluators to weigh the benefits from alternative courses of action, not merely the current path or choice considered in the analysis. Because the outcome is monetary, this study utilizes a Benefit-Cost analysis as the economic evaluation of choice. This method will produce the necessary information to gauge whether the FRD program examined is worth the university investment. It compares the program's benefit (i.e., total

grant dollars acquired by participants) to its' total cost of provision to determine its return-on-investment.

Conclusion

Supplementing the combined Kirkpatrick/Evans (i.e., quantitative/qualitative) evaluation approach with a sound economic evaluation based on the economic principle of opportunity cost and the ingredients method increases an evaluation's rigor. To this end, adding a fifth step to the Kirkpatrick/Evans model to incorporate an economic evaluation of talent development programs, as illustrated by Figure 2.6, provides for a holistic assessment of talent development efforts such as FRD programs at HEI. This three-pronged micro evaluation approach provides organizational leadership with resolutions to "What Happened?" during the program, "Why did it happen?", "How much did it cost?" the institution, and finally, "Was it worth it?" as illustrated in the figure. For example, in applying the CERTi model to evaluate an FRD program at HEI, levels 1-3 of the evaluation model qualitatively (i.e., through interviews with FRD program designers, participants, and institutional records) assess participant attitudinal, intellectual, and behavioral change. Levels 3 and 4 assess program results and impact quantitatively (e.g., grant submissions and acquisitions). Lastly, a comparison between the program's total cost (e.g., salaries, fringe, facilities) and its' total benefit (i.e., total grant dollars acquired by mentees) via data gathered during the same participant qualitative interviews would facilitate ascertaining its' return-on-investment.

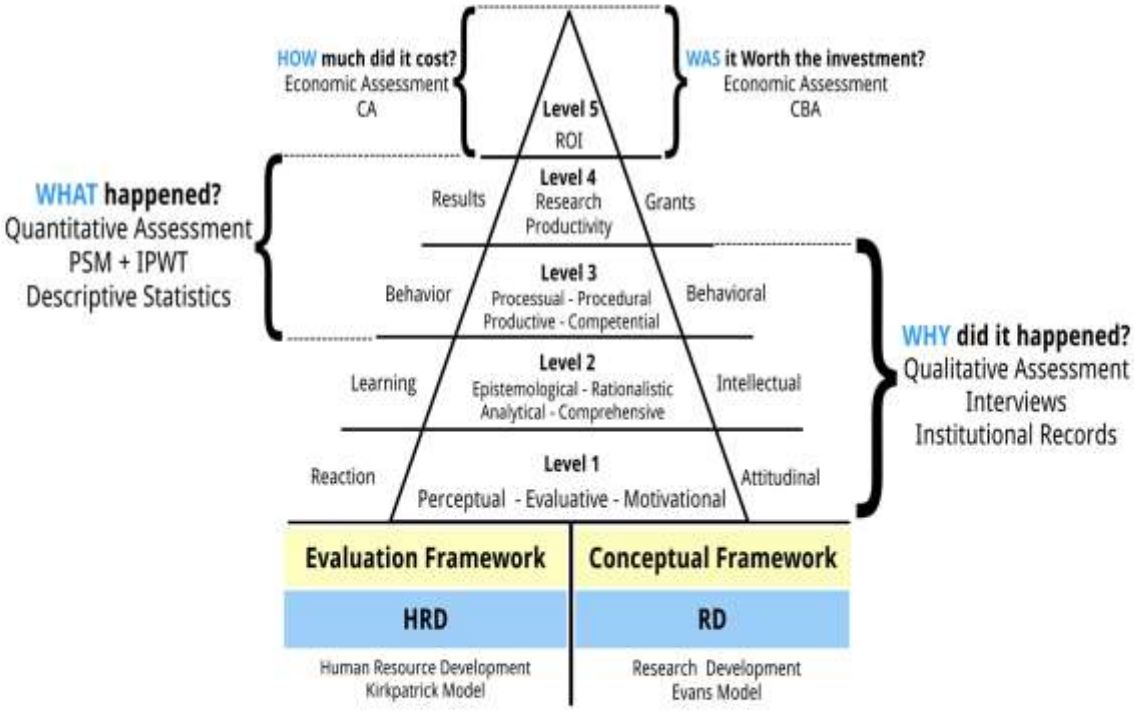

Figure 2.6 The Comprehensive evaluation framework

The Big Picture

The data and ensuing analyses from this three-pronged approach (i.e., Micro-Level) provide a realistic depiction of what transpired during the program's implementation, providing HEI leaders with a comparison between a talent development program's design and its' implementation in actuality to uncover incongruities. CERTi utilizes a five-step process mentioned by Stolovitch and Keeps (2006) that uses Logic Models (LM) as a systematic approach to operationalizing CERTi's macro-micro approaches to assess FRD program efficacy comprehensively. The approach consists of 1) Developing an LM representing the program-as-intended, 2) Identifying measures of key program indicators, 3) Developing an LM representing the program-as-implemented, 4) Comparing program-as-intended to program-as-implemented LM, and 5) Improving the program.

The findings resulting from this comparison may lead to one of the following conclusions; 1) The program was implemented as intended and was successful; good planning, proper implementation, positive result, 2) The program was implemented as intended and was not successful; poor planning, proper implementation, a negative result, 3) The program was not implemented-as-intended and was not successful; good planning, poor implementation, a negative result, 4) The program plan was not clear, poorly implemented, and was not successful; poor planning, poor implementation, negative result. Comparison data would then provide HEI leaders with rich feedback to develop a new and improved program that addresses the first program's shortcomings. Figure 2.7 illustrates both macro and micro-level evaluations comprising the CERTI model.

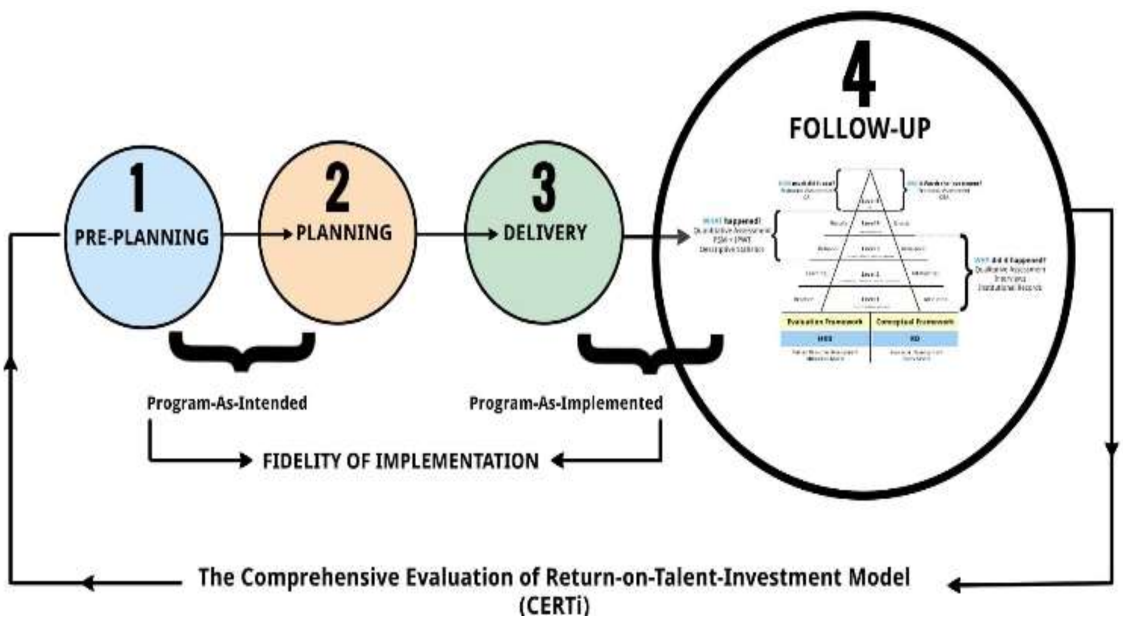

Figure 2.7 The CERTi model

Terms

For this study, several terms require definition.

1. *BCA*: Benefit-Cost Analysis is an analytical tool that compares alternatives based on the differences between their costs and a monetized measure of their effect.
2. *BCR*: Benefit-Cost Ratio is an indicator used in Benefit-Cost analyses that summarizes the overall financial value of a project relative to its cost.
3. *Cohort*: A group of individuals persisting together in a program from its' beginning to its' end. In doing so, these individuals develop a sense of community as they grow throughout the learning process by navigating the learning environment, experiencing the same stimulus material, and managing the challenges of their work environment (Ashworth & Goodland, 1990; Mather & Hanley, 1999; Sapon-Shevin & Chandler-Olcott, 2001).
4. *Cost Analysis*: A calculation of resource costs.
5. *Evaluation*: "Evaluation is the means to ascertain the worth or value of a performance improvement initiative. It can be used to improve a performance-improvement process or to decide to discontinue the effort. It is also useful in judging the relative worth of performance-improvement alternatives" (Stolovitch & Keeps, 2006, p. 25).
6. *Economic Evaluation*: "A comparative analysis between alternative courses of action in terms of both their costs (i.e., resource use) and consequences (i.e., outcomes)" (Drummond et al., 2015, p. 89).
7. *HRD*: Human Resource Development is "the processes and activities that can develop …adults' work-based knowledge, expertise, productivity, and

satisfaction, whether for personal or group/team gain or the benefit of an organization, community, nation, or ultimately the whole of humanity" (McLean & McLean, 2001).

8. *IPTW*: Inverse Probability Treatment Weighting is a statistical method that uses propensity scores (i.e., probability of treatment) to create pseudo-populations to reduce or eliminate confounding or selection bias.

9. *NPV*: Net present value; represents the discounted (i.e., present) value of the benefit minus the discounted (i.e., present) value of the costs.

10. *PS*: Propensity score is the probability of treatment assignment conditional on measured baseline covariates.

11. *PSM*: Propensity Score Matching is a statistical method used to reduce or eliminate selection bias and move towards more causal estimates.

12. *Mentor*: Senior (Professor) rank faculty.

13. *Mentee*: Junior (i.e., Clinical, Assistant, or Associate) level faculty.

14. *National Institutes of Health (NIH):* U.S. federal health sciences funding agency.

15. *Peer Mentorship*: A relationship in which a senior), experienced faculty commits to providing developmental help and guidance to a less experienced faculty (Kram, 1988).

16. *Research Development (RD):* "encompasses a set of strategic, proactive, catalytic, and capacity-building activities designed to facilitate individual faculty members, teams of researchers, and central research administrations in attracting extramural research funding, creating relationships, and developing and implementing strategies that increase institutional competitiveness" (NORDP, 2019).

CHAPTER THREE: METHODOLOGY

The Case

This dissertation presents a case study of an FRD program for grant acquisition to demonstrate the CERTi model's applicability in comprehensively assessing the efficacy of talent development efforts (e.g., FRD programs) at HEI. This section provides an overview of this FRD program (i.e., the case) to illustrate the CERTI model's applicability for practice and scholarship and provides methodological context for the dissertation's following chapters. It begins by contextualizing organizational financial context and the reasons leading to the program's creation and describes its design, timeline, and participants. It then elucidates the rationale undergirding the study's site and participant selection criteria and justifies both. Lastly, it provides context relating to the researcher's reflexivity (i.e., subjectivity and positionality), discusses study trustworthiness and validity strategies, and summarizes each articles' methodology.

Institutional Context

Facing state divestment in HEI and a recent decline in grant acquisitions, HEI leadership at the College of Public Health at a Southeastern R1 research-intensive university (i.e., university engaged in the highest research activity levels) implemented an FRD program to increase its faculty's grant acquisition skills. The college faces a leveling of federal grant acquisition due to reduced grant submissions and a decline in funded grant proposals. The program relies on senior level (i.e., Professor and Associate rank) faculty with a demonstrable record of grant acquisition to mentor a cohort of their junior

level (i.e., Research, Clinical, Assistant, and Associate rank) counterparts. The program was one year in length, and its completion coincided with federal agency proposal submission deadlines culminating in grant proposal submissions to that agency.

The college, which administered the FRD program, is housed at a Carnegie classification R1 research-intensive (i.e., university engaged in the highest levels of research activity) southeastern university. State divestment in higher education institutions is evident at this university (A 12.1 decrease as a percent of the total budget from 2005-06-2019-20). The university has primarily compensated for this decline by increasing student tuitions (a 16% increase as a percent of the total budget from 2005-06-2019-20). Table 3.1 highlights this relationship, showing a decline in external grant revenue at the university for the same period. Federal grants comprised 19.2% of the university's budget in 2005-06, while the current figure stands at 11.5% (i.e., 2019-20). State grant revenue as a percent of the total university budget has also declined over the same period, from 6.3% to 0.6%, while local grant revenue has been steady at 0.1%. Although private and foundation grant revenue at the university has increased from 3% to 11.9% over the same period, an 8.9% increase does not compensate for the combined decline in federal and state grant revenue, which stands at 13.4%. Additionally, private and foundation grants seldom provide indirect cost revenue, as is the case with federal grants.

Table 3.1 Revenue type as a percent of university total budget FY 2005-2019

Type of Revenue	2005-06	2019-20	Change
Tuition	33.2%	50.1%	+16.9%
State Appropriations	23.7%	11.6%	-12.1%
Federal Grants	19.2%	11.5%	-7.7
State Grants	6.3%	.6%	-5.7
Local Grants	0.1%	0.1%	0
Private Grants	3%	11.9%	+8.9%
Educational Sales	2.3%	3%	+0.7%
Enterprise Sales	12.2%	11.2%	-1.0%

College Context

The college that administered the program employs over 155 faculty who published 577 peer-reviewed journal articles in 2017 and secured $31 million in funding awards, of which $24 million was external grant research funding. The NIH, the predominant funder for public health schools, funded $12 million of these external grants. The college focuses on attaining federal grants because they provide indirect cost revenue for the support of general administration and facilities cost incurred for research, as opposed to private foundational grants that seldom pay for such expenses (Ammons & Salterio, 1999; Canizares, 2008; Ledford, 2014; Noll & Rogerson, 1997; Sale & Sale, 2010; Zuiches & Vallely, 1987). Moreover, earning a competitive and nationally recognized grant like those funded by the NIH brings the institution prestige, an institutional effectiveness metric for many research universities (Ali et al., 2010; Devine, 2009).

The college secures a large sum of funding through grant acquisition via its faculty's grant submissions. This funding is crucial for the financial well-being of the college in a time when state divestment in higher education is evident (Mitchell et al., 2015). Grant acquisition is an expectation communicated to every faculty by the college and emphasized by its' leadership. Faculty job advertisements highlight this expectation, position descriptions explicitly communicate it, and the college expects faculty candidates to present a working plan to acquire these funds during their job interview presentations. The college also provides financial start-up packages to new faculty to launch research projects that they hope leads to efficacious grant proposals to funding agencies.

The college faces a leveling off of federal grant funding (i.e., $19.9 million in 2014, $20.8 million in 2018, and $21.2 million in 2019), raising concern among the college's leadership. Especially worrying is a trend highlighting a leveling off in NIH grant submissions and a decline in funded grant proposals. Lastly, NIH grants' grant funding success rate at the college is substantially below the national average, 9% versus 22%, respectively (NIH, 2020). NIH grant proposal submissions declined from 2016 to 2017, and grant acquisitions are half of what they were in 2013. The faculty submitted 62 NIH grant proposals in 2015, 69 in 2016, and 67 in 2017. The NIH funded eleven of the 62 proposals in 2015, six in 2017, and four in 2017, as illustrated by Figure 3.1

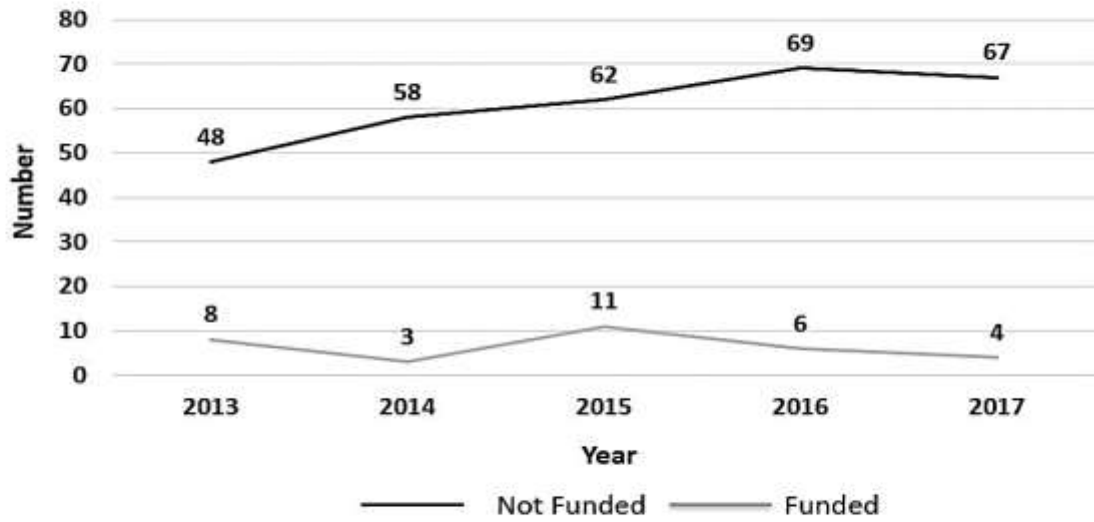

Figure 3.1 College Grant proposal submissions and acquisitions

The decline in grant proposal submissions and acquisitions led the college to implement an FRD program to reverse the trend. The cohort-based peer-led mentorship program aimed to develop new and junior faculty's grantsmanship skills (i.e., grant opportunity identification, proposal development, budget development, and grant submission) to increase their research productivity (i.e., grant acquisition). The dean's office funded the program, while the office of research administered it. The leadership chose seven senior rank faculty with a demonstrable record of grant acquisition to mentor their less experienced peers. Two of the mentors, the dean of research and a professor rank faculty, were tasked with planning and administering the program. Twenty-two junior-level (i.e., clinical, research, assistant, and associate) faculty comprised the cohort of mentees. The research office developed the program under the supervision of the dean of research and program director by benchmarking a successful FRD program at another university.

Program Design

Based on the financial benefits derived from federal grants specifically (i.e., providing indirect cost revenue to universities to support research administration and facilities costs) and the prestige associated with acquiring such grants, this FRD program's focus was acquiring NIH grants. Specifically, the focus was on the acquisition of large-scale R03, R21, and R01 grants. This focus emanates from the fact that these are long-term and high-dollar grants that align well with the types of research conducted by faculty in public health schools. On average, R03 grants carry $50,000 in direct costs budget and span up to two years of work, R21 grants carry up to $275,000 in direct costs and span up to two years of work, and the R01 level grants are $500,000 or more in direct costs and can span up to five years (NIH, 2019). The program's stated purpose is to "leverage the school's skills, knowledge, and expertise to maintain and increase institutional competitiveness." The advertisement also mentions the "competitive NIH funding climate" at the root of administering the program. The program announcement stated that it was a "strategic, proactive, and capacity building investment in one of the school's most valued assets, the faculty" that is guided by a cohort-based and peer-led mentor-mentee structure, which the leadership of the school deemed a vital component of the program" due to its well-documented benefits. Finally, such programs have been administered on a smaller scale within the college in individual departments informally, which led the leadership to formalize and support such efforts at the college level as a strategic investment.

Program Timeline

The inaugural program lasted 11-months, beginning in November 2017 and ending in September 2018. The program timeline coincided with NIH submission deadlines to have each mentee submit a grant proposal by the end of the program. A pre-program training for mentors occurred on November 7, 2017, to prepare the mentors for their upcoming duties, followed by a half-day (i.e., 8:30 am-12 pm) group meeting on December 8, 2017, to introduce the cohort participants to each other, highlight program expectations, and provide a timeline for program activities. Subsequently, the participants met two more times, as a group, for half-day sessions on February 16, 2018, and May 25, 2018, for program-specific training. Mentors and mentees also met one-on-one between these dates for individual mentorship at a date and time deemed convenient for both. Figure 3.2 summarizes this program's timeline and activities associated with the NIH grant proposal development and submission. At the same time, it outlines a summary of the program's 11-month progression as its developers envisioned it at the college.

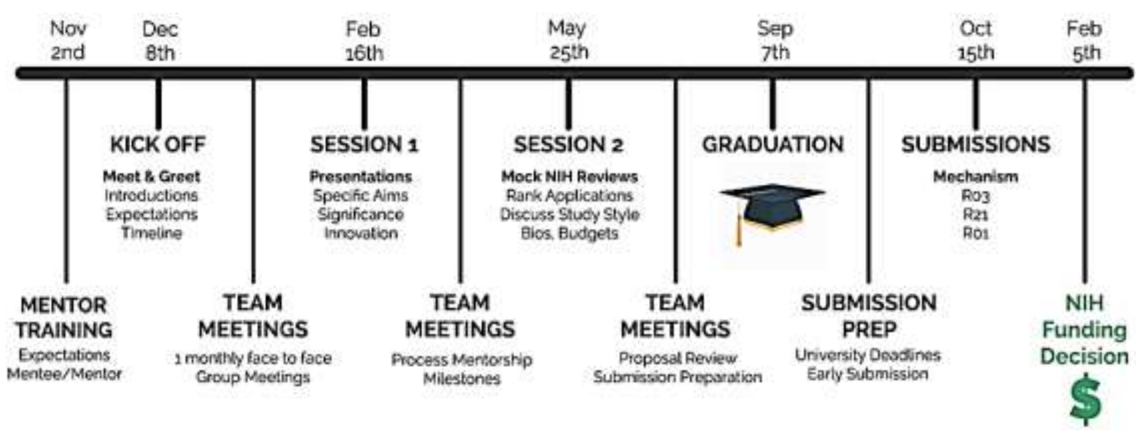

Figure 3.2 FRD program timeline

The focus of the mentorship revolved around NIH grant proposal development and focused on the following sections:

1. NIH Funding Mechanisms
2. Co-investigators and Collaboration
3. Proposal Specific Aims
4. Proposal Significance and Innovation
5. Proposal Preliminary Studies and Approach
6. Proposal Budget, Budget Justification, and Subcontracts

The specific aim of this FRD program was the acquisition of R-level NIH research grants. These grants represent NIH grants' pinnacle and are characterized by high-dollar and long-range funding, making them the most desired.

Participant selection, criteria, and justification

The participant selection strategy is “Criterion” in nature. The “Criterion” method of participant selection is purposive/subjective and relies on predetermined characteristics deemed essential to the study (Patton, 2001). The study participants must exhibit the requisite characteristics necessary to be included in peer-led cohort-based FRD programs. Not all university faculty are research faculty, and few are involved in RD activities requiring them to acquire NIH federal grants. Moreover, this study examined new and early career faculty (i.e., various rank clinical, research, assistant, and associate professors).

Table 3.2 illustrates the criterion and corresponding justification for study participant selection. The first criterion is research faculty, and the justification is that to acquire grant funding, the participants need to be research faculty. The second criterion is that the participants must be new or early-career faculty, and the justification is that this is the time of most need in terms of research mentorship. The third criterion is that the

participant must be enrolled in an FRD program aiming to increase participant research productivity, and the justification is attaining knowledge that yields NIH federal grants.

Table 3.2 Study participant selection criteria & justification

Criterion	Justification
Research Faculty	Required to seek and acquire NIH federal grant funding
New or Early-Career	Most eligible for research development mentorship
Program Enrollment	Enrolled in a research development mentorship program

Seven mentors who are senior-level faculty (i.e., Associate and Professor rank) and twenty-two mentees who are new and junior faculty (Clinical, Research, Assistant, and Associate professor rank) comprised the cohort-based and peer-led FRD program. One of the mentors served as program director, and another was the associate dean of research in the college. The program was advertised on the college website and through faculty meetings. Both mentors and mentees self-selected into the program. In total, 26 faculty applied to the first inaugural FRD program. After receiving three R21 grants post-enrollment, one participant withdrew from the program to focus on his newly acquired funding. Of the remaining 25, the leadership selected 23 applicants to participate in the program. Those not chosen were deemed too early in their career track (i.e., having no grant proposal development experience) to be applying for large-scale NIH grants and excluded. Before starting the program, an additional applicant withdrew, citing health issues, leaving 22 participants in the program., Table 3.3 illustrates program participant rank by frequency and percentage.

Table 3.3 FRD program participants

Rank	**N**	**%**
Professor	6	20.69%
Associate	7	24.14%
Assistant	13	44.83%
Clinical Associate	1	3.45%
Research Assistant	2	6.89%
Total	29	100%

Mentors and mentees self-selected into the program, mentors agreeing to participate when asked by the college leadership, and mentees applying to the program advertised on the college website and through faculty meetings. White female professors represented most mentorship program participants (45%), followed by equal distribution of male and female Asian professors and white male professors (14%). Female professors comprise the majority of mentees (73%), as illustrated in Table 3.4.

Table 3.4 Participant demographics information

	Mentors				Mentees			
	Male		Female		Male		Female	
	#	%	#	%	#	%	#	%
African American	0	0%	0	0%	0	0%	2	9%
Asian	0	0%	0	0%	3	14%	3	14%
White	4	57%	3	43%	3	14%	10	45%
Hispanic	0	0%	0	0%	0	0%	1	5%
Totals	4	57%	3	43%	6	27%	16	73%
Grand total			7	100%			22	100%

Site selection, criteria, and justification

This dissertation relied on a case study methodological approach. Specifically, the case study is an intrinsic, retrospective, and explanatory one. Stake (1994) defined a case study as "an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident" (p. 13). The FRD program is a "*contemporary real-life phenomenon*" (i.e., a Faculty Development Program) revolving around a "*social unit*" (i.e., School of Public Health) in a "*complex and unique*" (i.e., aiming for financial sustainability through NIH grant acquisitions) manner, bound by a single "*event*" (i.e., one-year mentorship program) during a specific "*time*" (i.e., 2017-2018) and in a single "*place*" (i.e., University Setting). This dissertation examined how such FRD programs' implementation practices impact their effectiveness through the empirical lens of an RD faculty-centric

adult learning theoretical framework lending to its' "intrinsic" nature. The fact that the program ended in September of 2018 makes it "retrospective." The rationale underlying the choice of an "explanatory" study type is grounded in the goals of this research study: an intensive, holistic, empirical, all-inclusive examination of the implementation practices of a group of people involved in a unique event with a complex context revolving around the financial sustainability of HEI.

Based on an interest in university investment in RD and return-on-investment of such efforts, the site selection strategy is "Criterion" in nature. The "Criterion" method of site selection is purposive/subjective and relies on predetermined characteristics deemed critical to the study (Patton, 2001). The choice is rooted in that the site must reflect an environment capable of producing the examined outcomes. Not all universities are research-oriented, and few conduct investments in research while being research-driven. Moreover, this study examined investment in FRD programs aiming to increase NIH federal grant acquisition productivity as a response to state divestment in higher education institutions, which limits the study sites to public research universities, and for this study, a Carnegie R1 classification public research university (i.e., higher education university in the United States involved in extensive research activity).

The study site selection criterion and corresponding justification is "Criterion" in nature. The first criterion is "Public University" (i.e., a university that is publicly owned and or receives significant public funding from its corresponding state), and the justification for this choice is that in recent years state divestment for these institutions has caused them to invest in RD in order to compensate for the decline in state funding by competing for federal grants funding. The second criterion is "R1 Carnegie

classification" (i.e., they receive a minimum of $40 million in federal support), and the justification for this criterion is that for a university to achieve this status, it must be competitive in seeking federal grants and in doing so employing RD and support programs as it aims to achieve this status. The third criterion is "Research Faculty" (i.e., faculty expected to seek after and secure NIH federal grants), and the justification for this is that in order for a university to secure NIH federal grants, it must employ research faculty with the expectation of seeking and securing those NIH federal grants. The fourth criterion is that the site must have an "Office of Research" (i.e., research infrastructure), and the justification for this is that the university of choice must have offices of research on the macro (i.e., university) and micro (i.e., college/department) levels tasked with supporting faculty in their attempts to secure federal grants. The final criterion is "Research Development" (i.e., grant acquisition training), and the justification is that the university, through its offices of research, invests in developing its faculty's grant acquisition skills, as illustrated by Table 3.5.

Table 3.5 Study site selection criteria & justification

Criterion	Justification
Public University	Impacted by state divestment
R1 Carnegie Classification	Expected to secure a minimum of $40 million in federal support annually
Research Faculty	Employs faculty with the expectation to secure NIH federal grants
Office of Research	Support for research is available on the university college level
Research Development	Administers grantsmanship skills training

The larger locale for the research site is the southeastern part of the United States. Specifically, the site's locale is a southern state. The specific locale of the study site is the

metropolitan capital city in that state. This city is home to the university, which is the specific site for the study. This university is inclusive of all criteria, as mentioned earlier, required for the site selection. It is a public R1 Carnegie classification research university with university and college-level offices of research and multiple research development initiatives. Specifically, the college administers a peer-led cohort-based FRD program, which is the case for this study. Furthermore, the researcher's proximity to the site facilitated the completion of the study.

Researcher Reflexivity

Subjectivity

The researcher’s subjectivity manifests on ethnic, cultural, and religious grounds. Although the United States government identifies anyone from the Middle East and North Africa as White, the researcher is a first-generation immigrant to the United States who still identifies with his former community (i.e., Middle Eastern patriarchal culture). The bulk of the participants in this study are female (68.18%), most are young early-career professors, and the majority are White and Asians. The researcher had a patriarchal upbringing and gave credence to female participant experiences to guard against interpreting their experiences through a patriarchal social lens characterized by subordination, oppression, power, dominance, hierarchy, and competition.

Specifically, the researcher had to be mindful of how his social constructs could explain away oppressive experiences women encounter due to the potential lack of resources or decision making. This awareness is crucial given that Waisbren et al. (2008) and Eloy et al. (2013) detected a gender gap in grant acquisition between female and male researchers where female researcher success rates were 41% compared to their male

counterparts at a 45% success rate. Female researchers also received less funding on average per application ($115,325) than males ($150,000). Ginther et al. (2011) also highlight a disparity in which Asian and African American researchers were 4 and 13 percentage points less likely to receive NIH funding than their white counterparts, respectively. This disparity was also manifest as Asian and African American researchers resubmitted their grant applications more often than their white counterparts before being awarded an NIH grant. The disparity is especially true for large-scale R01 level grants.

Positionality

Researchers must also reflect upon their position concerning their participants and the data they collect when conducting research (Yao & Vital, 2018). The researcher's employment at the research site and college that conducted the FRD program put him in a unique position to conduct the study. He works for the director of the FRD program, who was also a mentor. He was involved in the program since its inception as its' administrator. He is well acquainted with both mentors and mentees and has interacted with them on many occasions, and in some instances, was the administrator for the faculty search committees that initially hired them. He organized the program's events and was present for all meetings. He also had access to participant's past performance data regarding grant submission rates and their post mentorship data, which assessed their grant acquisition performance. This insider status gave him unique access to information about every aspect of the program and its participants. The participants promptly answered his requests because they are aware of this. After all, every email and request he sent had the dean of research and director of the mentorship program copied on it. This insider status proved advantageous in gathering pertinent data, yet he had to be

cognizant of this power dynamic. Rapport, the guarantee of confidentiality, and balancing between an insider and outsider positionality played a crucial role in eliciting candid feedback from the participants who were well aware of his insider position.

To minimize potential weaknesses arising from the researcher's subjectivity and positionality, he was mindful of his subjectivity by utilizing notetaking to document feelings and reactions during meetings with participants. He ensured participant voices' resonance by including their verbatim quotes, clarifying follow-up questions, and focusing on the study's purpose. His educational background (i.e., Human Resource Development), professional duties as an RD professional at an HEI, and role as the administrator for the FRD program during its implementation made him best suited to evaluate it, especially given that he observed the program in real-time and had access to all the data necessary to conduct the study.

Trustworthiness and Validity

First, adequate familiarity with the FRD program ensured the study's trustworthiness and validity by conducting an exhaustive review of the literature and program-related artifacts. This knowledge communicated to participants a fiduciary responsibility towards evaluating the program. Secondly, this deep understanding provided a detailed description of the research context for readers to enhance the results' transferability to similar contexts (Lincoln & Guba, 1985; Trochim, 2006). Thirdly, the purposeful sampling of the two participants (i.e., Dean of Research and mentorship program director) who possessed firsthand knowledge and involvement in the program's life span (i.e., preplanning to delivery) provided relevant context relating to its evaluation. Fourthly, preceding each interview, the participants were informed of the

research's purpose, assured confidentiality, and allowed to decline participation. This process enhanced the findings' confirmability and credibility by ensuring that participants shaped the findings rather than the researcher (Trochim, 2006). Fifthly, an iterative approach to the interview protocol (i.e., asking follow-up questions, paraphrasing responses, and summarizing participant statements) ensured detailing, mitigated ambiguity, and minimized discrepancies.

Triangulation, a method characterized by the use of multiple data sources (Patton, 1999) to strengthen a study (Patton, 2001), is used in qualitative research to test the validity and reliability of those data (Mathison, 1988). Denzin (1978) categorized triangulation into four types: across data sources (i.e., participants), theories, methods (i.e., interview, observation, documents), and among different investigators. The study utilized multiple data sources about the FRD program, such as administrative records, historical performance data, and program artifacts (e.g., Blackboard communications, presentations) to corroborate codes and themes that emerged from participant interviews. Often called the "Devil's advocate" of validity and reliability, peer debriefing relies on researcher peer review to elevate rigor by challenging researcher assumptions (Lincoln & Guba, 1985).

Methodological Approach

This dissertation employed a sequential explanatory design. The methodological approach unfolded in two stages. The first stage involved collecting and analyzing quantitative data, while the second involved collecting and analyzing qualitative data. Quantitative data and their subsequent analysis provided insight into whether the examined program was efficacious (i.e., facilitated participants acquiring grants). Next,

qualitative data and their analysis expounded the program's results (i.e., explained the effect) and provided insight into improving its delivery and maximizing its potential effectiveness. Last, the same qualitative data and analyses tabulated the total cost associated with the FRD program provision and compared it to the program's total benefit (i.e., total grants acquired by all participants) to ascertain program return-on-investment (i.e., cost out the effect) (Rossman & Wilson, 1985; Tashakkori, Teddlie, & Teddlie, 1998). The following chapters of the dissertation will include articles that include methods, findings, discussion sections, and pertinent sub-sections. The first article addresses program effectiveness (i.e., quantitative assessment), the second article addresses program implementation (i.e., qualitative assessment), and the third article finally addresses program return-on-investment. The dissertation closes with a comprehensive conclusion chapter that addresses the dissertation's purpose and conclusions, concerting the research questions.

Article 1 – Assessing Program Effectiveness

The first article employs the CERTi Model's (Aziz & Tran, in press) quantitative evaluation approach (i.e., level 3 behavior and level 4 results) embedded in the follow-up stage to assess a talent development program’s effectiveness. The article includes the use of the propensity score matching statistical technique, Inverse Probability Treatment Weighting (IPTW), to control for potential selection bias (i.e., non-representative sample of the target population), and confounding (i.e., influence on the outcome by unaccounted for extraneous independent variables). IPTW is a recommended technique to best estimate causality when one cannot conduct a controlled experiment but has observed data to model the relationship between those observed data and the desired outcome

(Thoemmes & Ong, 2016). IPTW performs well with small sample sizes (Pirracchio et al., 2012), which is beneficial, given that the case study of the examined faculty FRD program for grant acquisition only included 80 participants. Simply put, this article aims to and ‘what” transpired due to the program's administration.

Article 2— Assessing Program Implementation

The second article employs the CERTi Model's qualitative evaluation approach to assess a talent development program’s fidelity of implementation by appraising its’ pre-planning, planning, delivery, and follow-up practices through an adult-learner faculty-centric theoretical framework lens to understand better how the execution and delivery of a talent development program explain its effectiveness or lack thereof. It compares program intent to its’ implementation to uncover incongruities to “further understanding of effective researcher development by revealing the specific dimensions of change that collectively makeup individuals' development experiences" (Evans, 2011a, p. 29). This approach is essential for informing future iterations of talent development programs and facilitates continuous program improvement. Simply put, this article aims to answer “why” things transpired as they did as due to the program’s administration

Article 3— Assessing Program Return-on-investment

The third article employs the CERTi Model's economic evaluation approach (i.e., level 5 return-on-talent-investment—ROTI), which relies on the principles of economic evaluations to conduct a formal cost analysis (CA) that depends on the ingredients method, which is rooted in the economic principle of opportunity cost to account for all costs associated with the provision of talent development programs. It is the recommended cost analysis method by experts in the field (Levin & Belfield, 2015). The

total cost derived from the cost analysis will then be compared to the financial benefits derived (e.g., total grant dollars acquired by all participants) from the talent development program (e.g., FRD program) to ascertain its' return-on-investment in a formal cost-benefit analysis (CBA). Assessing program return-on-investment is essential given the new financial norm (i.e., organizational climate) that higher education institutions face, characterized by reduced state appropriations, increased competition for external grant funding, and demands for efficient use of their grant awards.   Simply put, this article aims to answer if the program's administration was "worth it?" (i.e., worth the investment).

CHAPTER FOUR: ARTICLE 1 – ASSESSING PROGRAM EFFECTIVENESS

Methods

Research Question

"Are faculty recipients of FRD programs more likely to increase their grant acquisition?

Grant Submissions Data Collection and Analysis

NIH Grant Application Review Process

The specific aim of the FRD program was to equip new and junior faculty with the knowledge and skills that facilitate the acquisition of NIH grants. The goal of the program was to assist mentees in identifying appropriate NIH study sections to submit proposals to, assembling a solid research team of co-investigators and collaborators, and developing a meritorious grant proposal comprised of the Specific Aim, Significance and Innovation, Preliminary Studies, and Approach, Budget, Budget Justification, and Subcontracts sections, for submission to the NIH by the end of the program. Research has shown that these sections are essential to successful funding bids (Liu et al., 2016). Once submitted to the NIH, a grant proposal goes through a series of steps that will culminate in either approved or denied funding. The process starts with an assessment of scientific merit; if deemed meritorious, an application goes before a peer review panel, which decides if it is worth further discussion. Next, the peer review panel discusses each meritorious application and assigns it an impact score and a percentile rank. Grant applications are compared based on these Impact scores and percentile ranks, which leads to funding and denial decisions. Reviewers decline to fund all applications lacking merit.

Scientific Merit—Once a grant application is received and is deemed compliant with NIH submission policies and guidelines, it goes to the Division of Receipt and Referral (DRR). The DRR checks the application for completeness, determines areas of research and which specific NIH institute or center to assign the application for possible funding, assigns the application an identification number, and assigns the application to a specific study section known as a "Scientific Review Group" (SRG) comprised of experts who evaluate the scientific and technical merit of the application.

Peer Review—If the application meets NIH scientific and technical merit guidelines, it moves to peer review, where it undergoes a rigorous two-stage review. A "Scientific Review Group" (SRG) comprises expert non-federal scientists in specific scientific disciplines and research areas relating to the grant application conducts the first level review. The SRG forwards meritorious applications for the second stage review, conducted by Institute and Center (IC), National Advisory Councils, or Boards, all scientific and public representatives chosen for their expertise and interests or activities relating to the grant proposal. Both SRG and the Advisory council must recommend a grant proposal for funding to be funded.

Discussion and Impact Score—Upon reviewing grant proposal applications, reviewers provide an overall impact score reflecting their assessment of the application's merit, considering five review criteria (i.e., Significance, Investigator(s), Innovation, Approach, and Environment). Non-meritorious grant applications are reported as "ND" (not discussed), while those with merit are coded "D" (Discussed) and are assigned an impact score. Reviewers then utilize a 9-point rating scale (1 = exceptional; 9 = poor) to assign each application an impact score. Each grant application's final impact score

derives from the mean of all the individual reviewer scores, which is then multiplied by ten, resulting in a final impact score that ranges from 10 (High Impact) to 90 (Low Impact).

Percentile Rank—Each application also receives a percentile rank, which is the approximate percentage of applications assigned a better overall impact score during the past year. Percentile ranks stem from the ordering of impact scores of grant applications. The lower a percentile rank score, the more meritorious an application; an application ranked in the 10th percentile, for example, would be considered more meritorious than 90% of the reviewed applications.

Funding Decisions—The NIH considers scientific merit, overall impact score, percentile ranks, and Institute/Center goals and needs in making funding decisions. Naturally, grant applications with the best scientific merit, best impact scores, meritorious percentile rank, and those aligning with Institute/Center goals and needs receive funding. Funding amounts vary per Institute/Center guidelines. Figure 4.1 provides an overview of this process that each application undergoes. The process might vary slightly from one Institute/Center to another, but in general, this is the process that the NIH utilizes in its evaluation of grant proposal applications.

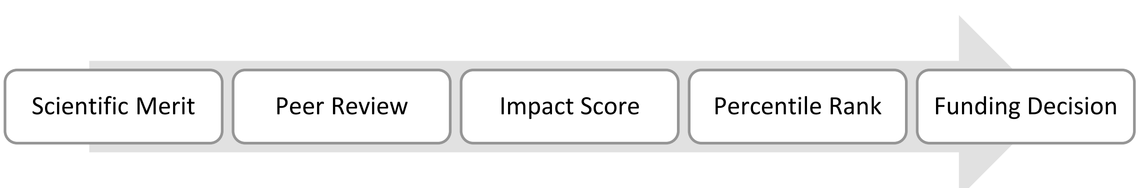

Figure 4.1 NIH Grant application review process

Grant Submissions Data Collection & Analysis

University electronic research administration records to provided data on mentee NIH grant proposal application submissions. Additionally, eCommons, an NIH funding decision portal, and NIH RePORTER, a research database on NIH-funded research projects, provided data on whether reviewers discussed submissions (i.e., assigned "D" for discussed or "ND" for not discussed), impact scores, percentile ranks of each submission, and funding decisions. R statistical software scrutinized these data to report descriptive statistics (i.e., Mean, Median, Mode, Standard Deviation, and Range) on grant proposal submission rates and ensuing progression to the final funding decision to assess mentee behavioral development.

Grant Acquisitions Data Collection and Analysis

Observational Studies (Non-RCTs)

Randomized control trials (RCTs), unlike observational studies, ensure that on average treated subjects do not differ systematically from their untreated counterparts based on measured and unmeasured baseline covariates (i.e., characteristics) (Austin et al., 2015). Covariates may affect the outcome (i.e., accuracy) of a study and can be of direct interest (i.e., independent variable) or unwanted (i.e., confounding variable). RCTs facilitate accounting for confounding variables (i.e., influence on the outcome by unaccounted for extraneous independent variables) and reduce the probability of selection bias (i.e., non-representative sample of the target population). Conversely, in observational (i.e., non-randomized) studies, the effect of treatment on outcome may be subject to treatment selection bias wherein receiving treatment or not based on shared covariates differs. Consequently, a simple comparison between these groups' outcomes

often is an insufficient method of estimating treatment effect (Rosenbaum & Rubin, 1984). Lack of randomization can lead to an unbalanced probability (i.e., propensity) of receiving treatment or not, conditional on baseline covariates, or omitted variables (i.e., selection bias), and not accounting for confounding variables (i.e., influence on the outcome by unaccounted for extraneous independent variables). This condition opens the door for oversampling in either direction. Figure 4.2 provides an example of such imbalance as there is a higher probability in this example of receiving treatment (P=.9) as opposed to not receiving it (P=.1) based on a shared baseline covariate (X).

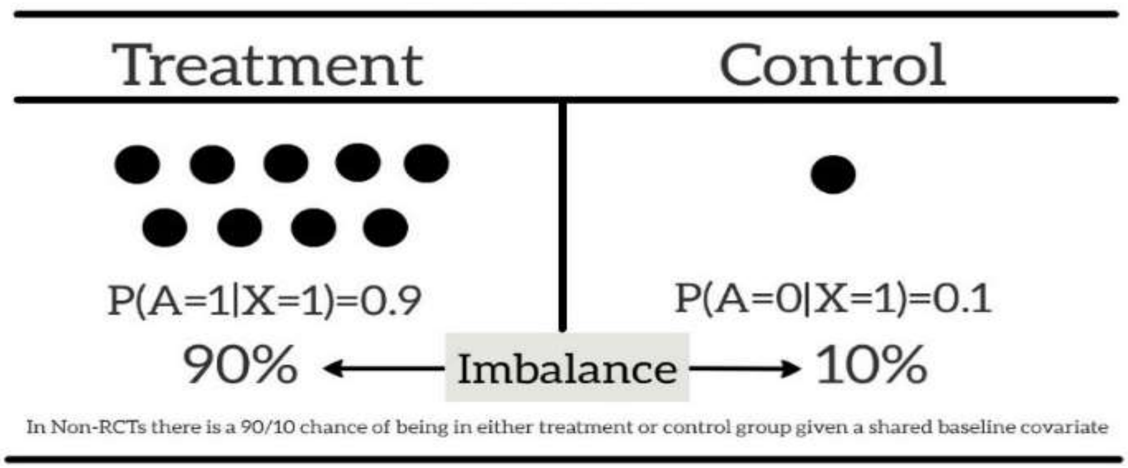

Figure 4.2 Non-randomized observational studies

Addressing Confounding

Although considered "the gold standard" approach for estimating treatment effects of interventions on outcomes (Berk et al., 2005; Torgerson, 2008), randomized control trials are expensive to administer, consume researchers valuable time, and are often impractical to implement, which explains the prevalence of observational studies. Researchers are increasingly employing statistical methods to mimic randomized control trials to increase their studies' rigor in the absence of randomization (Austin & Stuart, 2015), which is this study's aim. One such method increasingly used for addressing

confounding and moving towards more causal estimates is propensity score matching to balance observable baseline covariates between treatment and control groups. Researchers defined a propensity score as the probability of treatment assignment conditional on measured baseline covariates (Austin, 2011; Rosenbaum & Rubin, 1983, 1984). Propensity scores allow for reducing or eliminating the confounding effects when using observational data (Austin, 2011).

Pan and Bai (2015) outlined four steps to estimating a propensity score:

1. Estimate propensity score
2. Match
3. Evaluate the quality of matching
4. Evaluate outcomes

The first step can entail estimating the likelihood or probability that an individual unit in the data experiences the treatment given a set of characteristics (i.e., covariates). The second step involves matching the scores of those individual units within the data set (i.e., treatment group) to those of individuals outside of the data set (i.e., control group) with a similar propensity score (i.e., probability of receiving the treatment given the same set of covariates) to have a more convincing comparison group. Keeping in mind that the purpose of the matching is to have treatment and comparison groups similar based on observable covariates, the third step involves evaluating match quality (i.e., the balance of covariates). Operating under the assumption that the match's quality is good, the fourth and final step entails evaluating outcomes and estimating causal effects.

Step 1—Estimating Propensity Score. Estimating the propensity score is typically done using a logistic or probit regression of the treatment condition on a covariates vector

(Curtis et al., 2007). Logistic or probit regression is appropriate when conducting predictive analyses on binary/dichotomous dependent variables. It involves regressing all the covariates (i.e., control variables predicting the treatment) in the Model to ascertain whether an individual unit in the data set received treatment (i.e., the independent variable of interest).

Step 2—Matching. The most common matching methods are Nearest neighbor matching (i.e., matching the treatment unit to the nearest comparison unit), Caliper matching (i.e., matching the treatment unit to a comparison unit within a range or "caliper"), and Radius matching (i.e., matching a treatment unit to multiple comparison units within a band). Additionally, Greedy versus optimal matching (i.e., the first entailing choosing the first match even if a better match is found later in the data, whereas the latter entails matching based on closer propensity scores). Finally, with and without replacement matching, keeping an observation in the pool of match observations once matched versus discarding it.

Step 3—Evaluating Quality. Various approaches ensure matching quality (i.e., achieving a balance between the treatment and comparison groups on observable covariates). These include but are not limited to comparing means (e.g., t-test) to see if the difference is statistically significant, calculating standardized bias for each covariate, and estimating percent bias reduction (PBR), done through graphical comparisons (e.g., histograms, boxplots). Depending on the match's quality, one can estimate the treatment effect or return to step 3 to address the quality issue.

Step 4—Evaluating the Outcomes. The final step involves conducting an outcome analysis to evaluate the intervention in question. Methods for such an evaluation include

comparing the means of matched samples, running a regression on the matched sample while controlling for unbalanced covariates, or conducting a propensity score adjustment (i.e., running a regression controlling for the propensity score itself).

In summary, if an HEI leader is to utilize propensity score matching, good practice would include identifying an appropriate data set, defining the treatment, control, and outcome, selecting appropriate covariates, estimating the propensity score to 'match' the groups, assessing the 'matching' using balance techniques, and conducting an analysis of the outcome on the propensity score-adjusted sample. One-to-one matching based on specified baseline covariates allows for one "treated" subject to be matched to one "control" subject and, in doing so, creating balance. Figure 4.3 visually illustrates this process as subjects are matched based on shared baseline covariates. In this example, there is oversampling in the treatment group, where subjects have a higher probability (P=0.9) of being treated based on specific shared covariates instead of not being treated (P=0.1), illustrating oversampling of the treated group.

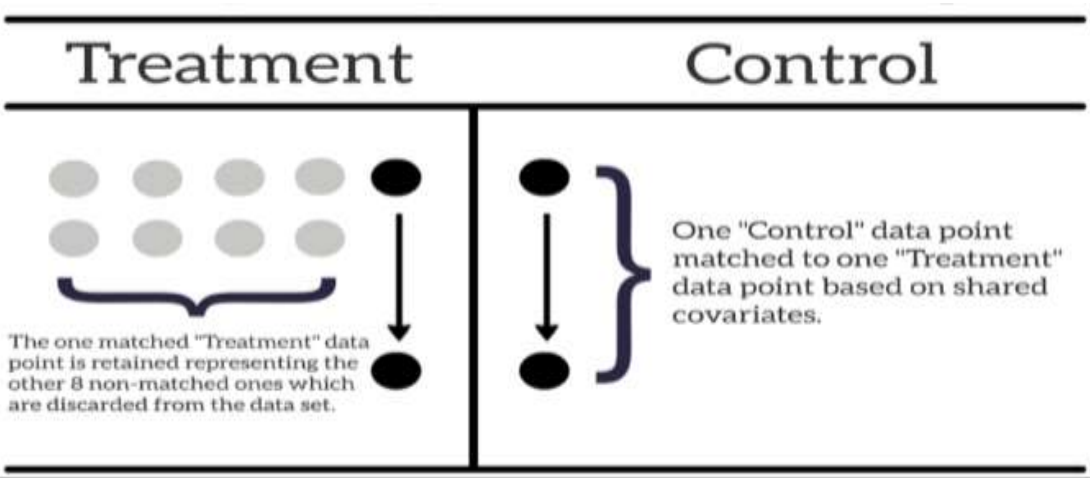

Figure 4.3 Propensity score matching

Addressing Selection Bias

The statistical literature describes four methods of using propensity scores to address selection bias: stratification (Rosenbaum & Rubin, 1983, 1984), adjustment (Rosenbaum & Rubin, 1983, 1984), matching (Austin, 2008; d'Agostino, 1998; Rosenbaum & Rubin, 1983, 1984) and more recently, inverse probability of treatment weighting (IPTW) (d'Agostino, 1998; Joffe et al., 2004; Kurth et al., 2005; Lunceford & Davidian, 2004; Robins et al., 2000; Rosenbaum, 1987). Among all these methods, both matching and IPTW have demonstrated the greatest efficiency in reducing imbalance in baseline covariates (Pirracchio et al., 2012).

Austin and Stuart (2015), in proposing best practices when using IPTW with propensity scores to estimate causal treatment effects in observational studies, systematically reviewed the literature and found that "the use of IPTW has increased rapidly in recent years" (p. 3664). The increased use of IPTW is because this method creates pseudo-populations that are non-confounded. In such cases, there is oversampling of treated or control groups based on specific covariates. For example, in the case of the FRD program under review, there may be an oversampling of faculty that exhibit qualities that predispose them to acquire NIH grants (e.g., previous grant acquisition, publishing record), hence their selection to participate in the program. Weighting counters this oversampling to achieve balance. Figure 4.4 illustrates such a situation in which there is oversampling in the treated group compared to the control group. Nine out of ten subjects, in this case, are treated, which creates an imbalance. This oversampling of subjects in the treatment group relative to the control group must be adjusted and can be rectified by up-weighting the control group by the inverse of the probability of being in

the control group and down-weighting the treatment group by the inverse of the probability of being in the treatment group, which creates balance.

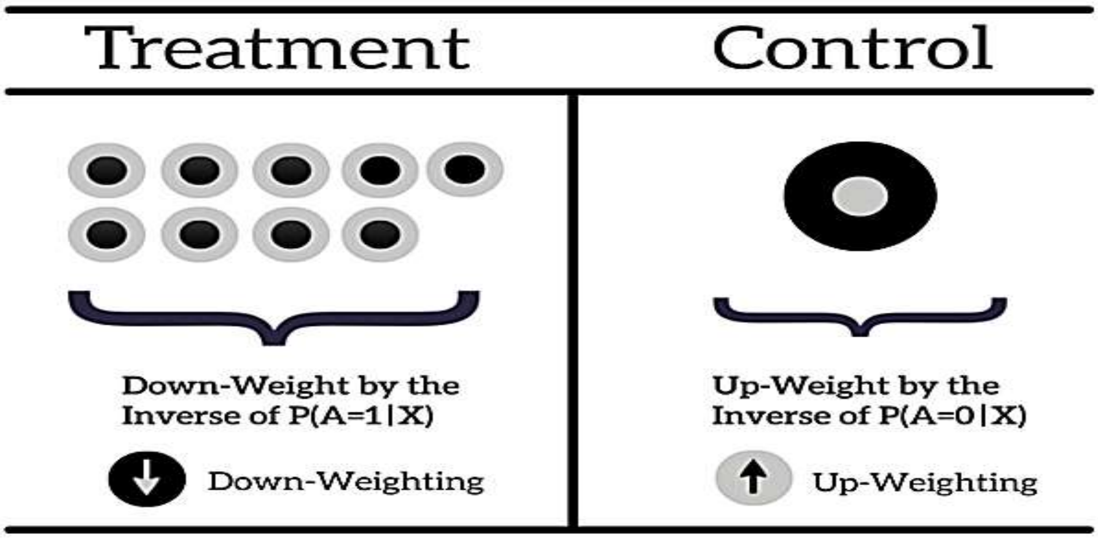

Figure 4.4 Inverse probability treatment weighting

Achieving this balance results in the creation of a pseudo-population that is balanced based on observable baseline covariates. This balance ensures that, on average, treated subjects do not differ systematically from their control counterparts based on measured and unmeasured baseline characteristics, allowing for direct comparison between the groups to estimate treatment effect as suggested by Austin and Stuart (2015) and illustrated by Figure 4.5. Each subject in the treatment group counts as nine-tenths of a subject (i.e., down-weighted) while the one person in the control group counts as ten subjects (i.e., up-weighted), achieving balance. As a consequence of this weighting, what is absent in this new population is the oversampling that was present in the original one. In the original group, subjects had a higher probability of receiving treatment instead of not based on shared baseline covariates, while in the new one, that probability is equal. Although this does not rise to the rigor of randomization, this, in essence, mimics the desired characteristic of randomization present in RCTs.

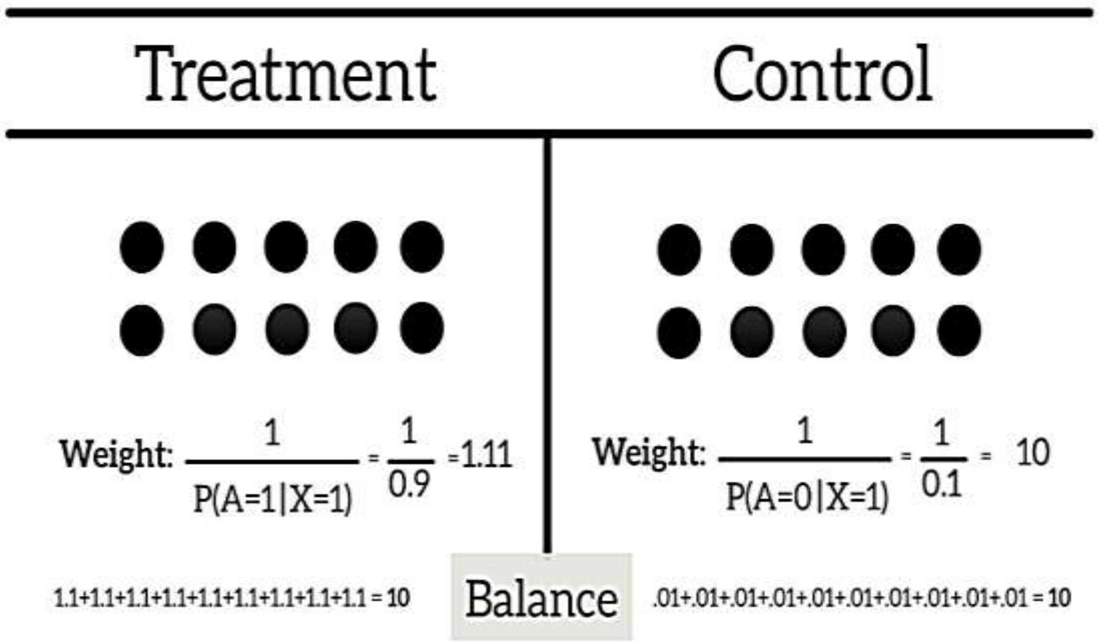

Figure 4.5 IPTW pseudo-population

Estimating The Causal Effect

Estimating the FRD program's causal effect on participant grant acquisitions contingent on observable baseline characteristics is possible using propensity score matching and IPTW. The process entails identifying 1) an appropriate data set, 2) defining the treatment and control groups, 3) selecting appropriate covariates, 4) developing an appropriate propensity score model, 5) defining matching method, 6) evaluating match quality, weighting for balance, and 7) estimating causal effect. Figure 4.6 graphically illustrates this process of estimating treatment effect based on the propensity score matching and IPTW techniques. This study employed this step-by-step process to control selection bias and confounding variables to best estimate a causal link between the FRD program and its' outcome (i.e., grant acquisition) to demonstrate this process's applicability.

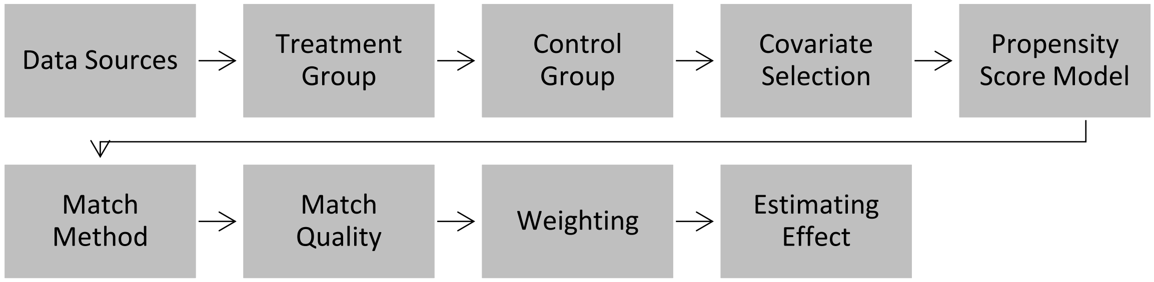

Figure 4.6 Causal effect estimation process

Data Sources

University administrative records provided participant data (i.e., Race, Gender, Rank, Experience, Tenure, Publishing, citations, *h*-index, and Grants). NIH Research Portfolio Online Reporting Tools (*RePORTER*) and university electronic research administration records supplied grant acquisition data. Data spanned the years 2013-2017 to provide for a better assessment of program effectiveness. This data provided better context on differences between treatment and control group participants relating to such covariates as publication, citation, *h*-index, and past grant acquisition records.

Treatment Group

In total, 26 faculty applied to the first inaugural FRD program. After being awarded multiple R-level grants after enrolling in the program, one participant withdrew from the program to focus on his newly acquired funding. Of the remaining 25 applicants, 23 participated in the program. Those not chosen were deemed too early in their career track (i.e., not enough experience with grant proposal development) to be applying for large-scale NIH grants and excluded from the program by the leadership team. Before starting the program, one applicant withdrew, citing health issues, leaving 22 participants in the program at its inception.

Comparison Group

Participant enrollment into the FRD program exhibits a lack of randomization and selection biases (i.e., self-selection and exclusion), which makes it impossible to state with certainty that the results of the program (i.e., grant acquisitions) were due to the treatment (i.e., mentorship). For example, most participants self-selected into the program, while the programs' designers excluded others. Self-selection into the program by participants and the subsequent exclusion of some by program leadership necessitates causal estimate statistical techniques such as IPTW to address such biases. While only a select group of faculties received mentorship in the program, data is available on all 155 faculty at the college that administered the program. These faculty, are spread across the six departments of the College (Communication Science Disorder, Environmental Health Sciences, Epidemiology/Biostatistics, Exercise Science, Health Promotion & Behavior, and Health Services Policy & Management), serve as the comparison group to match with the treatment group based on the aforementioned selected baseline observable covariates in order to address the selection biases.

Covariate Selection

Key covariates include variables related to the probability of selection into the program and the outcome of interest. Research has shown that specific researcher characteristics such as race (Ginther et al., 2011), gender (Waisbren et al., 2008), rank (Pagel & Hudetz, 2015), experience (Pagel & Hudetz, 2015), tenure (Conn et al., 2005), and publishing, citations, *h*-index, and grants (Ali et al., 2010) are predictive of the outcome of interest in this study (i.e., research productivity). For example, Ginther et al. (2011), in investigating the association between the probability of receiving a grant and

self-identified race or ethnicity, highlight a disparity in which Asian and African American researchers were 4 and 13 percentage points less likely to receive NIH funding as compared to their white counterparts, respectively. Furthermore, this disparity was also manifest as Asian and African American researchers resubmitted their grant applications more often than their white counterparts before being awarded an NIH grant. The disparity is especially true for large-scale R01 level grants. Meanwhile, Waisbren et al. (2008) and Eloy et al. (2013) detected a gender gap in grant acquisition between female and male researchers where female researcher success rates were 41% compared to their male counterparts which had a 45% success rate. Further exacerbating this funding gap, female researchers received less funding on average per application ($115,325) as compared to males ($150,000).

Pagel and Hudetz (2015), while investigating the relationship between both rank and experience, found that:

> The number of R-series, but not K-series, NIH grants received, total NIH grants, years of funding, and amount of support were also dependent on faculty rank (P < 0.0001). The number of individuals receiving NIH support was also faculty rank dependent (e.g., 57.6% of Professors vs. 27.9% of Assistant Professors; Table 3) …number of R-series NIH grants, and total amount of NIH funding increased in a time-dependent manner between groups with less than or equal to 10, 11 to 20, and more than 20 yr of experience. Individuals with more experience were also more likely to have NIH funding (53.9% of those with >20 yr of activity) compared with their colleagues with less experience (13.8% in those with ≤10 years of activity; Table 4). These data indicate that the productivity of FAER grant recipients consistently increases over time. (pp. 684-685)

Both researcher experience and rank are influential in their endeavor to attain NIH grants. Conn et al. (2005) draw our attention to the pressure associated with achieving tenure included the expectation to secure extramural grants. This pressure usually subsides upon tenure and focus shifts on service. Conversely, Anderson and Slade (2016)

recently investigated researcher time spent to pursue grants to respond to institutional pressure to secure these funds and found that promotion mechanisms weigh heavily on research productivity. They state that "junior and untenured faculty members, in response to institutional pressures related to promotion, are likely to allocate more time to research and grant writing" (Anderson & Slade, 2016, p. 103). Specifically, they found that non-tenured researchers spent an average of 1.68 more hours writing grants than their tenured counterparts. This finding indicates a higher motivation to seek and apply for competitive research grants among non-tenured researchers as opposed to their tenured counterparts.

Ali et al. (2010) found that past grant acquisition history to be associated with future success rates in acquiring grants as they state, "The percentage of faculty at an institution having already won competitive research grants, having journal publications, and whose publications have garnered citations, are also positively correlated with the dollar amount awarded" (p. 171). They further investigated the influence of researcher scholarly productivity (i.e., publications) on the dollar amount and the total number of federally funded competitive research grants. They observed a positive correlation between a researcher's publication and citation records and federally funded competitive research grants. Svider et al. (2014) also found a strong relationship between NIH funding and a researcher's h-index, a measure of research productivity (i.e., publications), and citation impact.

These researchers' findings collectively outline pre-treatment participant characteristics (i.e., Race, Gender, Rank, Experience, Tenure, Publishing, citations, h-index, and Grants) that could influence the outcome of the FRD program (i.e., grant acquisition), which makes accounting for these characteristics necessary. This influence

is especially pronounced in the absence of randomization and the presence of self-selection and exclusion biases, as is the case with the FRD program.

Propensity Score Model

The propensity score is the probability of a faculty member's selection for treatment (i.e., selected into the FRD program) based on the pre-treatment observable covariates. The dichotomous (0,1) variable Z will indicate treatment, and X will be the vector of available pre-treatment covariates to depict the propensity score better. The propensity score $p(\mathbf{X})$ for a faculty is the conditional probability of being treated (i.e., selected into the FRD program) given their covariates $\mathbf{X}$: $p(\mathbf{X}) = \Pr(Z = 1|\mathbf{X})$. The most commonly used method for creating propensity scores is logistic regression (Austin, 2011; Stuart, 2010). Propensity scores for treatment (i.e., selection into the FRD program) stem from the fit of a logistic regression model that considers the aforementioned observed baseline characteristics of race (i.e., American Indian or Alaska Native, Black or African American, Asian, Hispanic, or Latino, Native Hawaiian or Other Pacific Islander, and White), gender (i.e., Male, Female), rank (Clinical Assistant, Clinical Associate, Research Assistant, Assistant, and Associate), experience (i.e., years of experience at the college), tenure (i.e., Tenured, Non-Tenured), publishing (i.e., number of publications), citations (i.e., number of citations), *h*-index (i.e., productivity and citation impact of publications), and grants (i.e., number and amount of acquired grants a for the last 5.5 years 2014-2019).

The equation below reflects the logistic regression model:

$$\boldsymbol{L = Log\ \left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9}$$

Where L represents the dependent variable (i.e., propensity score) of inclusion into the FRD program, $\boldsymbol{\beta_0}$ represents the intercept, $\boldsymbol{\beta_1 X_1}$ represent the coefficient magnitude and value of the independent race variable, $\boldsymbol{\beta_2 X_2}$ represent the coefficient magnitude and value of the gender independent variable, $\boldsymbol{\beta_3 X_3}$ represent the coefficient magnitude and value of the rank independent variable, $\boldsymbol{\beta_4 X_4}$ represent the coefficient magnitude and value of the independent experience variable, $\boldsymbol{+\beta_5 X_5}$ represent the coefficient magnitude and value of the independent tenure variable, $\boldsymbol{\beta_6 X_6}$ represent the coefficient magnitude and value of the independent publications variable, $\boldsymbol{\beta_7 X_7 + \beta_8}$ represent the coefficient magnitude and value of the independent citations variable, $\boldsymbol{\beta_8 X_8}$ represent the coefficient magnitude and value of the independent *h*-index variable, and $\boldsymbol{\beta_9 X_9}$ represent the coefficient magnitude and value of the independent past grants variable. Simply put, the logistic regression will compute the probability that a faculty received the intervention (i.e., was included in the FRD program).

Match Method

Utilizing propensity scores, researchers can employ multiple approaches to create balanced treatment and control groups (e.g., exact, the nearest neighbor with replacement, and nearest neighbor with caliper matching). Exact matching entails matching treatment and control participants based on the same covariate value. Exact matching is simpler to accomplish on categorical variables such as race, gender, rank, and tenure versus continuous variables such as years of experience, the number of grants acquired, and the number of publications because of the limited choices inherent in categorial variables. Nearest neighbor matching with replacement allows for multiple matches between one control participant's propensity score and multiple treatment participants. This method is

the least used because data are not independent (Austin, 2009). A more widely used method that ensures high-quality matching is nearest neighbor matching (NN) that uses caliper adjustments (Austin, 2011). As Harris and Horst (2016) explain:

> When using NN with caliper adjustment, the researcher specifies a distance within which matches are considered acceptable. Using a caliper adjustment, cases are only matched when propensity scores fall within the designated distance, typically a fraction of a standard deviation of the logit of the propensity score (e.g., .2 sd) (p. 4).

Exclusion from the final set of matched samples occurs when a possible match is outside the caliper distance. The appropriate distance to set the caliper can be challenging to determine a priori. Researchers often do not know the distribution of possible covariates, nor the composite used to create the propensity score) before conducting analyses (Smith & Todd, 2005). This study utilized an R statistical software matching command (i.e., method =) to conduct exact matching as a preliminary approach followed by NN matching with a caliper, which is usually a fraction of the standard deviation of the logit of the propensity score (e.g., .2 sd) (Austin, 2009), in the absence of exact matching for all treatment participants. Executing this command will result in a probit regression model predicting the independent variable (i.e., probability of inclusion in the FRD program) based on the specified covariates. The results will also indicate the level of common support (i.e., alignment of propensity scores between treated and control subjects). The command will also estimate the average treatment on treated (ATT) and the standard error outputs.

Match Quality

The quality of the match is verified both numerically and visually. A T-test (e.g., unpaired t-test) comparing the means of both the treatment and comparison groups ensures match quality (i.e., achieving a balance between the treatment and comparison

groups on the observable covariates). Statistical significance of difference will determine match quality. If the difference is statistically significant, reassessment of the match will ensue; otherwise, the match is of good quality. Since the sample size is less than 30 ($n<30$), the study utilized a T-test (Bland & Altman, 2009). Graphical comparisons (e.g., QQ plots, jitter plots, and histograms) provide match quality. Harris and Horst (2016) state, "QQ plots display covariate scores across a probability distribution that is divided into quantiles." They add that "The QQ plot allows the researcher to visually compare how similar each group is at each quantile in the group's distribution on each of the covariates for the total sample and after creating matches" (p. 6).

Weighting

Although propensity score matching has demonstrated unbiased estimations of treatment effects with small sample sizes (Pirracchio et al., 2012), as is the case with this study, it reduces the dataset post matching. Instead, IPTW has been recommended and balances the oversampling of subjects in the treatment group relative to the control group while preserving the entire dataset. Any imbalance found in the data will be adjusted by up-weighting the control group by the inverse probability of being in the control group and down-weighting the treatment group by the inverse probability of being in the treatment group, creating balance. Calculating propensity score weights can be done using the following formula: $\boldsymbol{Weight} = \frac{\boldsymbol{T}}{\boldsymbol{P}} + \left(\frac{\boldsymbol{1-T}}{\boldsymbol{1-P}}\right)$. T is a binary treatment in which the treated (T=1) and the controlled are (T=0), and P represents the propensity score. Put merely; weights are 1/P for the treated and 1/(1-P) for the controlled (Guo & Fraser, 2015).

Estimating Effect

After calculating propensity score, matching treatment and comparison participants, verifying match quality, and applying weights to create a pseudo control population, the final step entails comparing outcome variables between the two groups because hopefully bias has been mitigated. Estimating the average treatment effect (ATE) (i.e., NIH grant funding) for the FRD program can be ascertained by comparing the mean outcome of the treatment group with that of the comparison group. A T-test utilizing R statistical software will determine whether the difference is statistically significant between the treatment and comparison groups and, in doing so, answering the research question, which sought to determine whether faculty recipients of FRD programs are more likely to increase their grant acquisition. R statistical software provided the estimated propensity scores. The commands checked covariate mean differences between the treated and control group and created comparison plots. The codes created weights, assigned weights to data points, created a weighted table to compare standardized differences, and checked the weighing balance.

Findings

Grant Submissions Descriptive Analysis

Mentee Proposal Submission Rates—Notwithstanding the program's attrition, most mentees (n=14) submitted a proposal to the NIH by the end of the program representing 63.64% of participants. The figure is even higher upon removing dropouts, 88.23%. This rate is significant given that the program's aim to culminate in a grant proposal submitted to the NIH by the end of the program as was outlined in the program announcement, "designed to provide mentees the tools and knowledge they need to be

successful at developing a high-quality first submission proposal to the NIH." Table 4.1 summarizes the frequency and distribution of mentee proposal submission rates.

Table 4.1 Mentee submissions

Proposal Submission	**N**	**%**
Yes	14	63.64%
No	8	36.36%
Total	22	100%

NIH Review Discussion Rates—Once a grant proposal is submitted, NIH subject matter experts review it for scientific merit. Meritorious grant applications are assigned a "D" (Discussed), while non-meritorious ones receive an "ND" (Not discussed) code. Six mentee grant proposals received a "D", while eight received an "ND" coding representing 42.86% and 57.14% of the total, respectively, as shown in Table 4.2.

Table 4.2 Mentee discussion rates

Discussion	**N**	**%**
Yes	6	42.86%
No	8	57.14%
Total	14	100%

Proposal Impact Scores—Grant proposals designated a "D" (Discussed) are then assigned an impact score by NIH reviewers based on a 9-point rating scale (1 = exceptional; 9 = poor). Each grant's total impact score is derived from the mean of individual reviewer scores and multiplied by ten, resulting in a final impact score. Impact

scores for mentee grant proposals were 20, 20, 23,27, 49, and 64. Table 4.3 summarizes the mentee NIH reviewer discussed grant proposal impact scores with a minimum score of 20.00 and a maximum of 64.00, the first quartile of 20.75, and the third quartile 43.50, and a median and mean at 25.00 and 33.83, respectively.

Table 4.3 Mentee impact scores

Minimum	**1st Quantile**	**Median**	**Mean**	**3rd Quantile**	**Maximum**
20.00	20.75	25.00	33.83	43.50	64.00

Mentee Proposal Percentile Ranks—Each grant proposal submitted to the NIH also receives a percentile rank, calculated by ordering impact scores of grant applications. Percentile ranks derive from the approximate percentage of applications assigned a better overall impact score during the past year. The lower the percentile rank, the more meritorious a grant application. The impact scores for the six discussed mentee grant proposals were 4,7, and 15. Two grant proposals did not receive a percentile rank due to their impact scores being non-meritorious (i.e., 49 and 64), and one was part of an NIH study section that does not issue percentile ranks. Table 4.4 summarized mentee grant proposal percentile ranks with a minimum of 4.00, first quantile of 5.50, a median of 7.00, mean of 8.67, third quantile of 11, and a maximum of 15.

Table 4.4 Mentee grant percentile ranks

Minimum	**1st Quantile**	**Median**	**Mean**	**3rd Quantile**	**Maximum**
4.00	5.50	7.00	8.67	11	15

Funding Decisions & Awards— As presented in Table 4.5, four mentee grant proposals received funding from the NIH, representing 28.57% of the total grant proposals submitted, making it higher than the NIH grant proposal success rate (20%) for

2019 (NIH, 2020). The grant awards were \$284,910, \$439,500, \$453,813, and \$3,429,123 totaling \$4,607,346. Figure 4.7 illustrates mentee grant proposal submissions, discussions, impact scores, percentile ranks, and funding decisions.

Table 4.5 Mentee funding

Funding Decisions	**N**	**%**
Yes	4	28.57%
No	10	71.43%
Total	14	100%

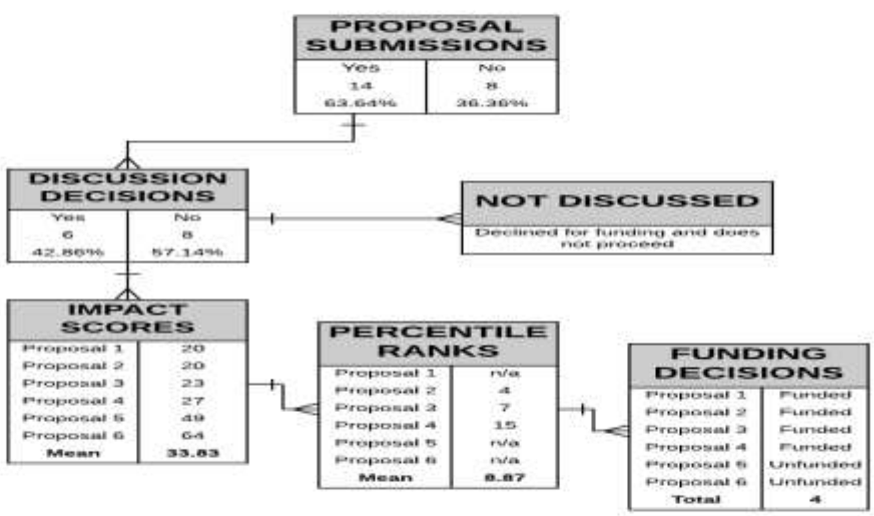

Figure 4.7 Descriptive analysis outcomes

Grant Acquisition Causal Analysis

The fundamental problem of causal inference is that individual-level effects are impossible to observe because there are no counterfactuals. Attempting to know the causal effect of an individual faculty's grant acquisitions based on mentorship requires measuring the results if they were in the FRD program and if they were not simultaneously and then subtracting those outcomes (i.e., $\delta_i = Y_i^1 - Y_i^0$). However,

empirically, only observing one of those conditions is possible (i.e., $\delta_i = Y_i^1 - ?$). The alternative to this would be to take the average outcome given the FRD program taking place minus the average outcome given it does not take place (i.e., $\delta = (\bar{Y}|P = 1) - (\bar{Y}|P = 0)$, which only works if those who participated in the FRD program do not systematically differ on measured and unmeasured baseline characteristics from their nonparticipating counterparts. This methodology is critical since comparing the two groups where one self-selected into the program while the other did not yield inaccurate and skewed results. Conversely, if they are systematically the same across those shared measured and unmeasured baseline characteristics, then the causal effect estimation is more accurate. RCTs using sizable samples facilitate comparison between groups. A well-randomized treatment assignment coupled with large enough sample sizes (i.e., scaled experiments) enables accurate causal estimation of treatment effect because this process preserves unmeasured baseline characteristics. RCTs afford researchers the ability to choose a large enough sample size that facilitates detecting causal effects. A large enough sample size is critical contextually because it allows for detecting effects by reducing variation, potentially found in the general population. Appropriately large sample sizes contribute to a study's accuracy by moving towards the null hypothesis by detecting an effect if it is present and not overestimating it if it is not. Conversely, insufficient sample sizes contribute to a study's inaccuracy by not detecting an effect if it is present or overestimating it when it does exist.

An RCT would have achieved these conditions via three steps regarding the FRD program. First, it would have a priori defined participant eligibility (e.g., research, clinical, Assistant, Associate professors) from the general population of interest. Second,

it would have selected an evaluation sample from that defined population. This step strengthens external validity in that it isolates a group of individuals (i.e., sample) from that identified population that met the eligibility criteria for participation in the FRD program. Third, an RCT would have divided that population-representative sample into treatment and control groups randomly. This step strengthens internal validity in that assignment into either the treatment or control group is random, eliminating self-selection bias or exclusion bias.

Assuming these three steps were satisfied, measuring a randomized FRD program's causal effect would be relatively straightforward. First, the researcher would check the balance of critical characteristics (i.e., confounders) between the treatment and control groups. This process ensures proportional equalization of those characteristics between the two groups, eliminating confounding (e.g., same average years of experience and publication record). Second, the researcher would estimate differences in average grant acquisition between treatment and control groups (i.e., average grant acquisition of the treatment group minus average grant acquisition of the control group).

This observational study aimed to analyze the effect of FRD programs on grant acquisition to answer the research question, "*Are faculty recipients of FRD programs more likely to increase their grant acquisition?*". Although the ideal research design to answer this question is a randomized experiment that arbitrarily assigns participants in both treatment and control groups to link the program to its potential effect, HEI interventions seldomly do so. This lack of randomization is prevalent across RD efforts (Freel et al., 2017a; Gardiner et al., 2007; Kulage & Larson, 2017; Libby et al., 2016; Paul et al., 2002; Santucci et al., 2008; Steiner et al., 2004; Steiner et al., 2002). This lack

of randomization is because an FRD program's underlying premise is enhancing participant grant acquisition skills; hence, depriving them of receiving such benefits by assigning them to a control group is counterproductive. This situation is precisely why development programs, especially educational ones, do not randomize their interventions nor separate their participants into treatment and control groups. A key component missing in observational studies that threaten internal and external validity is randomization. Specifically, what is lacking is a control group comprised of faculty that do not systematically differ from those that participated in the FRD program on baseline characteristics. This group's existence would reduce or eliminate confounding effects and strengthen both the study's external and internal validity.

Educational researchers often aim to determine the effects of interventions or non-randomized factors, such as race, gender, and experience, to determine an unbiased estimate of the causal relationship between a sample’s outcome and these nonrandomly assigned factors. They do this because non-experimental (i.e., non-randomized) interventions create potential biases where the effect of treatment on outcome may be subject to treatment selection bias wherein receiving treatment based on shared covariates differs. A simple comparison between these groups' outcomes becomes an insufficient method of estimating treatment effect (Rosenbaum & Rubin, 1984). Lack of randomization can lead to an unbalanced probability (i.e., propensity) of receiving treatment or not, conditional on baseline covariates, or omitted variables (i.e., selection bias), and not accounting for confounding variables (i.e., influence on the outcome by unaccounted for extraneous independent variables).

This study employs a multistep systematic methodology to estimate the unbiased treatment effect of the FRD program in question. The first step entails highlighting the differences between the treatment and control groups, which undermines any direct comparison between the two groups' outcomes and necessitates a statistical methodology to balance the groups based on shared observed covariates to estimate the program's causal effect. The second step conducts a univariate analysis to test each of those chosen baseline observable covariate's significance to a potential logistic regression model for estimating treatment effect. The third step utilizes directed acyclic diagrams to develop a logistic regression model to isolate confounding variables and estimate the treatment effect. The fourth step tests the model's parsimony and predictive power. The fifth step utilizes propensity score matching to match mentored and non-mentored faculty on the identified baseline covariates (i.e., confounders) to estimate the treatment effect. The sixth step utilized weighting on the identified baseline covariates (i.e., confounders) to estimate the treatment effect. The seventh step applies inverse probability weighting (IPW) to estimates the program's treatment effect.

1. Treatment and response group differences
2. Univariate analysis
3. Logistic regression model development
4. Logistic regression model evaluation
5. Treatment effect Estimation via Matching
6. Treatment effect Estimation via Weighting
7. Treatment effect Estimation via Inverse probability weighting

Treatment Condition Group Differences

Race Covariate—Although there were two African American faculty in each treatment condition (i.e., mentored and non-mentored faculty), the proportions are different. The two African American non-mentored faculty represented 3.45% of the total 58 non-mentored faculty, while the two mentored ones represented a higher, 9.09% of the total 22 mentored faculty. While there were 12 Asian faculty in the non-mentored group compared to 6 in the mentored group, proportionally, the percent of non-mentored Asian faculty, 20.69%, was lower than the mentored ones, 27.27% within each group. The number of non-mentored and mentored Hispanic faculty was equal, one in each group, but the proportion was higher within the mentored group, which was 4.55%, compared to the non-mentored one, 1.72%. The proportion of White faculty is also different between the groups. In total, there were 56 White faculty in the college at the time of the FRD program. Thirteen of the faculty participated in the FRD program representing 59.09% of the total faculty participating in the program compared to 43 who did not participate, representing 74.14% of non-participating faculty. The differences between mentored and non-mentored faculty highlight the imbalance between the two groups.

Gender Covariate—The proportional differences are evident for both genders. 15 female and seven male faculty members participated in the FRD program, while 27 females and 31 male faculty did not participate. Comparatively, there was a higher proportion of female faculty in the FRD program, while there was a lower proportion of female faculty than non-participants in the college. Female faculty constituted 68.18% of FRD program participants compared to 46.55% of non-participating faculty. Conversely, male FRD program participants comprised 31.82% of those partaking in the program

compared to 53.45% of those who did not, highlighting an imbalance regarding the gender covariate between the two groups.

Rank Covariate—The differences between non-mentored and mentored faculty based on their rank are clear. Non-mentored assistant faculty comprised 34.48% of non-mentored faculty while comprising a much larger 63.64% within the mentored group. Associate rank non-mentored faculty constituted 34.48% of total non-mentored faculty compared to 22.73% of total mentored ones. There were no clinical assistant nor professor rank faculty represented in the FRD program while accounting for 3.45% and 20.69% of program participants, respectively. Two clinical associate professors were in the non-mentored group compared to one in the mentored group representing 3.45% and 4.45% of each group's total. Although there were two research assistant professors in each group, the proportion of research assistant professors was higher in the latter group, 9.09% compared to 3.45%.

Tenure Covariate—Non-tenured faculty comprised most FRD program participants at 81.82%, with tenured participants representing the minority at 18.18%. There was an even split between non-tenured and tenured faculty within the non-mentored group at 29 each. The predominance of non-tenured faculty in the FRD program could have potentially resulted from the program’s leaders’ focus on recruiting new and lesser experienced faculty into the program. Figure 4.8 on the next page illustrates treatment condition group differences based on race, gender, rank, and tenure covariates.

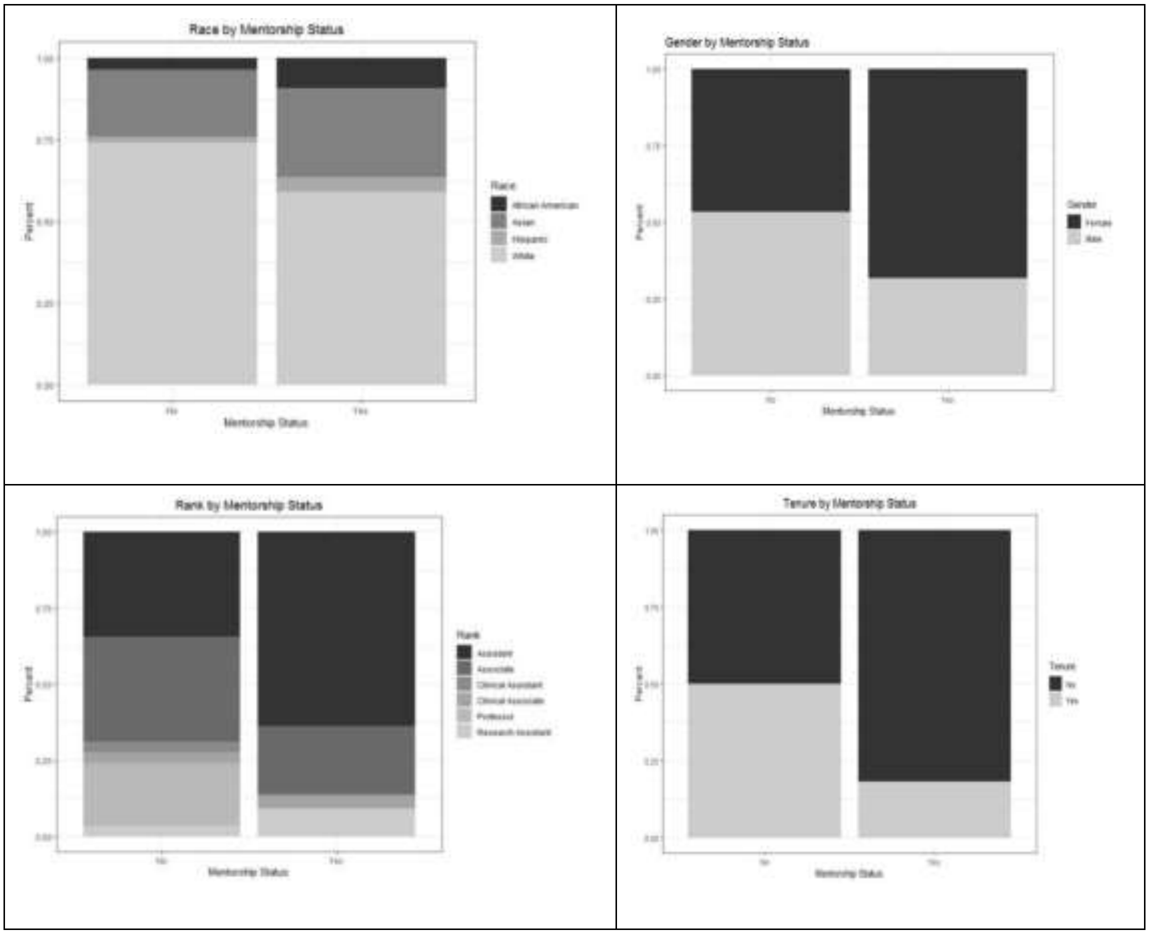

Figure 4.8 Race, gender, rank, and tenure covariate treatment condition group differences

Experience Covariate—On average, non-mentored faculty had more years of experience, given that the mean value for FRD program participants was 3.83 years, while it was 7.31 years for the non-mentored faculty. The standard deviation of the non-mentored faculty is 6.74, while it is 3.43 for those mentored. The higher mean years of experience among non-mentored faculty potentially resulted from the program's focus on new and early career faculty and those faculty's likely desire to enroll in the program due to perceived benefit.

Publication Covariate—On average non-mentored faculty publish more than their mentored peers, 92.36 compared to 51.32 publications. The variation in standard

deviation between the groups is 111.31 for the non-mentored faculty and 45.07 for the mentored ones denoting an imbalance between the groups. The higher mean publication among non-mentored faculty could have resulted from a higher proportion of higher rank and tenured faculty among non-mentored faculty who possess more years of experience and have more time to publish. For example, professor rank faculty within the non-tenured group comprised 20.69% of that group's faculty compared to zero among the mentored group. Additionally, non-tenured (i.e., lower rank and lesser experienced) faculty comprised 81.82% of the non-mentored faculty compared to only 50% of the no-mentored group.

Citation Covariate—. On average, non-mentored faculty have an exponentially higher citation rate than those who participated in the FRD program, 3201.1, compared to 1,171.1. The citation standard deviation of the non-mentored faculty is 6,136.41 compared to 1,977.49 for FRD program participants. The higher mean citations among non-mentored faculty (3201.1) could have resulted from the same reasons mentioned for the higher mean publications, a higher proportion of higher rank, and tenured faculty among non-mentored faculty.

h-index *Covariate*—Faculty who did not participate in the FRD program, on average, have a higher mean *h*-index than those who participated in the program, 21.71 versus 14.18. The *h*-index standard deviation of the non-mentored faculty is 15.91 compared to 9.11 for FRD program participants. The higher mean *h*-index among non-mentored faculty could be a byproduct of an *h*-index measuring productivity and citation impact. Non-mentored faculty, on average, had higher publications and citations than

their mentored counterparts, 92.36 publications compared to 51.32, and 3201.1 citations compared to 1171.7, respectively.

Past Grants Covariate—On average FRD program, participants acquired fewer grants than the faculty who did not partake in the program, .091 compared to .045. The standard deviation in past grant acquisitions is .075 for non-mentored faculty compared to .29 for the mentored ones. At face value, this difference potentially emanates from the fact that the mentored group being less experienced (i.e., early-career faculty) than their non-mentored counterparts. Table 4.6 shows treatment condition group mean and standard deviation differences based on race, gender, rank, and tenure covariates.

Table 4.6 Experience, publication, citations, h-index, and grants covariate group differences.

	Non-Treatment		**Treatment**	
	Mean	*SD*	*Mean*	*SD*
Experience	7.31	3.31	3.43	6.74
Publication	92.36	51.32	45.07	111.31
Citations	3201.1	1977.49	1171.7	6136.41
h-index	21.71	15.91	14.75	9.11
Grants	0.45	0.75	.091	0.29

Response Condition Group Differences

Race Covariate—One African American, one Asian, no Hispanic, and five White non-mentored faculty received funding after the FRD program. No African American, Hispanic, or Asian FRD program participants obtained funding, while four White faculty

did. Similar differences are evident in the unfunded category. One of the non-mentored African American faculty did not secure grant funding representing 1.96% of non-mentored unfunded faculty, while two African American mentored faculty did not secure grant funding representing 11.11% of the mentored unfunded group. The difference between Asian non-mentored and mentored unfunded faculty is 11.57% proportionally. Eleven Asian non-mentored faculty did not secure funding among the non-mentored group, while a higher 33.33% proportion did not secure funding among mentored unfunded faculty. Although there was an equal number of Hispanic faculty in both the non-mentored and mentored unfunded groups, one in each, the proportion of mentored unfunded faculty is much higher, 5.56%, than that of the non-mentored unfunded faculty at 1.96%. The proportion of White unmentored unfunded faculty was much higher, 74.51%, than that of the mentored unfunded faculty, 50% within each group.

Gender Covariate—Three female and four male un-mentored faculty received funding compared to an even split among those who participated in the FRD program, two each. More male un-mentored faculty received funding than their female counterparts, 57.14% compared to 42.86%, while there were equal proportions (i.e.,50/50) among both genders mentored in the group. There were similar differences between un-funded FRD program participants and their un-mentored counterparts; 24 female and 27 male non-mentored faculty did not secure funding representing 13 females and five males among the FRD program participants. The proportional difference among the non-mentored unfunded faculty was 47.06% female versus 52.94% males, while it was much larger, 72.22% females versus 27.78% males in the mentored group. The higher proportion of female faculty in the mentorship program, 68.18%, than the non-

mentored group, 46.55, could explain the higher proportional difference in the mentored group.

Rank Covariate—Assistant rank faculty who secured funding represented 57.15% of the non-mentored group while representing a much higher 75% of the mentored group. Two associate rank faculty secured funding in the non-mentored group compared to none in the mentored group. One clinical assistant professor secured funding in the non-mentored group compared to none obtained in the mentored group. Neither group's clinical Associate nor professor rank faculty secured any funding, while one mentored research assistant professor acquired funding compared to none in the non-mentored group.

Tenure Covariate—Five non-tenured and two tenured faculty within the non-tenured non-mentored group acquired funding compared to 4 non-tenured mentored faculty acquiring 100% of the same category's grants. Non-mentored tenured and non-tenured faculty were almost split evenly in the un-unfunded category, 47.06% and 52.94%, respectively, while there was a higher proportion of non-tenured unfunded mentored faculty, 77.78% compared to their tenured mentored counterparts at 22.22%. Figure 4.9 illustrates race, gender, rank, and tenure covariate response condition group differences.

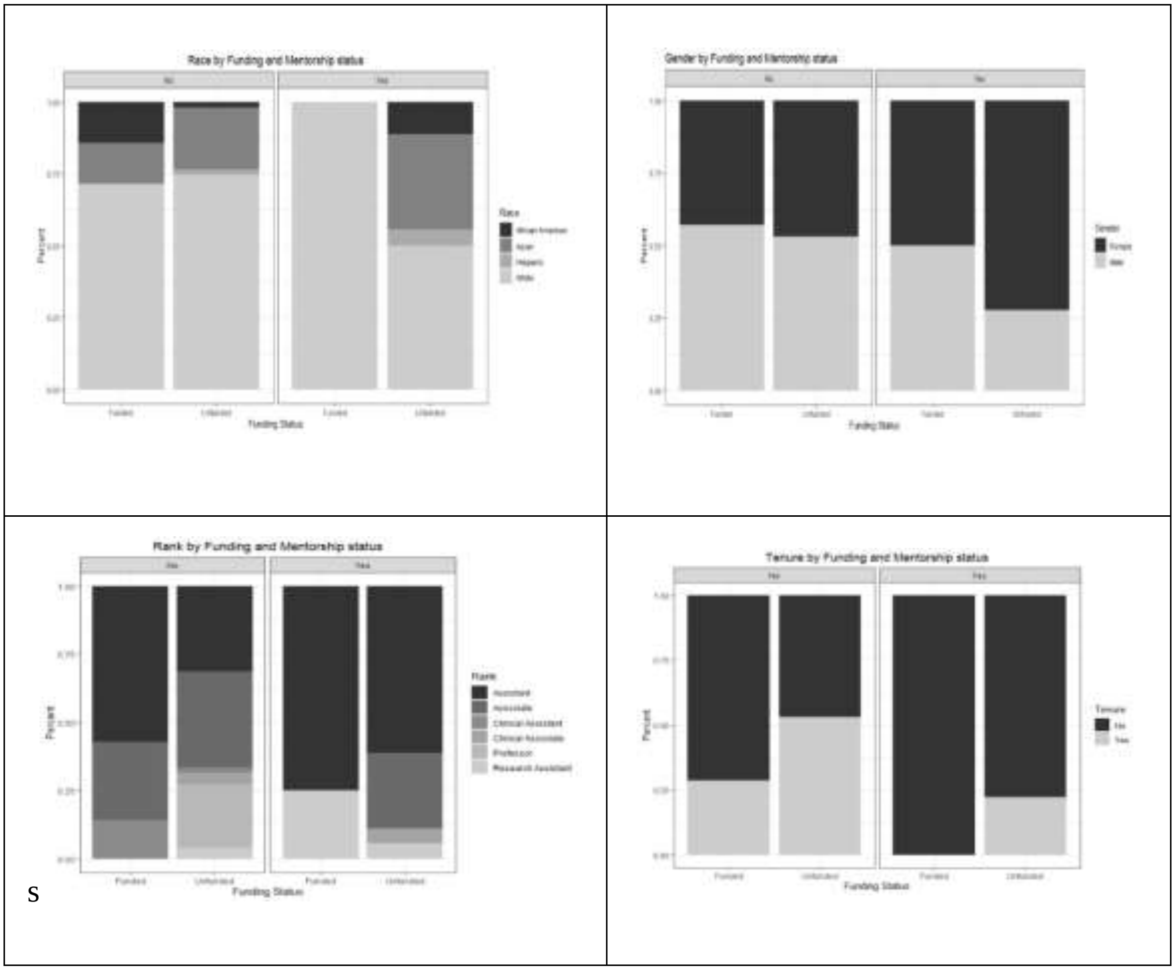

Figure 4.9 Race, gender, rank, and tenure covariate response condition group differences

Experience, Publication, and Citation Covariates—Mean years of experience for non-mentored funded faculty was almost quadruple mentored faculty mean years of experience, 4.54 compared to 1.18. The mean years of experience for un-funded non-mentored faculty are also higher than that of unfunded mentored faculty, 7.70 than 4.42. The average number of funded non-mentored faculty publications is more than double that of funded mentored faculty, 102.9 compared to 43.0 publications. Similarly, the average number of un-funded non-mentored faculty is more significant, 90.92 compared to 53.17. The mean number of citations for funded un-mentored faculty is slightly higher

than their funded mentored counterparts, 1890.6 compared to 1515, while the difference in the un-funded category is much higher, 2727.5 citations on average for non-mentored faculty compared to 829.3 citations for mentored ones.

h-index and Grants Covariates—The mean difference in h-index between funded non-mentored and mentored faculty is 7.3 points, 20 compared to 13.3, respectively. The mean difference between unfunded non-mentored and mentored faculty is slightly higher at 10.7 points, 25.5 compared to 14.8, respectively. There is also a higher degree of dispersion between the funded and unfunded group differences with a 2.8 difference in standard deviation among the funded group compared to 7.1 among the unfunded one. On average funded non-mentored faculty acquired more grants than their funded mentored counterparts, mean 0.29 for the first and zero for the latter. Among the unfunded category, the mean number of past grants was 0.47 among non-mentored faculty compared to .011 for those who participated in the FRD program. Table 4.7 highlights experience, publication, citations, h-index, and grants covariate response condition group differences.

Table 4.7 Experience, publication, citations, h-index, and grants covariate group differences

	Funded		**Un-Funded**	
	Mean	*SD*	*Mean*	*SD*
Experience				
Un-Treated	4.54	4.22	7.70	4.42
Treated	1.18	0.86	6.96	3.51
Publication				
Un-Treated	102.9	85.9	90.92	114.98
Treated	43.0	18.5	53.17	49.28
Citations				
Un-Treated	1890.6	1927.8	3380.9	6496.8
Treated	1515.0	2261.7	1095.4	1973.3
h-index				
Un-Treated	20.0	11.5	21.9	16.5
Treated	13.3	8.7	14.4	9.4
Grants				
Un-Treated	0.29	0.49	0.47	0.78
Treated	0.00	0.00	.011	0.32

Univariate Analysis

The complexity of treatment assignment into the FRD program is evident in a combination of self-and administrator selection. Most participants self-selected into the

program, while the administrators excluded others because they were deemed too early in their career track for inclusion. When estimating causal effects using nonexperimental data, the fundamental assumption is that treatment assignment is unconfounded, given the covariates used in the matching process (Rosenbaum & Rubin, 1983). For this assumption to come to fruition, it is essential to include any covariates related to treatment assignment and the outcome. However, since matching could typically occur for many outcomes, the most critical covariates to include are those related to treatment assignment. Covariates serve as the predictors of participation in the FRD program (0/1) and facilitate generating propensity scores via logistic regression. The probability of treatment (i.e., propensity score) facilitates balancing between non-mentored and mentored faculty groups conditional upon the covariates' multivariate distribution (Stuart & Rubin, 2008a). Therefore, the inclusion or exclusion of critical covariates affects the accuracy of inferences regarding the FRD program's effects (Brookhart et al., 2006; Steiner et al., 2010).

Critical covariates include those related to self-selection into the program (i.e., mentorship) and the outcome of interest (i.e., Funding) (Stuart, 2010). For example, if self-selection into the FRD program is related to race, gender, or rank, these factors are potentially influential covariates. Conversely, variables unrelated to self-selection into the program or its’ expected outcome are not significant unless they serve as proxies for related covariates. Consequently, Stuart and Rubin (2008) recommend including a broad set of covariates despite some having association with self-selection and proxy covariates, while others not necessarily the outcome. This process is especially crucial because past research by Austin et al. (2007) indicated that including covariates related to

both the intervention and the outcome resulted in the least bias while excluding significant covariates related to both intervention and treatment resulted in bias.

It is also crucial in deciding on covariates to consider theoretical relevance for self-selection into the FRD program (Brookhart et al., 2006; Stuart, 2010). For example, faculty rank may be essential to include as a covariate if the primary purpose of an FRD program is to develop new and early-career faculty, or that new and early-career faculty feel socially pressured to enroll in the program due to perceived notions that the college's leadership expect them to do so. Accounting for covert covariates such as faculty predisposition for enrollment into the FRD program is essential to reducing bias while ignoring them potentially increases it. Ignoring covariates' nature and theoretical considerations opens the door to different comparison groups based on unmeasured covariates.

Another critical consideration in covariate selection is covariate measurement reliability. Steiner et al. (2011, p. 230) state, "In general, the less reliable the measurement of constructs, the less they are collectively able to reduce bias." Covariate reliability is vital to model stability, which influences inference about the FRD program's effects on its desired outcome. However, "Although less reliable covariate scores are not ideal, such scores from a measure strongly related to selection bias may be more effective at reducing bias than highly reliable scores from a measure unrelated to selection bias" (Harris & Horst, 2016, p. 3). Including unreliable and theoretically irrelevant covariates consumes valuable degrees of freedom (df's), which increases the standard error and reduces precision. Simply put, the more covariates, the higher the chance some are significant due to random chance. For example, introducing ten irrelevant covariates

would still result in a 40% probability of one of them being significant due to random chance (Dranove, 2012). Similarly, the correlation of an irrelevant covariate to a valid one may be due to random chance, making it appear insignificant, causing the researcher to remove it from the model.

Building the logistic regression model requires judicious consideration of covariate selection. A balance between complexity (i.e., good fit of the data) and simplicity (i.e., guarding against overfitting) facilitates a parsimonious model that is easy to interpret. Hence, this study employed a purposeful covariate selection process by utilizing univariate analyses to identify critical covariates, one covariate at a time. Each analysis fitted a logistic regression with one covariate and analyzed the fits by examining the estimated coefficients, standard errors, and the likelihood ratio test for the coefficient's significance. As a rule of thumb, the covariates whose p-value is <0.25 and ones with known theoretical relevance were selected. The reasoning behind using a <0.25 cut-off point is that the more traditional 0.05 cut-off point can fail to identify known critical covariates (Bendal & Afifi, 1977; Mickey & Greenland, 1989). The analyses utilized both Wald and likelihood ratio tests since the former tend to be less reliable with smaller sample sizes than the latter.

The p-values for the Wald and Likelihood ratio tests for the covariate "race" slightly exceeded the <0.25 threshold (i.e., 0.26) on the treatment condition while exceeding it at a higher level on the response condition; 0.34 on the Wald test, and 0.31 on the Likelihood ratio test. Although the response condition univariate analysis exceeded the <0.25 threshold value, Ginther et al. (2011) investigated the association between investigators' self-identified race and the probability of obtaining grant funding

and found that Asian and African American applicants were 4 and 13 percentage points less likely to be funded by the NIH, respectively, giving credence to including the race covariate in the model. Similarly, the p-values for the “gender” covariate treatment condition were much less than the <0.25 threshold; 0.09 for the Wald test and 0.08 for the Likelihood ratio test, while exceeding the response threshold condition 0.62 on both tests. Including the gender covariate, albeit exceeding the <0.25 threshold on the response condition, is prudent since research has suggested its essential. For example, Waisbren et al. (2008) found significant gender differences in the mean number of grant submissions (women 2.3, men 2.7), success rates (women 41%, men 45%), and award rates (women $98,094, men $125,000) between 2480 female and male investigators when evaluating differences in the acquisition of research grant based on 6319 grant submissions.

The p-values for the Wald and Likelihood ratio tests for the covariate “rank” did not exceed the <0.25 threshold on both the treatment and response conditions. The Wald test p-values were 0.12 and 0.09 for treatment and response conditions, while there were slightly lower for the Likelihood ratio, 0.11 and 0.10, respectively. Further strengthening the inclusion of the rank covariate in the model are findings by Pagel and Hudetz (2015), which pointed to how NIH grants and support amounts depend on academic rank. Their findings also highlighted a statistically significant relationship ($P < 0.0001$) between investigator years of activity (i.e., experience level) and both NIH grant acquisition and amount of support. Their findings strengthen this univariate experience covariate analysis, which did not exceed the <0.25 threshold on both treatment and response conditions. Treatment condition Wald and Likelihood ratio p-values were 0.02 and 0.01 for each, while they were 0.06 and 0.02 respectively for the response condition.

The p-values for the Wald and Likelihood ratio tests for the covariate "tenure" did not exceed the <0.25 threshold. Wald and Likelihood ratio test p-values were 0.01 for the treatment condition and 0.11 and 0.08 for the response condition. Tenure covariate inclusion, irrespective of not exceeding <0.25 the threshold, is also warranted theoretically. Research by Conn et al. (2005) found that tenure and promotion were contingent on submitting grant proposals internally and externally and successfully acquiring such grants. For example, the college's leadership in that study incentivized grant submission and acquisition by increasing or decreasing teaching loads. The leadership reduced teaching and service responsibilities proportional to external salary support (i.e., grant funding) and developed "An explicit workload model calls for less-protected research time when faculty persistently demonstrated non-productivity." (Conn et al., 2005, p. 228).

Although the "publication" covariate p-values did not exceed the <0.25 threshold for the treatment condition, 0.11 Wald test, and 0.04 for the Likelihood ratio test, they did for the response condition coming in at almost 1 for all. Similarly, the "citations" covariate treatment condition Wald and Likelihood p-values did not exceed the <0.25 cut-off point at 0.11 and 0.03, respectively, while exceeding it on the response condition threshold 0.57 for the Wald test and 0.48 for the Likelihood ratio test. Likewise, the "*h*-index" covariate response condition p-values exceeded the <0.25 threshold at 0.61 for both the Wald test and 0.59 for the Likelihood ratio test. Conversely, the *h*-index covariate's treatment condition p-values did not exceed the <0.25 threshold, equaling 0.05 for the Wald test and 0.02 for the Likelihood ratio test. These mixed results, which might warrant the exclusion of these covariates at face value, are not warranted since past

research such as Ali et al. (2010); Pagel and Hudetz (2015) found faculty with greater scholarly output (i.e., publication, citations, and *h*-index) secured more grants than those with lesser scholarly productivity.

The p-values for the “grants” covariate treatment condition were 0.05 for the Wald test and 0.02 for the Likelihood ratio test, not exceeding the <0.25 threshold. Contrarywise, the p-values for the response condition did exceed the <0.25 threshold; 0.52 for the Wald test and .049 for the Likelihood ratio test. Including the grants covariate is essential given past research that showed that previous research grant acquisition played a pivotal role in securing competitive research grants (Ali et al., 2010; Pagel & Hudetz, 2015). Specifically, they found that “The percentage of faculty at an institution having already won competitive research grants, having journal publications, and whose publications have garnered citations, are also positively correlated with the dollar amount awarded" (p. 171). Table 4.8 below summarizes the Wald test and Likelihood ratio test p-values for treatment and response conditions on the Race, Gender, Rank, Experience, Tenure, Publication, Citations, *h*-index, and grants covariates. The asterisk indicates significance at the <0.25 level, as discussed earlier.

Table 4.8 Covariate P-Value results

Covariate	Wald-Test P-value		Likelihood Ratio P-value	
	Treatment	Response	Treatment	Response
Race	0.26	0.34	0.26	0.31
Gender	0.09*	0.62	0.08*	0.62
Rank	0.12*	0.09*	0.11*	0.10*
Experience	0.02*	0.06*	0.01*	0.02*
Tenure	0.01*	0.11*	0.01*	0.08*
Publication	0.11*	0.99	0.04*	0.99
Citations	0.11*	0.57	0.03*	0.48
H-index	0.05*	0.61	0.02*	0.59
Grants	0.05*	0.52	0.02*	0.49

* Met the <0.25 threshold

Logistic Regression Model Development

DAGs—Identifying a causal effect requires isolating the association between treatment and outcome. The directed acyclical diagram (DAG) in Figure 4.10 illustrates the confounded relationship between the treatment (i.e., Mentorship) and outcome (i.e.,

Funding) variables. The arrows in the DAG transmit associations between a participant's race, gender, rank, years of experience, tenure status, number of publications, number of citations, *h*-index, past grant acquisition (i.e., confounders), and the treatment and outcome variables. The DAG indicates that adjusting or conditioning for these confounders in the regression model is necessary to isolate the causal link between the treatment and outcome variables.

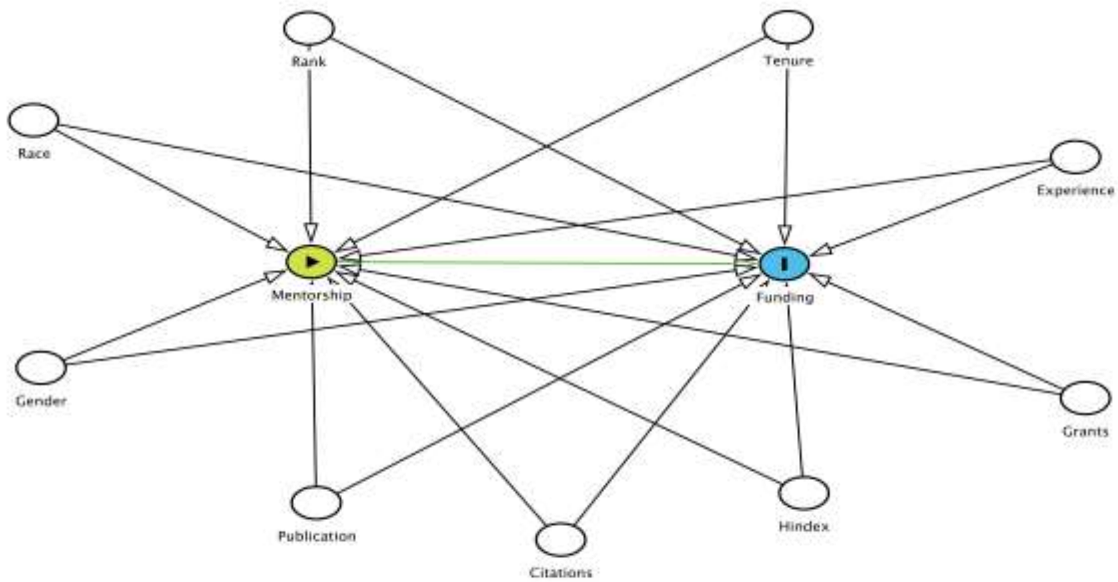

Figure 4.10 Directed acyclical diagram (DAG)

Given all the arrows (i.e., relationships), the model implies conditional independencies (i.e., no correlation) should exist between the confounding covariates when examining the existing data. However, Figure 4.11, which provides a heatmap of the spearman correlation tests' matrix results between the presumed covariate implied conditional independencies, paints a different picture. According to Swinscow (1997), for absolute values of r; 0-0.19 denotes a very weak correlation, 0.2-0.39 denotes a weak one, 0.40-0.59 denotes a moderate one, 0.6-0.79 denotes one that is strong, and 0.8-1 denotes a very strong correlation. As shown in the figure, several testable implications

fail the independence test, indicating a moderate to very strong correlation. There is a moderate correlation between rank and experience covariates (ρ=0.59), the rank and publication (ρ=0.52) covariates, the rank and citation (ρ=0.51) covariates, and the rank and *h*-index (ρ= 0.52) covariates. Similarly, there are moderate correlations between the experience and tenure (ρ=0.55) covariates; and the tenure and publication (ρ=0.49), citation (ρ=0.44), and *h*-index (ρ= 0.45) covariates. There is a strong correlation between rank and tenure (ρ=0.74) covariates, and lastly, there is a very strong correlation between the publication and citation (ρ=0.82), *h*-index (ρ= 0.88) covariates, and the citation and *h*-index (ρ= 0.96) covariates.

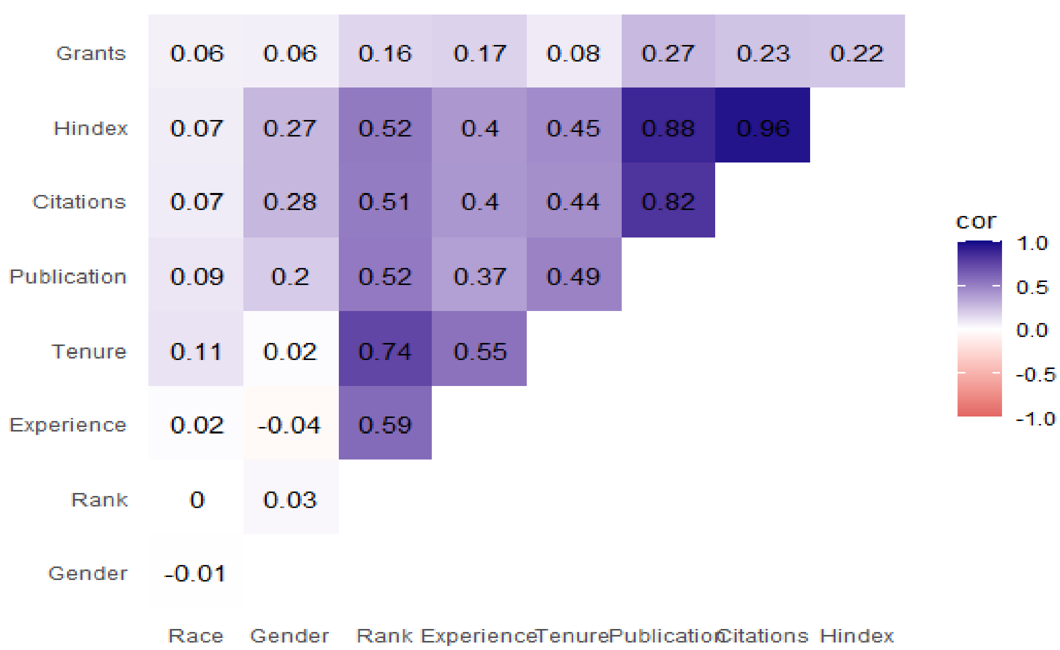

Figure 4.11 Covariate conditional independencies heatmap

These moderate to strong correlations between the covariates are in line with past research. Ence et al. (2016) detected a correlation between faculty rank and experience (i.e., career duration), where they reported that a longer career duration correlated independently with an increased probability of senior academic rank ($p<0.001$). They

also reported that a faculty member's years of experience play a role in the promotion and tenure decisions. The moderate correlation between rank record and publication is in line with past research (Amara et al., 2015; Emden, 1998; Hesli et al., 2011; McGrail et al., 2006), which has consistently demonstrated that publication record is a strong predictor of attaining promotion to the following faculty rank. Similarly, research (Ence et al., 2016; Orhurhu et al., 2020) detected an association between faculty rank and citation records, and h-indices. The correlation between tenure and publication, citation, and h-index has merit in the literature (Delgadillo, 2016; McGrail et al., 2006; Vannini, 2006). The adage “Publish or Perish” is still valid. These metrics are a quantifiable demonstration of research productivity and tied tenure and promotion. Educational attainment, seniority, research, teaching, and service are some of the criteria used in determining a faculty member’s rank. Customarily, the typical faculty entry-level rank on the tenure-track is Assistant Professor, albeit dependent on the field of study and institution. Promotion to Associate Professor rank or termination occurs after spending 6-8 years in the Assistant professor rank (Lackritz, 2004), which would explain the strong correlation between faculty rank and tenure. Lastly, the very strong correlation between publication records and both citations and h-index is logical. Publications lead to citations, and both are factors in the calculation of the *h*-index, which is a function of the total number of publications and the total number of cited papers an author has.

The substantial correlations between confounders assumed to be independent in the first DAG highlighted issues with the original DAG specification. This issue required generating a revised DAG by refining the theory underlying its’ development. The revised DAG illustrates the moderate to very strong failed conditional independencies.

Figure 4.12 illustrates a revised DAG, which includes arrows illustrating these moderate to very strong correlations. The new DAG highlights the original associations that met the conditional independence threshold and those that did not. The curved lines in the figure denote the failed testable implications of independence between covariates; rank and experience, publication, citations, and *h*-index; experience and tenure, tenure and publication, citations, and *h*-index; rank and tenure; publication and citations and *h*-index; and citations and *h*-index.

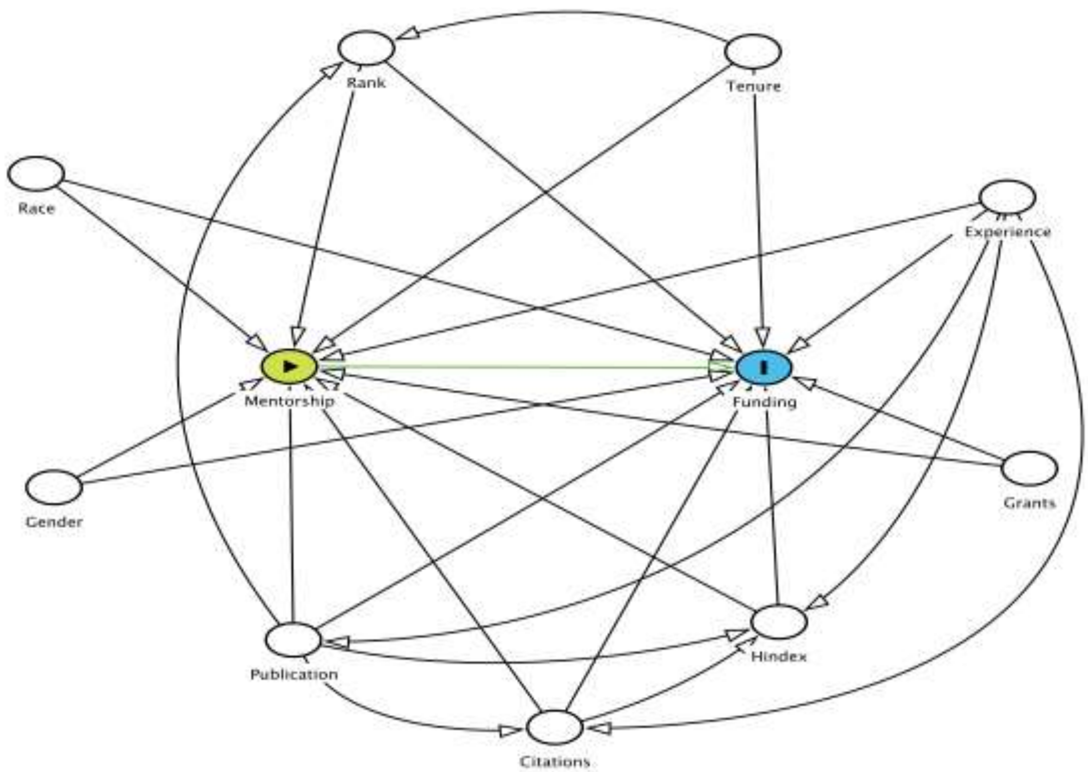

Figure 4.12 Revised testable implication DAG

The revised DAG requires testing for three more conditional independencies; rank's association to experience given publications, rank's association to citations given publications, and rank's association to *h*-index given publications, which required testing via logistic regression models.

- Rank ⊥ Experience | Publication

- Rank ⊥ Citations | Publication
- Rank ⊥ *h*-index | Publication.

All logistic regression models yielded statistically insignificant results, confirming conditional independence. Model 1 (i.e., Rank ⊥ Experience | Publication), which tested the association between the rank and experience covariate given the publication covariate, yielded insignificant (t=1.73), (p=0.08) results, which means the two are independent. Model 2 (i.e., Rank ⊥ Citations | Publication), which tested the association between the rank and citations covariate given the publication covariate, yielded insignificant (t=0.05), (p=0.96) results, which means the two are independent. Model 3 (i.e., Rank ⊥ *h*-index | Publication), which tested the association between the rank and *h*-index covariate dependence, given the publication record, yielded insignificant (t=1.26), (p= 0.21) results, which means the two are independent. The new DAG testable implications imply that nodes are statistically independent of each other and do not transfer associational information, indicating conditional independence between all confounding covariates.

Logistic Regression Model Evaluation

Randomization allows the isolation of each DAG arrow (i.e., relationship) between the treatment and outcome covariates and confounders. Using the DAG to isolate a pathway between the treatment and outcome, in essence, mimics RCTs' characteristics on the measured baseline characteristics identified in the DAG (i.e., Race, Gender, Rank, Experience, Tenure, Publication, Citations *h*-index, and Grants). However, searching for a model that explains the most outcome variation with the fewest variables is desirable for prediction. Decisions about the inclusion or exclusion of covariates often

hinge on if the covariate significantly improves or reduces model fit. The significance level is usually arbitrarily defined by some cutoff α-level for the coefficient p-value. As discussed earlier in the univariate analysis, some researchers usually set this cutoff point at 0.05, but it should preferably be set much higher (e.g., 0.25) for confounder selection (Dales & Ury, 1978; Greenland, 1989). Such criteria are equivalent to assessing whether the covariate explains a significant proportion of the residual variation or the outcome variation remaining given the model's established variables. To ensure that the logistic regression model is as simple as possible (i.e., parsimonious) while providing an adequate fit of the data that produces the most precise (i.e., least specious) effect estimate attainable from the available data, a likelihood ratio test for the logistic regression model using all identified confounders to compare the goodness of its's fit ensued. Using the *test="LRT"* command in R statistical software allows for comparing two hierarchically nested models (i.e., comparing the complex model to a simpler one) to determine whether adding complexity to the logistic regression model makes it significantly more accurate. In essence, this test informs if it is beneficial to reduce parameters to the model or continue with existing ones.

The Akaike information criterion (AIC) is a mathematical method for evaluating how well a model fits the data generated. It compares different possible models and determines which one is the best fit for the data. According to the AIC metric, the best-fit model explains the most significant variation using the fewest possible independent variables. Each model's maximum likelihood estimate (i.e., how well the model reproduces the data) and the number of independent variables used in the model calculate the AIC, a relative measure. Lower AICs are indicative of a more parsimonious model

relative to a model fit with a higher one. The Akaike's Information Criterion correction (AICc) statistic corrected for the small sample size since the FRD program included only 22 participants. According to the likelihood ratio test, the best covariates to include in the model are the experience, publication, and citations covariates, as illustrated by Table 4.9.

Table 4.9 The likelihood ratio test AIC and AICc statistics

Model	**AIC**	**AICc**
Race+Gender+Rank+Experience+Tenure Publication+Citations+Hindex+Grants	68.67	71.86
Race+Rank+Experience+Tenure+Publication+Citations+Hindex+Grants	66.71	69.28
Race+Experience+Tenure+Publication+Citations+Hindex+Grants	65.43	67.46
Race+Experience+Tenure+Publication+Citations+Hindex	64.35	70.11
Race+Experience+Tenure+Publication+Citations	63.05	71.86
Experience+Tenure+Publication+Citations	62.61	68.11
Experience+Publication+Citations	62.48	63.01

The likelihood ratio test indicates that only the experience, publication, and citations covariates should comprise the model. However, since including covariates possessing theoretical relevance is recommended (Brookhart et al., 2006; Stuart, 2010), and because the univariate analysis demonstrated the statistical significance of covariates, an all covariate logistic regression model comprised the one used in the analysis. A sensitivity analysis using only the experience, publication, and citations covariates will follow the initial analysis accounting for all confounders to test the findings' robustness.

Treatment Effect Estimation Via Matching

Confounder imbalance between the treated and control group opens the door to a higher probability of model dependence. A regression models' causal estimates can fluctuate markedly depending on a researcher's discretion (i.e., different specifications and assumptions for the model) (King & Zeng, 2007). Researcher discretion is critical because "When researchers ignore the uncertainty due to model dependence, scholarly works tend to have the flavor of merely showing that it is possible to find results consistent with ex-ante hypotheses." (King & Zeng, 2007, p. 231). Hence, ignoring the uncertainty inherent in model dependence in estimating causal effects can lead to bias. Researchers are increasingly using propensity scores (i.e., probability of treatment) to correct the imbalance in covariate distributions in observational studies in the absence of randomization to address the problem of model dependence (Pirracchio et al., 2012). This method allows researchers to move towards unbiased estimates of causal effects and facilitates making the un-confoundedness assumption (i.e., controlling all variables affecting both treatment and outcome). We can use matching techniques to pair similar treated and control observations and make the un-confoundedness assumption that if we examine faculty that are pretty much identical based on the confounders, and one participated in the FRD program while the other did not, then that participation choice was random. Because the DAG indicated that Race, Gender, Rank, Experience, Tenure, Publication, Citations, *h*-index, and Grants could help cause both mentorship and funding, and confound that relationship, attempting to find faculty with similar values on these characteristics (i.e., confounders) that both participated and did not participate in the FRD program is prudent.

As a baseline measure, a logistic regression model that did not account for the identified confounders examined the relationship between the explanatory (i.e., mentorship) and response (i.e., funding) variables. As listed in Table 4.10, this estimate compares the average funding difference between those who participated and did not participate in the FRD program. Although the mentorship coefficient indicates that the likelihood/odds of acquiring grant funding are five times greater (i.e., e^1.62 = 5.05) with mentorship (i.e., participation in the FRD program), this estimation is statistically insignificant given the (0.48) p-value (i.e., there is a 48% probability that this result is due to random chance). This model, which did not control DAG confounding covariates, provides a baseline measure for comparison purposes with models that will take those confounders into account (i.e., control them).

Table 4.10 The confounded logistic regression model

Predictor	**B**	**SE**	***t***	**p**	**95% CI**	
(Intercept)	0.14	0.40	-4.93	0.00	0.057	0.28
Mentorship	1.62	0.68	0.70	0.48	0.39	6.04

Using the *matchit()* function from the *MatchIt* R package in R statistical software, treatment, and control faculty were matched based on the identified confounders. Exact matching yielded no matches between the treated and control groups, leading to the utilization of the nearest-neighbor matching method, known as greedy matching, and the most commonly used (Thoemmes & Kim, 2011). This method requires running through the list of treated data points and pairing them with the datasets nearest qualifying control data unit. Pairing takes place disregarding other unit pairings; it is greedy because it does

not seek to optimize any criterion. The Mahalanobis distance, a useful multivariate distance metric that efficiently detects multivariate anomalies in highly imbalanced data sets (Chen, 2013), matched mentored and non-mentored faculty.

To enable multiple matching (i.e., one to many matching), the *replace = TRUE* argument was included, which increases the probability of matching between the treated and control data units. This argument is a logical value signifying the possibility of matching each control unit to more than one treated unit with the default being replace=FALSE. The *replace*=FALSE argument means that each control unit is only matched once (i.e., sampling without replacement). Using the argument *replace = TRUE* in the matched logistic regression model resulted in only 16 control units matching treated ones while setting the argument, *replace*=FALSE resulted in 22 control units matching treated ones as illustrated by Table 4.11. The lower match rate (i.e., 16 versus 22) was a function of the one-to-many matching, which caused multiple pairing between each untreated data point with more than one treated data point when using the *replace = TRUE argument*.

Table 4.11 One-to-one versus one-to-many matching

	***replace*=TRUE**		***replace*=FALSE**	
	Control	**Treated**	**Control**	**Treated**
All	58	22	58	22
Matched	16	22	22	22
Unmatched	42	0	36	0
Discarded	0	0	0	0

Table 4.12 lists the pre-post matching balance for each covariate. Nearest neighbor matching improved the balance for most covariates. The percent mean difference improvement for the covariates were race (53.55%), gender (100%), rank (57.85%), experience (90.55%), tenure (100%), publication (62.46%), citations (85.12%), *h*-index (79.46%), and grants (100%). The greatest improvement was in the gender, experience, tenure, and grants covariates and, to a lesser degree, in the citations and *h*-index covariates. The least mean difference improvement was evident in the race and rank covariates. Figure 4.13 provides a visualization of the covariate pre-post matching covariate balance. It provides a distributional balance of the adjusted (i.e., matched) and unadjusted (i.e., unmatched) data for each covariate.

Table 4.12 Matching percent balance improvement

	μ Diff	**eQQ Med**	**eQQ Mean**	**eQQ Max**
Race	53.53	0.000	69.44	33.33
Gender	100.0	00.00	100.0	100.0
Rank	57.85	100.0	11.61	-100.0
Experience	90.55	60.00	69.84	90.88
Tenure	100.0	0.000	80.36	0.000
Publication	62.46	61.29	73.44	74.61
Citations	85.12	79.52	77.78	83.79

h-index	79.46	45.45	55.18	47.22
Grants	100.0	0.000	72.50	0.000

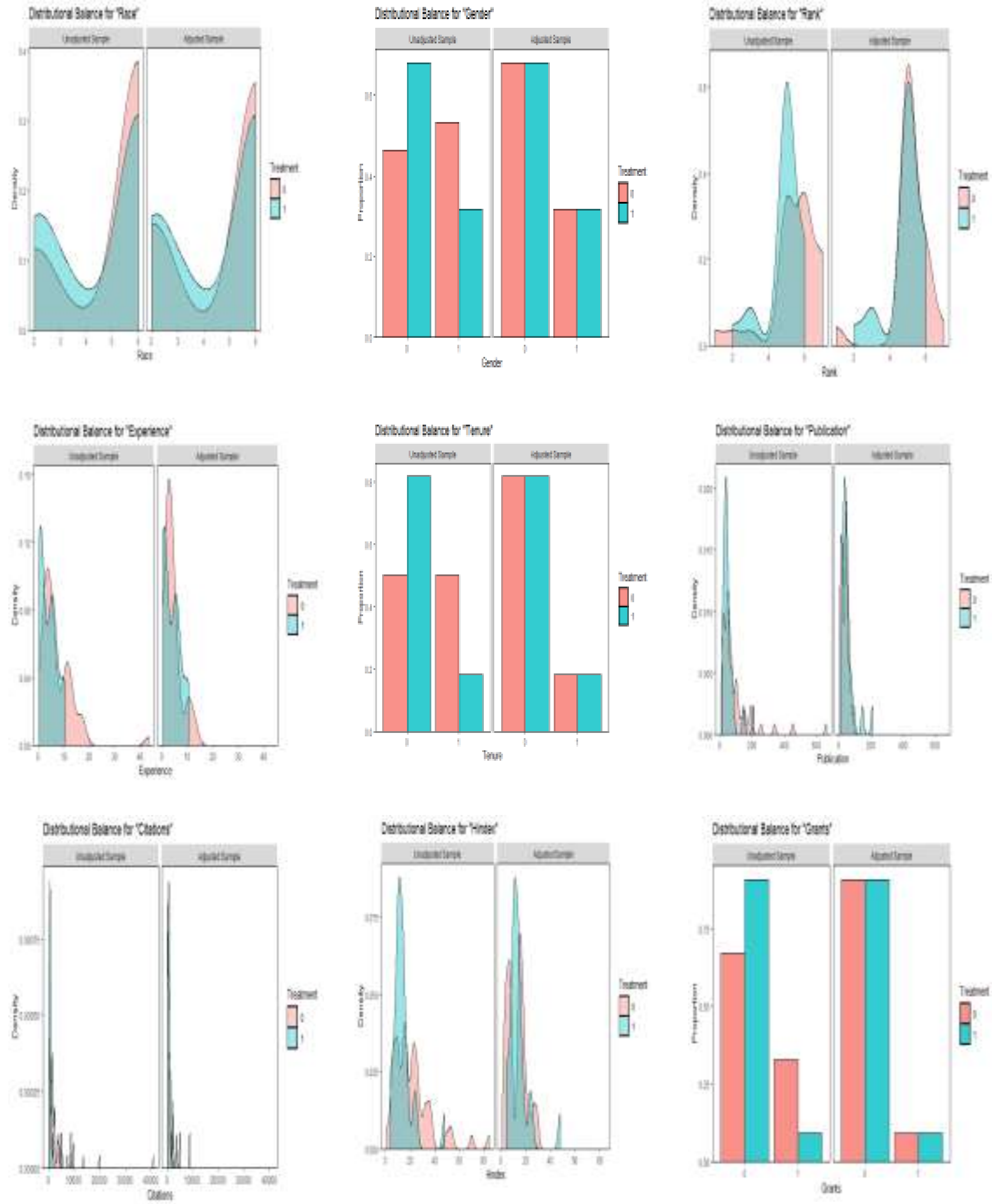

Figure 4.13 Covariate balance

The *match.data* argument was used to create a new trimmed dataset comprised of nearest neighbor matched data with replacement (i.e., *replace*=TRUE) considering the confounders. A logistic regression model estimated the treatment effect using the

matched dataset. The model estimated the likelihood of each faculty acquiring grant funding based on their mentorship status, given the confounders as represented by this equation.

$$Funding = Log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X_{Mentorship}$$

Although the results coefficient illustrated in Table 4.13 indicates that the likelihood/odds of acquiring grant funding are 4.76 times greater (i.e., e^1.56 = 4.76) with mentorship (i.e., participation in the FRD program), this estimate is statistically insignificant given the (0.64) p-value (i.e., there is a 64% probability that this result is due to random chance than attributable to the FRD program).

Table 4.13 Matched regression model

Predictor	**B**	**SE**	***t***	**p**	**95% CI**	
(Intercept)	0.14	0.76	-2.57	0.01	0.02	0.51
Mentorship	1.56	0.94	0.47	0.64	0.26	12.4

The likelihood of funding given mentorship status in the matched regression model is slightly less than the initial confounded model that did not account for any confounding covariates (1.56 versus 1.62). The difference may be because the multiple matched (i.e., one to many matching) observations are being over-counted and have too much importance in the model or remaining imbalance in the data. Fortunately, the *matchit()* package provides a weights column allowing the scaling down of over-matched observations when running the model. The weights generated from the replacement argument (i.e., *replace*=TRUE) can reflect the frequency with which each control unit was matched and provide better matching between treated and control data points. The

sole purpose of these weights is to facilitate the scaling down of the imbalance occurring from overmatching.

Treatment Effect Estimation Via Weighting

The *weights=weights* argument was used in a new regression model to account for matching imbalance (i.e., over-and-under weighting). Applying the weights in the model yields a likelihood/odds of acquiring grant funding 9.2 times greater (i.e., e^2.22 = 9.20) with mentorship (i.e., participation in the FRD program). However, this result is also statistically insignificant at a p-value of 0.43, as illustrated in Table 4.14. Even after matching and weighing to improve balance, the likelihood of funding given mentorship while controlling for all confounders is still implausible (i.e., there is a 43% probability that this result is due to random chance versus due to the FRD program).

Table 4.14 Weighted regression model

Predictor	**B**	**SE**	***t***	**p**	**95% CI**	
(Intercept)	0.10	0.87	-2.65	0.00	-4.56	-0.89
Mentorship	2.22	1.03	0.77	0.43	-1.09	3.23

One drawback to matching, in general, is the discarding of a considerable portion of data. Hence, unmatched data exclusion from the final matched dataset, as evident in the matched regression model. This situation arises because the control group decreases to the same size as the treatment group. Potentially, more efficacious use of the propensity scores is keeping all observations in the dataset by weighting them according to their propensity scores as described by (Austin, 2011). This method is known as Inverse probability treatment weighting (IPTW).

Treatment effect Estimation via Inverse Probability Weighting

Imbens (2000) proposed IPTW as an alternative to matching. In this method, treated observations receive a $\frac{1}{P}$ weight assignment, while non-treated ones receive a $\frac{1}{1-p}$ weight assignment, where p is the treatments' probability. Each observation in this approach is assigned a weight of the inverse probability of the treatment received. Innately, treated data points resembling control ones receive more weight, and control data points resembling treated ones are also assigned more weight. Analysis can then proceed via weighted average values or regression with explanatory variables (which may or may not be the same variables as those used in the propensity model for treatment). These inverse probability weights generate via a two-step process: (1) generating propensity scores or the probability of receiving treatment given all the confounders, and then (2) using this unique formula, $\frac{Treatment}{Propensity}$ - $\frac{1-Treatment}{1-Propensity}$, to convert those propensity scores into weights. Once the inverse probability weights weights are generated, they can be incorporated for causal estimation via a logistic regression model.

The *mutate()* argument facilitated creating a new column in the dataset for the inverse probability weights shown in Table 4.15. The Table, which illustrates the first six data points (i.e., mentored and non-mentored faculty), highlights how the inverse probability weight (IPW) corrects the matching imbalance and preserves all data. For example, the first and second faculty had a 6% probability of being mentored based on the confounders and were not mentored (i.e., the expected result), resulting in a low IPW (1.0). The third and fourth faculty had a 26% and 33% probability of being mentored and were not, a slightly higher probability of mentorship, resulting in a higher IPW (1.3 and 1.5). The fifth faculty had a 15% probability of mentorship yet was mentored (i.e., an

unexpected result), resulting in a much higher IPW (6.6). The sixth faculty had a 51% probability of mentorship and was not, resulting in a (2.0) weight due to the 50/50 probability of mentorship.

Table 4.15 Weighted regression model sample

Faculty	**Mentorship Status**	**Propensity Score**	**IPW**
1	0	0.06	1.0
2	0	0.06	1.0
3	0	0.26	1.3
4	0	0.33	1.5
5	1	0.15	6.6
6	0	0.51	2.0

Figure 4.14 provides a density plot of the distribution of inverse probability weights. At the bottom of the graph, the black lines are the data points (i.e., mentored and non-mentored faculty). Most of the weights are between 1 and 4, with much fewer larger weights between 4 and 10. Non-mentored faculty weights ranged on average ranged between 1 and 4 while mentored ones ranged much higher (i.e., 4-10), with most of the overlap in weights taking place between one and three.

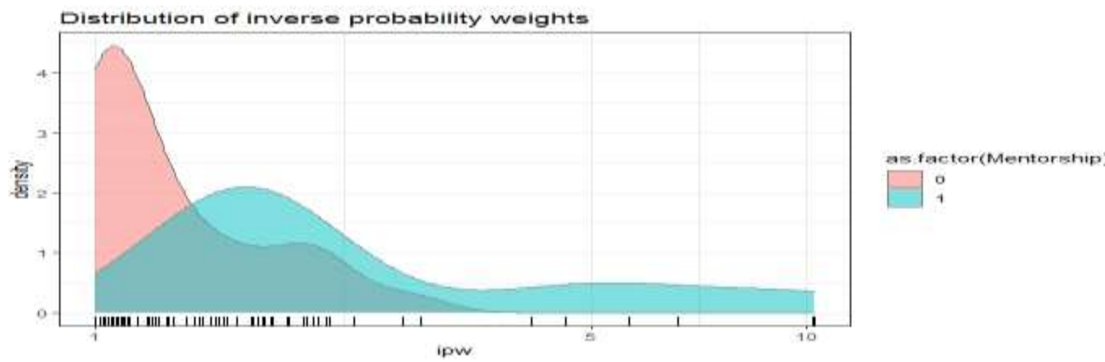

Figure 4.14 Distribution of inverse probability weights

IPTW preserves the entire dataset compared to the discarding of data, which happens in matching. The IPW column, which accounted for the confounders, was used in a logistic regression model to estimate the FRD program's treatment effect. The results of this model yield a likelihood/odds of acquiring grant funding 1.02 times greater (i.e., e^ 0.02= 1.02) with mentorship (i.e., participation in the FRD program). The model's results are not statistically significant given a p-value of (0.98) (i.e., there is a 98% probability that the results of the FRD program are due to random chance), as illustrated in Table 4.16.

Table 4.16. IPTW regression model

Predictor	**B**	**SE**	***t***	**p**	**95% CI**	
(Intercept)	0.14	0.05	2.64	0.01	0.03	0.25
Mentorship	0.00	0.08	0.02	0.98	-0.15	0.16

Sensitivity Analysis

A new analysis conducted with just the experience, publication, and citations confounders tested the finding's robustness. The analysis examined how results were affected by a change to the regression models (i.e., including only the experience,

publication, and citations confounders). This analysis employed the same steps as in the previous one, which started with a naïve regression model that did not control confounders, followed by a matched regression model that used nearest neighbor matching followed by a matched regression model with weights concluding with a regression model utilizing IPTW. The same data matching process (i.e., matching with replacement) and inverse probability weight generation ensured comparison reliability. The results listed in Table 4.17 confirm the initial analysis results, in that there was no statistically significant effect for the FRD program. Altering the assumptions made by the initial analysis (i.e., the inclusion of all confounders) to one that only included the experience, publication, and citations confounders in the analysis did not lead to different final interpretations or conclusions, which confirms the robustness of the initial analysis. Based on these findings, there was insufficient evidence to reject the null hypothesis (i.e., there is no significant difference in grant acquisition between FRD program participants and those who did not participate in it).

Table 4.17 Model Comparisons

	Matching		Matching+Weights		IPW	
	All.Cov	**Exp.Pub.Cit**	**All.Cov**	**Exp.Pub.Cit**	**All.Cov**	**Exp.Pub.Cit**
Intercept	0.14	-2.56	0.10	-2.30	0.14	0.14
SE	0.76	1.04	0.87	0.93	0.05	0.05
t	-2.57	-2.47	-2.65	-2.48	2.64	2.49
p	0.01	0.01	0.00	0.01	0.01	0.01
Mentorship	1.56	1.06	2.22	0.80	0.00	0.00
SE	0.94	1.18	1.03	1.08	0.08	0.08
t	0.47	0.90	0.77	0.74	0.02	0.00
p	0.64	0.37	0.43	0.46	0.98	0.99
Num. Obs.	38	36	38	36	80	80
AIC	36.9	32.1	37.5	32.7	80.4	73.4
BIC	40.2	35.2	40.8	35.9	87.6	80.6
Log.Lik	-16.459	-14.034	-16.752	-14.354	-37.212	-33.707

One potential limitation of this study is the small sample size (n=22), which potentially could lead to null findings due to insufficient power in the study to detect an effect. The strength of treatment, background noise, and experimental design are

determinants of highly powered studies. The first two are usually out of the researchers' control, while the last is in their control. First, most program evaluation researchers, for example, are not in control of the treatment. They are not anteriorly involved in the design or administration of the program they aim to evaluate. This situation is prevalent in educational research, which rarely randomizes its interventions. Second, background noise affects outcomes across participants, increasing outcome variability. Income, for example, is a more variable outcome than a rare disease. It would be hard to detect the effect of an intervention that increased participant income by only 1% because of income's inherent high variability. Third, researchers can use study design to increase their studies' power by increasing the number of participants (i.e., large sample size), judiciously randomizing participants into treatment and control groups, and controlling confounding variables.

To detect an actual non-null treatment effect, researchers emphasize the statistical power of study design. Researchers conducting RCTs routinely assess potential studies' statistical power to justify resources' expenditure, given the probability of detecting a meaningful and statistically significant effect, and provide context relating to interpreting potentially null conclusions (Austin et al., 2015). RCT statistical power methodologies are easily performed and well documented (Rosner, 2015; Schoenfeld, 1983). Conversely, although described for observational studies, such methodologies are frequently oversimplified and necessitate readily unavailable information (Cornell, 1992). Complicating a priori sample size and power analysis on observational studies using IPTW is that weights are a function of the observed data, which are unknown before conducting the analysis. Hence, it prevents the estimation of standard errors requiring a

priori knowledge of the weights. The only option remaining is to conduct a post hoc power analysis.

Some researchers (Thomas, 1997) encourage conducting retrospective or post hoc power analysis on observed results attempting to interpret results, especially negative ones (i.e., null treatment effect). However, experts discourage such efforts and label them as inappropriate. Hoenig and Heisey (2001) showed that power is mathematically directly related to the p-value; therefore, calculating power once the p-value associated with a statistic is known adds no new information. Additionally, Levine and Ensom (2001) decry the inappropriateness of post hoc power analysis because of the inapplicability of probability to an observed result. They state that post hoc power analysis:

> can lead the reader to the incorrect conclusion that the β probability, based on 1 - the calculated power, is the probability that the observed result was a false-negative one. Thus, the stated or implied conclusion is that the effect may be real but that there were too few subjects in the study. This interpretation of the β value is incorrect because, like the p-value, it must be interpreted as both a conditional and a frequency probability (p. 3).

They also add that post hoc power analysis on observed data "will always lead to a low value for power. As there is no decision criterion, the investigator and reader are therefore no further ahead in interpreting the negative result." (Levine & Ensom, 2001, p. 407). They and other researchers (Goodman & Berlin, 1994; Smith & Bates, 1992) suggest using confidence intervals instead of conducting post hoc power analyses because the latter provides no meaningful method for evaluating negative results. Examining confidence intervals provides a range of consistent values with the data. If this includes a critical effect size, then a researcher would know that their study was uninformative. Calculating power retrospectively using the effect size seen in the study

almost guarantees a low power. This process provides circular argumentation for resurrecting one's hypothesis and concluding that the experiment was not big enough.

Based on the knowledge that experts discourage the use of retrospective post hoc power analysis and their recommendation to use confidence intervals to assess findings' vigor instead, an analysis ascertained whether the initial model, which included all confounders, and the one that only took the experience, publication, and citations confounders into consideration were robust. Since actual non-null treatment effects (i.e., statistically significant results) yield confidence intervals that exclude zero, a plot assessed if any model's confidence intervals did so. Figure 4.15 illustrates each model's coefficient and corresponding 95% confidence intervals. It starts with the naïve model that did not account for any confounders, then the initial analysis models (i.e., matching, matching+weights, and IPW) that accounted for all confounders, and the sensitivity analysis models (i.e., matching_EPC, matching+weights_EPC, and IPW_EPC) that only accounted for the experience, publication, and citations confounders. As shown, all model confidence intervals cross zero, signifying an implausible actual non-null treatment effect (i.e., statistically significant result). Since the weighted datasets are generally bigger than the matched dataset (i.e., preserves all the data), the variance and the confidence intervals associated with the IPW estimations are expectedly smaller, as illustrated in the figure.

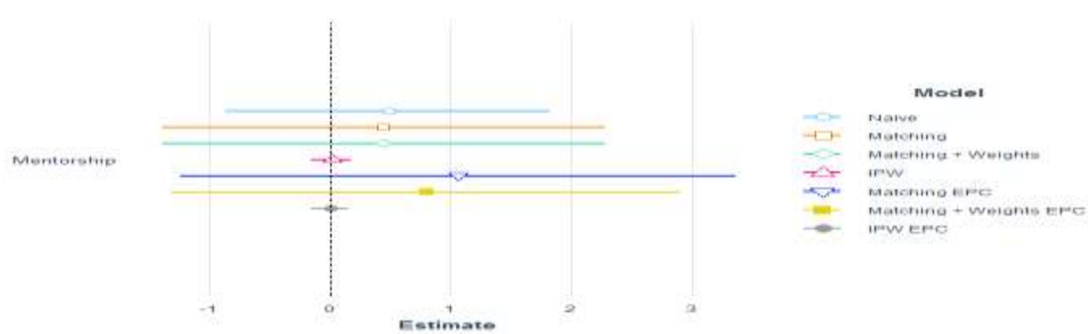

Figure 4.15 Model comparisons

Discussion

This article aimed to assess whether faculty recipients of FRD programs are more likely to increase their grant acquisition to demonstrate the CERTi model's applicability. The article's analysis determined that in the case of the examined FRD program, there was no statistically significant difference in grant acquisition between the program's participants and their non-participating counterparts within the college. This result was consistent across the many models used to assess the program's efficacy (i.e., propensity score matching, weighting, and inverse probability weighting). Moreover, to ensure the article's findings' robustness, a sensitivity analysis that only included statistically significant confounders corroborated the initial analysis across the same models. Lastly, an analysis that examined confidence intervals for all models (i.e., with all confounders and only significant ones) confirmed the null treatment effects of the program.

Assessing FRD program efficacy is critical in light of the hyper-competitive grant acquisition institutional environment many HEI leaders face. Such assessments provide these leaders with an answer to *what* transpired due to administering their FRD programs. However, what is even more critical for them is to understand *why* the result transpired. Such an understanding facilitates identifying unobservable confounders that potentially influence the desired outcome (i.e., grant acquisition). Researchers are increasingly linking the lack of implementation fidelity to poor intervention outcomes (Durlak & DuPre, 2008; Dusenbury et al., 2003; Lipsey, 2009; Mihalic, 2004). An intervention's implementation fidelity is crucial to its' validity. Since fidelity is closely related to the

statistical power of outcome analyses. “Failure to establish fidelity can severely limit the conclusions that can be drawn from any outcome evaluation” (Dumas et al., 2001, p. 39). Many researchers suggest measuring fidelity to ensure internal validity, circumvent compromising external validity, and maximize statistical power (H.-T. Chen & Rossi, 1983; C. E. Cook, 1998; Dumas et al., 2001; Maynard et al., 2013). More importantly, research (Lakin & Shannon, 2015) suggests that evaluating implementation fidelity leads to better future program design, leading to higher program fidelity. Supplementing analyses highlighting the robustness and validity of research with implementation fidelity measures is an efficacious approach to assuring research results and better inform HEI leaders’ decisions.

Assessing implementation fidelity is foundational to the CERTi comprehensive evaluation approach. The importance of assessing fidelity is precisely why the CERTi model relies on a sequential explanatory design, which unfolds in two stages. First, quantitative data is collected and analyzed, followed by the collection and analysis of qualitative data. This quantitative assessment aimed to assess whether faculty recipients of FRD programs are more likely to increase their grant acquisition (i.e., the effect). The assessment concluded that the examined FRD program had no statistically significant effect. This conclusion aimed to mitigate statistical test assumption violation, fishing and p-hacking, spurious statistical significance, and insufficient statistical power to derive its conclusion based on a systematic approach. However, what remains unexplored is what can be learned from implementing these programs to improve their delivery and maximize their potential effectiveness (i.e., explaining the effect), which the following article of the dissertation aims to do.

Chapter Five: Article 2 – Assessing Program Implementation

Methods

Research Question

What can be learned from implementing FRD programs to improve their delivery and maximize their potential effectiveness?

The CERTi Assessment Approach

This article explains why the FRD program may not have met its potential. To do so, it starts with presenting a logic model to show how it should have worked (i.e., program-as-intended logic model). Then it compares how it should have happened to how it actually happened (i.e., program-as-implemented logic model) and, through differences in implementation relative to the plan, highlight the missed opportunities that potentially explain the null effect found in the first article. The use of logic models facilitates explaining the program's theory of change (Armstrong & Barsion, 2006). A theory of change is "a description of how an intervention is supposed to deliver the desired results. It describes the causal logic of how and why a particular program, program modality, or design innovation will reach its intended outcomes" (Gertler et al., 2016, p. 32). Theories of change are critical to the assessment of HEI talent development efforts. They allow HEI leaders to hypothesize how and why their talent development efforts cause a change (i.e., impact) or why they did not. They do so by sequentially linking inputs, activities, outputs, and outcomes. Inputs consist of all human, financial, organizational, and public resources. Activities comprise actions (i.e., processes, tools,

events, and technology) that convert inputs into outputs, the direct and tangible products of program activities. Outcomes are the specific changes in program participants' attitudes, knowledge, and behavior and result from participant interaction with the outputs. An impact represents intended or unintended change arising from a program or intervention and informs the intended change's achievement. Theories of change "depict a sequence of events leading to outcomes; they explore the conditions and assumptions needed for the change to take place, make explicit the causal logic behind the program, and map the program interventions along logical causal pathways (Gertler et al., 2016, p. 32).

Theories of change are a critical tool to HEI leader decision-making relating to their talent development efforts. They describe program change strategy, define the problem, quantify requisite asset scope for addressing the problem, surface factors that could potentially affect the ability to exact change, facilitate employing best practices to support probable solution for the identified problem, and most importantly, build on existing knowledge of efficacious strategies. HEI leaders can clarify talent development program theory by first describing its problem to address and, second, specifying needs and assets, and third, identifying desired results (i.e., expected change). Fourth, documenting confounding factors that could influence the desired change. Fifth, researching success strategies that could facilitate the achievement of the desired change. Sixth, articulating assumptions undergirding how and why the program's strategies will facilitate change within the organizations' environmental context (Kellogg, 2004). Theories of change are often implicitly articulated rather than explicitly documented, which is ideal (Cook, 2000).

Implicit theories of change represent what is assumed by program designers relating to their hypothesis regarding program effectiveness at exacting desired change, usually after the fact. Conversely, explicit theories of change articulate and document program designer assumptions (i.e., claims and predictions regarding desired change) before program administration. Additionally, explicit theories of change attempt to explain the program's causal logic between its’ inputs, activities, outputs, and desired outcomes and impact.

Logic models are a systematic method widely used to link program activities with their outcomes (Kellogg, 2004). They holistically describe and illustrate how and why desired change happens within a particular context. They map out the "missing middle" between what a program does (i.e., its' activities) and how these lead to desired goals (i.e., its' impact). They also facilitate understanding theoretical assumptions undergirding a development effort by balancing its macro and micro-level suppositions (i.e., program theory of change). They are flowcharts that summarize a program's critical elements, such as inputs, activities, outputs, outcomes, and impact. 'Inputs' are resources needed to operate the program (i.e., human, financial, organizational, or material). 'Activities' are inputs' allocation or events, while 'Outputs' are activities' direct/immediate results. 'Outcomes' are short-term, intermediate, and longer-term results evidenced by specific changes in participant skills, knowledge, behavior, performance. 'Impact' is the ultimate change to the organization resulting from the program. (McLaughlin & Jordan, 2004).

Read from left to right, LM describes the program as it should work; inputs feed into activities yielding individual outputs resulting in specific outcomes and producing desired impacts. Read from right to left, they describe the program's theory; creating

individual impacts necessitates accomplishing particular outcomes resulting from specific outputs, emanating from critical activities, and requiring unique inputs. LM are essentially a graphic representation of change theory illustrating the linkages among resources, activities, outputs, audiences, and short-, intermediate- and long-term outcomes. Using logic models "provides a structure for the program to examine the degree that the desired learner outcomes, the program delivery methods, and the measurement approaches are aligned" (Armstrong & Barsion, 2006, p. 483). Researchers utilize them as a tool for deriving and articulating program theory (Savaya & Waysman, 2005). As a tool, they describe a path from a program's creation to its' implementation, providing HEI leaders with an assessment tool that is practical, reliable, and informative.

Understanding a talent development program's theory of change is essential to HEI leaders because it explains linkages between activities and outcomes and how and why the desired change is expected, based on past research or experiences. Program theory facilitates identifying "program resources, program activities, and intended program outcomes, and specifies a chain of causal assumptions linking program resources, activities, intermediate outcomes, and ultimate goals" (Wholey, 1987, p. 88), which elucidate transformative mechanisms expected to create or facilitate a desired change (Pawson & Tilley, 1997). A logic model is usually divided into two parts, as illustrated in Figure 5.1. The first part comprises program inputs and activities that program developers create and allocate while designing the program (i.e., pre-planning and planning stages). The second part comprises outputs, outcomes, and the desired impact (i.e., delivery and follow-up stages).

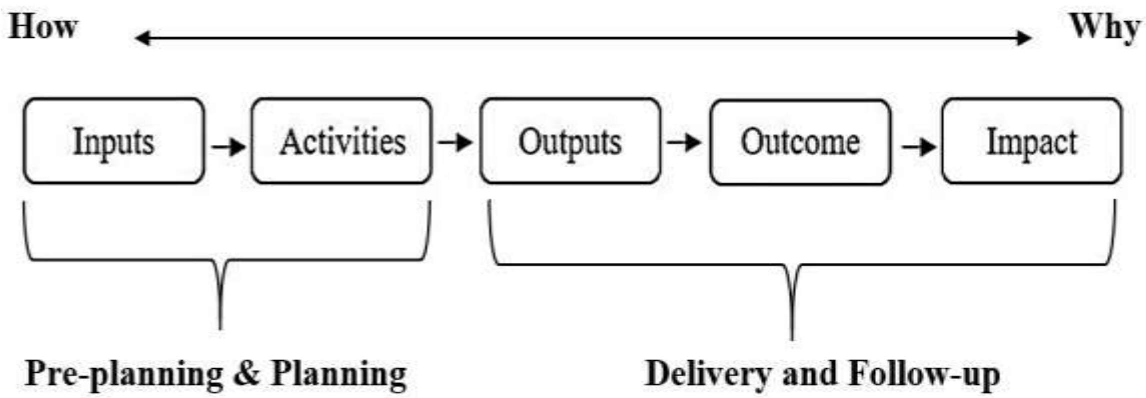

Figure 5.1 Logic model

The CERTi model utilizes a five-step process mentioned by Stolovitch and Keeps (2006) that uses LM as a systematic approach to operationalizing CERTi's macro-micro approaches to assessing the implementation fidelity of talent development programs. This approach consists of 1) Developing an LM representing the program-as-intended, 2) Identifying measures of key program indicators, 3) Developing an LM representing the program-as-implemented, 4) Comparing program-as-intended to program-as-implemented LM, and 5) Improving the program. This approach provides a roadmap for assessing the examined FRD program's implementation fidelity by first reconstructing the program as intended by its designers (i.e., program-as-intended logic model), then reconstructing the program as implemented in actuality (i.e., program-as-implemented logic model), then comparing the two to uncover any incongruities, and finally providing recommendations for improvement.

Study Sample

This article relies on semi-structured interviews with FRD program participants and program artifacts to gather data relating to program design and implementation. Two program designers, eleven mentees, and six mentors (i.e., totaling 19 participants)

partook in this study. The two program designers' interviews provided information about the program design (i.e., program-as-intended). The interviews with the eleven mentees and six mentors provided information about the program's implementation (i.e., program-as-implemented). Interviews with all program participants (i.e., two program designers, eleven mentees, and six mentors) and program artifacts such as attendance records, program timeline, mock review documentation, university submission portal records, and NIH grant proposal submission records provided information about program implementation fidelity (i.e., program implementation according to design).

Program-As-Intended Data Collection and Analysis

To better understand the program's theory of change as envisioned by its' designers, one-hour semi-structured interviews with its' two designers provided data on the preplanning and planning activities of the FRD program (i.e., its design, intent, and theory of change). The use of semi-structured interviews engaged the participants (n=2) in a conversation that elicited rich data, as Burgess (2002) suggested. The interviews explored select topics about what transpired during the programs' pre-planning and planning stages (i.e., its design and development) to elucidate its' theory of change. The semi-structured interviews took place at participant offices and via telephone. These two individuals were instrumental in launching the FRD program and were heavily involved in its preplanning and planning. Otter software recorded and transcribed participant interviews using the transcription key in Appendix A.

The interview protocol found in Appendix B elicited participant responses to the FRD program's design and structure. The questions were purposefully broad and included follow-up questions to facilitate conversation. Interview data coding

commenced in two cycles using the RQDA package in R statistical software. The first cycle relied on descriptive coding. Descriptive coding summarizes in a word a topic that the participant talked about and is appropriate for a variety of qualitative studies (Miles & Huberman, 1994; Saldaña, 2003; Wolcott, 1994). Each interview transcript was read independently and assigned one or multi-word descriptive codes that symbolized meaning to chunks of data (e.g., Funding, Program Timeline). This coding scheme allowed for patterns to emerge in each participant's statements. The second coding cycle relied on pattern coding, which allowed for a between-participants analysis to corroborate these emergent patterns. Pattern coding is explanatory and facilitates identifying emergent themes or explanations, and is appropriate for second cycle coding per Miles and Huberman (1994). FRD program records and artifacts (e.g., program timeline, the outline of events, session handouts, blackboard communications, and presentations), along with university institutional records, were also reviewed to identify key program indicators as recommended for case study research (Creswell & Poth, 2017; Yin, 2017) and to enable triangulation and validation of the information.

Program-As-Implemented Data Collection and Analysis

To better understand the FRD program's implementation in actuality, which facilitates comparison to its design and theory of change to uncover any incongruities, one-hour semi-structured interviews with program participants (i.e., mentees (n=11) and mentors (n=6)) provided data relating to the FRD program's implementation. Semi-structured interviews engaged the participants in a conversation to obtain rich data by utilizing open-ended questions and selecting topics about what transpired during the program's delivery.

Mentee and mentor interview data provided feedback on their experiences and allowed for cross-examination and verification of their interactions, providing a more holistic examination of the FRD program. These data provided thick descriptions of the minutia of the mentoring process. Mentor perceptions regarding their interactions with mentees and between-mentor comparisons served to add context to mentees' attitudinal, intellectual, and behavioral development, which provided a comprehensive picture of the development process. Additionally, an examination of FRD program records and artifacts (e.g., attendance records, blackboard communications, and presentations) and university institutional records provided rich data on program implementation, facilitating triangulation and validation of the information. Otter software recorded and transcribed the interviews using the same transcription key found in Appendix A.

The mentee interview protocol found in Appendix C assessed mentee RD development resulting from their participation in the FRD program. First, attitudinal development questions examined changes in mentee perceptions, values, and motivations towards the research aspect of their jobs. Second, questions concerning mentee intellectual development explored their epistemological, rationalistic, analytical, and comprehensive change or the lack thereof. Third, behavioral development questions analyzed potential changes in how they navigate grant proposal development and submission processes and procedures, assessed any competency gains, and examined changes in productivity (i.e., grant submissions and acquisitions). The interview protocol found in Appendix D garnered mentor perceptions and feedback on mentor/mentee interactions and mentee development.

Mentee interviews were coded in two cycles (i.e., once for magnitude and another for evaluation) using the RQDA package in R statistical software. According to Miles and Huberman (1994), magnitude coding indicates intensity, frequency, direction, presence, or evaluating content. Each data point was initially assigned a Magnitude code (i.e., +Positive, -Negative, ±Mixed, =Indifferent). The second cycle of coding utilized evaluation coding to analyze the data for patterns, interpret data for significance, make judgments on results, and offer actionable recommendations (Patton, 2008). This two-cycle process assessed the FRD program's efficacy at developing mentee RD skills.

Mentor interviews were coded in two cycles using the RQDA package in R statistical software. The first cycle relied on descriptive coding using the RQDA package in R statistical software by reading the interview transcripts independently and assigning one or multi-word descriptive codes that symbolize meaning to chunks of data (e.g., Mentorship, Research Development), which allowed for patterns to emerge in each participants' statements. Descriptive coding summarizes what participants talked about and is appropriate for various qualitative studies (Miles & Huberman, 1994; Saldaña, 2003; Wolcott, 1994). The second coding cycle relied on pattern coding, facilitating a between-participants analysis to corroborate these emergent patterns. Pattern coding is explanatory, facilitates identifying emergent themes or explanations, and is appropriate for second-cycle coding (Miles & Huberman, 1994).

Program Fidelity of Implementation Data collection and Analysis

Assessing the FRD program's fidelity of implementation by comparing its design to its' implementation in actuality facilitates understanding why the FRD program did not reach its' full potential and provided information on improving its delivery to maximize

its potential effectiveness, which facilitates answering the article's research question. An examination of environmental context, resource availability, adherence to program design, participant feedback, and the complexity of the research development process will provide thick data on potential contributing factors to the program's lack of effectiveness. The difference between program design, actual implementation, and participants' reflections will facilitate providing a recommendation for program improvement in future iterations.

Implementation research identified five components that influence implementations' success. These components, collectively referred to as "Fidelity of Implementation" (FOI), are Adherence, Dosage, Quality, Participant Responsiveness, and Program Differentiation, and are critical measures of implementation fidelity (O'Donnell, 2008). Some research argues that each of the five fidelity components represents an independent method of assessing fidelity of implementation (Mihalic, 2004). This research suggests using either adherence, dose, or quality as the only method of examining fidelity. Other research argues for assessing all five dimensions concurrently to portray implementation fidelity comprehensively (Dane & Schneider, 1998; Dusenbury et al., 2003). However, research advanced by (Carroll et al., 2007) proposes a newer methodology for assessing implementation's fidelity. They argue that overlap exists between these components and that adherence (i.e., the extent to which program components are delivered as prescribed by its' designers) is the bottom line. Hence, it serves as the main category for assessing FOI while including the subcategories of content, frequency, coverage, and duration (i.e., dose/exposure) within it.

Program content, methods, and activities within this framework are commonly regarded as indicators of adherence and generally reported as the proportion of program components delivered contrasted to the number prescribed (O'Donnell, 2008). The framework proponents posit variables such as the intervention's complexity, facilitation strategies, delivery quality, and participant responsiveness moderate (i.e., affected) the degree of adherence. Combining FOI evaluation with those assessing program outcomes/impact provides continuous improvement data to guide the re-development and re-implementation of more efficacious interventions (Dromgoole & Cummings). This approach better aligns with the CERTi model's comprehensive evaluation approach and is hence employed.

The benefits of implementation data are critical to HEI leaders concerning the efficacy of their talent development efforts. The collection, analysis, and interpretation of such data facilitate them understanding their talent development program's implementation, examining the development processes theoretical assumptions, contextualizing outcome findings within organizational and environmental contexts, providing pertinent feedback for continuous quality improvement, and affording their developers (i.e., trainers) insight into poorly implemented components of their programs. Researchers often refer to these data as differentiation or component analysis as referred to in FOI literature. These data can cyclically inform leadership decision-making regarding their talent development efforts.

Findings

Program-As-Intended

The college's financial sustainability and national prestige were significant drivers behind the creation of the FRD program. Although considering faculty needs when creating the program, the college's financial sustainability and national notoriety undergirded its' creation, as one of its' designers stated:

> Faculty is an investment; it costs us money, not just salary but sometimes a huge amount of money and startup. So, we want to make sure number one, they have what they need to be successful. And number two is that they're going to be guided in the right way. And that's where it's to our advantage to invest in both of those things…We want to bring somebody in that's going to, I hate to keep putting it in monetary terms, but it's going to be a net profit in terms of research productivity, dollars and cents, and publications and grants. It reflects how good your faculty is if they can get funded by the National Science Foundation (NSF), National Institutes of Health (NIH), and other federal agencies. It means they're as good as they can be.

The other program designer corroborated the first's sentiments and provided more detail on budgetary and nonbudgetary nuances, stating:

> It's important to the college because that's part of how the college operates in terms of its overall budget. You can't just have faculty that have substantial amounts of release time [from teaching] like we have in the [College] where we have a one-one or a one-two teaching load, which is quite low. You can't have people doing that without that [release] time being paid for through some type of grant funding. That's just a pure checks and balances kind of perspective, but on another perspective, at the school level, it brings prestige. So, when you have people that are successful on national level types of grant funding, consistently successful, that brings clout that allows you know the school too, for lack of a better word, bragging rights. It brings notoriety and brand; it allows you to attract better faculty because they see that good faculty are already there. It brings up the entire level of quality of everybody from the students to other faculty that come in to get a postdoctoral fellowship, and then it also helps out with getting additional grants. You know, so the rich get richer kind of thing.

University investment in faculty goes beyond the typical personnel investment (i.e., salaries and fringe benefits) by allocating substantial start-up packages to faculty. For example, these packages vary widely at the college administering the FRD program,

ranging from the low $100,000's to the million-dollar range. Faculty receive these packages to jump-start their research careers and pay for laboratory space, equipment, and supplies and are instrumental in faculty research productivity success. These expenditures are precisely why one of the program's designers characterized faculty as an "Investment." He hopes that the college would reap a return on its investment (i.e., grant funding), which simultaneously benefits both the institution and faculty members. His motivations and mindset towards the talent development process are similar to that of many HEI leadership. These individuals progressively face a transformed organizational landscape gripped by financial pressures, which their predecessors seldom contemplated (Gallop & Svare, 2016; Rabovsky & Ellis, 2014). The convergence of reduced state appropriations for higher education institutions programs due to the increasing cost of entitlement programs, flattening federal grants, and a hypercompetitive environment for those funds created this new environmental context (Cronan, 2012).

Although financial sustainability and prestige undergirded the program's creation, the program's designers communicated that many research development efforts, such as workshops and seminars, were being administered within the college informally. However, the designers communicated that these program's focus was the quantity of grant proposal submissions rather than the quality of each submission. The college's leadership wanted to shift the focus from quantity to quality to address the ever more competitive federal funding environment (Alberts et al., 2014; Beech et al., 2013; Ginther et al., 2011), as one of the designers communicated, "We were at the stage in our school that we had many proposals submitted. We didn't need more proposals submitted; we needed better ones." He added that the purpose of the program is to "train faculty to write

better proposals." The hyper-competitive federal grant funding environment was a primary driver behind a shift to formalize research development within the college. The program's designers felt that designing a year-long program that systematically trained faculty on the main aspects of NIH proposal development would provide faculty with a competitive advantage. One of the designers communicated this as he explained, "You can't learn a lot in a one-day workshop; it has to be step by step, " leading to the program's design.

The program's designers and program artifacts provided insight into its design and purpose to uncover its theory of change. The program's theory of change (i.e., hypothesis) as communicated by them posits that the program's inputs (e.g., mentorship, research development), activities (e.g., grant proposal development, mock NIH reviews), and outputs (e.g., high-quality proposals, proposal submissions) should facilitate positive reviews by NIH reviewers (i.e., discussed versus rejected). These positive reviews should then lead to reviewers assigning the proposals meritorious impact scores, which will lead to competitive percentile ranks, that result in funding (i.e., impact). Figure 5.2 depicts the FRD program's theory of change as communicated by its' designers. The blue part of the figure represents the FRD program (i.e., cause) and includes its' inputs, activities, and outputs. The orange part of the figure represents desired outcomes (i.e., high-quality grant proposals, proposal submissions, reviewer discussions, high impact scores, high percentile ranks). The green part of the diagram represents the desired impact (i.e., increased grant acquisition success rates.

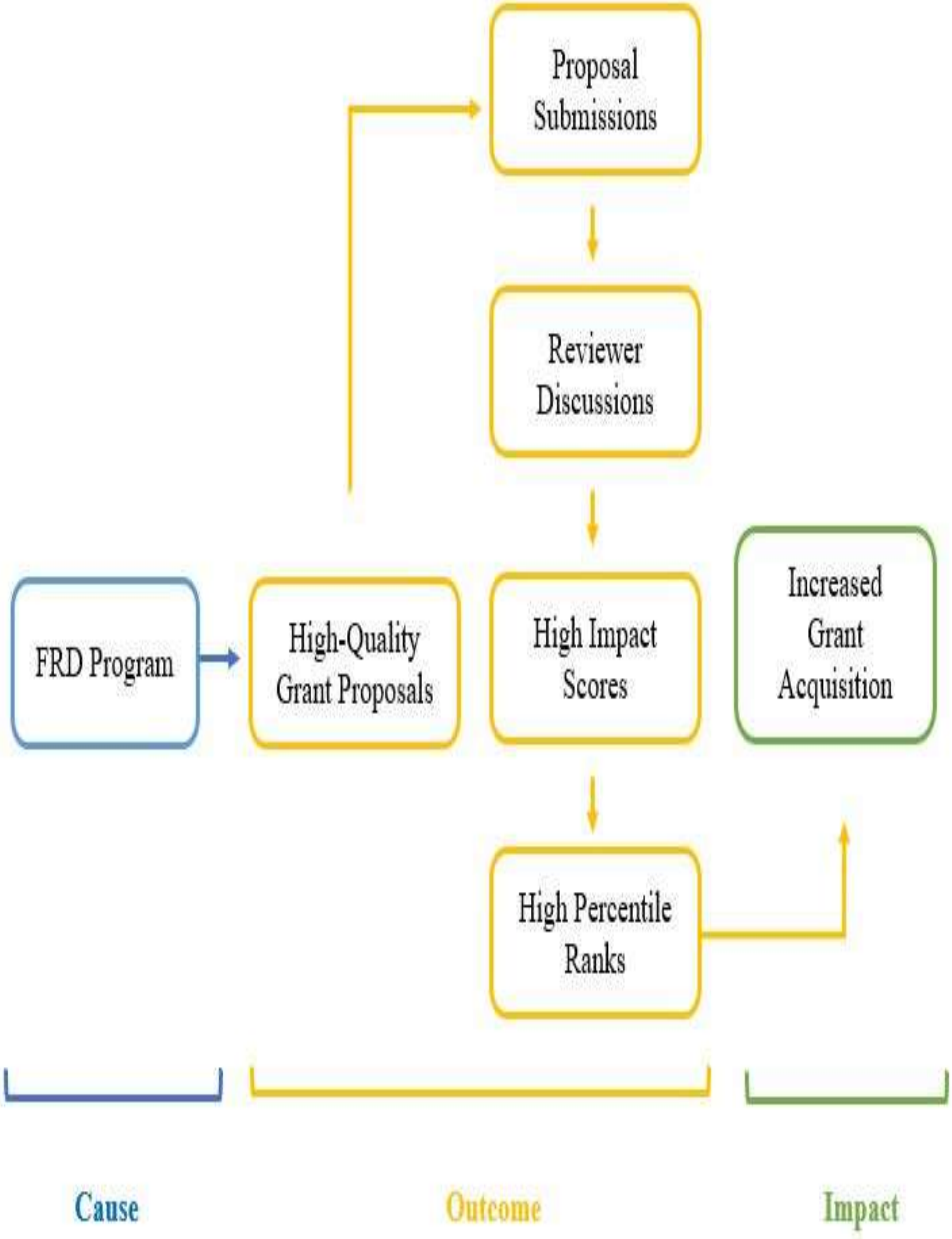

Figure 5.2 FRD program theory of change

Program Inputs

The program’s designers communicated that the college made available internal inputs such as meeting venues, conference rooms, participant offices, equipment, technology, and supplies at no cost to the program or its’ participants. They stated that the human resources comprised faculty who participated in the program (i.e., seven mentors, twenty-two mentees, and twenty-two subject-matter experts) and one administrative support staff member. The college budgeted $30,000 for the program's provision per the designers. Financial inputs included, first, external consultant fees (i.e., stipend, travel, lodging, and meal expenses). Second, financial inputs included external inputs (i.e., meeting venues, equipment, and supplies). Third, the financial inputs included $3,500 stipends to each mentor and a one-month salary equivalent stipend for the program director.

One of the program’s designers communicated that mentors would be “Process Experts, and not Content Experts,” that they were to “Help Mentees meet milestones for completing NIH applications,” and that they “Can provide feedback on the science but are not required.” The designers communicated that a quality grant proposal was both about the science aspect and about selling the science. Hence, they designed the program to provide mentees with content expertise (i.e., within college science subject-matter experts) and process expertise (FRD program mentors). The importance of the process expertise (i.e., selling) aspect of a quality proposal is evident in one of the program designers' statements. He shared that, “What makes a good proposal is that it has a consistent, coherent argument that logically develops through the narrative and is framed from impactful outcomes for both the scientific field and practical application.” He added

that many faculty “struggled to communicate why their ideas are something that we should invest in; they struggle with selling it.”

The designers communicated to mentees that their mentors' main task was to provide them with feedback on accomplishments and answer questions about developing a grant proposal (i.e., grant proposal process expertise instead of content or science expertise). Content expertise per the program’s designers what relegated to the subject matter experts that mentees would select from among their peers within the college to assist them with the scientific writing aspect of the grant proposal development instead of the process aspect of it. An examination of mentor training records (i.e., PowerPoint presentation) revealed that mentors’ main task was providing questions such as “What is the difference between Significance and Innovation,” “What does a quality Specific Aims page include,” “What does a good research team look like – putting together a team,” and “What is the process you (Mentor) go through to develop the science.” The focus on these sections per one program designer was to facilitate clear communication of proposed grant proposal ideas.

Activities

The FRD program’s product activities included advertising the program via the college's webpage on the university website, promoting it via emails, discussions with department chairs, and faculty meetings as was explained by one of the designers, “Mainly through emails and word of mouth, to a lesser extent.” He also conveyed the importance of participating in the program to the leadership within each department, adding:

> I went to each [department] chair, and I said, ‘you have new faculty coming in.’ They should participate in this [program]. We also have an administrative

> council meeting once a month; it's with the Dean, all the Chairs, Associate Dean's. So, I mentioned the program a lot during that time, and the dean had already communicated his input and just how important it is that we write better proposals, on average. We don't need more, just better.

Program promotion efforts also included guest appearances by the leadership and faculty with a proven track record of grant acquisition and mentor' presentations on grant proposal development. The leadership reinforced the expectation levied on faculty to acquire grants. One of the program's designers was communicating to the program participants the importance of grant funding to the college, saying, "we would *like* for you to get grant funding," only to be interrupted by the dean of the college, who corrected him saying, "We *expect* you to get grant funding," highlighting how institutional financial sustainability pressures trickle to the faculty (Musambira et al., 2012)

Service activities included four large-group events and workshops designed to provide program participants with the required research development to successfully develop a high-quality grant proposal. They included peer-group mentoring consisting of three faculty mentees and one mentor to share ideas, review proposals, receive constructive feedback on grant proposal development. These peer-group mentoring sessions per one program designer were "designed to leverage the skills, knowledge, and expertise within the school to maintain and increase institutional competitiveness." which one program designer stated was "a strategic, proactive, and capacity-building investment in one of the school's most valued assets, the faculty."

Capacity building and leveraging skills, knowledge, and expertise presuppose a competency level preceding the development process. The presupposed competency level relating to research productivity is evident by experience level requirements on many of the college's faculty position advertisements before and after the FRD program's

inception. They include terminology such as “a record of research.” An expectation of baseline research productivity level “will be expected to pursue external funding to support their research aggressively.” Additionally, some position advertisements require previous experience level specific to a faculty rank, such as evident in one such position description:

> Applicants at the Assistant Professor level must have at least nine months of experience by the beginning date of employment in faculty, post-doctoral fellow, or similar research or teaching position with potential for extramural grant funding, a record of grant-seeking, and publication of peer-reviewed original research.

The baseline competency expectation level was a critical program design aspect because some faculty applicants to the FRD program were excluded from participation because they did not meet this competency level requirement, as one of the program’s designers communicated. Other service activities included small work meetings between mentors and mentees to discuss progress on drafting grant sections (i.e., Significance, Investigator(s), Innovation, Approach, and Environment). Infrastructure activities included making facilities (e.g., meeting venues and conference rooms) available for program events, promoting collaboration between program participants and faculty content experts within the college, and capacity building (e.g., skill and competency development).

Outputs

Discussion with program designers and examining program artifacts revealed that the FRD program’s first output was a planned “Kick-off” group workshop scheduled for December 8th, 2017, for a half-day duration (i.e., 8:30 am-12 pm). The workshop's purpose was to address three topics; 1) NIH Mechanisms Overview (Pros/Cons), 2) Identification of Co-Investigators, and 3) Proposal Mechanisms: Specific Aims,

Significance, and Innovation. The first two topic presentations were the program's designers' responsibility, while a panel of expert faculty from among the college would direct the last. Mentors and mentees would utilize the time between this session and the next (i.e., December 8th – February 16th) to accomplish the program's first assigned milestone. This milestone included drafting the specific aims, significance/innovation sections of the NIH grant proposal, finalizing choice of mechanism, NIH institute, and funding opportunity announcement (i.e., FOA), as shown in Table 5.1.

Table 5.1 December 8th workshop

Date	**Location**	**Type**	**Topic**	**Speakers**
Dec. 8	External	Group 8:30am-12pm	Kick-off • Introduction • Expectations • NIH Mechanisms Overview (Pros/Cons)	• Program Designer • Program Designer
			• Identification of Co-Investigators	• Faculty Experts

The FRD program's subsequent output was a group workshop scheduled for February 16th, 2018, held at an external venue and lasting half-day (i.e., 8:30 am-12 pm). Three groups of mentees comprised of seven mentees each would split into meeting venues to present their work. The first 2.5 hours of the event would allow mentees to present their specific aims and significance and innovation sections they have been developing to receive feedback. The last 1.5 hours of the event would rely on invited

experts (i.e., college faculty) to train the mentees on the preliminary studies and approach sections of NIH grant proposals, as illustrated in Table 5.2. Mentors and mentees would utilize the time between this session and the next (i.e., February 16th – May 25th) to accomplish the program's second, third, and fourth assigned milestones. The second milestone revolved around updating the specific aims, significance/innovation, and drafting the NIH proposals' preliminary studies and approach sections by the end of March. The third milestone included drafting a completed research strategy by the end of April. The fourth milestone entailed sending a final draft proposal to the research office for a May group workshop review assignment by April 27th.

Table 5.2 February 16th workshop

Date	**Location**	**Type**	**Topic**	**Speakers**
Feb. 16	External	Group 8:30am-12pm	Part 1 (2.5 Hours) Presentations 3 groups of 7 mentees • Specific aims • Significance/Innovation	• Mentors • Mentees
			Part 2 (1.5 Hours) Proposal Mechanics 2 • Preliminary Studies • Approach	• Faculty experts

The FRD program’s subsequent output was a group workshop scheduled for May 25th, 2018, held at an external venue. It would start at 8:30 am, end at noon, and simulate

the NIH review process (i.e., mock review panel). In the first 2.5 hours of the event, participants would separate into three groups comprised of seven mentees and two mentors and provide each other with un-blinded (i.e., face-to-face) peer grant proposal reviews using NIH review guidelines. Mentors would facilitate the review process and provide feedback during the workshop. For the remaining 1.5 hours, invited expert faculty/grant administrators would train mentees on NIH study section selection (i.e., appropriateness of study section to submit a proposal through), understanding review panel nuances (e.g., what NIH reviewers consider quality proposals), biographical sketch development (e.g., highlighting investigator strengths), budgeting, and sub-contracts/consultant (e.g., highlighting co-investigator strengths) procedures as Table 5.3 details.

Table 5.3 May 25th workshop

Date	**Location**	**Type**	**Topic**	**Speakers**
Feb. 16	External	Group 8:30am-12pm	Part 1 (2.5 Hours) Mock NIH Reviews 3 groups of 7 mentees • Proposal review • Proposal feedback	• Mentors • Mentees
			Part 2 (1.5 Hours) Proposal Mechanics 2 • Preliminary Studies • Approach	• Faculty experts • Grant Administrators

Mentors and mentees would utilize the time between this session and the next (i.e., May 25th – September 7th) to accomplish the program's fifth, sixth, and seventh assigned milestones. The fifth milestone entailed revising grant proposals and sending them to internal and external subject matter experts by the end of June. Mentors comprised internal subject matter experts, while senior faculty within each of the colleges' departments comprised external ones. Internal subject matter expertise revolved around mechanistic grant proposal development (i.e., proposal development), while external revolved around content expertise (i.e., the science). The sixth milestone necessitated mentees to amend grant proposals based on internal and external subject matter expert feedback by the end of July. The seventh milestone required mentees to revise proposal biographical sketches, budget, sub-contract, and consultant sections based on internal and external subject matter expert feedback by the end of August.

The FRD program’s following output was a group workshop scheduled for September 7th, 2018, held at an internal venue. It would last from 8:30 am until 12 pm and would include all current FRD program participants and those selected for the second year’s program as guests. The leadership would update the mentees on revised NIH clinical trials policies and study section selection. The session would encourage faculty to finalize and submit their completed grant proposals by the designated October 15th, 2018, NIH deadline. The program leader would present a tale of two grants, a presentation focusing on resubmission strategies given NIH grant funding's competitive climate.

The presentation would highlight his research agenda, characterized by persistence in developing and submitting grants during each NIH funding cycle. Specifically, he would emphasize the high rate of decline on his grant submissions and

how he revises and resubmits his grant proposal cyclically via other funding opportunity announcements, NIH institutes, and study sections. He would explain that familiarity with funding opportunities, knowledge of study sections, and persistence revision would pay off. He would present a flow chart of several grant proposals that he revised and submitted on average six times through various funding announcements, NIH institutes, and study sections until receiving funding to highlight the importance of persistence in seeking NIH grants. Grant proposal resubmission is critical contextually given the stagnation of grant proposal resubmissions within the college.

Outcomes

The main aim of the FRD program was to develop its' participants capacity in developing a high-quality fundable NIH grant proposal. The dean of research stated that the program was a "capacity-building investment in one of the school's most valued assets, the faculty." Research capacity "'is conceived as the most and best research which could be done now if there were the political will and the necessary resources for it to be done" (Mcintyre & Mcintyre, 1999, p. para. 1.4). It encompasses all human capital terminology such as skills, knowledge, attitudes, understanding, competence, and procedures and includes external factors such as resources, academic freedoms, and professional status. The inclusion of these external factors provides an all-inclusive definition (Evans, 2012). Figure 5.2 presented the FRD program's theory of change, which posited that the inputs, activities, and outputs of the program (i.e., the cause) would result in its' desired outcomes (i.e., high-quality grant proposals, proposal submissions, reviewer discussions, high impact scores, and high percentile ranks). Program designers posited that a systematic year-long proposal development process

based on continuous feedback from subject matter experts, grant proposal process peer mentors, and mock NIH peer reviews facilitate mentees developing high quality, nationally competitive grant proposals leading to funding by the NIH (i.e., program theory of change). Hence, they designed the program with specific short, intermediate, and long-term outcomes to facilitate achieving the program theory of change.

The FRD program's short-term development outcome was attitudinal development, which encompasses enhancing participant perceptions, values, and motivation towards the research aspect of conducting their jobs, which is essential to developing a high-quality grant proposal. Attitudes can be inherited or learned via direct or indirect interactions (e.g., mentorship) (Stangor et al., 2017). They are belief, feeling, or behavior-based and represent evaluations used by people to make effortless behavioral engagement judgments. The principle of attitude consistency posits that behavior is more consistently guided by attitudes when ‘affect’ (i.e., feelings about an attitude object), ‘behavior’ (i.e., intention towards an attitude object), and ‘cognition’ (i.e., beliefs about an attitude object) are aligned.

Program designers aimed to achieve this short-term outcome by inviting faculty from the college with a successful track record in grant acquisition in the first workshop to highlight their success stories in acquiring competitive NIH grants hoping to enhance mentee perceptions about the research aspect of their jobs. They aimed to enhance mentee values towards the research aspect of their jobs via the program's cohort design, hoping it would facilitate collaboration, affording them all the necessary tools such as subject-matter expertise, proposal development process mentorship, and administrative support. They aimed to enhance mentee motivation towards the research aspect of their

jobs (i.e., grant acquisition) by highlighting how it takes senior faculty multiple attempts over many years to acquire grants, hoping to motivate mentees to persist in revising and resubmitting their proposals versus being demoralized when not receiving funding on the first attempt. One example of this is the program designer's presentation in the last workshop about his many attempts before acquiring a grant.

The FRD program’s intermediate development outcomes included intellectual development, which encompasses the “knowledge, understanding or reflective or comprehensive capacity or competence is modified” (Evans, 2012, p. 22). Intellectual research development involves ameliorative epistemological, rationalistic, analytical, and comprehensive change. Intellectual development involves altering knowledge structures, especially concerning researching (e.g., grant proposal development) and research-related knowledge structures. Research has consistently demonstrated the effect of epistemic beliefs on how individuals deal with a modern-day knowledge-based work environment's essential requirements, such as acquiring and evaluating knowledge. This topic is critical due to repeated findings that stress the beneficial effects of advanced epistemic beliefs (e.g., weighing and evaluating knowledge claims) (Kerwer & Rosman, 2018; Kienhues et al., 2016; Phan, 2008; Strømsø & Kammerer, 2016).

An examination of the program’s advertisement found in Appendix E shows that the program’s design aimed to “provide mentees the tools and <u>knowledge</u> they need to be successful in developing a high-quality first submission proposal” (i.e., research development). It also states that the program comprised “large group events and workshops to gain tools and <u>knowledge</u>, peer group activities for sharing ideas, reviewing proposals and providing constructive feedback, and small group work with mentors who

have established track records of securing external funding and a commitment to mentoring," and to "guide faculty in writing high-quality research applications." One of the program's participants' expressed excitement at the potential to gain invaluable knowledge when interviewed by the college's public relations director in anticipation of the program's launch. She expressed her excitement for the program and its potential impact to enhance her knowledge of grant-writing and the submission process by stating, "I'm looking forward to getting feedback on my grant proposal regarding both content and style." She underscored potentially gaining knowledge from participation in the program by adding:

> I think getting my work critiqued by both a mentor and colleagues will result in a really nice grant submission that's easy for reviewers to read and understand. I'm also excited about getting tips for grant writing from people there and being successful. Hopefully, I can learn from their successes and past mistakes.

The FRD programs' long-term outcomes included behavioral development, which encompasses the physical act of researching in all its forms. It involves processual, procedural, competential, and productive changes. Behaviors undergirded by ameliorative attitudinal and intellectual change signify genuine commitment compared to one emanating from compliance, pressures, or imposition, reflecting a lack of conviction. Desired behavioral development included attending all group workshops and individual mentor-mentee sessions, drafting and revising grant proposal sections (i.e., milestones), and submitting a grant proposal to the NIH by the designated funding deadline. Program designers posited that each participant would submit a grant proposal to the NIH, receive a favorable review by NIH reviewers, be given a meritorious impact score, and achieve a competitive percentile rank, leading to funding. These outcomes were communicated in the program's advertisement, "Meet, at minimum, once per month with their assigned

Mentor," "Achieve the FRD program Milestones as indicated in the timeline," and "Submit their NIH proposal developed during the FRD program for the NIH Grant Cycle immediately following graduation from the program."

Impact

The FRD program's desired impact was to "increase the success rate of [College] faculty applying for NIH Research Project Grants" (i.e., grant acquisition) per its advertisement and as communicated by its' designers. The figure found in Appendix G illustrates the FRD program's program-a-implemented logic model. The figure provides a pictorial representation of the program as intended (i.e., designed) by the college's leadership. It represents the program's theory of change based on its designer's conceptualization. It details all inputs allocated for the program's provision, the activities that put the program's inputs into use, leading to its' outputs. These three (i.e., inputs, activities, and outputs) are the program as envisioned by the program's designers. The outcome and impact detailed earlier and listed in the program-as-intended logic model represent what they desired to achieve. The hope is that these inputs, activities, and outputs representing research development would produce desired ameliorative mentee attitudinal, intellectual and behavioral development outcomes leading to research productivity (i.e., grant acquisition). Comparing this program-as-intended logic model based on specific desired outcome and impact to a logic model representing the program-as-implemented in actuality facilitates testing its theory as hypothesized by its' designers. The following section will examine program implementation to understand how the program was executed compared to its design as described in this section, facilitating explaining its null effects as gleaned in the first article.

Program-As-Implemented

The previous section elucidated the program's intent per its design and as envisioned by its' designers. This section expounds on potential factors contributing to the program's lack of effectiveness based on conversations with its' participants (i.e., 11 mentees and six mentors) and via examining its' artifacts. Mentee/mentor conversations revealed several themes that included factors that potentially contributed to the program not reaching its' full potential (i.e., ineffectiveness). Theme 1 *You cannot teach old-faculty new tricks* highlights how some more experienced mentees felt that their past research development impeded their development, perceiving the program to be too rudimentary based on their experience level. Theme 2 *Professional Learning Communities—PLCs* revealed that association with productive research teams within the college or exclusion from them helped or hindered mentee research development and productivity. Theme 3 *Gender Disparities* revealed perceived gender disparities among female mentees that potentially impeded research development and productivity. Theme 4 *Misalignment on Science Expertise* delved into how several mentees and mentors cited a misalignment in scientific expertise, which both felt stood in the way of mentee research development and productivity. Theme 5 *Knowing the Science* discusses mentors' sentiments that many mentees did not possess foundational knowledge of their respective scientific fields, which they felt stood in the way of their research development. Theme 6 *Preliminary data* exposed the importance of preliminary or pilot data to grant acquisition. Theme 7 *Selling the Science* highlights mentor feedback on mentees' inability to communicate (i.e., sell) the significance of their research ideas to the broader scientific community (i.e., NIH reviewers). Lastly, Theme 8 *It takes two to mentor* underscores that

mentors felt that a combination of mentees not taking advantage of their mentorship and conflicting mentoring from their departmental mentors mitigated their ability to be research productive.

Theme 1: You cannot teach old-faculty new tricks

Several mentees expressed that their past research development, whether via doctoral, postdoctoral, or professional development as faculty, stood in the way of their research development. Tim, for example, explained that the program came a little too late to benefit him, substantively:

> Prior to the program, I'd say I had a bit [of grant proposal writing experience] because, again, I was well in [as an assistant professor] by the program's time. So, I knew at least a decent amount about the development [of grant proposals] and have previously gone through some [trainings]. But again, had this come in much earlier [in my career], I could definitely see some of the benefits.

Charlotte's background and experience stood in the way of her research development. Like Tim, she was a more experienced assistant professor who came to the university from what she characterized as a "soft money" university where the expectation was to support one's salary entirely via grant acquisition. She also shared that she had extensive training in grant proposal development and primarily enrolled in the program to take advantage of the deadlines it offered to keep herself on track for the proposal submission deadline. She emphasized:

> Again, I feel like it was a decent program. I just feel like I personally didn't benefit, but I could see how many people potentially could have. I feel like I signed up just more like, 'well, I don't really have a grant, so I might as well.'

Sensing that there is more to that statement, I asked if she felt a sense of obligation or pressure to enroll in the program; she provided a delicately worded answer adding, "Kinda. But, again, I also was looking for the opportunity to learn. Until I actually get

one [large-scale grant] I feel like I always continue to get better. Learn some kind of new trick or whatever".

Her answer reveals a juxtaposition of dueling interest that potentially could be explained by FRD program promotion efforts. The leadership employed various methods to promote the program, such as emails., webpage ads, and faculty meetings. However, one method they employed was asking department chairs to encourage the faculty to enroll in the program, as one of the program designers shared. These promotion efforts could potentially explain why Charlotte felt obligated to enroll in the program. After all, if the department chairs perceived the program as valued by the dean and supported by him, they would have encouraged their faculty to attend, especially those who have not acquired large-scale grant funding.

Michael, an assistant professor with extensive postdoctoral research training, shared that the mentorship did not benefit him much, "I don't think I got help [from the mentorship] because like the most [my mentor] could do was just give me slight [writing] feedback." The mentorship's contribution to his grant proposal development was minimalistic as he described an example, saying, "Maybe at the beginning, I had three [specific] aims, and [my mentor would] say, 'It just seems like [that is] too many let's go down to two or something." He concluded by giving a summative assessment of the benefit he perceived to receive from the program, adding, "So for me, the thing I got the most out of the [program] was just the deadlines." Michael's mentor corroborated what he said regarding the minimal mentorship he received saying. "I didn't have to employ a lot of hand-holding or motivating. Everyone was on schedule. Everyone had everything done. I almost didn't have to mentor much at all. They were pretty advanced," lending

credence to his assertion and highlighting how past research development stood in the way of his research development.

Theme 2: Professional Learning Communities—PLCs

Tim, who communicated that the program came too late to benefit him based on his past research development, had a unique perspective on intradepartmental nuances that potentially hindered research development for some faculty as he explained:

> I sort of sat back and looked at things and looked at some of the success stories and some of those who haven't been successful across the school's departments. One of the things I've noticed [that] there is these [productive research groups within the college] that have those nice direct and inherent common research interests. Working together comes across naturally [within these groups], while others are [isolated] on an island. It's really tough when it comes to [new] hires to actually cover your needs in terms of teaching and bring in individuals who have good backgrounds [and are] likely to be successful. But [to] also get them into that real sort of tight knit [research] group to ensure that they will at least get a greater likelihood of being successful [in acquiring grants].

He was deriding the fact that a few productive research teams exist within the college, known in the talent development research as professional learning communities (PLCs). These PLCs form organically at university campuses via personal contacts among faculty seeking to build institutional relationships. These faculty value the face-to-face interactive nature of these groups, welcome their ground-up informality and appreciate them as safe places to reflect and innovate (Cherrington et al., 2018). In the case of what Tim was describing, a senior rank professor with a demonstrable record of research productivity usually leads such a team, comprised of three to four faculty, several postdoctoral fellows, graduate students, and staff. They are well-funded by the NIH via grant acquisition and can access many resources that other faculty lack due to that funding. Contrarily, the rest of the faculty are operating alone within their respective departments without such support networks. One of the program designers postulated this

situation during a one-on-one conversation, describing these faculty as "*lone wolves.*" Tim is one of those faculty, which explains his mixed perceptions of the program (i.e., good, but with shortcomings). Randy, another mentee, exemplifies the "lone wolf" faculty described by Tim and the program director. Once back on the job, he operated independently of any support structure that reinforced the development he was receiving via the FRD program, mirroring Tim's experience during and after the program. This sentiment was evident when asked to provide recommendations for program improvement as he critiqued the program's design addressing its designers, saying, "And if you have any communication with the previous participant or someone [it would be great], just keep in touch so that they are still motivated."

Contrarily, Michael, who acquired a grant, seemed to benefit from his association with one such productive research group within his department as he shared:

> The [program] really made me submit it. I don't know what I would have submitted otherwise. I probably would have sat on it some more. So, I'm glad that's the one thing I learned from my [fellow productive researchers in the department], just submit it, get some comments, then deal with it. It created that kind of collegiality where I didn't want to let the team down. Like me personally, what I think of, maybe my close friends happen to be, [two productive researchers], who are just cranking out grants like maniacs. So, then it starts to blur that line of what's expected from the department versus [trying to keep up with] them. I definitely try, and again, I was always hearing that tenure clock ticking for me personally. What do they want, what do they expect, so I thought I need to submit a couple of times a year.

The colleagues that drove Michael to submit his grant proposal, the ones he described as "cranking out grants like maniacs," are what Tim described as pockets of productive research teams within the college's departments. Michael confirmed the influence of these teams on motivating faculty and increasing their research productivity.

The lack of on-the-job support is crucial contextually in light of one program designer stressing the importance of grant proposals' resubmissions. His presentation to

the faculty highlighted the importance of what a mentee characterized as “perseverance” in the face of the current competitive NIH funding environment. Figure 5.3 illustrates the tale of two of his grant proposals. It highlights the process, timeline, and perseverance of his efforts in acquiring funding. The first grant proposal took approximately three years and five submission cycles until funded, while the second was in year three of resubmission as he was still retooling it to achieve funding. Juxtaposing the program designers' efforts to acquire grant funding to those who participated in the FRD program exposes an overwhelming challenge, begging continuous on-the-job support. This situation is critical contextually within an organizational environment wrought with fierce competition for limited resources (i.e., grant funding). After all, and based on his presentation, it takes a senior, well-published, well-funded, well-connected faculty such as himself, on average, five submissions, to multiple institutes, over multiple years to acquire an NIH grant. Incorporating his wealth of knowledge on grant proposal submissions, revisions, and resubmissions in the form of on-the-job post-development support would have potentially contributed to faculty post-development success.

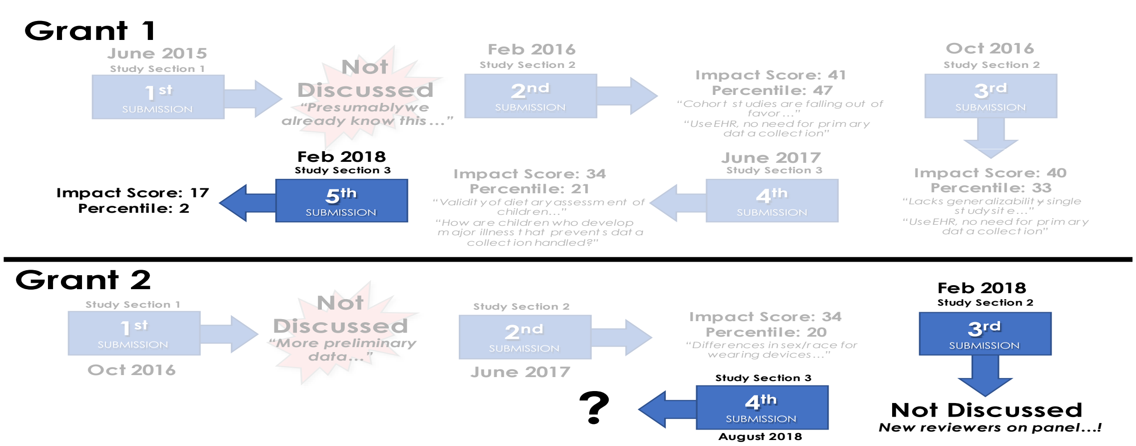

Figure 5.3 Program designer proposal submission and resubmissions

Theme 3: Gender Disparities

Gender disparities manifest in Amber's experience as she shared how new and female faculty face challenges otherwise not experienced by their more seasoned and male counterparts. She first delved into the results of her effort to acquire a grant recalling why the NIH did not fund her proposal, disappointingly reflecting about how her lack of experience played a role in the proposal's decline bemoaning:

> The [grant] that I submitted, I keep getting a good score' and then it is not discussed. And then the comments are like, 'She hasn't led a clinical trial.' So, the main comments are, well, she hasn't led a clinical trial or a [specific science] trial. And so, they comment on 'if, she can do it.' Then the other main one, this time, that sank me was someone who had a big problem with an accessory document in database management. He felt that I didn't say that there's going to be two people entering data. And then, I stated I would review the data every six months, and he thought I needed to do it every three months. And he gave me a six on my approach. Those were his only two comments. And then someone else gave me all 1's and a 3, like this fabulous score, and then this person tanked me.

She continued sharing how personal responsibilities might have influenced her research productivity. She frustratingly and in a lively manner (e.g., waiving hands all over the place) exacerbated:

> And then the schools are closed, and there's nothing, they're like "Oh we're gonna take a holiday." And you're like [exclamation], there's nothing you can do, you're out [of the office]. I can't get here before a certain time [gasp], and I have to leave at a certain time to watch [the kids]. And I get a lot done at night after they go to bed, which is not the best time for writing. I find I'm reviewing papers then or making figures or something [late at night].

Sophia shared Amber's sentiment on the challenges of balancing personal/professional responsibilities as a female faculty as she discussed her proposal's decline by the NIH. She highlighted personal expectations levied on female faculty that potentially hamper their research productivity:

> It's hard, especially when you have family, you have children. There are a lot of things [challenges], especially for those female faculty, I think, it is more

> competitive, compared to other, you know, male faculty because of family obligations. You don't expect the [male faculty] to cook and also to take care of the children, [but], you would expect me to, you know, to do all those things [as a female].

Their sentiments are contextually critical in light of research showing that parenthood transition puts a time squeeze on working married couples. This time squeeze creates a gender disparity that is not present at childbirth, as reported in the study that used longitudinal time diary and survey data:

> The transition to parenthood is a pivotal life course transition in which gender performances in the family may be cemented for the next several years. Our results suggest that gender disparities in the work of the family, including paid and unpaid work, were magnified across the transition to parenthood for the primarily highly educated dual-earner couples we studied. According to the time diary fixed effects results, the women in these families experienced a large increase of 3 hours a day in their total work (not including child engagement) across the transition to parenthood, whereas men increased their total work by about 40 minutes a day. This means that, over the course of a year, parenthood increased women's total workload by about 4½ weeks of 24-hour. days, whereas parenthood increased men's total workload by approximately 1½ weeks—a 3-week gender difference (Yavorsky et al., 2015, p. 12).

Emily, a mentor, expressed concern for the upcoming cadre of female faculty. Her perspective revolved around female faculty fortitude in internalizing criticism, overcoming rejection, and standing up for their ideas which would potentially impede their research productivity, as she explained:

> I wondered if you ever come across this [in the literature]. I found this somewhere that women handle rejection differently from men. So, I've been on teams where I'm trying to think if this has ever happened with a male lead, and I can't think of it. Where they, [female faculty] submit [a proposal], and it doesn't get funded, and they're like oh, okay, that didn't work, and it's like, instead of having an attitude like [the male faculty], that these people [i.e., NIH reviewers] don't know what they're talking about, they're idiots! The [female faculty] are like, well they don't like it, I'm not gonna do it again.

Emily's assessment of female faculty mentality towards research productivity is what Sophia had reflected about after the program. She mentioned that she would give up

on a grant proposal idea after one or two rejections in the past and how surprised she was that it takes many more rejections and resubmissions to acquire a grant after watching the program designer's presentation on his many efforts (e.g., five-six resubmissions) to acquire a grant finally. She lamented those missed opportunities recalling:

> I think it was scored 32, [unfundable impact score], or something. The second time it was around the 32, the impact score. So, after [submitting it] twice, we thought, oh, we just could not improve it anymore. We basically gave up. If that had happened after the program, we wouldn't have [given up]. We would have continued to revise and maybe even change the study section.

Emily reflected on the difference between her research development and the current and upcoming cadre of female faculty, saying:

> I've talked about [this problem] with female colleagues who came up with me during our Ph.D. career. Because there was this generation [of female faculty]. We were trained under the women who trained us. We were all coming up at a time where there were very few women in science, and so you couldn't be wishy-washy about getting a rejection, so it kind of weeded out [those who couldn't handle it]. The only women who were there are like [the] hardcore need to suck it up [type], so that's whom we were trained under. And probably what we are doing now is training people with a little more like, it's okay, you know rejection is hard or, you know, it doesn't matter. So yeah, I don't know. We're probably part of the problem.

Adult learning can be challenging because these individuals often have full-time jobs and can be exacerbated by family obligations (e.g., childcare). Hence, adult learning theory encourages the inclusion of adults in the planning of their development, which facilitates their program design based on these individuals' input. Although the gender disparities mentioned by some of the female faculty did not result from the program, not including faculty in planning their development as recommended by adult learning theory potentially contributed to female faculty's lack of grant acquisition (i.e., a program design a flaw). The all-male faculty program designers did not consider female faculty input when designing their development. Gender disparities exist and are well documented).

Female faculty are well aware of the challenges they face. Hence, their inclusion in planning their development as recommended by adult learning could have potentially facilitated a program design that addressed these disparities.

Despite the challenges, Amber did end on a positive note underscoring her resilience and perseverance, saying, “I do feel like there's progress. I got something scored, and the comments they gave me are pretty minor, honestly, so I dropped the ball, so that's good. But yeah, I don't feel good about it yet.” She took personal responsibility saying, “I think, honestly, it wasn't ready yet. When I read the comments, they were spot on. There is still some [work to do], and I think that's where I lacked having a subject matter expert here hurt me.” However, she closed the conversation by postulating that a better mentor match, expertise-wise, might have resulted in a better-quality grant proposal, leading to grant funding. Some of the mentees shared that such a misalignment on science expertise potentially contributed to their lack of research productivity as they communicated, and their mentors corroborated.

Theme 4: Misalignment on Science Expertise

Both mentees and mentors reported a misalignment in science expertise that potentially contributed to the program's ineffectiveness. Several mentees felt that this mismatch hindered their development and contributed to their lack of grant acquisition. Sophia, an assistant professor with a wealth of experience, described the mismatch frustratingly, saying:

> Different contexts, different [scientific] area. It's very hard to work with them, but I do think it is a great opportunity. I really think [that] I wasn't matched [correctly], though. I don't know who did the matching, but I was just not matched with the correct mentor. That's why my mentor just seems like [wasn't able] to help me.

Sophia’s mentor John confirm what she communicated regarding his inability to provide scientific expertise mentorship due to a divergence of research interests stating, “I wasn’t able to help her evaluate the topic. It was way outside of my [expertise] area, but that wasn’t part of my job. So, I couldn't evaluate the impact it would have on the field.”

Madelyn's feelings towards the program were mixed. She described it as "a great idea" while having reservations towards the actual mentorship process, adding, "I don't know, I would have preferred it differently." Her feelings were driven by specific feedback on her proposal development process, lamenting the lack of understanding of her specific research area, saying, "I think the mentor meant well, and feedback was substantial, but it didn't help. I don't think they grasped perhaps, on [their] part, of how difficult I see the process."

Asked to clarify how her mentor didn't grasp her difficulties with the process, she shared a statement [they] made to her "Science is Science," meaning she should know her science and rely on his proposal development process expertise. She frustratingly expressed that sure science is science and the process is the process, but when it comes to "grantsmanship…we, [your mentees], aren't there yet, [not at the mentor's level].”

John’s statement that scientific expertise being not part of his job and Joseph's statement that “Science is Science” highlight a disconnect between mentor and mentee perceptions on what each’s duties were. The program’s design dictated that mentees should secure subject-matter experts from within their respective departments to help them with the science aspect of their grant proposal while relying on the program’s mentors for process (i.e., grant proposal development) expertise. However, the program’s designers removed this requirement due to the mentee's inability to find such faculty from

their respective departments. This issue highlighted organizational limitations that program designers did not anticipate due to them not scanning organizational resources as encouraged by adult learning theory. Despite the misalignment on scientific expertise, many mentors felt that their mentees did not possess the requisite baseline scientific knowledge within their field to be research productive. They discussed this in detail, lamenting that many of the mentees were lacking knowledge.

Theme 5: Knowing the Science

Research productivity (i.e., grant acquisition) is the ultimate aim for any research development program. It was the stated goal of the FRD program and the reason for the nearly year-long mentoring. Mentors were chosen based on their demonstrable record of grant acquisition; hence their reflection on what makes a productive research faculty was paramount. John associated research productivity with a lengthy methodic process that culminates in a meritorious grant proposal submission as he explained:

> I think planning and time management. And that's part of what I learned from this [program] is time management. I know many people who write a proposal in a week, and that's basically how it's viewed when it gets up to Washington. And then they wonder why they're not funded. This [grant proposal development] is a process that needs to be written and rewritten, and you have to have time to forget about it and then pick it up and look at it fresh. Again, it should take seven or eight months to write a proposal. When you do it that way, it probably has a really good shot of being a good proposal.

This line of thinking undergirded the design of the FRD program, which required months-long purposeful proposal development based on expert feedback. Some mentees also cited it as contributing to their research development. For example, Samantha had praised the process for facilitating her grants' acquisition:

> I had never had the opportunity to think about a proposal and work on it for 6,7,8 months. That was totally different from the way that I was used to working on grants. So, I think that was helpful because it gave me a lot of time to think about each of the proposal's sections.

Joseph, a professor rank mentor, took a more faculty-centric perspective on research development. He deconstructed research development and subsequent productivity extensively when asked about his thoughts on mentee research productivity. He critically and somewhat frustratingly elaborated on this saying:

> I don't know if they really have a burning desire to consistently develop new knowledge and challenge existing paradigms and design studies. I don't know if they're inquisitive about their areas versus just wanting to [get funded]. I think it's pretty easy to design a course and teach it. It's easy to show up at meetings and make some cursory decisions from a service perspective. [However] I think it's challenging to have new ideas, and to have those ideas, not well received by your scientific peers. I think it's challenging to come up with new perspectives and paradigms that you apply to research. I think it's challenging in general just to conduct research, particularly with human beings. And so, I think all that stuff while I think it sounds like people want to do it. I think when they actually start doing it, they're like, oh geez, you know, this is not for me.

He delved deeper into what he believed to be essential characteristics required for a successful research faculty career that he felt some mentees did not possess, which impeded their development, potentially explaining the lack of grant acquisition and explaining the null results of the first article, adding:

> I just don't think people have [what it takes] to be a good scientist. You have to be inquisitive, and you have to have ideas, and I think that those are two commodities that are maybe in rare amounts. They have to be creative, and you have to be able to think abstractly because a lot of this stuff is totally abstract. And I think that a lot of faculty have a hard time thinking conceptually about stuff.

Anthony, another mentor, also spoke about this issue when reflecting on why some mentees acquired grants and others did not. Having been at the college for many years and mentored many faculty, he revealed a unique perspective about those faculty, noting:

> I see people applying for grants all the time, where they really haven't done much work in that area. If that's your case, your chance of getting that grant is really low. So, I just think that sometimes people think, writing and getting grants is somehow separate from what you do as a scientist. That's complete

> nonsense, and I think that reflects a certain misunderstanding. You get grants because you're a good scientist, just writing grants without having done any good science, that's gonna fail.

Emily also shared Joseph and Anthony’s sentiments reflecting on how many of her doctoral students and new/junior faculty she had known across the years struggled with formulating research agendas, saying:

> Some faculty have many [research] ideas in mind. They have too many ideas going around and ask me ‘which should I do,’ and it's like just pick one and do the other ones later. So, some people, I think, feel like they have to do everything at once.

She communicated that many new and junior faculty members enter the faculty roles underprepared for the research aspect of their jobs. She shared a suggestion she makes to every one of her doctoral students that, in her estimation, could potentially clarify for them whether they are cut out for a research career, adding:

> I tell all my students to do [a postdoc]. Many of my students are on the fence on whether or not they want to do academia, and I'm like that's precisely why you need to do a postdoc because that will tell you, ‘I hate this, or I want to move forward with it.’ So, I definitely encourage students to do it. I had one of my students went right into a faculty position. I think she's doing fine, but there's some teaching load to it. It's really hard to continue getting your research going if you're thrown right into that. If nothing else, a postdoc allows you to finish up papers, clean them up, and get them published. I almost think when you do the postdoc, you get done with the dissertation, you get all that out of the way, and you start actually doing on-the-job training, pretty much. You hone your skill before because I've talked to a couple of faculty that went straight into faculty roles, and it hit them. They said it's overwhelming to do research and grant proposal development with a teaching load, especially preparation for classes if they’ve never done it before. So, it takes so much time away from the research.

Such a transition (i.e., via postdoctoral training) to a research faculty role was instrumental to Michael’s grant acquisition, as he communicated. He was an assistant professor who acquired the most significant grant among his peers in the program (i.e., multi-million-dollar R01 grant). He explained that he spent several years as a

postdoctoral fellow at another university before his current position. He communicated that his postdoctoral fellowship shaped him as a researcher, explaining his extensive previous research development, saying:

> They're used to cranking out long hours. Let's say me, and you are the top postdocs. They would come up to us and say we've identified you two as the top two postdocs, so you're gonna compete for this [grant] opportunity. We had all these seminars weekly, and we need you to go to those and be asking high-level questions that each of these talks, we need to see your light on during the night, until midnight. These things were physically said to me, [and I said, yes] I will do it. So, I'm used to working long, hard hours because that's what everyone did there. Part of it is that fear. There is not that safety net that we have here like you're gonna have your job. For me, that stress wasn't fun, though. But now that I'm here, more stable, and I can apply that work ethic, it's hard to get out of it.

He attributed this extensive postdoctoral development and a wealth of preliminary data to facilitating his grant acquisition. He added, "I probably, no joke, had seven-plus years of preliminary data. It was ready. [My grant proposal] was cooked, fully cooked, ready to go," which revealed another potential confounder of research productivity that the first article did not address.

Theme 6: Preliminary data

The NIH cites the importance of preliminary data as a factor in successful grant acquisition, as evident in this National Institute of Allergy and Infectious Diseases (NIAID)'s funding news page, for example, "Preliminary research may be the difference between your application's scoring within or short of NIAID's pay lines. For R01s, the thoroughness of your preliminary research can be just as important for impressing peer reviewers as the Specific Aims it supports" (NIAID, 2018, p. np). The same institute also highlights the importance of bolstering grant applications with multidisciplinary co-investigators. Michael shared that his past grant acquisition afforded him opportunities to generate a wealth of preliminary data, boost his peer-reviewed publication record, and

facilitate his postdoctoral training, which was instrumental in him acquiring an NIH-funded grant post participation in the program. He shared that he held on to the grant proposal until transitioning to an assistant professor at the college. In essence, he wanted to acquire the grant and have it count for his tenure and promotion as an assistant professor. He credited his postdoctoral experience and efforts in shaping the grant proposal as he elaborated:

> The difference is that mine had about a million dollars' worth of preliminary data and seven-plus years [worth of work from previous grants]. The COBRE [grant] was $150,000 a year for four years. I had two $50,000 grants. A $100,000 postdoc grant, and a mentor with a thousand pubs, with the biggest exercise, I mean, literally, the biggest exercise [science] study ever published. So that's kind of the pedigree and knowledge I had. My grant proposal was marketed as translational, personalized medicine, and those buzzwords you think NIH reviewers want to see. So again, I'm not bragging; I'm completely lucky and have fallen to the right place. I really think it's about whom you know. Again, I had zero publications as a Ph.D. student. I had like 30 within a couple of years [during my postdoc]. I've worked crazy hard during my postdoc. I think you gotta be at the right place at the right time, but you need to know how to take advantage of that. I think part of that was my previous training; then the culture here, [with fellow productive researchers], helped me continue it.

The fact that Michael's grant, the largest one acquired in the program, was developed before participating in the program highlights how past research development confounded the first article's results and potentially strengthened its' null findings. The study controlled many potential confounders (e.g., race, gender, tenure) to estimate the treatment effect. However, as evident from revelations by Michael, past research development and robust preliminary data seem to influence grant acquisition. Some mentors understood the importance of preliminary data when they recommended that one mentee not submit her grant proposal, as she revealed. Caroline's mentor encouraged her not to submit her grant proposal due to the weakness of her study's preliminary data. She

expounded on the intricacies of the research development process and why the decision was made not to submit a grant proposal as required by the program outline, saying:

> This was a kind of a collective concern. So, as I was working with my small group [fellow mentees], the feedback I was getting was like these pilot [preliminary] data are not very strong to inform some of the grant proposal's ideas as I started drafting the study's specific aims and even the research plan. And it was just the feedback I was getting, not just from the small group. That was the concern of [my program mentor] and other mentors [in the department] working with me on research mentorship. [They] had the same kind of consensus. Because I was doing such a strong pilot study, it just made sense to delay and was the collective advice I got from a collective group of mentors, so not just the [program] mentors. I absolutely could have submitted a proposal that kept in line with those timelines. It just wouldn't have been good [as a grant proposal].

Although mentee scientific knowledge was lacking per the mentors, another concern was their mentee's ability to communicate the significance or importance of their research to the broader scientific community via their grant proposals. They stated that they struggled to sell the science as a byproduct of not knowing their science.

<u>Theme 7: Selling the science</u>

Several mentors discussed mentees struggling with "selling' their ideas to funding agency reviewers. This issue led to several mentors differentiating between a meritorious grant proposal and one that was non-meritorious. Joseph deconstructed the meritoriousness of a grant proposal in detail, first stating:

> What makes a good proposal is that it has a consistent, coherent argument that logically develops through the narrative and is framed from impactful outcomes for both the scientific field and practical application. It also just, visually, needs to be easy to read. It needs to correspond with figures and data that help support the document. So, I think taking all those together. That creates about 75% of it, and then the other part of it is, you know, hoping that you get good reviewers as well.

His second point highlights why there was a disconnect between him and his mentee Madelyn. She had communicated frustration with him when she asked for content

expertise assistance (i.e., the science), to which he replied, “science is science,” insinuating that she should know her science. However, per his perception and due to the program’s design, his purpose was to facilitate the grant proposal development process. However, she was not the only one that struggled with the science aspect of the proposal development process. Marshall, another mentee, had stated that “I think the science is still the hardest part,” which Joseph confirmed:

> I think the mentees struggled to communicate why their ideas are something that we should invest in. You know, they struggle with selling it. So, I think they just struggle with how to communicate it well. I think they struggle because they do not understand their own science, and what are the real challenges that your field is having collectively, and how can you design a study that may help solve some of those challenges. So, I think it makes it hard for them to put together a [sellable] narrative from that standpoint. It's like a short story that you're putting together. I think that many faculties don't know how to write. It's kind of like, ‘I got a couple of ideas. Which one do you think I should do?’ I'm like, I don't care what idea you do. You got to be the one to decide. I can't answer that for you.

Beth, a mentor with a long history of grant acquisition, had a perspective on grant proposal meritoriousness that she felt many mentees did not understand, potentially contributing to their lack of research productivity. First, she relates to knowing one's science and thinking abstractly, as discussed by her peers:

> So, I think any proposal that gets a really good score has convinced the reviewer that there's a real gap in the literature on whatever it is that you're proposing, and if the answer to their question were known, good things would happen. And I think grant writers are more or less skilled at making a really compelling case, both for the gap and the impact. And the impact side, I think, is tremendously hard for junior faculty. You often have to think outside of what you know or do daily to get at the salient impacts that are meaningful for society. So, I think that those two elements are absolutely critical. Without that, a grant will not fare well.

The second component revolved around the design of the proposed study or intervention, which further supports her peer’s notions about knowing one’s scientific field and

confirms the need for a content expert mentor, which many of the mentees communicated that they did not have:

> The other thing then is the design, the approach. It has to be the Cadillac of approaches that you can envision the best possible way to answer that question. You have to lay out the alternatives, the approaches you chose not to take, very clearly in your strengths and limitations section because there are always many different ways to approach the same question. Other reviewers will have other approaches in mind and will favor one or the other. It's really important to position your grant proposal such that a reviewer sees; ah, yes, this applicant actually has thought about alternatives and has made a really educated decision why they're going forward with this approach. And that actually prevents, oftentimes, reviewers from being able to fault an applicant on things they omitted to present, you know, or they omitted, or they chose not to measure or analyze or whatnot. Because if you say, I've intentionally chosen not to do X, Y, and Z, and I, as a reviewer, say, I really wanted them to do X, Y, and Z. But they're just giving me the reason why they chose not to, I'm in a much harder position to say, I still think they need to do it. So that's a very good strategy to stay on the positive side of things. So, the design has to be really excellent; you have to know the strengths and limitations. I think that's really important.

Beth’s third component of a meritorious grant proposal relates to Joseph’s statement about writing consistency, structural coherence, and the logical flow that frames its impactful proposed outcomes that are visually pleasing and easy to read that he felt was missing from many of the mentees grant proposals contributing to them not acquiring a grant:

> The last thing I want to say is that an overarching observation on the grant [proposal] is readability [and] structure, which is critical. Like the other elements, if you have a diamond in the rough as a grant, it will probably not fare well because it makes the reviewers work really hard to find the salient information and defend it. A grant that's super well written, super well organized so that a reviewer can flip back and forth and immediately find what they're looking for. That's what makes a reviewer really happy and facilitates an easy review.

John shared his fellow mentor assessment of mentee inability to communicate their research ideas, adding:

> Number one is your idea. Is it a very solid question that will help the field fill in a [research] gap? Then it's the clarity of each section you write. The [NIH]

> reviewers look at the strengths, weaknesses, those kinds of things. They're looking for specific things in the significance versus an innovation [sections]. Many faculty don't even know the difference because they sound similar, but we tried to illustrate [the difference] to the mentees, the solid differences between the two. The reviewers look at the significance section, and sometimes, [the mentees] are talking about some method. And, well, that turns them right off. Another major thing is that one-page specific aims page you always hear about. It's a single page, but it will turn a reviewer on or off. Sometimes it's even the title. They look at the specific aims page and decide whether to invest more time reading the rest of the proposal.

The mentors stated that they tried to communicate the importance of knowing one's research area, understanding their scientific field, and, most notably, the ability to communicate it to the respective broader scientific community to their mentees. However, some of the mentees did not reciprocate with the requisite effort to develop a meritorious grant proposal as some mentors perceived.

Theme 8: It takes two to mentor

Some of the mentors revealed that some mentees did not take full advantage of the mentoring opportunities that potentially hindered their development. John said:

> Everybody is different, and you have to approach them differently on how you suggest things. It's just a human interaction thing. But you also always have to have a deadline for somebody to reach. Some faculty get things to you a week or two ahead of time, and some people are writing it ten minutes before it's due. And some people live by deadlines; they are a little bit more planned.

Joseph bemoaned some mentees not taking advantage of the wealth of knowledge available to them via the mentors combined years of experience with grant proposal development and demonstrable record of grant acquisition sharing:

> Man! It takes two to mentor. Some people just don't follow through on their end of the bargain. If I had somebody willing to dedicate their time to me [as a mentor], and I knew that they were not just some self-proclaimed [exclamation], a person that they actually had a track record of success, I mean I would milk that for all it's worth. And I just don't think people do that; I just don't think they take advantage. And I think that's what dumbfounds me the most. If you look at the mentors' collective portfolio, and you have individual mentees that are not tapping into that resource like that, just that blows my mind. And at the end

> of the day, it's not my fault if you're not successful. I tell my mentees, I tell my doc students, I tell everybody, I'm like, look, if you want me to talk with you every day, read something every day, and give you feedback on it, I will do my best to make that happen. But if you choose not to do anything. I'll do that as well. It doesn't bother me. You cannot force people to do stuff. You can only provide an environment that offers that opportunity, and then they get to make their own choices.

John provided a politically correct assessment of one of his mentees' attitude towards the mentoring feedback summarizing the interaction, “You know, I definitely got the impression that she was further along than the newer people. I don't want to say not open to suggestions, but a little less. I think she felt she knew where she wanted to go with her proposal.” Sensing that there was more to the interaction than his carefully crafted answer, I gave him an example. I recalled scenarios from my previous career in the business field, where sales managers were apprehensive about hiring experienced salespeople because they were not open to feedback and hence closed to change, to which he retorted, "I think that could be a good generalization."

Similar to John, Deborah communicated that one of her mentees was not receptive to her feedback potentially due to divergence in scientific areas:

> She was probably the least receptive to the feedback. You always kind of got the sense that it was sort of like she was apart from the rest. I feel like she was more likely to come in and say that she was kind of still working on it. Do you know what I mean? I think she had said like she's going to take sort of a different direction with it, and she might have even stopped coming. It was something like that, and then she kind of fell off.

Additionally, she felt that this mentee was receiving conflicting mentoring from her departmental mentor, which she felt contributed to nonproductive research development as she explained:

> I think that's one thing before you go on. I think she was also getting sometimes conflicting mentorship from [her department mentor] and me. And that's kind of too bad because the reality is that it's just science. People have different

> perspectives. But I do wish they'd share that because I think both mentors could think a little bit more together.

Conflicting mentoring was evident in Randy's case as well. He did not submit a grant to the NIH post participation in the program as he shared, "I have not submitted it anywhere [with the NIH], I got a different idea and then submitted it two times to [a different funding agency]." Asked why he did so given the requirement to submit grant proposals to the NIH revealed apprehension on his part as he hesitantly replied, "Why? That is something that I don't want to discuss." He then hesitated for a few moments as if wanting to share the real reason for the lack of submission to the NIH; he anxiously said, "But honestly, I am very much limited about NIH submission due to my own little conflict with my principal investigator. I'm not encouraged to submit anything to the NIH," which explains why he did not submit his grant proposal developed in the program. Institutional records revealed that another faculty funded his position within his department (i.e., departmental mentor), per Randy, demanding he focus on other funding agencies that align with this research.

Program participant feedback shed light on potential contributing factors to its' ineffectiveness. Some mentees entered the development environment with a wealth of knowledge that stood in the way of their research development, while others did not possess the baseline knowledge required for developing a large-scale NIH grant proposal. Association with productive research teams either facilitated or hindered mentee research productivity. A mismatch between mentors and their mentees on scientific expertise contributed to less than productive mentoring sessions. Possessing robust preliminary data to support one's grant proposal was a potential facilitator of grant acquisition and an impedance to grant proposal submission explaining why some mentees did not submit

their grant proposal post-development. Lack of commitment to the mentorship process and pressure not to submit grant proposals to the NIH by departmental mentors could have contributed to the program's ineffectiveness.

This section highlighted many confounders that potentially explain the null results of the first article. These confounders would never have come to light unless employing the comprehensive approach of the CERTi model that complements quantitative assessments with qualitative ones to better understand the talent development process. The following section delves into program fidelity of implementation to gauge program adherence to its' intent (i.e., design), elucidate additional potential mediators of its' fidelity of implementation, and provide more information potentially explaining program ineffectiveness.

Program Fidelity of Implementation

Adherence

FRD Program artifacts, participant interviews, and direct observations by the researcher gleaned adherence to program guidelines. The program-as-intended logic model Appendix F included inputs (i.e., physical, human, and financial resources); activities (i.e., products, services, and infrastructure), output (i.e., workshops, peer mentorship, and milestones), outcomes (i.e., short, intermediate, and long-term), and impact (i.e., grant acquisition). FRD program inputs, activities, leader, mentee, and mentor interview data will serve as indicators of adherence. The active ingredients (i.e., content) that the program aimed to deliver to its' participants were four workshops and the peer-mentoring sessions between the workshops. The program did not specify the number of peer mentoring sessions per its' program design, with one being the minimum.

Moreover, mentors had the discretion to do as many as necessary to facilitate their mentees' grant proposal development. The workshop schedule called for workshops on December 8, 2017, February 16, 2018, May 25, 2018, and September 7, 2018. The peer mentoring sessions were to take place between the workshops. The date ranges were December 8, 2017—February 16, 2018, February 16, 2018—May 25, 2018, and May 25, 2018—September 7, 2018.

Workshop Adherence—The researcher's direct observations confirmed that all workshops were administered on their designated dates, as shown in Table 5.4. Program artifacts (i.e., workshop sign-in sheets) provided coverage detail. Interestingly the level of coverage declined as the program progressed. The first workshop garnered 100% adherence, which could be explained by it being the "Kick-off" session. All mentees attended the session, which would have been reasonable because it was the first session. Another reason potentially contributing to the perfect coverage is that the college dean was in attendance. The second workshop's high 95% coverage could have resulted from the program requiring mentees to present their grant proposal's first draft (i.e., specific aims, significance, and innovation). Mentees had a baseline draft of these sections based on having to provide a "One-page summary (Specific Aims page) for proposed grant topic to work on during [the program]" per the requirement to attend the program.

There was a moderate decline in the third workshop's (i.e., mock reviews) coverage (77.27%) from the second one partially explained by two mentees being absent due to teaching/research-related international travel. The sub-optimal 63.63% coverage on the last workshop potentially resulted from its "Graduation" title. Perhaps some faculty did not feel the need to attend it because there was no new content benefiting

them. However, this was the workshop that the programs' leader highlighted the importance of grant proposal resubmission, which some faculty (e.g., Amber, Caroline, Sophia) felt greatly benefited them. Another potential contributor to the lower attendance in this session is that it was administered during the summer when many faculty members receive a release from teaching and research obligations. Some mentees dropped out of the program towards its' end, further explaining the later sessions' lower coverage rate. The researcher's direct observations confirmed workshop frequency and duration.

Table 5.4 Program Coverage

Workshop	**Adherence**	**Coverage**	**Frequency**	**Duration**
1	Yes	100%	1/1	3.5/3.5
2	Yes	95%	1/1	3.5/3.5
3	Yes	77.27%	1/1	3.5/3.5
4	Yes	63.63%	1/1	3.5/3.5

Peer-Mentoring Adherence—Eleven of the twenty-two mentees and six out of the seven mentors, agreed to participate in interviews. Of those, nine mentees and five mentors provided detail that elucidated peer mentoring session adherence. These data were gleaned from interviews with the participants and via a questionnaire that they received via email. Mentees reported meeting with their mentors an average of 5.6 sessions for peer mentoring across the program's timeline, with each session lasting an average of 1.3 hours. However, mentors reported meeting with their mentees a higher 6.6 session average, with each session lasting about the same length of time, 1.4 hours per meeting. The difference potentially resulted from an outlier mentor who reported meeting

with his mentees 12 times than the rest who reported meeting with theirs 5.3 times. Two of the mentors' mentees reported meeting with him nine and seven times, respectively. Taking the average of these mentees' recollections into consideration brings the peer mentorship average to 5.8, bringing mentor and mentee recollections on the number of peer mentoring sessions into close alignment. Although there was no set number of peer mentorship pre-determined by the program's design, the mentor/mentee self-reports denote high adherence. Mentors reported that their mentees attended all mentor-prescribed mentoring sessions.

The FRD program exhibited low modification, which explains its relatively high adherence. However, a low modification does not detract from the impact of such modifications. For example, the FRD designers program dropped the "content expert" requirement. This modification arose due to the mentee's inability to secure content experts from within the college. Mentee's inability to secure content experts resulted from a limited pool of expert faculty within the college, as explained by one of the program's designers. This slight modification in the program was significant because it led to a mismatch between mentors/mentees on scientific expertise leading to non-productive mentoring sessions that potentially contributed to its' ineffectiveness.

Purposeful modifications can impact a program as much as unintentional ones as one of the program's designers had eluded. He believed that the program's timeline potentially hindered grant proposal submissions. He mentioned that the lag time between the May 25, 2018, mock reviews session and submission deadlines in October created a gap that he felt was unconducive to facilitating grant proposal submissions. He had mentioned that "there was too long of a lag [between proposal development and actual

NIH submission deadlines], “They were going to submit the next June or October. That was too long. It was dragging on, and they were losing their drive." In hindsight, the program's timeframe was more protracted than desired, potentially contributing to a lower adherence to grant proposal submission rates in his estimation. In total, fourteen mentees submitted proposals to the NIH post mentorship, representing a 63.36% rate of submission, compared to the program's 100% goal. It is essential to understand what potentially moderated these results to get a fuller picture of the development process. The following section delves into such potential moderators to provide a more robust analysis.

Moderators of Adherence

Participant responsiveness

According to Carroll et al. (2007), participant engagement in a program may hinge on the degree of relevance they perceive it offers when back on the job. Adult learners are problem-centered and value the real-life applicability of the information acquired in a talent development environment. Therefore, participant responsiveness may moderate the effects of such programs. Self-reports are the most common method of assessing participant responsiveness (McBride et al., 2002), which can gauge participant committal to the intervention, their perceptions concerning its' usefulness (Herzog & Wright, 2005), and their insight regarding its' environmental conduciveness to facilitating their development (Faw et al., 2005).

The researcher asked participants to explain why they decided to enroll in the program and how they felt about the program as a whole to gauge their reactions (i.e., responsiveness) towards it. Program participants' reactions to the program were predominantly positive. Eight of the mentees expressed positive views of the program

with varying intensity. Samantha lauded the program, saying, "I thought the program was amazing," while Amber emphatically exclaimed, "Loved it! I think it's great! Jackie thought it to be great while highlighting its support of junior faculty, adding, "I think it's great! I thought it was really helpful, and I think it's a great way to support junior faculty**.**" Marshall's overall assessment of the program was positive. Still, he felt some aspects of the program were more beneficial than others. Caroline provided a summative assessment of the program, saying, "Overall, it was positive for me," while Michael viewed it favorably due to its' flexible time commitment saying, "I liked it. I don't think it was that big of a time commitment." Randy and Madelyn expressed their favorable views of it, albeit to a lesser intensity stating, "The program is really good" and "I think it was a great idea," respectively. More experienced faculty had mixed and indifferent views towards the program. Charlotte shrugged her shoulders as if to communicate a sense of indifference, saying, "I mean good concepts; I wouldn't necessarily have anything bad to say about it.", while Tim guardedly said, "So, in general, I think it was a good program. There were some obvious shortcomings in certain ways that are just difficult at some level to work with." Sophia expressed that she would recommend the program for junior faculty. However, she frankly added, "I want to say that my mentor really did not help a lot. because his work is lab-related."

Although participant responsiveness was overly favorable towards the program, research cautions against using such a measure to make a linear relationship between it and learning transfer, knowledge application. (Tamkin et al., 2002). A convergence of literature discounts such a relationship showing a scant correlation between participant responsiveness and learning transfer and changed behaviors (Alliger & Janak, 1989;

Holton III, 1996; Warr et al., 1999). Tamkin et al. (2002) posited that the relationship between participant responsiveness and learning potentially results from the developee's conflating stylish yet straightforward presentations by developers (i.e., trainers) with good learning while associating complex ones with poor learning. Some research suggests that learning only occurs when training is unpleasant or challenging. For example, Tan et al. (2003) found that participants' negative evaluations were the best predictor of employee learning, explaining that "satisfaction" (i.e., responsiveness) is not necessarily related to good learning, strengthening the claim that sometimes discomfort is essential.

Additionally, participants' inaccurate feedback, based on biased views of the training's complexity, potentially leads to their trainers' unfair treatment. It can also trigger changes in a talent development program based on incorrect information (Ghodsian et al., 1997), causing some researchers to call for removing responsiveness evaluation altogether Holton III (1996). Mentor/mentee interviews exposed the complexity of the research development process, which lends to the notion that a program's complexity acts as a moderating factor. Hence considering the complexity of the talent development environment and its' potential moderation of program effects is essential to examine.

<u>Intervention complexity</u>

Interventions can be simple or complex with variable levels of detail. Some come with great detail, while others are nebulous. Intervention complexity classification as a moderator within this approach to assessing FOI emanates from research showing that detailed and specific interventions increase FOI (Grol et al., 1998). Well-planned

interventions with a high degree of specificity facilitate adherence (Mihalic & Director, 2009). However, research has also demonstrated that FOI is more likely to be achieved via simple rather than complex interventions (Dusenbury et al., 2003). Complex interventions comprise many components, which may increase the level of variability in their delivery (Greenhalgh et al., 2004). This variability makes implementing the many components more difficult, rendering them susceptible to lower FOI. Hence, researchers recommend balancing between simplicity and specificity of interventions to maximize FOI (Arai et al., 2005; Roen et al., 2006). The complexity of talent development efforts at HEI is evident in the examined FRD program. Although most participants in the program espoused favorable views towards the program, detailed discussions with them regarding their research development exposed the complex and multifaceted research development process's intricate nature, as shown in Table 5.5.

The Table provides magnitude codes gleaned from FRD program participant interviews across the research development spectrum. It highlights magnitude codes elucidating mentee responsiveness, attitudinal, intellectual, and behavioral development. The information provided in the Table seems to support earlier research discounting the causal link between participant responsiveness and learning transfer, knowledge attainment, and behavioral change of talent development program participants. The Table illustrates a complex talent development effort that resulted in varying participant views concerning their talent development. Nine of the eleven interviewed participants had favorable views of the program, one had mixed views, and the last was indifferent. Participant positive responsiveness alludes to a successful program that resulted in the

desired impact. However, the in-depth analysis of participant feedback concerning their development reveals a more complex talent development process.

Table. 5.5 Mentee development

	Attitudinal				Intellectual			Behavioral				
Mentee	Responsiveness	Perceptual	Evaluative	Motivational	Epistemological	Rationalistic	Analytical	Comprehensive	Processual	Procedural	Competential	Productive
Samantha	+	+	+	+	+	+	+	+	+	+	+	+
Tim	+	±	-	-	+	+	+	+	-	-	+	-
Jackie	+	=	+	+	+	+	-	-	-	-	-	+
Amber	+	+	+	+	+	+	+	+	-	-	+	-
Marshall	+	+	+	+	+	+	+	+	+	-	+	+
Caroline	+	+	-	+	+	+	-	+	+	-	+	-
Charlotte	=	-	-	-	-	-	-	-	-	-	-	-
Michael	+	-	-	-	-	-	-	-	-	-	+	+
Randy	+	±	+	+	+	+	+	+	+	-	+	-
Sophia	±	-	+	+	+	-	+	+	+	-	-	-
Madelyn	+	±	-	-	-	-	+	-	-	-	-	-

Faculty feedback on their development experience compared to their overwhelmingly favorable responsiveness supports previous research findings. Participants' baseline reactions to the development processes are not an accurate measure of actual program efficacy, as evident from the information in the Table. Despite viewing the program favorably, the nuances and minutia of the development process reveal that such views did not facilitate desired ameliorative attitudinal, intellectual and behavioral development in many of the participants. Past research development, lack of on-the-job support, a mismatch on scientific expertise, lack of preliminary data, and other factors hindered the development process.

Facilitation strategies

Although no empirical research has conclusively demonstrated the influence of facilitation strategy on FOI, some research posited such a relationship. Facilitation strategy inclusion as a moderator of FOI emanates from research demonstrating their ability for "optimizing fidelity and standardising what is being implemented is arguably even more important in the case of complex interventions, which can be multifaceted and therefore more vulnerable to variation in their implementation" (Carroll et al., 2007, p. 6). Such strategies include manuals, guidelines, training, monitoring, and feedback. However, more facilitation strategies do not necessarily translate to increased FOI; the quantity is dependent on the intervention's level of complexity; simple interventions require fewer facilitation strategies than their more complex counterparts.

The FRD program was straightforward and not complex and employed a few facilitation strategies. The program facilitation strategies revolved around providing its' participants with guideline documents such as the "Research Instructions For NIH And

Other PHS Agencies," "The Anatomy of a Specific Aims Page," "a grant proposal template, and multiple spreadsheet budget templates. The program also provided its' participants NIH review templates to conduct the mock reviews and provide impact scores based on NIH scoring guidelines. Despite efforts to mimic the NIH review environment, one mentor, Deborah, communicated dismay at how unprepared the mentees were for conducting the mock reviews, saying it was nothing like how NIH reviews happen in real life. This facilitation strategy's failure to mimic real-life NIH reviews could have resulted in inaccurate feedback, potentially leading to non-meritorious grant proposals, which warrants examination.

Table 5.6 details mentee mock review impact score, actual NIH impact scores, mentor mismatch, facilitation strategies, and grant proposal submission decision for grant proposals submitted by interviewed mentees. The first column lists interviewed mentee names (i.e., all except Caroline, who did not submit a grant proposal). The second column (i.e., Mock Review Impact Score) represents mentee-issued impact scores during the mock review workshop session. The third column (i.e., NIH Impact Score) represents the actual impact score given by NIH reviewers post submission. The fourth column lists whether that mentee communicated a scientific content mismatch between them and their mentor (Y=Yes, N=No). The fifth column communicates whether the mentor utilized facilitation strategies other than those mentioned earlier (e.g., The Anatomy of a Specific Aims Page) to aid their mentees' development (Y=Yes, N=No). The Sixth column represents the NIH funding decision (Y=Funded, N=Unfunded).

The NIH review process's first step entails reviewing grant proposals for merit. Meritorious grant applications are assigned a "D" (Discussed), while non-meritorious

ones receive an "ND" (Not discussed) code. Discussed proposals are then issued impact scores ranging from 1-9. Lower numbers represent better scores or more meritorious proposals. The average of the scores is then multiplied by 10, making the range between 10-90—for example, an average impact score of 2.7 reports as a 27 impact score. Hence, the review process begins with a baseline measure of whether the proposal warrants discussion in the first place based on its merit, and only those deemed meritorious then receive a score. This process represents the NIH review process's real-world applicability, not how the mock review sessions were structured. The mock review did not include the discussion step and proceeded immediately to the scoring one.

Table 5.6 Impact score comparison

Mentee	**Mock Review Impact Score**	**NIH Impact Score**	**Mentor Mismatch**	**Facilitation Strategy**	**Funding Decision**
Samantha	37	20	N	Y	Y
Tim	27	ND	Y	N	N
Jackie	20	20	Y	Y	Y
Amber	33	ND	Y	N	N
Marshall	27	23	Y	Y	Y
Charlotte	30	ND	N	N	N
Michael	27	27	Y	N	Y
Randy	23	ND	Y	N	N
Sophia	43	64	Y	N	N
Madelyn	70	ND	Y	N	N

A comparison between the impact scores issued by mentees and actual NIH reviewer ones yields mixed results. The mock review session's incongruent design concerning the actual NIH review process and Deborah's recollection of it not aligning with the actual NIH review process seem to have credence at face value. For example, NIH reviewers did not discuss four out of the five proposals given favorable and borderline fundable impact scores (i.e., 27, 33, 23). As mentioned, impact scores run from 10-90, and generally, those scoring 10-30 merit funding based on the NIH's fundable range (NIH, 2021). The fact that mentees gave these grant proposals meritorious impact scores while NIH reviewers did not even discuss them underscores a disconnect between mentees and NIH reviewer merit views. However, the comparison between mentee and NIH reviewer impact scores on the remaining discussed grants seems to align. Although there was a difference between mentee (37) and NIH reviewer (20) impact scores on Samantha’s proposal, both fell within the funding range. Mentees and NIH reviewers' impact scores were identical on Jackie's proposals, 20 and 27, respectively. Despite mentees scoring Sophia's proposal better (43) than NIH reviewers (64), both scores fell in the unfundable range. Although NIH reviewers did not discuss Madelyn's proposal (i.e., deeming it non-meritorious), mentees gave it an unfundable impact score of 70. The comparison between mentee and NIH reviewer impact scores on discussed proposals does not seem to support Deborah's assertion that incongruence between program mock review design and real-world NIH review process. Examining these mixed results in light of mentor mismatch and facilitation strategies could better elucidate the issue.

Eight out of the ten mentees listed in Table 5.6 reported a mismatch in scientific content expertise between them and their mentors (i.e., Tim, Jackie, Amber, Marshall, Michael, Randy, Sophia, and Madelyn). Tim, Amber, Michael, Randy, Sophia, and Madelyn did not receive facilitation strategies from either the program or their mentor to mitigate the mismatch on scientific grounds. Except for Madelyn, all were given fundable impact scores by their peers. However, none of their proposals were even discussed by NIH reviewers, except for Michael's. His past extensive research development seems to have mitigated the mentor mismatch issue. Although there was a divergence of scientific expertise between Jackie and Marshall and their mentors, both received facilitation strategies that seem to have potentially mitigated the mentor mismatch. Jackie mentioned that her mentor Anthony connected her with a content expert at another institution who facilitated her grant proposal development and funding success. Marshall also mentioned relying on his doctoral mentor (i.e., content expert) in developing his grant proposal. Moreover, his mentor Beth also mentioned that one strategy she employed to overcome the divergence of scientific content expertise between her, and her mentees was to delve into their field's literature to understand it better to give good quality feedback during the grant proposal development process.

The FRD program's mock review session's structure did not seem to mirror the NIH review process in its design. As evident in Table 5.6, all grant proposals were issued an impact score by the mentee reviewers, which means the merit discussion step did not occur, making that element of the program (i.e., proposal discussions) incongruent with the real-world environment. Perhaps a better design that aligns with actual NIH reviews would have required mentors to be the "merit" checkers before issuing impact scores.

This step would have elucidated whether each grant proposal was sufficiently meritorious even to receive an impact score. After all, that is how the process takes place in the real-world environment. A second expert examination of grant proposals by multiple mentors that did not facilitate its' development would have provided a more critical view of its's merit. Candid feedback was something that Tim, a mentee, had mentioned as lacking in the mock review sessions. He felt it did not reflect a real-world environment where proposals received blinded candid reviews by peer faculty as he had communicated:

> Sometimes I think we spend too much time with our white gloves on and not enough time saying no! This just isn't good enough. It has weaknesses here and here. If you submit this [proposal], it's not going anywhere. We like to be a little too positive, and you know why? The last thing I want to do is come across as the asshole in the crowd who is just always negative, but at times I think there just needs to be that constructive [feedback] with both [the] positive and negative, working with individuals to ensure that what comes out, really comes out at the quality that's needed. And that's sort of the one piece that I think is really tough because not everybody, thinking of the mentor-mentee relationships, [not] everyone knows [each other] well enough, and is comfortable with that [kind of] relationship.

Tim's recollections seem to align with the impact scores given during the mock review sessions. All were within the fundable range except one.

<u>Quality of delivery</u>

Quality exemplifies the developer's level of preparedness, enthusiasm in delivering the material, pertinent examples, proper interaction with participants, mutual respect, and confidence in answering questions relevant to the topic and contingent on those factors (Carroll et al., 2007). It also revolves around the appropriateness of the delivery and can be contingent on facilitation strategies' quality. Developers could deliver all components of a talent development program (i.e., high adherence) poorly (i.e., low quality). Hence, researchers cautioned not to confuse full adherence with good quality (Dusenbury et al., 2003). The FRD program was delivered with relatively high

adherence, as mentioned earlier. However, a primary component of the program's design, content (i.e., science) expertise, was not adhered to due to organizational limitations. Initially, the program required participants to secure content experts from within the college to provide feedback on the science aspect of mentee grant proposals. The exclusion of this requirement resulted from participants facing difficulty in finding such faculty within the college. One of the program's designers confirmed this and attributed the shortage of content experts to a limited number of college faculty members with specific science expertise.

Addressing such a problem via scanning institutional resources during the pre-planning stage of the program's development would have provided its' designers with ample opportunity to mitigate this challenge. Lawler and King (2000) encouraged organization leaders to reflect on their talent development efforts and view it as a reflection process during the pre-planning stage. They recommended them to understand their organization's culture better, to base the role of the developers (e.g., mentors) based on their talent needs, to assess talent development needs, to evaluate resources requisite to facilitating a successful development effort, and to establish goals that link program effort with their talent's developmental needs.

Several mentors and mentees highlighted the lack of content expertise as an issue contributing to unproductive mentoring sessions. For example, both Sophia and her mentor John corroborated this divergence. She went as far as comparing her mentoring sessions as having to explain her science to her mother. Another mentor, Beth, communicated that she had to research her mentees' scientific content to understand better what they were proposing to investigate to provide better feedback. Although there

was overlap in the scientific field between Deborah and two of her mentees, there was none with her third one. She communicated that that mentee felt like an outsider and eventually stopped showing up to the mentoring sessions.

Another aspect that potentially contributed to low-quality grant proposals was mentee preparedness for grant proposal development in the first place. Discussion with the mentees revealed that the emergence of two groups. The first group comprises faculty possessing extensive postdoctoral training before assuming their faculty roles, while the second group comprises faculty immediately assuming the role of faculty without such research development training. The former's postdoctoral training in research-intensive universities or under-productive research mentors facilitated their intellectual development. They understood their science and the field they operated within, which provided them with the requisite knowledge and expertise to develop a grant proposal. Moreover, these faculty's attitudes towards the research aspect of their careers skew to the positive; they enjoyed the pressure of seeking after grants. Such was the case for Amber, who could not understand why her peers even signed up for a research faculty position at an R1 research-intensive university if they could not handle the requirement of acquiring grants and the pressure associated with it, as evidenced by her recollections:

> I sometimes hear from people that they feel a lot of pressure with it, but I don't understand that because that is the job. They were like, 'Oh, I hate all this pressure I'm getting!' Well, that's the job. [At least] now we're getting support to do that; of course, there's pressure. Some people would say, 'Oh, I hate all the pressure to get all the big money!' I'm like, that's 100% the job.

Similarly, Michael communicated that his stint as a postdoctoral fellow in a hypercompetitive research institution was foundational to his attitude concerning his job's research aspect. He excitedly exclaimed, "My blood boils soft money" (i.e., supporting

one's career via grant acquisition) when discussing his research development as a postdoctoral fellow before assuming the role of assistant professor at the college.

Conversely, the other faculty members who did not progress to their faculty role via such research-intensive postdoctoral training or peer-mentorship exhibited apprehension and tepidness toward the research aspect of their jobs as faculty. They expressed anxiousness towards the grant proposal development process, trepidation in the face of administrator pressure to secure grant funding, and anxiousness of impending tenure and promotion requirements to secure grants. They used words such as "anxiety-provoking," "traumatized," "lack of support" and criticized their detachment from productive research teams within their respective departments (i.e., PLCs) and lack of follow-up post mentorship when describing their feeling toward the research aspect of their jobs.

For example, Marshall expressed both relief and angst at acquiring a grant. His relief revolved around the grant acquisition, providing him with a buffer from his job's research productivity aspect for at least a few years. It did not emanate from a deep-rooted motivation to acquire a grant or his blood boiling soft money, such as in Michael's case. His angst stems from the understanding that the pressure will only increase to acquire a larger scale grant based on his newly acquired one. He recognized this based on an understanding of the organizational context he operated within, driven by a revenue-generating mindset. The mismatch between Madelyn and her mentors' scientific backgrounds undergirded her traumatic experience in the program as she described it. A follow-up email with her in March of 2021 revealed that she did not, nor does she plan to resubmit her grant proposal confirming her negative sentiments toward research

development. Conversations with mentors supported the finding that the absence of content expert mentors was detrimental to some mentees' research development. However, the mentors expressed concern that some mentees lacked a basic understanding of their scientific field in general despite this issue.

Despite mentor sentiments regarding mentees' lack of expertise in their respective fields, content expertise research development was necessary. However, it did occur due to a deficiency in scanning institutional resources to ensure resource availability. Scanning of the institutional context would have highlighted the limitation concerning context experts within the college. As discussed earlier, the mentor mismatch potentially contributed to lower quality non-meritorious grant submissions. Five experienced faculty expressed a mismatch between themselves and their mentor. Of those, only Michael and Sophia's grant proposals received a "D" (Discussed) code, deeming them meritorious by NIH reviewers. Michael acquired a grant based on his submission, while Sophia did not. Her proposal received a non-meritorious (64) impact score leading to an unfunded application. However, discussions with Michael revealed that his success in acquiring a grant resulted from him developing his entire grant proposal during his tenure as a postdoctoral fellow at his previous job and a wealth of preliminary data to support it, as he noted. This information further supports the notion that his mentor did not significantly contribute to his success.

Tim, Charlotte, and Madelyn's proposals received "ND" (Not Discussed) codes, indicating non-meritorious grant proposal applications. The "ND" designation supports the notion that the divergence of scientific background between mentors and mentee potentially contributed to lower quality grant proposals. However, the lower quality

grants potentially emanate from a deeper issue in light of mentor suggestions that many mentees lack a burning desire to develop new knowledge consistently, as Joseph explained. It could also emanate from them not understanding that being a good scientist leads to grant acquisition, as Anthony explained. Their sentiments align with what Beth had shared regarding research productivity. She explained that in order to capture the attention of a grant reviewer, a faculty must first know their field, identify the gap which no other scientist is addressing, highlight why their approach is the best to address it, and most importantly, why their approach is the most innovative in light of what other scientists have proposed as an approach to address it.

Program As Implemented Logic Model

A program-as-intended logic model represented a picture of how the FRD program should work and provided its presumed theory and underlying assumptions. It linked the program's desired outcomes (both short- and long-term) with planned activities/processes and the theoretical assumptions/principles). It provided a conceptual map that outlined the program's components. It served as a diagram describing, theoretically, how it should work to achieve benefits for participants and impact for the institution. It listed an 'If-Then' sequence of proposed changes intended to set in motion via its inputs, activities, and outputs. The proposed sequence of change posited that specific resources are necessary to operate a program. The first sequence posited that if these resources are accessible, then planned activities are accomplishable. The next sequence hypothesized that if planned activities come to fruition, then intended outputs are producible. The third sequence then assumed that if the institution and its' participants benefit, then the intended outputs are produced. The last sequence then followed that if

participants and the organization benefit, specific (e.g., desired) change might occur (i.e., research development and grant acquisition). Per this research's proposed comprehensive evaluation methodology, a detailed assessment of the FRD program's implementation succeeded the program-as-implemented logic model's development. This assessment detailed program implementation in actuality, providing requisite data to compare program intent and implementation.

Program Inputs—The FRD program's inputs were physical, human, and financial resources. Internal inputs included meeting venues, conference rooms, participant offices, equipment, technology, and supplies at no cost to the program or its' participants. Seven mentors, twenty-two mentees, and twenty-two subject matter experts comprised the program's human resources. The college also budgeted $30,000 for the program's provision. Financial inputs included external inputs (i.e., meeting venues, equipment, and supplies) and $3,500 stipends to each mentor, and a one-month salary equivalent stipend for one of the program’s designers.

The examination of program implementation in actuality revealed that the program predominantly deployed all proposed human and financial inputs as planned. However, the program did not deploy one crucial human input (i.e., twenty-two subject matter experts) as the program intended. This discrepancy was due to a shortage of subject matter faculty experts within the college. The analysis revealed that the program conducted no organizational resources scan to determine if such individuals were available before program implementation as recommended by talent development research and adult learning theory (Lawler & King, 2000; Little, 2010; Malcolm & Elwood, 1998; Tran, 2020). This shortage led to a chain reaction starting with a science

mismatch between some mentees and mentors, leading to unproductive mentoring sessions, insufficient feedback, and ending with less than desired research development.

Activities— The FRD program's activities consisted of product, service, and infrastructure activities. The leadership advertised the program via the college's webpage on the university website, promoted it via emails, discussed it with department chairs, and highlighted it at faculty meetings. They also incorporated guest appearances by the leadership and faculty with a proven track record of grant acquisition and mentors' presentations. The program implemented four large-group workshops designed to successfully provide program participants with the required research development to develop a high-quality grant proposal. It facilitated peer-group mentoring between mentors and mentees to share ideas, review proposals, receive constructive feedback on grant proposal development (i.e., Significance, Investigator(s), Innovation, Approach, and Environment). The program implementation assessment showed that the program predominantly deployed all proposed activities. The program made facilities (e.g., meeting venues and conference rooms) available for program events and ensured requisite equipment, supplies, and technology availability for each of the activities.

Outputs— the program's outputs consisted of four group workshops, unspecified peer-mentoring sessions, with one being the minimum. The researcher's direct observation and program artifacts confirmed the adherence to all four group workshops. Coverage (i.e., participants' exposure to each workshop) declined over time, as discussed earlier, 100% for the first workshop, 95% for the second, 77.27% for the third, and 63.63% for the last workshop. Mentor and mentee self-reports confirmed that each group met on average 5-6 times, with each session lasting 1.4 hours, which exceeded the one

meeting minimum. The mentees had seven milestones to meet. First, mentees were to present a draft of the grant proposal's specific aims, significance, and innovation sections. Second, to update those sections based on peer and mentor feedback. Third, complete a research strategy. Fourth, the mentees had to present the previous sections for mock review. The fifth milestones entailed revising the sections based on mock review feedback. The sixth required amending the proposal based on internal and external expert feedback (i.e., mentor and subject matter expert). The seventh necessitated that mentees complete their bio-sketch, finalize the budget, and complete the grant proposal's subcontract section. A review of mentee blackboard submitted documentation revealed that eighteen out of the twenty-two participants developed a grant proposal for mock reviews. Moreover, fourteen of those who developed a grant proposal submitted it to the NIH, which was not congruent with the desired goal of the program of having every mentee develop and submit a grant proposal to the NIH.

Outcomes—The program's short-term desired outcomes entailed ameliorative attitudinal development, which entailed enhancing mentee perceptions, values, and motivation toward the research aspect of their jobs (i.e., positive perceptual, evaluative, and motivational change). Attitudes, lack of resources, policies, laws, regulations, and geography comprise limiting factors that could potentially act as barriers to the desired change (Kellogg, 2004). Table 5.7 illustrated interviewed mentee attitudinal development. As shown in the Table, 64% of the program's participants reported experiencing ameliorative perceptual change. 45% reported experiencing ameliorative evaluative change, and only 36% said they experience ameliorative motivational change. In total, 52% of the interviewed mentees reported attitudinal development, which derives

from averaging perceptual, evaluative, and motivational percentages. Reasons for lack of perceptual change included the program coming too late in one's career, a feeling of being obligated to participate in the program due to lack of research productivity, extensive past research development, and mismatch with the mentor on scientific expertise. Reasons for lack of evaluative change included an existing valuation of the research aspect of one's career per institutional requirement, work-life balance views, and mismatch with the mentor on scientific expertise. Reasons for lack of motivational change included lack of on-the-job support (i.e., lone wolf), extensive past research development, and valuing the teaching aspect of one's job more than the research aspect.

Table 5.7 Mentee attitudinal development

Change	**Attitudinal Development**			
	Perceptual	Evaluative	Motivational	Total
No change (-)	64%	45%	36%	48%
Change (+)	36%	55%	64%	52%

The program's intermediate-term desired outcomes entailed ameliorative intellectual development, which entailed enhancing mentees' epistemology, rationalism, analyticism, and comprehension toward the research aspect of their jobs (i.e., positive epistemological, rationalistic, analytical, and comprehensive change). Table 5.8 illustrates interviewed mentee intellectual development. As shown in the Table, 73% of mentees reported experiencing ameliorative epistemological change. 64% reported experiencing ameliorative rationalistic change, 64% reported experiencing ameliorative analytical change, and 66% said they experience comprehensive ameliorative change. In

total, 66% of the interviewed mentees reported intellectual development, which derives from averaging epistemological, rationalistic, analytical, and comprehensive percentages. Reasons for the lack of intellectual development and its' associated foci of the change (epistemological, rationalistic, analytical, and comprehensive change) mainly included extensive past research development and mentor mismatch on scientist expertise areas.

Table 5.8 Mentee intellectual development

Change	**Intellectual Development**				
	Epistemological	Rationalistic	Analytical	Comprehensive	Total
No change (-)	27%	36%	36%	36%	34%
Change (+)	73%	64%	64%	64%	66%

The program's long-term desired outcomes entailed ameliorative behavioral development, which entailed enhancing mentees' capacity to navigate internal (i.e., university) processes, external (i.e., NIH) procedures to enhance research competence and increase research productivity concerning the research aspect of their jobs (i.e., positive processual, procedural, productive, and competential change). Table 5.9 illustrated interviewed mentee behavioral development. As shown in the Table, 45% of mentees reported experiencing ameliorative processual change. Only 9% reported experiencing ameliorative procedural change, 64% reported experiencing ameliorative competential change, and 36% said they experience productive ameliorative change. In total, 39% of the interviewed mentees reported behavioral development, which derives from averaging epistemological, rationalistic, analytical, and comprehensive percentages. Reasons for the lack of intellectual development, especially in the processual and

procedural area, could be explained by expert grant submission administrative support in the department housing six of the interviewed mentees, which facilitated their internal and external grant proposal submissions. Mentor mismatch, lack of sufficient preliminary data, and internal department mentor discouragement were the underlying causes of the lack of productive change. As was mentioned, one mentee was discouraged from submitting their grant proposal until they collect additional preliminary data, and another was discouraged from submitting their proposal to the NIH by their departmental mentor due to unalignment with that mentors' research (i.e., non-NIH grant work).

Table 5.9 Mentee behavioral development

Change	**Intellectual Development**				
	Processual	Procedural	Competential	Productive	Total
No change (-)	55%	91%	36%	64%	61
Change (+)	45%	9%	64%	36%	39

Impacts—the program's primary aim (i.e., research productivity) emanated from the institutional context. The FRD program's desired impact derived from its' aim to "increase the success rate of [College] faculty applying for NIH Research Project Grants" (i.e., grant acquisition) per its advertisement. One of the program’s designers confirmed this aim as he had stated, "We were at the stage in our school that we had a lot of proposals submitted. We didn't need more proposals submitted; we needed better ones". His concern stems from college data highlighting grant acquisition decline compared to a relatively stable submission rate. The college's faculty submitted 62 NIH grant proposals in 2015, 69 in 2016, and 67 in 2017. However, the NIH funded eleven of the 62 proposals

in 2015, six in 2017, and four in 2017. Hence, the focus was on quality grant proposal submission rather than on their quantity. However, this aim could not come to fruition due to the college's lack of internal subject matter experts to support faculty grant proposal development. Scanning one's organizational environment for resource availability is a critical step to planning talent development efforts (Lawler & King, 2000). Such a scan could have highlighted such a shortage and allowed program designers to address this resource's absence. Moreover, the inclusion of faculty in planning their development would have also highlighted a critical need of theirs (i.e., science expertise), further stressing the need to address this issue. Lack of content (i.e., science) experts initiated a chain of events leading to unproductive mentoring sessions, potentially impeding program efficacy, as illustrated in the program-as-implemented logic model in Appendix G.

Logic Models Comparison

Stolovitch and Keeps (2006) suggested comparing a programs' design to its implementation in actuality via a five-step process. First, develop a logic model representing the program-as-intended. This step involves the documentation of inputs, activities, outputs, outcomes, and impact. The inputs are performance variables that the program aims to improve and are program-specific, emanating from the actual program's context. Second, develop measures of key program indicators. This step entails the development of quantifiable indicators for the program-as-intended logic model. These indicators are derived for each of the logic model's components (i.e., inputs, activities, outputs, outcomes, and impact) to compare to the model-as-implemented logic model in an attempt to ascertain program implementation fidelity. Third, develop a logic model

representing the program-as-implemented. Once the program-as-intended logic model is developed, and key program indicators identified, a detailed collection and analysis of data will enable evaluators to develop a program-as-implemented logic model. This model represents the program as it happened, which may or may not resemble the program-as-intended logic model. Fourth, compare program-as-intended to program-as-implemented logic models. This step allows an evaluator to compare the program as its designers intended to its' implementation in actuality to uncover incongruities. The findings resulting from such a comparison may lead an evaluator to one of the following conclusions:

1. The program was implemented as intended and was successful: good planning, good implementation, positive result.
2. The program was implemented as intended and was not successful: poor planning, good implementation, negative result.
3. The program was not implemented as intended and was not successful: good planning, poor implementation, negative result.
4. The program plan was not clear, program implementation was poor, and the program was not successful: poor planning, poor implementation, negative result.

The fifth and final step entails improving the model. The evaluator utilizes this final step when comparing the program-as-intended and program-as-implemented logic models' results in an unsuccessful program. Using comparison data as feedback, the evaluator can develop a new and improved program that addresses the first model's shortcomings.

The analysis of the FRD program showed its implementation was predominantly congruent with its design. However, several factors potentially contributed to it not

reaching its full potential. First, failure to scan the institutional environment led to a critical component in its' design (i.e., subject matter experts) going missing. This issue led to unproductive mentoring sessions between some mentors and mentees. Second, including faculty in the planning process would have revealed the diversity in the experience present among the program's participants. Specifically, more experienced faculty did not benefit from the program as much as those will lesser experience. Third, the program's protracted timeline could have contributed to attrition in grant submission rates, as one of the program’s designers mentioned. The program's last session was held at the end of May, while submission deadlines were from late September to mid-October, which he felt was too much lag time.

Fourth, the program's design did not take advantage of a critical strength within the college (i.e., productive research teams) to mitigate the lone wolf syndrome. Participants did not receive on-the-job support post mentorship, which some felt was not conducive to their continuous development. This issue is critical in light of research showing it takes a new faculty on average three years to acquire their first large-scale grant in the current hypercompetitive NIH funding climate and other research emphasizing the importance of post-development support. Fifth, the analysis highlighted the importance of preliminary data to successful grant acquisition. Some faculty were encouraged not to submit their proposal until they secure more robust preliminary data. This issue would have come to light with faculty inclusion in their development, as recommended by adult learning theory.

Discussion

This article aimed to understand what can be learned from implementing talent development programs such as the examined FRD program to improve their delivery and maximize their potential effectiveness (i.e., explaining the effect or lack of). Doing so elucidated the intricacies of talent development efforts at HEI that brought several implications to light that HEI and organizational leaders should consider. First, seeking talent development program participants' input in the pre-planning and planning stages of such program's development is critical to talent development program success. Such input is especially critical in light of adult learning theory's recommendation to do so. Adults, per the theory, need to understand what the proposed development means to them, which leads them to provide their input making them more invested in the development process. The Exclusion of the examined FRD program's participants, for example, created missed opportunities for the program's leaders to understand the needs of its' potential participants truly (e.g., science expertise, on-the-job support).

Second, scanning organizational context to ensure resource availability is vital to program success. Resource availability is essential to program theory coming into fruition. No talent development program successfully accounts for all requisite resources to administer such programs successfully. However, formulating facilitations strategies to address resource scarcity is fundamental to counter such issues. The theory of change for the examined FRD program postulated that subject-matter experts would facilitate participant research development. However, a limited supply of such individuals within the college hindered the program's theory of change from materializing. Scanning institutional context for resource availability is critical in the planning stages for HEI

leaders to consider when planning their talent development programs (Lawler & King, 2000).

Third, considering talent's backgrounds, characteristics, and motivations for partaking in the development effort (i.e., participant diversity), is critical to their individual development. Adult learning theory emphasizes that adults enter the learning environment with a diverse array of experiences, which provide a basis for much of their learning that developers should tie to the development process. HEI leaders should pay close attention to their employees' diversity. Special consideration of the diversity of faculty experiences and values should underlie planning and implementation development strategies for the entire talent pool to include those high performers as a comprehensive development strategy, which better aligns with organizational goals to meet current and future organizational needs (Garavan et al., 2012). Managing employee diversity is not optional. It is a critical priority. Many organizations pay close attention and cater to their employees' diversity from the demographic perspective (i.e., race, age, and gender). However, as evident by faculty feedback in this FRD program, the diversity of research experiences is also critical to the success of talent development efforts at HEI.

Fifth, although employing cohort design in the development process contributes to positive talent development program outcomes during the development, it is essential to facilitate such groups' interactions post-development. Research recommends that organizations identify, promote, and strengthen professional learning communities (PLCs) within their institutions (Brown & Duguid, 1991; Glowacki-Dudka & Brown, 2007; Maher, 2001; Steinert, 2010). Cherrington et al. (2018) highlighted the emergence of professional learning communities (PLCs) in higher education settings. These PLCs

form organically at university campuses via personal contacts among faculty seeking to build institutional relationships. These faculty value the face-to-face interactive nature of these groups, welcome their ground-up informality, and appreciate them as safe places to reflect and innovate. HEI leaders are encouraged to foster such PLCs, maintain their organic ground-up nature, create spaces for “champions” to emerge from among these groups to promote them, and, most importantly, make them more representative of faculty within the university. Specifically, they should use such communities to socialize new and junior faculty into the organization to mitigate the lone wolf syndrome's ill-effects, as described by some of this research’s participants. Some participants explained that most college faculty members operate as lone wolves within the college without PLCs' support hindering their research productivity.

In fostering research PLCs and making sure they are representative of all faculty, HEI leaders would address research development disparities detected in this research and confirmed by past research. The lack of grant acquisition experience and expertise disadvantages explicitly new, junior, non-white, and women faculty in a declining and ever more competitive grant acquisition field (Freel et al., 2017b). Research by Ginther et al. (2011); Waisbren et al. (2008) have shown that these groups submit fewer grant applications and ask for lesser funding amounts and years of funding than their white male counterparts. Beech et al. (2013); Ginther et al. (2011) found that it currently takes, on average, 4-5 years for new faculty to secure federal funding as compared to one year in 1980. In the same year, 16% of NIH grant awardees were 36 years of age or younger than only 3% presently (B. Alberts et al., 2014). Grant funding data from the NIH spanning from 2000 to 2006 showed that funding rates for African American faculty

were 10% lower than their white peers (Ginther et al., 2011). Moreover, research has also detected a gender disparity in grant awards as funding for female faculty was lower than that of their male counterparts (41% versus 45%). Women faculty also received fewer years of funding, and on average, $27,000 less funding than their male counterparts (Waisbren et al., 2008).

Sixth, research still shows that HEI lag behind other industry's talent management practices (Lynch, 2007), especially in that their talent development approach is primarily a deficiency remedying approach rather than a development-driven one. These institutions' research development efforts are complex and undergirded by financial pressures (i.e., a financial sustainability strategy), as this study confirmed. It is not a surprise that one of the examined FRD program designers referred to its' participants as "an investment" that begs a return in light of this context, which is not a learner-centric approach the violates the adult learning theory principle stating that adults learn best when the motivation is internal rather than external. (Knowles et al., 2012). HEI leaders should not operate under the assumption of fixing a deficiency (e.g., lack of grant acquisitions), as is the case with many developmental programs. Instead, their focus should be on a positive approach that leverages their faculty's internal motivation and values faculty members' input in the planning process. This learner-centric approach should consider faculty needs, interests, experiences, and capabilities. Of utmost importance in this stage is the choice of development topic, identification and selection of the presenters (i.e., trainers), preparation for delivering the training, the preparation that facilitates the transfer of learning, structuring events that facilitate participation, and including evaluation in the process (Lawler & King, 2000).

HEI talent development efforts should be quality-driven rather than a quantity-driven talent development strategy that transcends customary short-sighted deficiency fixing training approaches focused on short-term objectives. A talent development strategy guided by recognizing the critical importance of talent in facilitating organizations' addressing environmental challenges and achieving strategic goals is a talent-centric one (Tran, 2020). Talent management differs from training in that it is strategic, comprehensive, and continuous rather than short-term training endeavoring to improve knowledge, skills, or attitudes (Caplan, 2013). A holistic approach aligned with organizational strategies and goals focused on developing people within organizations to meet current and future organizational needs is a talent-centric one (Freidberg & Kao, 2008). It fosters an organizational culture that aligns its values with its employees' values (Rothwell et al., 2018). Such a strategy better serves HEI leaders to address an organizational climate that is ever more competitive in grant acquisitions, which is critical to HEI's financial sustainability.

As talent management efforts' complexity increases, the CERTi comprehensive evaluation approach becomes more critical to assess talent development program efficacy. Such an approach is even more essential for HEI, which faces economic turmoil, forcing them to reduce expenditures in talent development program areas such as faculty research development to enhance faculty grant acquisition skills. Before considering these cuts, HEI should conduct robust assessments of their efficacy, including whether they generate more revenue than they cost to administer. These assessments were critical contextually as state divestment in higher education and mounting public pressure against tuition hikes forced HEI to rely heavily on external

funding sources, which became essential in these economically uncertain times. This issue is precisely why the CERTi model incorporates the principles of economic evaluations within its framework to systematically account for total program cost associated with the provision of talent development programs such as FRD programs at HEI compared to their total benefits/effectiveness to determine program return-on-talent-investment (ROTI), which the following article addresses.

Chapter Six: Article 3 – Assessing Program ROTI

Methods

Study Sample

Program participants consisted of seven senior-level faculty mentors (i.e., Associate and Professor rank) and twenty-two mentees who are new and junior faculty (assistant, associate, and professor rank). In total, 26 faculty applied to the first inaugural FRD program. However, after receiving three R21 grants post-enrollment, one participant withdrew from the program to focus on his newly acquired funding. The leadership selected 22 applicants of the remaining 25 to participate in the program. Those not chosen were deemed too early in their career track to be applying for large-scale NIH grants and excluded. Hence, the program's total number was 29, twenty-two mentees and seven mentors: six professor rank, eight associate rank, and fifteen assistant rank faculty. Table 6.1 provides a summary of program participants by rank. It provides the number and percent for each category.

Table 6.1 FRD program participants

Rank	**N**	**%**
Assistant	15	51.72%
Associate	8	27.59%
Professor	6	20.69%
Total	29	100%

Although the total number of participants in the FRD program was 29 (i.e., 22 mentees and 6 mentors), only eleven mentees and six mentors agreed to participate in this dissertation. Of those, seven mentees and five mentors provided cost-relevant data beneficial for this study's completion. Hence, participant's (i.e., seven mentees and five mentors) cost data will extrapolate the study sample's (i.e., 29) total cost data. This process will be explained in detail later in the study. Table 6.2 provides a summary of those participants who provided study cost data by rank. In total, six assistant rank, two associate rank, and four professor rank faculty provided this cost-relevant data.

Table 6.2 Participant cost data providers

Rank	**N**	**%**
Assistant	6	50.00%
Associate	2	16.67%
Professor	4	33.33%
Total	12	100.00%

Research Question(s)

This study's research question is: *What is the cost of implementing an FRD program, and do the financial benefits derived from the program exceed the investment cost?* Subsumed in this broad research question are four sub-questions:

- What is the total monetary benefit of the FRD program?
- What is the total cost associated with the provision of the FRD Program?
- Does the total monetary benefit of the FRD program exceed its' total cost?
- What is the sensitivity of these estimates to different assumptions?

FRD Program Benefit Estimation Method

This Benefit-Cost article relies upon this dissertation's first article (i.e., assessing program effectiveness) in estimating FRD program benefit. That study estimated total grant dollars acquired by all mentees (i.e., program benefit). It employed the CERTi Model's quantitative evaluation approach, which relies on an HRD evaluation framework (Kirkpatrick, 1994) to assess program effectiveness. Specifically, that article utilizes steps three and four (i.e., Behavior and Results) embedded in the follow-up stage of the CERTi Model to calculate FRD program benefit. It also included the use of the propensity score matching statistical technique, Inverse Probability Treatment Weighting (IPTW) to control for potential selection bias (i.e., non-representative sample of the target population), and confounding (i.e., influence on the outcome by unaccounted for extraneous independent variables). The results section of this article will provide additional detail on FRD program benefit estimation and program effectiveness detail.

FRD Program Cost Estimation Method

The Ingredients Method

This study utilizes the ingredients method of cost estimation (Levin, 1975) to tabulate the FRD program's costs. The ingredients method entails identifying all resources associated with the program's provision, including accounting for their opportunity costs (i.e., the value of the resource used for a program estimated by the forgone next-best alternative use, usually captured in the market price). Levin and McEwan (2000) stated that “This approach begins not with a budget, but with the details of the intervention and its resource requirements. Budgets provide an inaccurate estimation of costs, usually understating them” (pp. 45-46). The ingredients method is a

straightforward and rigorous cost estimation method that provides a succinct level of detail about all ingredients (i.e., resources) required to replicate a program's implementation to achieve a given effect. The method entails the identification and specification of ingredients, valuation of ingredients (i.e., costing out and pricing), analyzing and reporting costs (i.e., adjusting costs (e.g., inflation)), calculating costs (e.g., total cost, the average cost per participant, marginal cost), and reporting costs (e.g., annually, aggregated, dis-aggregated).

Identification and Specification of Ingredients

Steps—According to Levin et al. (2017), "The first step in applying the ingredients method is to identify and specify the ingredients of the intervention, program, or reform needed to replicate the implementation (and hence the impact)" (p. 62). Simply put, the ingredients method aims to ascertain an answer to the question, "What resources are required to obtain the observed impacts?" (p. 62). The list of ingredients must be comprehensive to include both financed and in-kind (non-monetary) resources. A crucial aspect of ingredient identification and specification is deep-rooted familiarity with the program, which provides a considerable knowledge base on the resources utilized for provision of the program.

Assumptions—Three assumptions underly the process of identifying and specifying program ingredients. First, sufficient detail should be provided regarding the program ingredients to facilitate their valuation in the next stage of the ingredient's method. Second, consistency in categorizing (e.g., Personnel, Facilities) the ingredients must be ensured. Lastly, "the degree of specificity and accuracy in listing ingredients should depend upon their overall contribution to the total cost of the intervention" (Levin

et al., 2017, p. 63). Educational interventions such as the examined FRD program are notoriously labor-intensive; hence, this study pays exceptional attention to personnel costs. Levin et al. (2017) explain the importance of this by stating that an "error of 10% in estimating teacher time will have a relatively large impact on the total cost estimate because of its relative importance in the overall picture" (p. 63). Contrarywise, they add that "a 100% error in office supplies may create an imperceptible distortion because office supplies are usually an inconsequential contributor to overall cost" (Levin et al., 2017, p. 63). Personnel, training, facilities, equipment and materials, and other program inputs comprise the five classified ingredient categories.

Ingredient Categories—Personnel ingredients refer to all human resources required to provide a program and include but are not limited to full-time and part-time employees, consultants, and even volunteers and listed according to their roles and responsibilities (e.g., dean, program coordinator, mentor, mentee) and qualification (e.g., degree, years of experience in position). Time inputs (i.e., time spent on program-related activities) should also be accounted for and represented by the percentage of a full-time position and hours/days spent. As a rule of thumb, both roles and qualifications are generally qualitative, while time spent is quantitative. Training ingredients revolve around equipping trainers with the requisite instructional methodologies to deliver the instruction or training (e.g., Mentor train-the-trainer session). Identification of training ingredients should differentiate between a trainer's time spent on the training and how much training took place (i.e., preparation time) to gauge whether to distribute costs over time (Levin et al., 2017). Facilities, according to Levin et al. (2017), are "the physical space required for the intervention. This may include space for program delivery or

training or meetings or storage for materials" (Levin et al., 2017, p. 67). Equipment and materials include but are not limited to computers, software, audiovisual equipment, internet access, and paper. Some material's use spans multiple years. Hence it is essential to cost out the ingredients over the program's lifespan correctly. Moreover, identifying other program inputs such as financial incentives (e.g., Mentor stipends), transportation reimbursements, food, and scholarships is vital.

Ingredient Sources—Ingredients collection occurs through multiple sources such as direct observations of the program, interviews with administrators, reports, and program documents (i.e., program description, internal memos and communications, budgets, website). Levin et al. (2017) state that "The most comprehensive and detailed sources of information on ingredients are accessible from those involved in implementing the intervention. Those individuals might include the program designers, program directors, and administrative staff" (p. 71). They suggest a multi-faceted approach (i.e., interviews, observations, and reading of documents) to triangulate ingredients' sourcing.

Valuation of Ingredients

In general, the most common method for valuing (i.e., monetizing) ingredients is market prices because competition produces an equilibrium price representing the value of the good (Mitchell & Dorfman, 1967). The simplicity and availability of market pricing have contributed to their everyday use in the educational field. Several things must be taken into consideration when valuing ingredients, such as geographic location. National average pricing is good for generalizability, but local average pricing is sometimes advantageous, especially when addressing local constituents such as policymakers. The critical consideration in choosing between national and local average

pricing is transparency in detailing ingredient valuation. In case market prices are not available, shadow pricing, estimated as the value of the ingredients' next best use (e.g., the value of a park to the social well-being of a community when calculating the cost of a construction project), provides the value of ingredients. “Societal willingness to pay for a specific impact” is the basis of shadow pricing (Levin & McEwan, 2000, pp. 60-61). Various methods can calculate shadow pricing in the absence of market prices. One can use the market analogy method (i.e., using the market prices for comparable goods) or the defensive expenditure method (i.e., using estimates of society’s willingness to pay to avoid adverse outcomes). Additionally, economists have made use of the hedonic method (i.e., use estimates of how much people are willing to pay for personal gains) and the trade-off method, and the contingent method (i.e., surveying people about how much they would be willing to pay).

Personnel—Personnel ingredients typically constitute the bulk of total program cost; hence it is essential to dedicate serious effort to valuating their costs. The assumption underlying personnel ingredient valuation is that personnel positions “can be filled by attracting persons with the appropriate education, experience, and other characteristics at the prevailing salary and fringe benefits generally paid for such talent in the marketplace” (Levin et al., 2017, p. 80). Thus, this marketplace price becomes the expected prevailing price (i.e., salary and fringe benefits) to attract persons for such a position and represent their costs. Therefore, “ingredients data on the nature of a position and required qualifications and training can be used to estimate a national average salary for a position using available databases” (Levin et al., 2017, pp. 83-84). Despite this marketplace price, it is also essential to consider factors (e.g., working conditions, rural

vs. metropolitan settings) that might affect the valuation of personnel costs (Chambers, 1980). It is also crucial to include fringe benefits (e.g., employer contributions, social security, pension plans, health, and life insurance) in valuating personnel ingredients. Economists usually express fringe benefits as an overall percentage of salaries, never directly, including them in the salary, and are typically estimated at 30% (Levin et al., 2017). Valuation of personnel ingredients should be comprehensive, which entails accounting for volunteers. Volunteers are typically not paid, yet their time and effort carry an opportunity cost and must be valued based on the same process (i.e., role, qualification, and time) and utilizes market pricing (Levin et al., 2017).

Training—Many educational interventions require training in preparation for program delivery. These training can range from the short-term such as one-day workshops, to the long-term such as one-week retreats for certification (Hollands et al., 2013). Training ingredients can divert personnel ingredient time and effort away from primary duties representing an opportunity cost and must be valued. Additionally, in many instances, long-term training is considered an investment in human capital that might reap benefits over several years, which poses an issue of overestimating its value if accounted for just one year. Hence it should be given careful consideration. (Levin et al., 2017).

Facilities Equipment & Materials —Institutions can either rent/lease to administer their interventions or use their existing ones. In the case of the first, "the intervention will utilize rented or leased space so that its market value is evident–either from direct expenditures if using local or idiosyncratic prices or from national average prices for leasing similar facilities if using expected prices" (Levin et al., 2017, p. 87).

The second is very common in educational interventions, where the institution owns the facilities (e.g., conference rooms, auditoriums) used in the intervention. In this case, Levin et al. (2017) suggest using local leasing pricing by comparing the intervention facility to similar ones within the locale or relying on a real estate agent's expertise. Like facilities ingredient valuation, equipment values can be ascertained through either national or local averages or by relying on the equipment's rental values, and in the rare case of the absence of any of this information, one can use replacement values. Materials (e.g., pens and pencils, paper, toner cartridges) are generally difficult to estimate and can consume an evaluator's valuable time and typically constitute 5% of educational interventions' total cost (Levin et al., 2017). Therefore, Levin et al. (2017) state, “one might estimate the cost of supplies by simply adding the total expenditures on supplies to the estimated value of those that are constituted.” (p. 91).

Ingredients Data Collection & Analysis

Data collection—Program artifacts, administrative records, and participant feedback provided information on all Ingredients associated with FRD program administration (i.e., personnel, training, facilities, equipment and materials, and other inputs) spanning the length of the program (November 2017 – September 2018). Each ingredient was described, quantified, and assigned a cost based on the quantity described and priced data available as recommended by (Levin & McEwan, 2000). Table 6.3 categorizes the multiple ingredients associated with the provision of the FRD program.

Table 6.3 Program ingredient categories

Category	Type
Personnel	All Human Resources (e.g., Salaries and fringe benefits of mentors, mentees, administrative staff, and volunteers)
Training	Talent Development (e.g., Mentor training in preparation for the program)
Facilities	Physical spaces utilized for the program. (e.g., conference rooms and event venues)
Equipment and Materials	Instructional equipment and office supplies (e.g., computers, projectors, pens, paper, and printed materials)
Other Inputs	Financial Incentives (e.g., Mentor bonuses, attendance)

The FRD program's personnel ingredients included seven mentors and twenty-two mentees. Its training ingredients incorporated one mentor training session held in preparation for the program. Facilities ingredients associated with the program consisted of internal college conference rooms and offices, and external meeting halls. The equipment & materials comprised instructional equipment and office supplies (e.g., computers, projectors, pens, paper, and printed materials) to provide the mentorship program. Other inputs included financial incentives (e.g., mentor stipends) and group event catering (i.e., food supplies).

Analysis—The ingredient method's underlying assumptions require providing a detailed accounting of program ingredients, consistency of the categorization of ingredients, and a level of specificity of sourcing out program ingredients that ensure

replicability. The semi-structured interviews with the dean of research, program director, mentors, and mentees provided values (i.e., costs) for each ingredient utilized for the program's provision (i.e., personnel, training, facilities, equipment & materials, and other inputs). A review of program artifacts (e.g., blackboard page, program documents, timelines) facilitated the triangulation of values. The Center for Benefit-Cost Studies of Education's *CostOut* tool, an online tool kit designed to capture program cost data based on the ingredients method, was used to systematically account for all costs associated with the FRD program. *Costout*, developed using the U.S. Department of Education, Institute for Education Sciences grant funding to assist practitioners, researchers, and policymakers in conducting cost analyses, is a free online tool that utilizes a prompt system that aids in the specification of all ingredients used in administering programs and subsequently assigns applicable prices based on the ingredients' quantity and quality (CBCSE, 2021). It automatically adjusts prices for inflation, geographic location, and the number of years on investment to estimate the total cost of administering a program and cost per participant.

FRD Program Benefit-Cost Estimation

Benefit-Cost Analysis—According to (Levin et al., 2017), "Benefit-Cost analysis allows us to determine if an educational investment is socially efficient. This determination was made when the monetized benefits—resources accrued as a result of the investment exceeds the costs, which are all resources used to implement the investment" (p. 221). Two central economic metrics used in Benefit-Cost analyses are Net Present Value (NPV) and Benefit-Cost Ratio (BCR), which bring program benefits

and cost together to obtain an economic metric that informs as to the efficiency of educational investments (Levin et al., 2017).

NPV—Net Present Value (NPV) represents the discounted value of the benefit minus the discounted value of the costs. Discounting is a process of determining the present value of money since money is worth more today than worth tomorrow according to the time value of money (TVM) principle (Lokken, 1986). There are many methods for choosing a discount rate. One method is the consumer saving options (i.e., returns sacrificed by consumers in order to consume resources now instead of saving them), and the another is the average return-on-investment made by entrepreneurs in the private sector (i.e., sacrificing resources in one project instead of using them in another) (Levin et al., 2017). There are disagreements on a set discount rate; some suggest setting the discount rate between 0% and 11% (Barnett, 1996), and others argue for a rate closer to 3% (Lipscomb et al., 1996; Neumann et al., 2016). Weighing in on the issue, Levin et al. (2017) state, “The disagreement in the literature suggests that evaluators should choose an initial discount rate of 3% to 5%” as a baseline discount rate and then test for uncertainty by conducting a sensitivity analysis that varies the discount rate between 0% to 10% to check the robustness of their findings. This process allows for the adjustment of the TVM (Levin et al., 2017), as represented by the equations below. B_i and C_i are benefit and cost, *t* is the year in a series ranging from 1 to n, and *i* is the discount rate.

$$\mathbf{B_{PV}} = \sum_{t=1}^{n} = \frac{\mathrm{B}_i}{(1+i)^{t-1}} \qquad \Bigg| \qquad \mathbf{C_{PV}} = \sum_{t=1}^{n} = \frac{\mathrm{C}_i}{(1+i)^{t-1}}$$

The equation below calculates the NPV, where NPV=net present value, B=benefit, C=cost, and PV=present value.

$$\mathbf{NPV} = B_{PV} - C_{PV}$$

Interventions with an NPV amount of less than zero are assumed inefficient and rejected. According to Levin et al. (2017), “The NPV metric has the advantage of being the most straightforward to report and interpret” (p. 222).

BCR—Although the NPV is a straightforward method for ascertaining program return-on-investment, it does come with a tradeoff. The method’s simplicity makes it difficult for comparison across programs because a program’s scale makes such a difference to the total number. A simple adaptation to the NPV metric of dividing benefit present value by cost present value is one method of overcoming this shortcoming and dividing benefit present value by cost present value results in a Benefit-Cost ratio that can compare programs as illustrated by this equation.

$$\boldsymbol{BCR} = \frac{B_{PV}}{C_{PV}}$$

A BCR above 1 represents benefits exceeding costs, and contrarywise, a BCR lower than 1 represents costs exceeding benefits, allowing for a better comparison of program ROI between programs. The aim of the FRD program was the attainment of NIH large-scale R-level grants. Program benefit (i.e., total grant dollars acquired by all participants in the mentorship program) provided by the quantitative assessment of the FRD program (i.e., Program Effectiveness) or the cumulative dollar amount resulting from all grant submissions constituted benefit data. The cumulative costs associated with all ingredient categories (i.e., personnel, training, facilities, equipment & materials, and other inputs) comprised total program cost.

Accounting for Uncertainty—A sensitivity analysis will be conducted for the cost estimates to test for the estimates' robustness. For example, in cost analysis, market-level values are used to price out salary costs. A sensitivity analysis comparing local, regional, and national costs will determine whether salary figures' choice substantively affects program costs estimates. Furthermore, a test of how representative faculty who provided cost data are of the total participant in the FRD program will also provide an additional layer or sensitivity to the analysis.

Findings

FRD Program Benefit

What is the total monetary benefit of the FRD program?

The first article of the dissertation determined no statistically significant difference in grant acquisition between program participants and non-participants (i.e., null effects) based on multiple causal estimation analyses (i.e., propensity score matching, propensity score matching plus weighting, and inverse probability treatment weighting). However, to demonstrate the applicability of the CERTi model in ascertaining program Benefit-Cost, this analysis examines the programs' return on investment (i.e., ROI) in terms of inputs (i.e., the total cost of program administration) and outputs (i.e., total grants acquired by program participants).

The first article's analysis found that most mentees (n=14) submitted a proposal to the NIH by the end of the program, representing 63.64% of participants. The figure is even higher upon removing mentee dropouts (n=5) that did not persist in the program, 88.23%. Once a grant proposal is submitted, NIH subject matter experts review its' scientific merit. Meritorious grant applications are assigned a "D" (Discussed), while

non-meritorious ones receive an "ND" (Not discussed) code. Six mentee grant proposals received a "D," while eight received an "ND" coding representing 42.86% and 57.14% of the total, respectively. Grant proposals designated a "D" (Discussed) are then assigned an impact score by NIH reviewers based on a 9-point rating scale (1 = exceptional; 9 = poor). Each grant's total impact score is derived from the mean of individual reviewer scores and multiplied by ten, resulting in a final impact score. Impact scores for mentee grant proposals were 20, 20, 23,27, 49, and 64. The minimum impact score was 20.00, the maximum was 64.00, the first quartile impact score was 20.75, the third quartile was 43.50, and a median and mean at 25.00 and 33.83, respectively.

Each grant proposal submitted to the NIH also receives a percentile rank, which is calculated by ordering impact scores of grant applications and is derived from the approximate percentage of applications assigned a better overall impact score during the past year. The lower the percentile rank, the more meritorious a grant application is. The impact scores for the six discussed mentee grant proposals were 4,7 and 15. Two grant proposals did not receive a percentile rank due to their impact scores being non-meritorious (i.e., 49 and 64), and one was part of an NIH study section that does not issue percentile ranks. The minimum percentile rank was 4.00; the first quantile was 5.50. The median was 7.00, the mean was 8.67, the third quantile was 11, and the maximum was 15. Four mentee grant proposals received funding from the NIH, representing 28.57% of total grant proposals submitted, making it higher than the NIH grant proposal success rate (20%) for 2019 (NIH, 2020). The grant awards were $284,910, $439,500, $453,813, and $3,429,123 totaling $4,607,346, representing the total benefit of the FRD program.

FRD Program Cost

What is the total cost associated with the provision of the FRD Program?

Personnel Ingredient Costs—Personnel costs primarily drove FRD program costs. 29 faculty (seven mentors and 22 mentees) participated in the FRD program. However, only 17 (eleven mentees and six mentors) participated in this dissertation, of which 12 (seven mentees and five mentors) provided cost-relevant data. These 12 participant's cost data extrapolated the study sample's (i.e., 29) total cost data. Workshop attendance, peer-mentorship, and mentor/mentee time spent on grant proposal development (e.g., writing, giving feedback, revisions) comprised total program time (i.e., time spent on program-related activities) by participants. The following questions elicited cost data from program participants; How many workshop sessions did you attend? How many individual meetings in total did you have with each mentor/mentee? How long did those meetings usually last? Where were those meetings held? How much time did you spend reviewing and giving feedback to your mentor/mentees? Were you paid any bonuses/stipends as part of your involvement in the program beyond your usual university pay (i.e., salary & fringe benefits)? How much, and how often?

Among those who provided cost data (i.e., cost data participants), assistant rank faculty (n=6) averaged 97.13 hours, associate rank faculty (n=2) averaged 58.50 hours, and professor rank faculty (n=4) averaged 28 hours in the program as illustrated by Table 6.4.

Table 6.4 Average hours per participant rank

Ingredient	n	Avg. hours
Assistant Professor	6	97.13
Associate Professor	2	58.50
Professor	4	28.00

Table 6.5 provides the total hours for all faculty participants in the program based on the extrapolated average hours provided in Table 6.4. There were 15 assistant rank faculty, seven associate rank faculty, and six professor rank faculty, among those who participated in the program for a total of 29. The number of faculty multiplied by the average hours yielded total hours for each group. Assistant rank faculty total hours were 1456.95, associate rank faculty total hours were 470.00, and professor rank faculty total hours were 168.

Table 6.5 Average hours per participant rank

Ingredient	n	Avg. hours	Total hours
Assistant Professor	15	97.13	1456.95
Associate Professor	8	58.75	470.00
Professor	6	28.00	168.00

Table 6.6 illustrates personnel ingredient costs. Assistant professor rank faculty total hours in the program were 1456.95. Their adjusted price was $51.47, bringing their total cost to $74,984.85 (i.e., $2585.68 per participant), or 42.75% of the total program cost. Associate professor rank faculty total hours in the program was 470.00. Their

adjusted price was $91.82, bringing their total cost to $43,155.40 (i.e., $1488.18 per participant), or 24.60% of the total program cost. Professor rank faculty total hours in the program was 168. Their adjusted price was $139.32, bringing their total cost to $23,405.93 (i.e., $807.10 per participant), or 13.34% of the total program cost. Total personnel per participant cost was $4,880.90, representing 80.70% of the total program cost. All prices were expressed in 2018 (i.e., year of program administration) and at a 3.5 discounted rate.

Table 6.6 Personnel ingredient costs

Ingredient	**Quantity**	**Adj. Price**	**Total Cost**	**Per Participant**	**% Of Total**
Assistant Professor	$1,456.95	$51.47	$74,984.85	$2,585.68	42.75%
Associate Professor	$470.00	$91.82	$43,155.40	$1,488.12	24.60%
Professor	$168.00	$139.32	$23,405.93	$807.10	13.34%
Total	**$2,094.95**	**$282.61**	**$141,546.17**	**$4,880.90**	**80.70%**

Facilities Ingredient Costs—Workshop venue rentals comprised the total cost of facilities ingredients. There were three external workshops, and one internal, at a different college within the university. The total facilities cost was $1311.47 or $45.22 per participant, representing less than 1% (0.75%) of the total program cost. Workshop 4 was the most expensive facility ingredient at $950.00. The significant difference in cost between the external and internal workshop venues was that the college had a standing contract with that external venue that reduced the cost considerably compared to the internal one, as listed in Table 6.7.

Table 6.7 Facilities ingredient costs

Ingredient	Quantity	Adj. Cost	Total Cost	Per Participant	% Of Total
Workshop 1	1	$61.47	$61.47	$2.12	0.04%
Workshop 2	1	$150.00	$150.00	$5.17	0.09%
Workshop 3	1	$150.00	$150.00	$5.17	0.09%
Workshop 4	1	$950.00	$950.00	$32.76	0.54%
Total	**4**	**$1,311.47**	**$1,311.47**	**$45.22**	**0.75%**

Equipment & Materials Ingredient Costs—Presentation equipment (i.e., audiovisual equipment) comprised program equipment & materials. Audiovisual equipment and materials cost $179.28 for the first workshop, $175.00 for the second, and $40 for the fourth. Program participants used the equipment and materials for presentations as listed in Table 6.8. The program director utilized them to present training pieces, while the mentees used them to present their progress on various grant proposal sections as they developed them. In total, equipment and materials represented a low investment for the college. This category represented less than one percent of total program investment (i.e., .22%).

Table 6.8 Equipment & Materials ingredient costs

Ingredient	Quantity	Adj. Cost	Total Cost	Per Participant	% Of Total
Workshop 1	1	$179.28	$179.28	$6.18	0.10%
Workshop 2	1	$175.00	$175.00	$6.03	0.10%
Workshop 3	1	$40.00	$40.00	$1.38	0.02%

Total	**3**	**$394.28**	**$394.28**	**$13.60**	**0.22%**

Other Inputs Ingredient Costs—Stipends, catering, and parking costs comprised program other inputs ingredient costs. Five mentors received a one-time $3,402.20 stipend for their roles as mentors in the program. The Dean of research communicated that he did not receive a stipend, while the program director's compensation was a one-time $10,000 payment for organizing and managing the program and mentoring faculty during the program. Catering costs were $1,307, $1,529.04, $1,518.24 and $642.16 and workshop three's parking costs were $140. In total, other inputs cost the college $32,147.58 or 18.33% of the total program cost, as Table 6.9 shows.

Table 6.9 Other Inputs ingredient costs

Ingredient	**Quantity**	**Adj. Cost**	**Total Cost**	**Per Participant**	**% Of Total**
Mentor Stipends	5	$3,402.20	$17,011.02	$586.59	9.70%
Director Stipend	1	$10,000.00	$10,000.00	$344.83	5.70%
WS-1 Catering	1	$1,307.12	$1,307.12	$45.07	0.75%
WS-2 Catering	1	$1,529.04	$1,529.04	$52.73	0.87%
WS-3 Catering	1	$1,518.24	$1,518.24	$52.35	0.87%
WS-3 Parking	1	$140.00	$140.00	$4.83	0.08%
WS-4 Catering	1	$642.16	$642.16	$22.14	0.37%
Total	**11**	**$18,538.76**	**$32,147.58**	**$1,108.54**	**18.33%**

Total Ingredient Costs—Personnel, facilities, equipment & materials, and other input ingredients comprised the total program cost. Personnel ingredient cost comprised the bulk of program cost totaling $141,546.17**,** or $4,880.90 per participant**,** representing

80.70% of the total program cost. Facilities ingredient costs totaled $1311.47, or $45.22 per participant, representing less than 1% of the program's total cost (i.e., .75%). In total, the equipment & materials ingredient costs were $394.28, representing $13.60 per participant, or .22% of total program costs. The other inputs ingredients represented the second-largest cost (18.33%), totaling $32,147.58, representing $1,148.12 per participant. The program's total cost was $175,399.50 or $6,048.26 per participant, as illustrated by Table 6.10. Comparing total program cost to its' total benefit as gleaned in the first article of the dissertation will determine if the program was worth the investment. The next section utilized the Benefit-Cost ratio to determine the program's ROI and tests the cost estimate sensitivity.

Table 6.10 Other Inputs ingredient costs

Ingredient	**Total Cost**	**Per Participant**	**% Of Total**
Personnel	$141,546.17	$4,880.90	80.70%
Facilities	$1,311.47	$45.22	0.75%
Equipment & Materials	$394.28	$13.60	0.22%
Other Inputs	$32,147.58	$1,108.54	18.33%
Total	$175,399.50	$6,048.26	100.00%

Benefit-Cost Results

Does the total monetary benefit of the FRD program exceed its' total cost?

The program's total monetary benefit (i.e., output) in net present value is $4,607,346.00, while its' total cost (i.e., input) in net present value is $175,399.50 using a 3.5% discount rate. The programs total benefit greatly exceeded its' total cost with a net present value

(NPV) of $ 4,431,946.5 (i.e., $4,607,346.00 minus $175,399.50). The program's Benefit-Cost ratio is 26.27 (i.e., $4,607,346.00 divided by $175,399.50), which is much higher than one, indicating that the program has a positive return-on-investment, meaning it generated much more output (i.e., total grant dollars acquired by its' participants) than inputs (i.e., the total cost of its' administration).

Sensitivity Analysis

What is the sensitivity of these estimates to different assumptions?

Sensitivity of costs—The cost estimates in this study relied on prices for ingredients based on a 3.5 discount rate and expressed in national (i.e., all states) geographic location context. Three sensitivity analyses provided a robustness check of cost estimates. The first sensitivity analysis tested various discount rates to check the robustness of the cost estimates. Varying the discount rate from the original 3.5% to 5%, 7%, and 10% yielded no difference in the price estimates. The second sensitivity analysis varied the geographic location from the national level (i.e., all states) to the state level (i.e., the state where the FRD program took place) at the 3.5% discount rate considering all areas of the state (i.e., metropolitan and non-metropolitan), which resulted in a lower total price of the FRD program, $161,718.94 or $13,680.56 less than the national level total price estimate. To further check the sensitivity of the cost estimates, adjusting ingredient costs estimates for metropolitan area designation resulted in a total price of $164,173.91 for metropolitan areas within the state, while non-metropolitan areas' total price was $150,493.36. The lower estimates supported the primary analysis's conclusion that the program's total benefit exceeded its total cost. Table 6.11 details differences

between the national and state pricing for each ingredient considering all areas of the state.

Table 6.11 Other Inputs ingredient costs

Ingredient	**National**	**State**	**Difference**
Personnel	$141,546.17	$130,506.18	$11,039.99
Facilities	$1,311.47	$1,209.17	$102.30
Equipment & Materials	$394.28	$363.52	$30.76
Other Inputs	$32,147.58	$29,640.07	$2,507.51
Total	$175,399.50	$161,718.94	$13,680.56

The third sensitivity analysis relied on findings from the previous study of this dissertation (i.e., study 2) to test the robustness of this study's cost estimates. Michael, who acquired the most considerable sum of funds via his grant acquisition among the mentees, communicated that he developed his entire grant at a different institution and held on to it until securing his current position to ensure it counted towards his research productivity and tenure metrics. Samantha also communicated that the grant she acquired was not the one she worked on during the program. Hers was a resubmission from an earlier attempt, a K-level grant that was not the program's type aimed for its’ participants to acquire (i.e., R-level grant). Michael’s grant totaled $3,429,123 while Samantha’s totaled $439,500 for a combined total of $3,868,623. Reducing this amount from the total benefit of the program ($4,607,346) yields a total benefit of $738,723.00, which still exceeds the program’s total cost of $175,399.50.

Discussion

This study aimed to demonstrate the applicability of the CERTi model in economically evaluating the worth of talent development programs such as FRD programs at HEI. It utilized the fifth level of its' micro-level analysis (i.e., ROI) to answer the question, "*What is the cost of implementing an FRD program, and are the financial benefits derived from the Program worth the investment cost?"* along with its' subsumed questions (i.e., What is the total monetary benefit of the FRD program? What is the total cost associated with the provision of the FRD Program? Does the total monetary benefit of the FRD program exceed its' total cost? What is the sensitivity of these estimates to different assumptions?)

The study employed a systematic approach that utilized the ingredients method of cost estimation to calculate the examined FRD program's total cost. It did so by identifying and specifying program ingredients, costing the ingredients, adjusting the costs, aggregating the costs, and reporting the costs. The analysis utilized market national pricing to cost out each ingredient (i.e., personnel, facilities, equipment and materials, and other inputs) at a 3.5 discount rate. It calculated the costs per ingredient, per participant, in total and reported the costs in aggregate (i.e., total program cost) and disaggregated (i.e., per ingredient). The study determined that the examined FRD program yielded a positive return-on-investment, meaning that its' total monetary benefit ($4,607,346) exceeded its' total monetary cost ($175,399.50). The study conducted a sensitivity analysis of the cost estimates to test for the estimates' robustness. The study's findings withstood cost estimate sensitivity testing whether by varying discount rate,

changing the geographic location, or excluding funding incongruent with program goals. Despite these positive results, several study limitations present themselves.

The first limitation of this study is that it relied on program benefit from the first study of this dissertation that determined that there was no statistically significant difference between FRD program participants and their non-participating peers within the college in grant acquisition. Hence, utilizing the $4,607,346 in acquired grant funding as a basis for the total program poses a limitation for the study, especially that we cannot reject the null hypothesis in the first study (i.e., there is no significant difference in grant acquisition between FRD program participants and those who did not participate in it). The second limitation of the study presents itself that only a handful of faculty provided cost data. In total, 29 faculty participated in the program, while only 12 provided cost-relevant data. The third limitation revolved around elapsed time. Three years had elapsed since program participants partook in the program, potentially posing a problem concerning their recollection of actual time and effort involved in the program. Despite these limitations, this study provided a methodic approach to assessing the ROI of talent development programs that organizational leaders can employ to assess such programs' worth.

CHAPTER SEVEN: CONCLUSION

Summary

This dissertation's purpose was to introduce the CERTi Model, an innovative, comprehensive, and interdepended approach combining quantitative, qualitative, and economic methodologies, and demonstrate its' applicability in assessing talent development efforts' efficacy. It did so by holistically assessing the efficacy of an FRD program at an HEI aiming to increase the grant acquisition skills of its' participants, which demonstrated the applicability of this evaluative framework for practice and scholarship. The first article of the dissertation determined no statistically significant difference in grant acquisition between FRD program participants and non-participants (i.e., null effects). The second article highlighted several factors that potentially contributed to the first articles null effects, such as participant past research development, science expertise mismatch between mentees and their mentors, gender role responsibilities, scientific knowledge, preliminary data, administrative support, and association with productive research teams (i.e., PLCs) within the college. The third article determined that despite the program generating more output (i.e., grant dollars acquired) than its' total input (i.e., total program administration cost), the article could not directly link that total output to the program due to the first article's null effects.

Such an approach is critical contextually as HEI faces an organizational environment wrought with declining public financial support and accountability demands characterized by increased pressures for financial return-on-investment expectations. As

the economic crisis resulting from the great depression (i.e., 2008) intensified, higher education leadership grappled with identifying alternate sources of revenue to support university operations. They increased tuitions and accepted higher proportions of out-of-state students who pay higher tuition rates. Revenue-deprived universities engaged in competition with their peers to obtain external grant funding to alleviate budgetary pressure. Because funds have become increasingly limited (Hourihan & Parkes, 2016; Howard & Laird, 2013; NIH, 2019b), this competitive environment continues to exist today and has become the new normal for a growing number of these institutions. However, as state divestment towards higher education institutions becomes the norm, tuition increases abate, and out-of-state student ratios level off, federal grant funding has become crucial for public research university financial sustainability (Gallop & Svare, 2016).

Universities seek all forms of external funding (i.e., private, foundation, local, state, and federal) to compensate for state divestment. However, federal grants are at the forefront of these efforts because they provide indirect cost revenue to universities for the support of general administration and facilities cost incurred for research, as opposed to the other grants that seldom pay for such expenses (Ammons & Salterio, 1999; Canizares, 2008; Ledford, 2014; Noll & Rogerson, 1997; Sale & Sale, 2010; Zuiches & Vallely, 1987). Indirect costs are also known as facilities and administration costs (i.e., overhead) and are the actual costs of university operations, such as the cost of operating and maintaining buildings and capital depreciation, and are not easily identifiable with specific research projects but incurred for the joint objectives related to all research projects at an institution of higher education (Ledford, 2014). The federal government

reimburses universities for these funds to encourage research, alleviating researcher concerns about overhead chipping away at their total research dollars. They also serve as a dual benefit since students also take advantage of these facilities for learning (Sale & Sale, 2010).

Beyond financial benefits, earning a competitive and nationally recognized grant brings the institution prestige (i.e., national recognition, rankings, and publications), an institutional effectiveness metric for many research universities (Ali, Bhattacharyya, & Olejniczak, 2010; Devine, 2009). Federally funded research grants play a critical role in developing and disseminating new knowledge and enhancing faculty recipients' reputations (Ali et al., 2010). For example, many universities clamor for membership in the prestigious Association of American Universities (AAU), composed of the nation's leading research universities. The AAU is an elite membership by invitation only organization whose members accounted for 58 percent of U.S. universities' research grants, 52 percent of all doctorates awarded in the United States, and 43 percent of all Nobel Prize winners (AAU, 2020). Among the primary metrics for inclusion into this prestigious organization is competitively funded federal research support (i.e., federal grants), highlighting the importance of these funds to higher education institutions. As the world copes with the financial implications of the COVID-19 global pandemic, HEI, already under fiscal strain, is sure to divest in talent development efforts such as FRD programs or potentially eliminating such programs. Hence, comprehensively assessing the efficacy and ROI of such programs is ever more crucial before such divestment takes place.

The dissertation employed a single-study interdependent or recursive format which permits utilization of data from a similar sample population across multiple articles facilitating the exploration of unique variables and applying various methodological approaches. Chapter one of the dissertation articulated the problems it planned to address, communicated its purpose, highlighted its significance, listed its research questions, and detailed its study design. Chapter two provided a detailed review of the literature progressively detailed the varying elements comprising the CERTi model both on the macro and micro levels. Chapter three introduced the examined FRD program serving as an example to demonstrate CERTi’s applicability to provide methodological context for all three dissertation articles. Chapters four, five, and six represented each of the dissertation articles, each comprised of methods, findings, and discussion sections. The first article examined the program’s effectiveness, the second its’ implementation, and the third its return-on-talent-investment (ROTI).

Findings

The first article’s research question was, “*Are faculty recipients of FRD programs more likely to increase their grant acquisition?”* The article's methodology utilized propensity score matching and inverse probability treatment weighting (IPTW) to create a synthetic comparison control group based on shared baseline covariates (e.g., gender and race) to control confounding in the absence of randomization. The comparison in grant acquisition between FRD program participants and the comparison group determined whether there was a difference in grant acquisition between FRD program participants and comparison group faculty based on matching, matching with weighting, and inverse probability treatment weighting. The article's findings determined no

statistically significant difference in grant acquisition between FRD program participants and those who did not participate in it.

The second article's research question was "*What can be learned from implementing FRD programs to improve their delivery and maximize their potential effectiveness?"* The article's findings determined a science expertise mismatch between mentees and their mentors, past research development (e.g., postdoctoral fellowships), gender role responsibilities, scientific knowledge, preliminary data, administrative support, and association with productive research teams (i.e., PLCs) within the college potentially affected participant research development and productivity. Although the examined program design called for mentees to secure science subject-matter exerts from within their respective department, that requirement did not come to fruition due to the lack of such experts within the college. Some experienced faculty did not benefit from the program due to their extensive past research development (i.e., doctoral and postdoctoral training), making program content rudimentary and unbeneficial. Several female faculty shed light on how gender-role responsibilities (e.g., childcare and home duties) potentially hampered their research productivity (i.e., grant acquisition). Knowledge of one's scientific field and robust preliminary or pilot data emerged as contributors to successful bids to acquire grants. Intra-departmental research-related departmental support also emerged as a mitigating factor in grant proposal submission and acquisition. Lastly, association with productive research groups and research productive faculty members facilitated grant proposal development and submission.

The second article reconstructed the examined FRD program as its designers intended to compare to its implementation in actuality to determine its fidelity of

implementation. Four possible results could have emerged from the analysis. 1) The program was implemented as intended and successful: good planning, good implementation, and positive results. 2) The program was implemented as intended and was not successful: poor planning, good implementation, negative result. 3) The program was not implemented as intended and was not successful: good planning, poor implementation, negative result. 4) The program plan was not clear, program implementation was poor, and the program was unsuccessful: poor planning, poor implementation, and negative results. The analysis determined that the program was implemented as intended and was not successful: poor planning, good implementation, a negative result, based on the findings. The lack of organizational environmental scanning and not involving faculty in planning their development as encouraged by adult learning theory contributed to the program's negative results.

The third article’s research question was, “*What is the cost of implementing an FRD program, and are the financial benefits derived from the program worth the investment cost?”* The article’s findings determined that the total cost of program provision was $175,399.50, and the financial benefit was $4,607,346. A sensitivity analysis tested cost estimate robustness by varying discount rates from the original 3.5% to 5, 7, and 10%, changing geographic location from the national to the state level, and deducted two grant funds from the program benefit data. The sensitivity analyses yielded no difference in program estimates. The determination was that these findings are tenuous given that the first study determined no statistically significant effect in grant acquisition between program participants and their non-participating peers. Hence, the article only provided a descriptive analysis of FRD program inputs (i.e., costs) and

outputs (i.e., grant dollars acquired) without claiming causality between the program and its hoped-for benefit.

Implications

Several implications emerged that are pertinent to organizational leaders in general and HEI leaders specifically. These leaders need to involve their program's participants, the faculty who are at the forefront of seeking after and expected to acquire these grants, in planning their development. This involvement elucidates valuable information that facilitates accounting for these individuals' needs relating to the grant acquisition requirement of their jobs, facilitating better talent development programs, such as FRD programs. Parallel to seeking faculty input in planning their development, scanning organizational context ensures the availability of requisite resources necessary for program success. Such environmental scanning is especially critical regarding human resources such as subject-matter experts facilitating learning and transfer, leading to behavioral changes. Moreover, these scans afford organizational leaders and talent program designers opportunities to formulate facilitation strategies that would mitigate the ill effects of resource inadequacy. Although it is essential to account for participant backgrounds, characteristics, and motivations when planning talent development efforts, the findings of this dissertation suggest that accounting for their diversity of experience is critical, especially in the higher education research development field.

The findings of this dissertation also corroborate past research underscoring the benefits of cohort design. Although this design contributes to positive talent development program outcomes and facilitates such groups' interactions post-development, the findings showed that cohorts tend to continue without organizational support post-

development, which begs a formalization of organizational support. Organizational leaders should identify, promote, and strengthen these cohorts and facilitate their development into long-lasting professional learning communities (PLCs) within their institutions. Moreover, such communities should serve as institutional mechanisms that socialize new and junior employees into their organization, mitigating the lone wolf syndrome's ill-effects as described in this dissertation.

Many complex financial pressures undergird HEI institutions' research development efforts as they vie for financial sustainability (e.g., state divestment, pushback against tuition hikes, resistance to higher out-of-state student ratios). These pressures are worsening as these institutions reel from the financial ramifications of the COVID-19 global pandemic, which has exacerbated financial pressures on HEI beyond traditional state divestment. Hence, the most significant implication of this study for these institutions' leaders and their peers in other industries is that the complexity of talent development efforts begs a comprehensive approach that provides a broader picture of such effort's efficacy. This approach is especially critical since research has shown that HEI lags behind other industries' talent management practices (D. Lynch, 2007). A holistic (i.e., Macro/Micro) approach that provides thorough and pertinent information allows them to make informed decisions regarding their talent development programs within a new financial environment, allowing for a cyclical refinement approach to provide them with a competitive talent advantage. The Comprehensive Evaluation of Return on Talent Investment Model (CERTi) is one such approach that these leaders can utilize in comprehensively assessing the efficacy and ROTI of their talent development efforts.

HEI leaders can utilize this model in systematically assessing the efficacy of their talent development efforts by following this research’s systematic approach in evaluating such HEI talent development programs. This research provided for an innovative (i.e., new process), successfully tested (i.e., evidence-based), deliberate (i.e., guided process), and intuitive (i.e., easy to follow) method that evaluates institutional financial sustainability strategies. It is also credible, observable, relevant, provides a relative advantage, easy to understand, compatible with HEI program assessment methods, and most importantly, testable, as evidenced by this research. HEI leaders can provide their talent development professionals training on utilizing the CERTI model, making them certified in the method. These individuals would then be available to assess the efficacy of talent development efforts such as FRD programs at the university, college, and departmental levels to assess such programs' efficacy.

Limitations and Future Research

This study utilized a three-article design and a case study methodology to introduce and demonstrate the applicability of the CERTi model for research and practice. However, all studies have limitations. The following areas pose limitations and potential for future research: The dissertations scope and its’ timeframe:

First, the dissertation scope posed a limitation in that it only examined the FRD program's first cohort. Multiple cohorts have progressed through the same program since its’ inception. Examining multiple cohorts provides more data points that could strengthen study findings, especially in the dissertations' first article. That article determined no statistically significant difference in grant acquisition between FRD program participants and those who did not participate in it. Although the article's

analysis controlled for confounding and employed varying causal estimation methods (i.e., propensity scores, propensity scores + weighting, and IPWT), which yielded actual null treatment effects (i.e., statistically insignificant results) and confidence intervals that included zero, it is still reasonable to question its' findings in light of the small sample size. The potential for inaccuracy still exists in either skewing the results towards the null hypothesis (i.e., not detecting an effect when present) or away from it (i.e., overestimating the effect when not present). Future research should examine multiple cohorts of the same talent development programs to increase the robustness of such analyses findings and increase rigor. The null findings of the first article are critical because it limits the ability to causally link the program to its benefit (i.e., grant dollars acquired), which impedes determining program return on investment as discussed in the third article.

Second, the dissertations time-lapse potentially poses a limitation for its findings and an opportunity for future research. Three years had elapsed since the inception of the program. This time potentially hampered more accurate recollections on the part of some of its' participants. Several mentioned this and said that the information they provided was to the best of their recollections, which is especially critical concerning cost data. Participants could not recollect exact time commitments to the varying aspect of the program and requested more time to provide cost-relevant information. Notwithstanding several attempts to secure these data, some participants did not oblige the requests, while others did. Moreover, despite having more time to check their records to provide more accurate information, the potential for inaccuracy increased due to the time-lapse issue.

Third, the dissertation underscored the importance of resubmitting the grant proposal to securing funding through the NIH. One of the program's designers dedicated a half session of one of its' workshops to drive home the importance of retooling and resubmitting proposals to the NIH and gave examples from his own experience where he showed participants that it takes several attempts over multiple years to acquire a grant. Other mentors, such as Anthony, also stressed the importance of resubmissions, especially in light of the new hyper-competitive environment there described in the literature, where Beech et al. (2013); Ginther et al. (2011) found that it currently takes, on average, 4-5 years for new faculty to secure federal funding as compared to one year in 1980. Future research should employ a longitudinal approach to examine these cohorts of faculty participating in FRD programs to understand the lasting effects of their development better.

References

AAAS. (2019). R&D at Colleges and Universities. American Association for the Advancement of Science. https://www.aaas.org/programs/r-d-budget-and-policy/rd-colleges-and-universities

Alberts, B., Kirschner, M. W., Tilghman, S., & Varmus, H. (2014). Rescuing US biomedical research from its systemic flaws. Proceedings of the National Academy of Sciences, 111(16), 5773-5777.

Ali, M. M., Bhattacharyya, P., & Olejniczak, A. J. (2010). The effects of scholarly productivity and institutional characteristics on the distribution of federal research grants. The Journal of Higher Education, 81(2), 164-178.

Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: Thirty years later. Personnel psychology, 42(2), 331-342.

Amara, N., Landry, R., & Halilem, N. (2015). What can university administrators do to increase the publication and citation scores of their faculty members? Scientometrics, 103(2), 489-530.

American Academy of Arts and Sciences. (2016). Public Research Universities: Serving the Public Good. https://www.amacad.org/multimedia/pdfs/publications/researchpapersmonographs/PublicResearchUniv_PublicGood.pdf

Ammons, J., & Salterio, S. (1999). The Effects of State Budget Policy on Research University Indirect Cost Recovery Rates and Management Accounting.

Anderson, D. M., & Slade, C. P. (2016). Managing institutional research advancement: Implications from a university faculty time allocation study. Research in Higher Education, 57(1), 99-121.

Arai, L., Roen, K., Roberts, H., & Popay, J. (2005). It might work in Oklahoma but will it work in Oakhampton? Context and implementation in the effectiveness literature on domestic smoke detectors. Injury Prevention, 11(3), 148-151.

Armstrong, E. G., & Barsion, S. J. (2006). Using an outcomes-logic-model approach to evaluate a faculty development program for medical educators. Academic Medicine, 81(5), 483-488.

Ashworth, C., & Goodland, M. (1990). SSADM: A practical approach. McGraw-Hill Book Company Limited.

Association of Public & Land-Grant Universities. (2016). How does public university research and community engagement benefit society at large? In A. o. P. L.-G. Universites (Ed.), (pp. 2): Association of Public & Land-Grant Universites.

Austin, P. C. (2008). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. Statistics in medicine, 27(12), 2037-2049.

Austin, P. C. (2009). Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and Monte Carlo simulations. Biometrical Journal: Journal of Mathematical Methods in Biosciences, 51(1), 171-184.

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate behavioral research, 46(3), 399-424.

Austin, P. C., Grootendorst, P., Normand, S. L. T., & Anderson, G. M. (2007). Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study. Statistics in medicine, 26(4), 754-768.

Austin, P. C., Schuster, T., & Platt, R. W. (2015). Statistical power in parallel group point exposure studies with time-to-event outcomes: an empirical comparison of the performance of randomized controlled trials and the inverse probability of treatment weighting (IPTW) approach. BMC Medical research methodology, 15(1), 1-12.

Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Statistics in medicine, 34(28), 3661-3679.

Aziz, M., & Tran, H. (in press). Enhancing Institutional Competitiveness: The CERTi Approach to Assessing Faculty Research Development Efforts In Higher Education Journal of Research Administration.

Barnett, B. G., Basom, M. R., Yerkes, D. M., & Norris, C. J. (2000). Cohorts in educational leadership programs: Benefits, difficulties, and the potential for developing school leaders. Educational Administration Quarterly, 36(2), 255-282.

Barnett, W. (1996). Lives in the balance: Benefit-cost analysis of the Perry Preschool Program through age 27. Monographs of the High/Scope Educational Research Foundation. Ypsilanti, MI: High/Scope Press.

Baum, S., Ma, J., Bell, D. W., & Elliott, D. C. (2014). Trends in College Pricing, 2014. Trends in Higher Education Series. College Board.

Beech, B. M., Calles-Escandon, J., Hairston, K. G., Langdon, M. S. E., Latham-Sadler, B. A., & Bell, R. A. (2013). Mentoring programs for underrepresented minority faculty in academic medical centers: a systematic review of the literature. Academic medicine: journal of the Association of American Medical Colleges, 88(4).

Bendal, R., & Afifi, A. (1977). Comparison of stopping rules in forward regression. J Am Stat Assoc, 72(357), 46-53.

Berk, R. A., Berg, J., Mortimer, R., Walton-Moss, B., & Yeo, T. P. (2005). Measuring the effectiveness of faculty mentoring relationships. Academic Medicine, 80(1), 66-71.

Bland, J. M., & Altman, D. G. (2009). Analysis of continuous data from small samples. Bmj, 338.

Boardman, A. E., Greenberg, D. H., Vining, A. R., & Weimer, D. L. (2017). Cost-benefit analysis: concepts and practice. Cambridge University Press.

Boucher, B. A., Chyka, P. A., Fitzgerald Jr, W. L., Hak, L. J., Miller, D. D., Parker, R. B., Phelps, S. J., Wood, G. C., & Gourley, D. R. (2006). A comprehensive approach to faculty development. American Journal of Pharmaceutical Education, 70(2), 27.

Bourke, V. J. (1962). Rationalism. Dictionary of Philosophy, 263.

Bradley, D. N., Crawford, E. P., & Dahill-Brown, S. E. (2016). Defining and assessing FoI in a large-scale randomized trial: Core components of values affirmation. Studies in Educational Evaluation, 49, 51-65.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., & Stürmer, T. (2006). Variable selection for propensity score models. American journal of epidemiology, 163(12), 1149-1156.

Brown, J. S., & Duguid, P. (1991). Organizational learning and communities-of-practice: Toward a unified view of working, learning, and innovation. Organization science, 2(1), 40-57.

Browne-Ferrigno, T., & Maughan, B. (2014). Cohort development: A guide for faculty and program developers. Pittsburgh, PA: Carnegie Project on the Education Doctorate, Duquesne University.

Brutkiewicz, R. R. (2012). Research faculty development: an historical perspective and ideas for a successful future. Advances in health sciences education, 17(2), 259-268.

Bryant, S. E. (2005). The impact of peer mentoring on organizational knowledge creation and sharing: An empirical study in a software firm. Group & Organization Management, 30(3), 319-338.

Callan, P., & Immerwahr, J. (2008). What colleges must do to keep the public's good will. Chronicle of Higher Education, 54(18), 1. http://www.csun.edu/pubrels/clips/Jan08/01-07-08R.pdf

Canizares, C. (2008). A Primer on Indirect Costs. MIT. http://web.mit.edu/fnl/volume/205/canizares.html

Cantwell, B., & Mathies, C. F. (2012). Expanding research capacity at United States universities: A study of academic research and development investment from 1990–2005. Higher Education Quarterly, 66(3), 308-330.

Caplan, J. (2013). Strategic talent development: develop and engage all your people for business success. Kogan Page Publishers.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. Implementation science, 2(1), 40.

CBCSE. (2021). CostOut. In: Center for Benefit-Cost Studies in Education.

Chambers, J. G. (1980). The development of a cost of education index: some empirical estimates and policy issues. Journal of Education Finance, 5(3), 262-281.

Chen, H.-T., & Rossi, P. H. (1983). Evaluating with sense: The theory-driven approach. Evaluation review, 7(3), 283-302.

Chen, L.-F. (2013). An improved negative selection approach for anomaly detection: with applications in medical diagnosis and quality inspection. Neural Computing and Applications, 22(5), 901-910.

Cherrington, S., Macaskill, A., Salmon, R., Boniface, S., Shep, S., & Flutey, J. (2018). Developing a pan-university professional learning community. International Journal for Academic Development, 23(4), 298-311.

Christensen, C. M., & Eyring, H. J. (2011). The innovative university: Changing the DNA of higher education from the inside out. In: John Wiley & Sons.

Chun, M. B. (2010). Building a Research Administration Infrastructure at the Department Level. Journal of Research Administration, 41(3), 77-84.

Chun, M. B. J. (2010). Building a research administration infrastructure at the department level. Journal of Research Administration, 41(3), 77.

Conn, V. S., Porter, R. T., McDaniel, R. W., Rantz, M. J., & Maas, M. L. (2005). Building research productivity in an academic setting. Nursing Outlook, 53(5), 224-231.

Cook, C. E. (1998). Lobbying for higher education: How colleges and universities influence federal policy. Vanderbilt University Press.

Cook, T. D. (2000). The false choice between theory-based evaluation and experimentation. New directions for evaluation, 2000(87), 27-34.

Cornell, R. G. (1992). Statistical Methods in Cancer Research, Volume II: The Design and Analysis of Cohort Studies. Journal of the American statistical Association, 87(418), 590-591.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. Handbook of mixed methods in social and behavioral research, 209, 240.

Creswell, J. W., & Poth, C. N. (2017). Qualitative inquiry and research design: Choosing among five approaches. Sage publications.

Cronan, M. (2012). Grant strategies in a difficult funding climate. Research Development & Grant Writing News, 2(9).

Curs, B. R., & Jaquette, O. (2017). Crowded Out? The Effect of Nonresident Enrollment on Resident Access to Public Research Universities. Educational Evaluation and Policy Analysis, 39(4), 644-669.

Curtis, L. H., Hammill, B. G., Eisenstein, E. L., Kramer, J. M., & Anstrom, K. J. (2007). Using inverse probability-weighted estimators in comparative effectiveness analyses with observational databases. Medical care, 45(10), S103-S107.

d'Agostino, R. B. (1998). Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Statistics in medicine, 17(19), 2265-2281.

Dales, L. G., & Ury, H. K. (1978). An improper use of statistical significance testing in studying covariables. International journal of epidemiology, 7(4), 373-376.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control? Clinical psychology review, 18(1), 23-45.

Delgadillo, L. M. (2016). Quantifying the impact of research for tenure and promotion in family and consumer sciences. Family and Consumer Sciences Research Journal, 44(3), 255-263.

Denzin, N. K. (1978). The research act: A theoretical orientation to sociological methods.(1978). In: New York: McGraw-Hill, SecondGoogle Scholar.

Devine, E. B. (2009). The art of obtaining grants. American Journal of Health-System Pharmacy, 66(6), 580-587.

Dranove, D. (2012). Practical regression: Building your model-What variable to include. In: Kellogg Case Publishing USA.

Dromgoole, D. A., & Cummings, S. Program Implementation Fidelity Matters!

Drummond, M. F., Sculpher, M. J., Claxton, K., Stoddart, G. L., & Torrance, G. W. (2015). Methods for the economic evaluation of health care programmes. Oxford university press.

Dumas, J. E., Lynch, A. M., Laughlin, J. E., Smith, E. P., & Prinz, R. J. (2001). Promoting intervention fidelity: Conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. American journal of preventive medicine, 20(1), 38-47.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. American journal of community psychology, 41(3-4), 327.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. Health education research, 18(2), 237-256.

Eloy, J. A., Svider, P. F., Kovalerchik, O., Baredes, S., Kalyoussef, E., & Chandrasekhar, S. S. (2013). Gender differences in successful NIH grant funding in otolaryngology. Otolaryngology--Head and Neck Surgery, 149(1), 77-83.

Emden, C. (1998). Establishing a'track record': research productivity and nursing academe. The Australian journal of advanced nursing: a quarterly publication of the Royal Australian Nursing Federation, 16(1), 29-33.

Ence, A. K., Cope, S. R., Holliday, E. B., & Somerson, J. S. (2016). Publication productivity and experience: factors associated with academic rank among orthopaedic surgery faculty in the United States. JBJS, 98(10), e41.

Evans, L. (2000). The effects of educational change on morale, job satisfaction and motivation. Journal of educational Change, 1(2), 173-192.

Evans, L. (2011a). The scholarship of researcher development: Mapping the terrain and pushing back boundaries. International Journal for Researcher Development, 2(2), 75-98.

Evans, L. (2011b). What Research Administrators Need to Know about Researcher Development: Towards a New Conceptual Model. Journal of Research Administration, 42(1), 15-37.

Evans, L. (2012). Leadership for researcher development: What research leaders need to know and understand. Educational management administration & leadership, 40(4), 423-435.

Faw, L., Hogue, A., & Liddle, H. A. (2005). Multidimensional implementation evaluation of a residential treatment program for adolescent substance abuse. American journal of evaluation, 26(1), 77-94.

Files, J. A., Blair, J. E., Mayer, A. P., & Ko, M. G. (2008). Facilitated peer mentorship: a pilot program for academic advancement of female medical faculty. Journal of women's health, 17(6), 1009-1015.

Forero, D. A., & Moore, J. H. (2016). Considerations for higher efficiency and productivity in research activities. BioData mining, 9(1), 35.

Fox, G. J., Benedetti, A., Mitnick, C. D., Pai, M., Menzies, D., & MDR-TB, C. G. f. M.-A. o. I. P. D. i. (2016). Propensity score-based approaches to confounding by indication in individual patient data meta-analysis: non-standardized treatment for multidrug resistant tuberculosis. PloS one, 11(3), e0151724.

Freel, S. A., Smith, P. C., Burns, E. N., Downer, J. B., Brown, A. J., & Dewhirst, M. W. (2017a). Multidisciplinary Mentoring Programs to Enhance Junior Faculty Research Grant Success. Academic medicine: journal of the Association of American Medical Colleges, 92(10), 1410-1415. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5589470/pdf/nihms-845093.pdf

Freel, S. A., Smith, P. C., Burns, E. N., Downer, J. B., Brown, A. J., & Dewhirst, M. W. (2017b). Multidisciplinary Mentoring Programs to Enhance Junior Faculty Research Grant Success. Academic Medicine, 92(10), 1410-1415.

Freidberg, M., & Kao, T. (2008). The State of Talent Management: Today's Challenges, Tomorrow's Opportunities'. Hewitt's Human Capital, 1-37.

Gallop, G. G., & Svare, B. B. (2016). The undesirable consequences of the growing pressure on faculty to get grants (essay) | Inside Higher Ed. Inside Higher Ed. https://www.insidehighered.com/views/2016/07/25/undesirable-consequences-growing-pressure-faculty-get-grants-essay

Gallup, G., & Svare, B. (2016, July 25, 2016). Hijacked by an External Funding Mentality. Inside Higher Education.

Garavan, T. N., Carbery, R., & Rock, A. (2012). Mapping talent development: definition, scope and architecture. European journal of training and development.

Gardiner, M., Tiggemann, M., Kearns, H., & Marshall, K. (2007). Show me the money! An empirical analysis of mentoring outcomes for women in academia. Higher Education Research & Development, 26(4), 425-442.

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. (2016). Impact evaluation in practice. The World Bank.

Ghodsian, D., Bjork, R. A., & Benjamin, A. S. (1997). Evaluating training during training: Obstacles and opportunities. Training for a rapidly changing workplace: Applications of psychological research, 63-88.

Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F., Haak, L. L., & Kington, R. (2011). Race, ethnicity, and NIH research awards. Science, 333(6045), 1015-1019.

Glowacki-Dudka, M., & Brown, M. P. (2007). Professional development through faculty learning communities. New Horizons in Adult Education and Human Resource Development, 21(1-2), 29-39.

Goodman, S. N., & Berlin, J. A. (1994). The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. Annals of internal medicine, 121(3), 200-206.

Graczyk, P. A., Domitrovich, C. E., & Zins, J. E. (2003). Facilitating the implementation of evidence-based prevention and mental health promotion efforts in schools. In Handbook of School Mental Health Advancing Practice and Research (pp. 301-318). Springer.

Greenhalgh, T., Robert, G., Bate, S., Kyriakidou, O., & Macfarlane, F. (2004). How to spread good ideas: A systematic review of the literature on diffusion, spread and sustainability of innovations in health service delivery and organisation.

Greenland, S. (1989). Modeling and variable selection in epidemiologic analysis. American journal of public health, 79(3), 340-349.

Greenlee, B. J., & Karanxha, Z. (2010). A study of group dynamics in educational leadership cohort and non-cohort groups. Journal of Research on Leadership Education, 5(11), 357-382.

Greenstone, M., & Looney, A. (2012). Regardless of the cost, college still matters. Brookings on Job Numbers.

Grol, R., Dalhuijsen, J., Thomas, S., Rutten, G., & Mokkink, H. (1998). Attributes of clinical guidelines that influence use of guidelines in general practice: observational study. Bmj, 317(7162), 858-861.

Guerci, M., Bartezzaghi, E., & Solari, L. (2010). Training evaluation in Italian corporate universities: a stakeholder-based analysis. International journal of training and development, 14(4), 291-308.

Guo, S., & Fraser, M. W. (2015). Propensity score analysis. Sage.

Harris, B. A. (2006). The importance of creating a “sense of community”. Journal of College Student Retention: Research, Theory & Practice, 8(1), 83-105.

Harris, H., & Horst, S. J. (2016). A brief guide to decisions at each step of the propensity score matching process. Practical Assessment, Research & Evaluation, 21.

Haycock, K., Lynch, M., & Engle, J. (2010). Opportunity Adrift: Our Flagship Universities Are Straying from Their Public Mission. Education Trust. https://files.eric.ed.gov/fulltext/ED507851.pdf

Hebel, S., & Blumenstyk, G. (2014). From public good to private good: How higher education got to a tipping point. The Chronicle of Higher Education, 60, 25-29.

Hedayati Mehdiabadi, A., & Li, J. (2016). Understanding talent development and implications for human resource development: An integrative literature review. Human resource development review, 15(3), 263-294.

Herman, J. H. (2012). Faculty development programs: The frequency and variety of professional development programs available to online instructors. Journal of Asynchronous Learning Networks, 16(5), 87-106.

Herzog, S., & Wright, P. (2005). Addressing the fidelity of personal and social responsibility model implementation. Res Q Exerc Sport, 76.

Hesli, V. L., Lee, J. M., & Mitchell, S. M. (2011). Predicting Rank Attainment in Political Science: What Else Besides Publications Affects Promotion?

Hintikka, J. (1973). Logic, language-games and information: Kantian themes in the philosophy of logic. Clarendon Press Oxford.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: the pervasive fallacy of power calculations for data analysis. The American Statistician, 55(1), 19-24.

Hollands, F. M., Pan, Y., Shand, R., Cheng, H., Levin, H. M., Belfield, C. R., Kieffer, M., Bowden, A. B., & Hanisch-Cerda, B. (2013). Improving early literacy: Cost-effectiveness analysis of effective reading programs. New York: Center for Benefit-Cost Studies of Education, Teachers College, Columbia University.

Holton III, E. F. (1996). The flawed four-level evaluation model. Human resource development quarterly, 7(1), 5-21.

Hoover, E., & Keller, J. (2011). More students migrate away from home. The Chronicle of Higher Education, 58(11), A1-A6.

Hourihan, M., & Parkes, D. (2016). FY 2017 R & D appropriations so far: a roundup. Available at:) Date accessed: September, 14.

Howard, D. J., & Laird, F. N. (2013). The new normal in funding university science. Issues in Science and Technology, 30(1), 71-76.

Imbens, G. W. (2000). The role of the propensity score in estimating dose-response functions. Biometrika, 87(3), 706-710.

Jacobi, M. (1991). Mentoring and undergraduate academic success: A literature review. Review of educational research, 61(4), 505-532.

Jaquette, O. (2017). State university no more: Out-of-state enrollment and the growing exclusion of high-achieving, low-income students at public flagship universities.

Jaquette, O., & Curs, B. R. (2015). Creating the out-of-state university: Do public universities increase nonresident freshman enrollment in response to declining state appropriations? Research in Higher Education, 56(6), 535-565.

Joffe, M. M., Ten Have, T. R., Feldman, H. I., & Kimmel, S. E. (2004). Model selection, confounder control, and marginal structural models: review and new applications. The American Statistician, 58(4), 272-279.

Johnson, N., Oliff, P., & Williams, E. (2011). An update on state budget cuts. Center on Budget & Po-licy Priorities. Washington, DC at, 1, 10-12.

Kaufman, R., & Keller, J. M. (1994). Levels of evaluation: beyond Kirkpatrick. Human resource development quarterly, 5(4), 371-380.

Kellogg, W. K. (2004). Using logic models to bring together planning, evaluation, and action: logic model development guide. Michigan: WK Kellogg Foundation.

Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., Zhang, J., Rathbun, A., Wilkinson-Flicker, S., Diliberti, M., Barmer, A., Bullock Mann, F., & Dunlop Velez, E. (2016). The Condition of Education 2016.

Kerwer, M., & Rosman, T. (2018). Mechanisms of Epistemic Change–Under which circumstances does diverging information support epistemic development? Frontiers in Psychology, 9, 2278.

Keyser, D. J., Lakoski, J. M., Lara-Cinisomo, S., Schultz, D. J., Williams, V. L., Zellers, D. F., & Pincus, H. A. (2008). Advancing institutional efforts to support research mentorship: A conceptual framework and self-assessment tool. Academic Medicine, 83(3), 217-225.

Kienhues, D., Ferguson, L., & Stahl, E. (2016). Diverging information and epistemic change. Handbook of epistemic cognition, 318-330.

Kim, M. M., & Ko, J. (2015). The impacts of state control policies on college tuition increase. Educational Policy, 29(5), 815-838.

King, G., & Zeng, L. (2007). Detecting model dependence in statistical inference: A response. International Studies Quarterly, 51(1), 231-241.

Kirkpatrick, D. (1959). Techniques for evaluation training programs. Journal of the American Society of Training Directors, 13, 21-26.

Kirkpatrick, D. (1994). Evaluating Training Programs: The Four Levels. 1994 San Francisco. Calif Berrett-Koehler.

Kirkpatrick, D. (1998). Evaluating Training Programs: The four levels. 2nd Editio. In: Berett-Koehler, editor. San Fransisco.

Kirkpatrick, D., & Kirkpatrick, J. (2006). Evaluating training programs: The four levels. Berrett-Koehler Publishers.

Kirkpatrick, D. L. (1960). Techniques for evaluating training programs: Part 3-Behavior. Journal of ASTD, 14(1), 13-18.

Kirkpatrick, D. L. (1998). Another look at evaluating training programs: Fifty articles from training & development and technical training: Magazines cover the essentials of evaluation and return-on-investment. American Society for Training & Development Alexandria, VA.

Knowles, M. S., Holton III, E. F., & Swanson, R. A. (2012). The adult learner. Routledge.

Kompridis, N. (2000). So we need something else for reason to mean. International journal of philosophical studies, 8(3), 271-295.

Kram, K. E. (1988). Mentoring at work: Developmental relationships in organizational life. University Press of America.

Kulage, K. M., & Larson, E. L. (2017). Intramural Pilot Funding and Internal Grant Reviews Increase Research Capacity at a School of Nursing. Nursing Outlook.

Kurth, T., Walker, A. M., Glynn, R. J., Chan, K. A., Gaziano, J. M., Berger, K., & Robins, J. M. (2005). Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. American journal of epidemiology, 163(3), 262-270.

Lackritz, J. R. (2004). Exploring burnout among university faculty: Incidence, performance, and demographic issues. Teaching and teacher education, 20(7), 713-729.

Laderman, S., & Carlson, A. (2017). SHEEO's State Higher Education Finance, FY 2017 (State Higher Education Executive Officers, Issue. http://www.sheeo.org/projects/shef-%E2%80%94-state-higher-education-finance

Lakin, J. M., & Shannon, D. M. (2015). The role of treatment acceptability, effectiveness, and understanding in treatment fidelity: Predicting implementation variation in a middle school science program. Studies in Educational Evaluation, 47, 28-37.

Lawler, P. A., & King, K. P. (2000). Refocusing faculty development: The view from an adult learning perspective.

Lawrence, R. L. (2002). A Small Circle of Friends: Cohort Groups as Learning Communities. New directions for adult and continuing education, 95, 83-92.

Ledford, H. (2014). Indirect costs: Keeping the lights on. Nature News, 515(7527), 326.

Levin, H. M. (1975). Cost-effectiveness analysis in evaluation research. Handbook of evaluation research, 2, 89-122.

Levin, H. M., & Belfield, C. (2015). Guiding the development and use of cost-effectiveness analysis in education. Journal of Research on Educational Effectiveness, 8(3), 400-418.

Levin, H. M., & McEwan, P. J. (2000). Cost-effectiveness analysis: Methods and applications (Vol. 4). Sage.

Levin, H. M., McEwan, P. J., Belfield, C., Bowden, A. B., & Shand, R. (2017). Economic evaluation in education: Cost-effectiveness and benefit-cost analysis. SAGE publications.

Levine, M., & Ensom, M. H. (2001). Post hoc power analysis: an idea whose time has passed? Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy, 21(4), 405-409.

Libby, A. M., Hosokawa, P. W., Fairclough, D. L., Prochazka, A. V., Jones, P. J., & Ginde, A. A. (2016). Grant success for early-career faculty in patient-oriented research: difference-in-differences evaluation of an interdisciplinary mentored research training program. Academic medicine: journal of the Association of American Medical Colleges, 91(12), 1666.

Lincoln, Y. S., & Guba, E. G. (1985). Establishing trustworthiness. Naturalistic inquiry, 289, 331.

Lipscomb, J., Weinstein, M., & Torrance, G. (1996). Ch. 7-Time Preference. Gold MR.

Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. Victims and offenders, 4(2), 124-147.

Little, B. (2010). Best practice talent management. Training Journal, 53(3), 45-49.

Liu, J. C., Pynnonen, M. A., St John, M., Rosenthal, E. L., Couch, M. E., & Schmalbach, C. E. (2016). Grant-writing pearls and pitfalls: maximizing funding opportunities. Otolaryngology–Head and Neck Surgery, 154(2), 226-232.

Lokken, L. (1986). The Time Value of Money Rules. Tax L. Rev., 42, 1.

Lunceford, J. K., & Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Statistics in medicine, 23(19), 2937-2960.

Lunsford, L. G., Baker, V., Griffin, K. A., & Johnson, W. B. (2013). Mentoring: A typology of costs for higher education faculty. Mentoring & Tutoring: Partnership in Learning, 21(2), 126-149.

Lynch, D. (2007). Can higher education manage talent. Retrieved June, 11, 2009.

Ma, J., Baum, S., Pender, M., & Welch, M. (2017). Trends in College Pricing. New York: College Board. College Board.

Ma, J., Matea, P., & Libassi, C. (2020). Trends in College Pricing and Student Aid 2020 (CollegeBoard, Issue. https://research.collegeboard.org/trends/college-pricing

Maher, M. A. (2001). Professional Living Situations: Cohorts as Communities of Living and Learning.

Maher, M. A. (2005). The evolving meaning and influence of cohort membership. Innovative Higher Education, 30(3), 195-211.

Malcolm, K., & Elwood, H. (1998). The Adult Learner: The Definitive Classics in Adult Education and Human Resource Development. 2-5. In: Gulf Publishing, Houston, Texas.

Mandzuk, D., Hasinoff, S., & Seifert, K. (2005). Inside a student cohort: Teacher education from a social capital perspective. Canadian Journal of Education/Revue canadienne de l'education, 168-184.

Mason, E., & Learned, L. (2006). The role of" development" in a research administration office. Journal of Research Administration, 37(1/2), 23.

Mather, D., & Hanley, B. (1999). Cohort grouping and preservice teacher education: Effects on pedagogical development. Canadian Journal of Education/Revue canadienne de l'education, 235-250.

Mathison, S. (1988). Why triangulate? Educational researcher, 17(2), 13-17.

Maynard, B. R., Peters, K. E., Vaughn, M. G., & Sarteschi, C. M. (2013). Fidelity in after-school program intervention research: A systematic review. Research on Social Work Practice, 23(6), 613-623.

McBride, N., Farringdon, F., & Midford, R. (2002). Implementing a school drug education programme: reflections on fidelity. International Journal of Health Promotion and Education, 40(2), 40-50.

McFarland, J., Hussar, B., Wang, X., Zhang, J., Wang, K., Rathbun, A., Barmer, A., Cataldi, E. F., & Mann, F. B. (2018). The Condition of Education 2018. NCES 2018-144. National Center for Education Statistics.

McGrail, M. R., Rickard, C. M., & Jones, R. (2006). Publish or perish: a systematic review of interventions to increase academic publication rates. Higher Education Research & Development, 25(1), 19-35.

Mcintyre, D., & Mcintyre, A. (1999). ESRC Teaching and Learning Programme Capacity for Research into Teaching and Learning.

McLaughlin, J. A., & Jordan, G. B. (2004). Using logic models. Handbook of practical program evaluation, 2, 7-32.

McLean, G. N., & McLean, L. (2001). If we can't define HRD in one country, how can we define it in an international context? Human Resource Development International, 4(3), 313-326.

McNamara, G., Joyce, P., & O’Hara, J. (2010). Evaluation of adult education and training programs. International Encyclopedia of Education, 3, 548-554.

Mickey, R. M., & Greenland, S. (1989). The impact of confounder selection criteria on effect estimation. American journal of epidemiology, 129(1), 125-137.

Mihalic, S. (2004). The importance of implementation fidelity. Emotional and Behavioral Disorders in Youth, 4(4), 83-105.

Mihalic, S., & Director, B. (2009). Implementation fidelity. Unpublished manuscript.

Miles, M. B., & Huberman, M. (1994). Qualitative data analysis: An expanded sourcebook. sage.

Miller, A. S. (2007). Students that Persist: Caring Relationships that Make a Difference in Higher Education. Online Submission.

Mitchell, M., Leachman, M., & Masterson, K. (2016). Funding down, tuition up. Center on Budget and Policy Priorities, 15.

Mitchell, M., Palacios, V., & Leachman, M. (2015). States are still funding higher education below pre-recession levels. Journal of Collective Bargaining in the Academy(10), 71.

Mitchell, W. C., & Dorfman, J. (1967). Types of economic theory from mercantilism to institutionalism.

Moncher, F. J., & Prinz, R. J. (1991). Treatment fidelity in outcome studies. Clinical psychology review, 11(3), 247-266.

Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. Nursing research, 40(2), 120-123.

Musambira, G., Collins, S., Brown, T., & Voss, K. (2012). From “Publish or Perish” to “Grant or Perish” Examining Grantsmanship in Communication and the Pressures on Communication Faculty to Procure External Funding for Research. Journalism & Mass Communication Educator, 67(3), 234-251.

NASBO. (2019). NASBO State Expenditure Report 2017-2019 (State Expenditure Report, Issue. https://higherlogicdownload.s3.amazonaws.com/NASBO/9d2d2db1-c943-4f1b-b750-0fca152d64c2/UploadedImages/SER%20Archive/2019_State_Expenditure_Report-S.pdf

Neumann, P. J., Sanders, G. D., Russell, L. B., Siegel, J. E., & Ganiats, T. G. (2016). Cost-effectiveness in health and medicine. Oxford University Press.

NIH. (2019). Comparing Popular Research Project Grants: R01, R03, or R21. NIH NIAID Retrieved from https://www.niaid.nih.gov/grants-contracts/research-project-grants

NIH. (2020). NIH Data Book. National Institutes of Health. https://report.nih.gov/nihdatabook/

NIH. (2021). Grants and Funding: Extramural Programs (EP). Retrieved from https://www.nlm.nih.gov/ep/FAQScores.html#:~:text=Impact%20scores%20run%20from%2010,than%2046%20are%20rarely%20funded.

Noll, R. G., & Rogerson, W. P. (1997). The economics of university indirect cost reimbursement in federal research grants. http://www-siepr.stanford.edu/workp/swp97039.pdf

NORDP. (2019). What is Research Develompment? National Organization of Research Development (NORDP). https://www.nordp.org/what-is-research-development-

Norris, C. J., & Barnett, B. (1994). Cultivating a New Leadership Paradigm: From Cohorts to Communities.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. Review of educational research, 78(1), 33-84.

Orhurhu, M. S., Orhurhu, V., Salisu, B., Abimbola, A., & Cohen, S. P. (2020). Factors associated with academic RANK among chronic pain medicine faculty in the USA. Regional Anesthesia & Pain Medicine, 45(8), 589-596.

Pagel, P. S., & Hudetz, J. A. (2015). Scholarly Productivity and National Institutes of Health Funding of Foundation for Anesthesia Education and Research Grant RecipientsInsights from a Bibliometric Analysis. Anesthesiology: The Journal of the American Society of Anesthesiologists, 123(3), 683-691.

Pan, W., & Bai, H. (2015). Propensity score analysis: Fundamentals and developments. Guilford Publications.

Partners, K. (2018). The Kirkpatrick Model. (06/05/2019). https://www.kirkpatrickpartners.com/Our-Philosophy/The-Kirkpatrick-Model

Patton, M. (2001). Qualitative research and evaluation components. In: Sage, Thousand Oaks, CA.

Patton, M. Q. (1999). Enhancing the quality and credibility of qualitative analysis. Health services research, 34(5 Pt 2), 1189.

Patton, M. Q. (2008). Utilization-focused evaluation. Sage publications.

Paul, S., Stein, F., Ottenbacher, K. J., & Liu, Y. (2002). The role of mentoring on research productivity among occupational therapy faculty. Occupational Therapy International, 9(1), 24-40.

Peck, S. L. (2008). Science Suffers when Getting a Grant Becomes the Goal. Chronicle of Higher Education, 55(7).

Pettigrew, J., Graham, J. W., Miller-Day, M., Hecht, M. L., Krieger, J. L., & Shin, Y. J. (2015). Adherence and delivery: Implementation quality and program outcomes for the seventh-grade keepin'it REAL program. Prevention Science, 16(1), 90-99.

Phan, H. P. (2008). Predicting change in epistemological beliefs, reflective thinking and learning styles: A longitudinal study. British journal of educational psychology, 78(1), 75-93.

Phillips, J. J. (1996). ROI: The search for best practices. Training & Development, 50(2), 42-48.

Pingel, S. (2018). Postsecondary tuition capping and freezing. In.

Pirracchio, R., Resche-Rigon, M., & Chevret, S. (2012). Evaluation of the propensity score methods for estimating marginal odds ratios in case of small sample size. BMC Medical research methodology, 12(1), 70.

Rabovsky, T. M., & Ellis, W. C. (2014). Higher Education and Congressional Influence on Administrative Decisions: An Examination of NSF and NIH Research Grant Funding to Four-Year Universities. Social Science Quarterly, 95(3), 740-759.

Raymond, B. C., & Kannan, V. R. (2014). A survey of faculty mentoring programs in AACSB schools of business. Journal of Management Education, 38(6), 818-842.

Reio, T. G., Rocco, T. S., Smith, D. H., & Chang, E. (2017). A critique of Kirkpatrick's evaluation model. New Horizons in Adult Education and Human Resource Development, 29(2), 35-53.

Reiser, R. A., Moore, A. L., Bradley, T. W., Walker, R., & Zhao, W. (2015). Supporting Faculty Efforts to Obtain Research Funding: Successful Practices and Lessons Learned. The Journal of Faculty Development, 29(3), 43.

Reynolds, K. C., & Hebert, F. T. (1998). Learning achievements of students in cohort groups. The Journal of Continuing Higher Education, 46(3), 34-42.

Rizzo, M., & Ehrenberg, R. G. (2004). Resident and nonresident tuition and enrollment at flagship state universities. In College choices: The economics of where to go, when to go, and how to pay for it (pp. 303-354). University of Chicago Press. http://www.nber.org/chapters/c10103.pdf

Robbins, L. C., & Robbins, B. (1935). An Essay on the Nature & Significance of Economic Science... Revised and Extended. London.

Robins, J. M., Hernan, M. A., & Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. In: LWW.

Roen, K., Arai, L., Roberts, H., & Popay, J. (2006). Extending systematic reviews to include evidence on implementation: methodological work on a review of community-based initiatives to prevent injuries. Social science & medicine, 63(4), 1060-1071.

Rosenbaum, P. R. (1987). Model-based direct adjustment. Journal of the American statistical Association, 82(398), 387-394.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. Biometrika, 70(1), 41-55.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. Journal of the American statistical Association, 79(387), 516-524.

Rosner, B. (2015). Fundamentals of biostatistics. Nelson Education.

Rossman, G. B., & Wilson, B. L. (1985). Numbers and words: Combining quantitative and qualitative methods in a single large-scale evaluation study. Evaluation review, 9(5), 627-643.

Rothwell, W. J., Lee, J. Y., & Macko, P. (2018). The Changing Role of the Corporate Trainer: The Shift from "Training" to "Talent Development". Handbook of Vocational Education and Training: Developments in the Changing World of Work,

Saldaña, J. (2003). Longitudinal qualitative research: Analyzing change through time. Rowman Altamira.

Sale, M. L., & Sale, R. S. (2010). Indirect cost rate variation determinants in university research, an empirical investigation. Research in Higher Education Journal, 6, 1. http://s3.documentcloud.org/documents/1239466/martha-sale-indirect-costs-paper-2009.pdf

Santucci, A. K., Lingler, J. H., Schmidt, K. L., Nolan, B. A., Thatcher, D., & Polk, D. E. (2008). Peer-mentored research development meeting: a model for successful peer mentoring among junior level researchers. Academic Psychiatry, 32(6), 493-497.

Sapon-Shevin, M., & Chandler-Olcott, K. (2001). Student cohorts: Communities of critique or dysfunctional families? Journal of Teacher Education, 52(5), 350-364.

Savundranayagam, M. Y. (2014). The Role of Sustained Mentorship in Grant Success. Educational Gerontology, 40(4), 272-277.

Schoen, J. W. (2015, 2015-06-16T10:22:13-0400). Why college costs are so high and rising. https://www.cnbc.com/2015/06/16/why-college-costs-are-so-high-and-rising.html

Schoenfeld, D. A. (1983). Sample-size formula for the proportional-hazards regression model. Biometrics, 499-503.

Shand, R., Muroga, A., Rodriguez, V., & Levin, H. M. (2018). REACH Cost Analysis Report.

Sigritz, B., Cummings, L., Gashaw, B., Mazer, S., Wavrunek, L., & White, K. (2018). State Expenditure Report FY 2018 (State Expenditure Report, Issue. https://www.nasbo.org/mainsite/reports-data/state-expenditure-report

Smith, A. H., & Bates, M. N. (1992). Confidence limit analyses should replace power calculations in the interpretation of epidemiologic studies. Epidemiology, 449-452.

Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? Journal of econometrics, 125(1-2), 305-353.

Smith, S. D. (2016). Factors that Motivate Faculty to Pursue External Funding at a 4-Year Public Institution of Higher Education.

Stake, R. E. (1994). Identification of the case. Handbook of qualitative research, 236-247.

Stangor, C., Jhangiani, R., & Tarry, H. (2017). Principles of Social Psychology: 1st International Edition.

Steiner, J. F., Curtis, P., Lanphear, B. P., Vu, K. O., & Main, D. S. (2004). Assessing the role of influential mentors in the research development of primary care fellows. Academic Medicine, 79(9), 865-872.

Steiner, J. F., Lanphear, B. P., Curtis, P., & Vu, K. O. (2002). Indicators of early research productivity among primary care fellows. Journal of General Internal Medicine, 17(11), 854-860.

Steiner, P. M., Cook, T. D., & Shadish, W. R. (2011). On the importance of reliable covariate measurement in selection bias adjustments using propensity scores. Journal of Educational and Behavioral Statistics, 36(2), 213-236.

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. Psychological methods, 15(3), 250.

Steinert, Y. (2010). Faculty development: from workshops to communities of practice. Medical teacher, 32(5), 425-428.

Stokking, K. M. (1996). Levels of evaluation: Kirkpatrick, Kaufman and Keller, and beyond. Human resource development quarterly, 7(2).

Stolovitch, H. D., & Keeps, E. J. (2006). Handbook of human performance technology: Principles, practices, and potential. John Wiley & Sons.

Strømsø, H. I., & Kammerer, Y. (2016). Epistemic cognition and reading for understanding in the internet age. Handbook of epistemic cognition, 230-246.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. Statistical science: a review journal of the Institute of Mathematical Statistics, 25(1), 1.

Stuart, E. A., & Rubin, D. B. (2008). Best practices in quasi-experimental designs. Best practices in quantitative methods, 155-176.

Svider, P. F., Husain, Q., Folbe, A. J., Couldwell, W. T., Liu, J. K., & Eloy, J. A. (2014). Assessing National Institutes of Health funding and scholarly impact in neurological surgery. Journal of neurosurgery, 120(1), 191-196.

Swinscow, T. (1997). Statistics at Square One. 9th (online) edition. Southampton: BMJ Publishing Group, 27.

Tamkin, P., Yarnall, J., & Kerrin, M. (2002). Kirkpatrick and Beyond: A review of models of training evaluation. Institute for Employment Studies Brighton, England.

Tan, J. A., Hall, R. J., & Boyce, C. (2003). The role of employee reactions in predicting training effectiveness. Human resource development quarterly, 14(4), 397-411.

Tandberg, D. A. (2008). The Politics of State Higher Education Funding. Higher Education in Review, 5, 1-36.

Tashakkori, A., Teddlie, C., & Teddlie, C. B. (1998). Mixed methodology: Combining qualitative and quantitative approaches (Vol. 46). Sage.

Thoemmes, F., & Ong, A. D. (2016). A primer on inverse probability of treatment weighting and marginal structural models. Emerging Adulthood, 4(1), 40-59.

Thoemmes, F. J., & Kim, E. S. (2011). A systematic review of propensity score methods in the social sciences. Multivariate behavioral research, 46(1), 90-118.

Thomas, L. (1997). Retrospective power analysis. Conservation Biology, 11(1), 276-280.

Thyer, B. A. (2011). Harmful effects of federal research grants. Social Work Research, 35(1), 3-8.

Torgerson, D. (2008). Designing randomised trials in health, education and the social sciences: an introduction. Springer.

Tran, H. (2020). Revolutionizing school HR strategies and practices to reflect talent centered education leadership. Leadership and Policy in Schools, 1-15.

Tran, H., & Smith, D. A. (2020). Designing an employee experience approach to teacher retention in hard-to-staff schools. NASSP bulletin, 104(2), 85-109.

Trochim, W. M. (2006). Qualitative validity. In.

Tsen, L. C., Borus, J. F., Nadelson, C. C., Seely, E. W., Haas, M. A., & Fuhlbrigge, A. L. (2012). The development, implementation, and assessment of an innovative faculty mentoring leadership program. Academic medicine: journal of the Association of American Medical Colleges, 87(12), 1757.

Twale, D. J., & Kochan, F. K. (2000). Assessment of an alternative cohort model for part-time students in an educational leadership program. Journal of School Leadership, 10(2), 188-208.

Van der Weijden, I., Belder, R., Van Arensbergen, P., & Van Den Besselaar, P. (2015). How do young tenured professors benefit from a mentor? Effects on management, motivation and performance. Higher Education, 69(2), 275-287.

Vannini, P. (2006). Dead Poets' Society: Teaching, publish-or-perish, and professors' experiences of authenticity. Symbolic interaction, 29(2), 235-257.

Waddell, J., Martin, J., Schwind, J. K., & Lapum, J. L. (2016). A Faculty-Based Mentorship Circle: Positioning New Faculty for Success. The Canadian Journal of Higher Education, 46(4), 60.

Waisbren, S. E., Bowles, H., Hasan, T., Zou, K. H., Emans, S. J., Goldberg, C., Gould, S., Levine, D., Lieberman, E., & Loeken, M. (2008). Gender differences in research grant applications and funding outcomes for medical school faculty. Journal of women's health, 17(2), 207-214.

Warr, P., Allan, C., & Birdi, K. (1999). Predicting three levels of training outcome. Journal of Occupational and Organizational Psychology, 72(3), 351-375.

Watkins, R., Leigh, D., Foshay, R., & Kaufman, R. (1998). Kirkpatrick plus: Evaluation and continuous improvement with a community focus. Educational Technology Research and Development, 46(4), 90-96.

Webber, D. A. (2017). State divestment and tuition at public institutions. Economics of Education Review, 60, 1-4.

Werner, J. M., & DeSimone, R. L. R. L. (2011). Human resource development. Cengage Learning.

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

Yao, C. W., & Vital, L. M. (2018). Reflexivity in international contexts: Implications for US doctoral students international research preparation.

Yavorsky, J. E., Kamp Dush, C. M., & Schoppe-Sullivan, S. J. (2015). The production of inequality: The gender division of labor across the transition to parenthood. Journal of Marriage and Family, 77(3), 662-679.

Yerkes, D. M. (1995). Using Cohorts in the Development of Educational Leaders.

Yin, R. K. (2017). Case study research and applications: Design and methods. Sage publications.

Zuiches, J. J., & Vallely, R. (1987). Accounting for the Full Cost of Research. A Study of Indirect Costs.

APPENDIX A

Transcription Key

Table A.1: Transcription Key

Symbol	Explanation
AD	Speaker Initials
(???)	Inaudible
…	Pause
Italics	Noise other than talk (i.e., laughter, cough, and unusual notice)
(Word?)	Uncertain of the word
[ ]	Overlapping Speech
(5 Second)	Five-second pause (Approximation)
↗	Rising intonation
↘	Falling intonation

APPENDIX B

Leadership Interview Protocol

<u>Preplanning Stage Questions (Focus: Organizational Context)</u>

1. First, I am interested to know your thoughts about the purpose of faculty development in general. What does 'faculty development mean' to you?
2. The NIH Bootcamp mentorship program is focused on research development (i.e., aims to increase grant acquisition skills). Why is that important to the college and its' faculty?
3. Who all were involved in the discussion leading to the creation of the program, and why?
4. What resources were allocated to support the development and implementation of the program?
5. What would you change about the preplanning process, in hindsight?

<u>Planning Stage Questions (Focus: Faculty Centric)</u>

6. Who all were involved in the planning of the program?
    a. Why were those individuals chosen?
    b. What was their contribution?
    c. Did you involve the faculty in the planning process? Why or Why not?
7. How were the needs of the faculty taken into consideration when designing the mentorship program?

8. What resources were made available to support this initiative by the college's leadership (e.g., funds, stipends, facilities)?
9. Walk me through why you chose this particular timeline and these session topics for the program? (Provide participant with Appendix A)
    a. Were the faculty included in the planning of the timeline and session topics? Why, Why not?
10. How did you promote the program to the faculty?
11. How could you have better promoted it to increase participation?
12. How do you feel about the delivery of the program in relation to your preplanning and planning for it?
    a. How did you handle and adjust for unexpected issues?
    b. How did you track progress (e.g., attendance to group sessions, mentor/mentee individual meetings)?
13. What did you do to ensure that the program design was implemented in accordance with how you designed it?
14. How did you plan to assess the following?
    a. Faculty satisfaction with the program.
    b. Faculty knowledge gains.
    c. Faculty application of knowledge.
    d. Impact of the mentorship program.
15. What would you change about the planning process in hindsight?

Did I miss anything that you would like to share with me?

APPENDIX C

Mentee Interview Protocol

Level 1 – Attitudinal Development

❖ NIH Bootcamp Mentorship Program

1. Why did you decide to participate in the program?
2. How do you feel about the program as a whole?

❖ Perceptual Change (Viewpoints, beliefs, and mindsets)

3. How do you feel about having to seek after and secure NIH grants?
4. How has that changed as a result of your participation in the NIH Bootcamp?

❖ Evaluative Change (What is important and what matters)

5. How important is it for you to secure NIH grants? Why?
6. What are your feelings about the NIH grant funding process?
7. How have those feelings changed as a result of participating in the program?
    a. What is the biggest contributor to this change?

❖ Motivational Change (Morale and job satisfaction)

8. How do you feel about your job as a faculty at the college?
    a. How has the program affected those feelings?

Level 2 – Intellectual Development

❖ Epistemological Change (Knowledge Structures)

9. How much did you know about developing an NIH grant proposal prior to the program?

a. Where did your knowledge of this come from?

b. How has that knowledge changed as a result of the program?

c. What contributed most to this change (if any)?

d. What was your mentor's contribution?

❖ Rationalistic Change (Reasoning in relation to conducting research)

10. How did you develop your grant proposal idea, the actual research?

b. To what extent did your mentor assist in formulating this idea?

❖ Analytical Change (Breaking down research into workable parts)

11. Could you please walk me through how you went about putting together your grant proposal?

12. How has this changed as a result of the mentorship you received?

❖ Comprehensive Change (Overall increase in knowledge)

13. How has participation in the NIH Bootcamp changed your overall understanding of NIH grant proposal development?

a. How so, in what areas, and if not, why?

14. To what extent did your mentor contribute to this change?

Level 3 – Behavioral Development

❖ Processual Change (Developing the grant proposal)

15. Walk me through the process of writing an NIH grant proposal prior to the mentorship program?

a. How has that process changed as a result of the program?

b. To what extent did your mentor contribute to this change?

❖ Procedural Change (Submitting the grant proposal)

16. Once you wrote a grant proposal, how did you go about actually submitting it to the NIH, and what procedures did you have to navigate?
    a. How comfortable were you with those procedures before the mentoring?
    b. How do you feel about navigating those same procedures post mentorship?

- Competential Change (Competence gains)

17. How do you feel in terms of your grantsmanship abilities after the program?
    a. What competencies do you feel you gained, if any?
    b. What was your mentor's contribution to this?

- Productive Change (Securing funding)

18. The records show that you did not submit a grant proposal. Why?
    a. What would have facilitated you doing so?
19. The records show that you submitted a grant proposal, but it wasn't funded.
    a. Why do you think it wasn't funded?
    b. What do you think would have facilitated it being funded?
20. The records show that you submitted a proposal that was funded.
    a. Why do you think it was funded?
    b. To what extent did the mentorship you received contribute to it being funded?

- Final Thoughts (Feedback)

21. What recommendations would you give the leadership as they seek to improve future iterations of this mentorship program?

22. Any final thoughts? Did I miss anything you’d like to share?

APPENDIX D

Mentor Interview Protocol

Pre-Mentorship

1. How did you hear about the NIH Bootcamp mentorship program?
2. What made you decide to become a mentor?
3. What are your thoughts about the program and its structure?
4. How did you prepare for your mentoring duties?
5. Walk me through the strategies you employed to mentor your peers?

Mentorship

6. If I was a fly on the wall during one of your mentorship sessions, what would I hear and see?
7. What are your thoughts about each of your mentee's attitudes towards grantsmanship work?
    a. How did these change over the course of your mentorship?
8. How did you go about ensuring that your mentees were able to submit an NIH grant proposal by the program deadline?
    a. Could you share some positive and negative experiences of your interactions with the mentees?
    b. What in your estimation were the factors that made the difference between those who were able to submit a grant versus those that didn't?

c. As a faculty with a demonstrable record of grant acquisition, what in your opinion was the difference between funded and unfunded grant proposals?

<u>Post-Mentorship</u>

9. How do you feel about the whole mentoring experience?
10. What did you learn from mentoring your peers?
11. What recommendations would you give the leadership as they seek to improve future iterations of this mentorship program?
12. Any final thoughts? Did I miss anything you'd like to share?

APPENDIX E

Program Advertisement

The FRD program is a mentoring program funded by the Dean's Office of the [College] that is designed to support and increase the success rate of [College] faculty applying for NIH Research Project Grants (R03, R21, R01).

The FRD program is an 9-month program comprised of the following:

- **Large group events and workshops** designed to provide mentees the tools and knowledge they need to be successful at developing a high-quality first submission proposal to the NIH
- **Peer group activities** of approximately 3 faculty mentees and 1 faculty mentor. Mentees will share ideas and review proposals and receive constructive feedback from peers and mentors
- **Small group work with Mentors** who are faculty members with established track records of external funding and a commitment to mentoring, who will meet with and advise their assigned mentee peer group. Mentors will meet with their assigned mentees, at minimum, once per month to discuss progress on drafting grant sections and provide feedback individual sections. The primary responsibility of the mentors is to facilitate the process of the mentee meeting deadlines and submitting an NIH application by the required date.

Program Participation Requirements

Participants agree that, if selected to be part of the FRD program, they are required to:

- Attend the half-day workshops, AND…
- Meet, at minimum, once per month with their assigned Mentor, AND…
- Achieve the FRD program Milestones as indicated in the timeline, AND…
- Submit their NIH proposal developed during the FRD program for the June NIH Grant Cycle immediately following graduation from the FRD program

Figure E.1: Program Advertisement

APPENDIX F

Program-As-Intended Logic Model

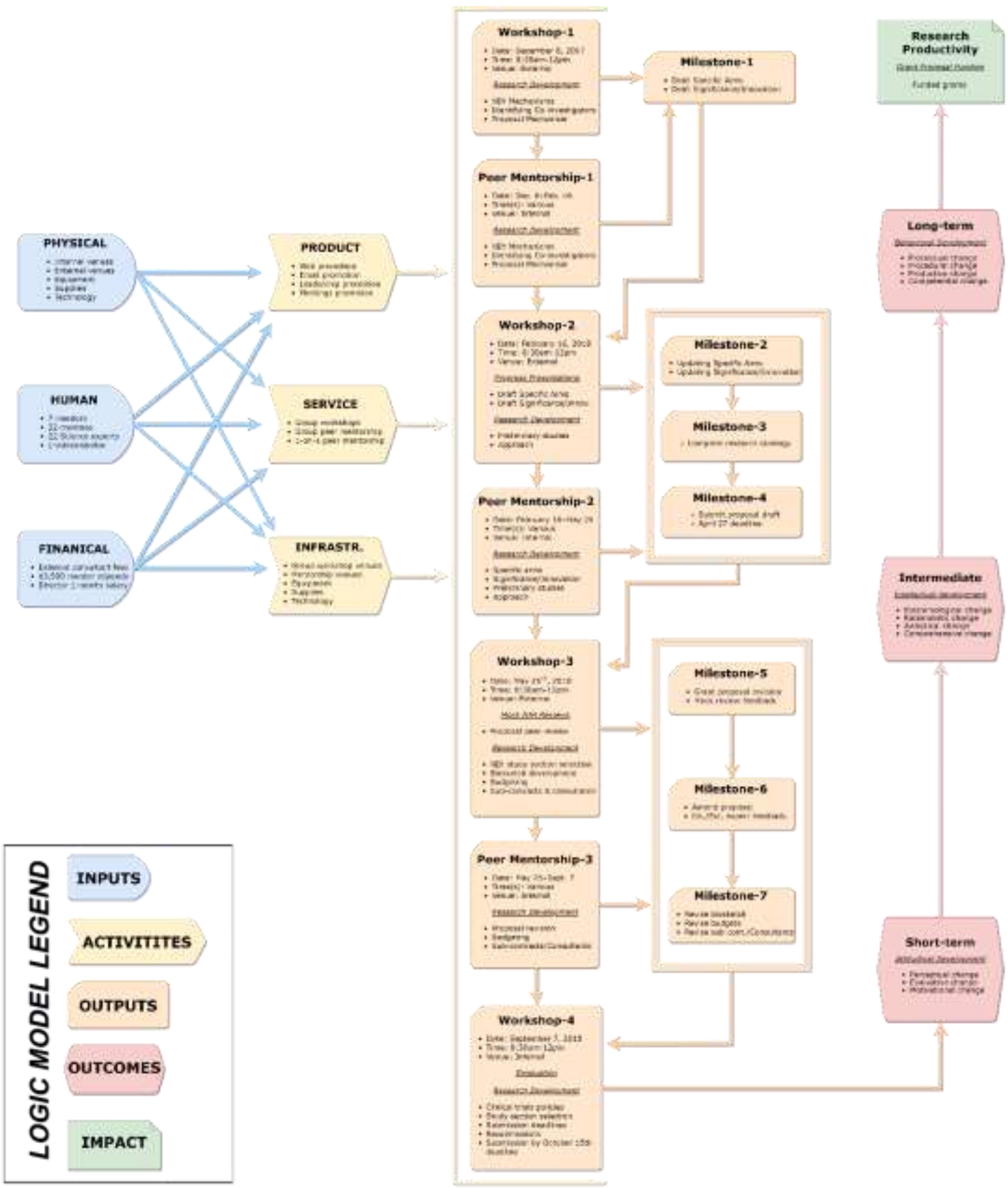

Figure F.1: Program-as-intended logic model

APPENDIX G

Program-as-Intended Logic Model

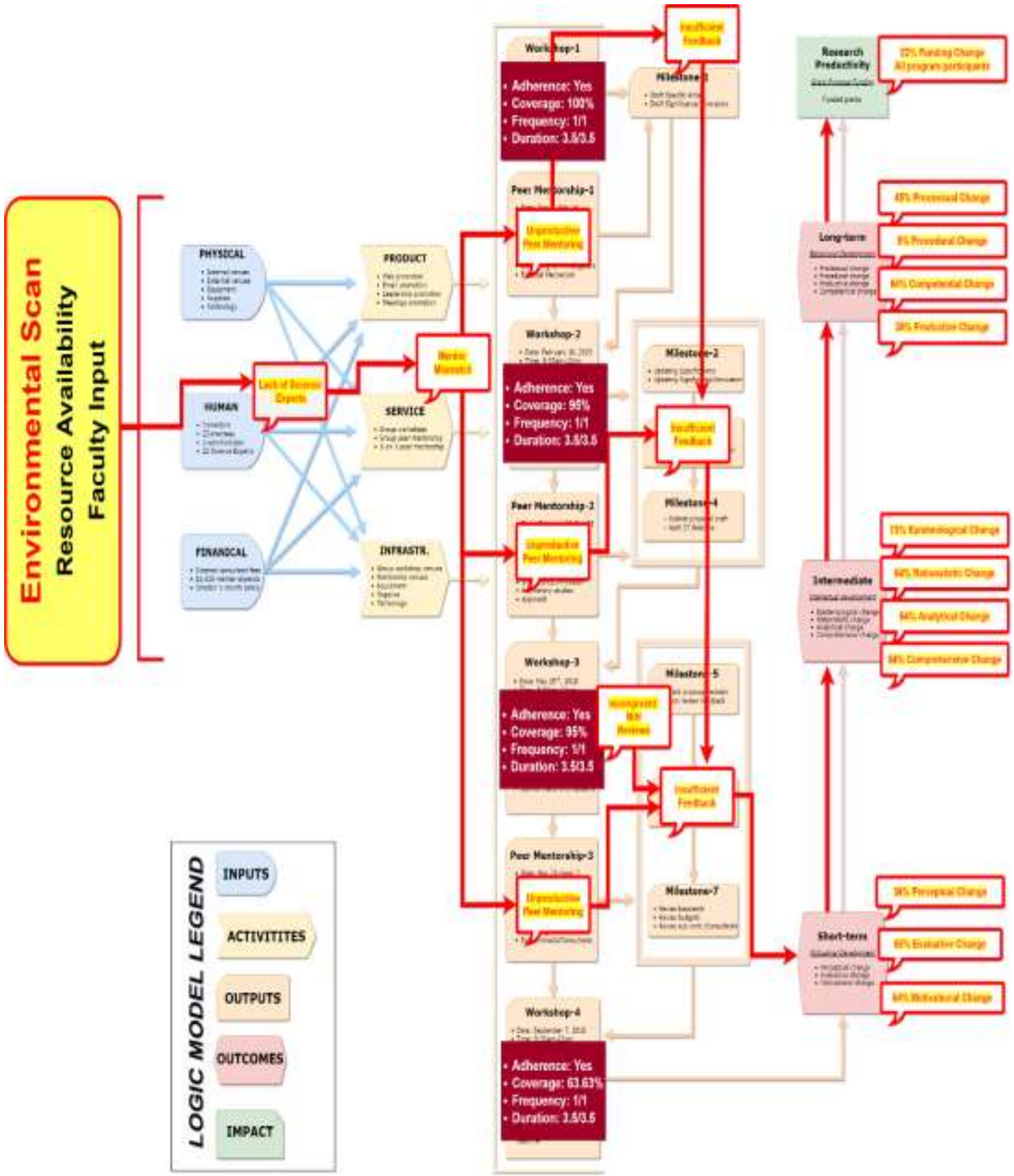

Figure G.1: Program-as-intended logic model