

# Enhancing Institutional Competitiveness: The CERTi Approach to Assessing Faculty Research Development Efforts In Higher Education

**Mazen Aziz, Ph.D.**

University of South Carolina

**Henry Tran, MPA, PHR, SHRM-CP, Ph.D.**

University of South Carolina

**Abstract:** *Faculty Research Development (FRD) in higher education institutions (HEI) is often implemented haphazardly and rarely evaluated. In this paper, we introduce a robust assessment framework (CERTi) that utilizes an overarching (Macro-level) adult-learner faculty-centric theoretical framework which incorporates using qualitative, quantitative, and economic evaluations (Micro-Level) to assess FRD efforts at HEI conjointly. The framework's cyclical approach begins by assessing FRD program effectiveness, followed by an in-depth examination of implementation practices to assess FRD program efficacy, then measures program return-on-investment (ROI), ultimately repeating the process for continuous improvement.*

**Keywords:** *Research Competitiveness, Strategic Research Development, Researcher Talent Development, Faculty Research Development, Research Financial Management, Research Development Evaluation*

## Introduction

Recent economic turmoil has forced higher education institutions (HEI) to consider reducing expenditures in faculty research development areas (FRD). Research development (RD) represents "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, para. 3). Before implementing such cuts, HEI should conduct robust assessments of their efficacy, including whether they are likely to bring in more revenue than they cost to operate. These assessments were critical in the context of governmental divestment in HEI and mounting public pressure against tuition hikes that forced HEI to rely on external sources of funding more heavily (Cronan, 2012) and are even more critical during economically uncertain times. This theoretical paper critically examines existing evaluation methodologies of FRD programs. It builds on the scholarship to propose a new comprehensive faculty-centric evaluation model known as *The Comprehensive Evaluation of Return on Talent Investment Model (CERTi)*.

The paper begins with a robust literature review to understand existing measurement and evaluation methodologies used to assess FRD programs' efficacy. It then presents a new unique approach that combines multiple evaluation frameworks from varying scientific disciplines into a comprehensive approach to evaluation that advances theory on adult professional development (PD) in a higher education setting. This holistic assessment approach relies both on macro and micro levels by utilizing an overarching (Macro-level) adult-learner faculty-centric theoretical framework that incorporates using 1) qualitative, 2) quantitative, and 3) economic evaluations (Micro-Level) to jointly assess RD efforts at HEI. Specifically, it begins with Kirkpatrick (1994) seminal Human Resource Development (HRD) framework. It then includes Evans (2011) (RD) conceptual framework that elucidates what can be learned from implementing FRD programs to improve their delivery and maximize their potential effectiveness. Finally, it utilizes principles of economic evaluations (i.e., CBA-Cost-Benefit Analysis) to measure FRD program ROI. To demonstrate the model's utility, we present a case study of an FRD program for grant acquisition to illustrate the applicability of the evaluative framework for practice and scholarship. As HEI face an era of declining public financial support, an atmosphere wrought by accountability demands, and increased requests for financial ROI, *CERTi's* approach is ever more critical to evaluating FRD programs' efficacy and advancing scholarship.

## Background

One of the most critical resources that HEI possesses is its faculty. Developing faculty represents an investment in institutional human capital. This investment, whether in the areas of teaching, research, or service, bears returns in many forms (i.e., better student outcomes, more publications, and higher rates of research grant acquisitions) (Freel et al., 2017; Haras, Taylor, Sorcinelli, & von Hoene, 2017; Morrison et al., 2014). 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. This shift from a "publish or perish" to a "grant or perish" measure of performance is manifest as the ability to obtain external funding became a core criterion for hiring and evaluating faculty (Musambira, Collins, Brown, & Voss, 2012). Competition between universities for limited federal grant funds and reduced funding for federal agencies (AAAS, 2019) created a need for FRD. As research productivity becomes a standard measure of performance for faculty, FRD has manifested itself on the scene in HEI as a field concerned with developing faculty research skills. This new facet of faculty development has taken many forms (e.g., grant writing workshops, seminars, and professional training). However, increasing in popularity among HEI is the use of cohort-based, peer-led faculty mentorship programs designed to leverage the expertise and experience of senior faculty with successful track records of grant acquisition, to mentor new and junior faculty as they seek external grant funding (Van der Weijden, Belder, Van Arensbergen, & Van Den Besselaar, 2015).

## FRD Evaluations

Recently, several studies have examined FRD program effectiveness, implementation, and return-on-investment. The larger share of that research examined FRD program effectiveness (i.e., whether faculty recipients of FRD programs are more likely to increase their chances of securing external funding) (Feldman et al., 2012; Gardiner, Tiggemann, Kearns, & Marshall, 2007; Jagsi, Griffith, Jones, Stewart, & Ubel, 2017; Longo et al., 2011; Newgard et al., 2018; Paul, Stein, Ottenbacher, & Liu, 2002; Steiner, Curtis, Lanphear, Vu, & Main, 2004). In comparison, not many researchers have scrutinized their implementation practices to ascertain what can be learned from implementing such programs to improve their delivery and maximize their potential effectiveness (Tsen et al., 2012). Lastly, few researchers delved into assessing their ROI (e.g., CBA and CEA) (Kulage & Larson, 2017; Lunsford, Baker, Griffin, & Johnson, 2013; Villar & Strong, 2007; Wingard, Garman, & Reznik, 2004).

HEI invests in FRD programs on the premise of a positive ROI. Examining current literature on the effectiveness, implementation practices, and ROI of these programs highlight limitations. First, most studies examining these programs' effectiveness lack randomization (RCT) or control measures (CM) for confounding, rendering findings suspect about the program's actual effect (Fox et al., 2016). Second, failing to account for the moderating influence of program fidelity of implementation (FOI) per program guidelines can skew results (O'Donnell, 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 CBA cannot produce necessary information to determine if the program was financially worth university investment (Levin, McEwan, Belfield, Bowden, & Shand, 2017). Although a rigorous examination of each area is crucial, the literature's most noticeable gap is that past researchers did not examine FRD programs' robustness concurrently, comprehensively, and in totality, which this research addresses. Table 1 lists a sample of previous FRD program evaluation research and highlights their non-concurrent and non-comprehensive evaluation methodology. Additionally, they were non-RCTs, lacked control measures to control for confounding, did not account for FOI's moderating influence, and failed to conduct a sound economic evaluation to measure ROI.

Table 1. Literature

Study	Effectiveness		Implementation	ROI
	RCT	CM	FOI	CBA
(Paul, Stein, Ottenbacher, & Liu, 2002)	x	x	x	x
(Steiner, Lanphear, Curtis, & Vu, 2002)	x	x	x	x
(Gardiner, Tiggemann, Kearns, & Marshall, 2007)	x	√	x	x
(Santucci et al., 2008)	x	x	x	x
(Brown et al., 2008)	x	√	x	x
(Longo et al., 2011)	x	x	x	x
(Tsen et al., 2012)	x	x	x	x
(Kulage et al., 2015)	x	x	x	√
(Libby et al., 2016)	x	√	x	x

## CERTi Development

### Macro-Level

The Comprehensive Evaluation of Return-on-Talent-Investment Model (CERTi) is grounded in an adult learner, faculty-centric theoretical framework. Introduced by Lawler and King (2000), the four-stage Adult Learning Model for Faculty Development illustrated in Figure 1 serves as the overarching macro-level theoretical framework for CERTi. The model incorporates multidisciplinary adult learner-centric approaches from various scientific disciplines (e.g., adult learning, program development) to comprehensively guide adult learning PD evaluation.

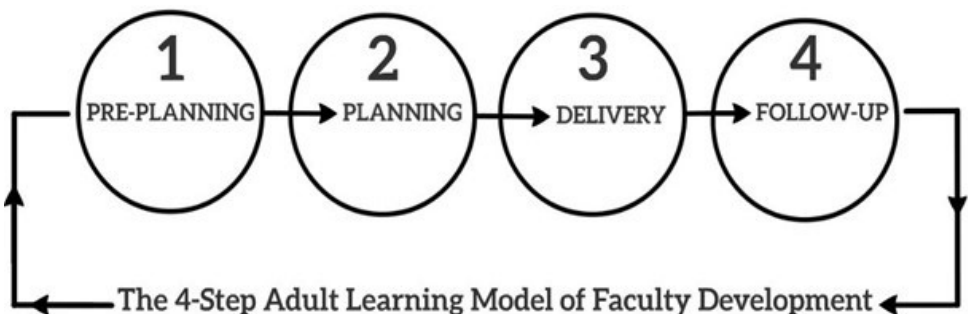


Figure 1. The Lawler & King Model

[Click here for larger image](#)

The *Pre-planning* stage is undergirded by an insight into the purpose underlying the development process, how it relates to institutional mission, and resources required to support development efforts. The *planning* stage stipulates that the development process should be faculty-centric in that faculty interests, experiences, and capabilities should underlie the development process. The *Delivery* stage is dependent on a successful assessment of adult learner needs (i.e., pre-planning) and preparation (i.e., planning). Delivery is presumed under this theory to stem from a need for development that considers adult learner inputs, institutional context, best practices, implementation processes, and progress monitoring. The final stage, *Follow-up*, assumes that development does not end with one adult learner program. Adult learning within the context of higher education faculty development is a cyclical process that should consider faculty feedback to produce a more improved development program for continuous performance improvement. Faculty empowerment underlies this stage by applying newly acquired knowledge on the job; hence, a rigorous evaluation process post-implementation to improve the development program's future iterations is encouraged.

This adult learner-centric model provides an overarching macro-level approach that guides an evaluation of adult learning efforts within a higher education context. Developers of FRD programs in HEI are encouraged to employ multiple assessment methods to comprehensively analyze faculty feelings towards development events, knowledge attainment, and learning transfer. However, the model does not provide an actionable plan for operationalizing such efforts. Embedding a multi-pronged micro-level evaluation framework within this adult learning model, as recommended by its' authors, is essential to an efficacious and holistic evaluation approach of FRD programs. To this end, *CERTi* extends this adult learning theoretical framework by advocating for a three-pronged (Micro-Level) approach consisting of 1) qualitative, 2) quantitative, and 3) economic evaluations to jointly assess FRD programs at HEI.

### *Micro-Level*

Guided by the adult-learner, faculty-centric macro-level theoretical framework, the *CERTi* Model relies on a multidisciplinary micro-level assessment approach. First, it begins with the Kirkpatrick (1994) pivotal four-stage HRD evaluation framework for FRD 1) program effectiveness (i.e., quantitative assessment). It then supplements that framework with the Evans (2011) RD conceptual model that elucidates what can be learned from 2) the implementation of FRD programs to improve their delivery and maximize their potential effectiveness (i.e., qualitative Assessment). Lastly, the framework is broadened by the principles of economic evaluations (Levin et al., 2017) to systematically account for total program cost associated with the provision of an FRD program in comparison with its total benefits to determine 3) program ROI (i.e., economic Assessment).

1. *Program Effectiveness (Quantitative Assessment)*— PD can facilitate the attainment of HEI goals through numerous mechanisms, including the improvement and preparation of the current and future job performance of the workforce. The benefits of PD are widespread, ranging from improved student outcomes (Yoon, Duncan, Lee, Scarloss, & Shapley, 2007) to faculty retention (Kena et al., 2016). Unfortunately, HEI rarely evaluates PD programs

despite their importance, and when they do, the rigor is often questionable (Astin, 2012). As FRD programs become essential to HEI financial sustainability, HRD evaluations of such programs become more critical. While there is often a debate about the disconnect of academic scholarship with field practice, HRD work can serve as a decision and explanatory science to provide actionable information to support practitioners. It does so by solving field-based problems and advancing scholarship, simultaneously addressing rigor and relevance.

Almost every mention of employee evaluation begins with Donald Kirkpatrick's seminal work. Kirkpatrick developed the most well-known and used evaluation models in the field, commonly referred to as the "four steps to evaluation." Illustrated in Figure 2, Level 1 (Reaction); assesses participants' PD favorability, engagement, and relevance to their jobs. Level 2 (Learning); evaluates the change in knowledge, skills, attitudes, confidence, and commitment based on participation in PD. Level 3 (Behavior); gauges changes in job behavior resulting from PD to identify learning transfer. Level 4 (Results); appraises targeted outcomes of PD.

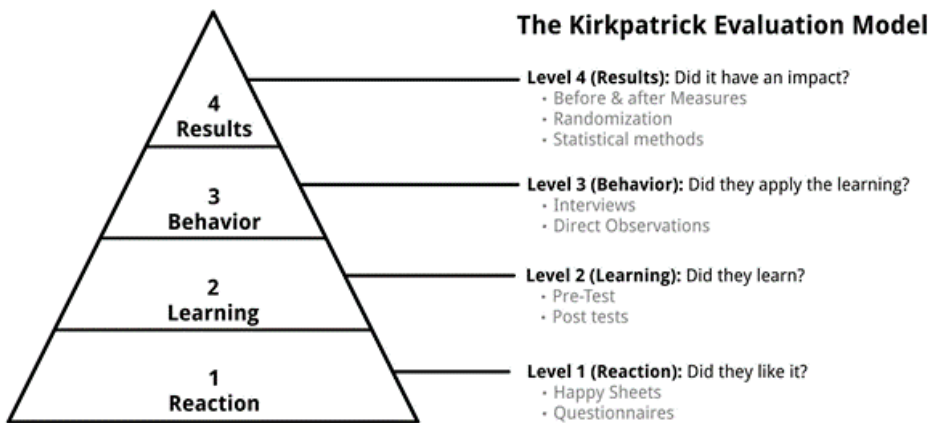


Figure 2. The Kirkpatrick Evaluation Model

[Click here for larger image](#)

Kirkpatrick's Model's simplicity and pragmatism made it the most widely used model by practitioners in the evaluation field and most cited in the literature. While Kirkpatrick's evaluation framework provides programmatic quantitative (e.g., surveys) insight regarding FRD program effectiveness (i.e., measures of participant reactions, knowledge attainment, and transfer, and organizational impact), it fails to broaden our understanding theoretically about the 'why.' For example, why did some participants have a favorable reaction to training vis-a-vis their counterparts? Why did some participants attain 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? It is essential to understand what underlies these attitudinal, intellectual, and behavioral changes to further our understanding.

2. *Program Implementation (Qualitative Assessment)*—RD is defined 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, 2011, p. 21). Evans introduced a qualitative conceptual assessment model of RD, as illustrated in Figure 3. She presents three developmental components: attitudinal, intellectual, behavioral, and their respective subcomponents or foci of change.

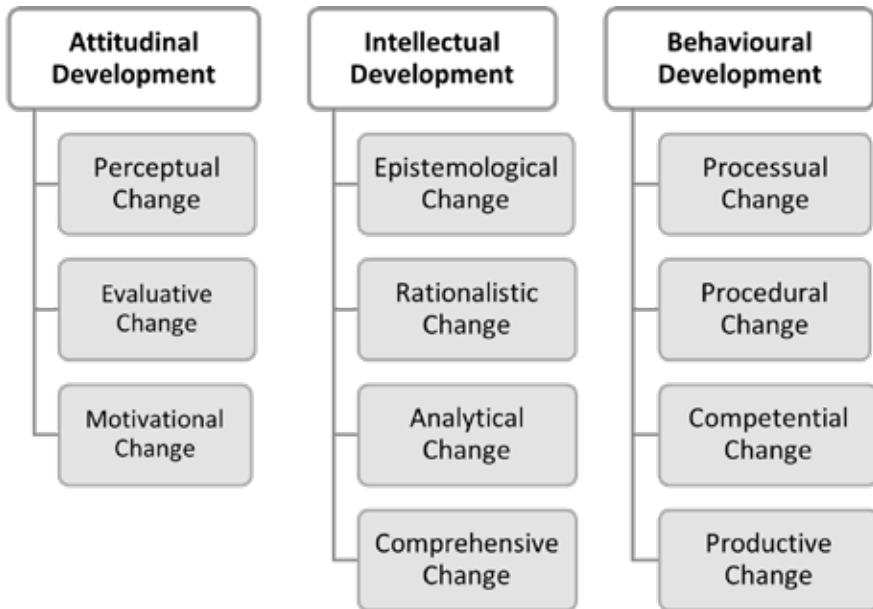


Figure 3. The Evans RD Model

[Click here for larger image](#)

*Attitudinal* development is the process by which people's attitudes are modified and has three subcomponents of change: a) Perceptual—change to perceptions, viewpoints, beliefs, and mindsets concerning research as a component of one's work; b) Evaluative—change to research related values, including the minutiae of what they consider important or what matters to them about doing research; and c) Motivational—change to the levels of job-related morale and job satisfaction relating to their research activity. *'Intellectual development'* is the process by which knowledge is modified and has four subcomponents of change; a) Epistemological—change to research related knowledge structures, b) Rationalistic—change to the extent and nature or reasoning applied to research, c) Analytical—change to analyticism (i.e., ability to break research into workable parts), and d) Comprehensive—change relating to grasping new and previously untenable research-related concepts. *'Behavioral development'* is the process by which performance is modified and has four subcomponents of change; a) Processual—change to research practice (i.e., conducting the various elements or research-related activities), b) Procedural—change in the

capacity to manage procedures with research-related practice, c) Componential—change involving enhancement of research-related skills and competencies, and d) Productive—change in research output (e.g., grant acquisitions).

Evans postulated that positive modifications in these areas would yield greater research productivity (e.g., grant acquisitions) through research capacity enhancement. These developmental components, along with their foci of change, provide a detailed accounting of why change might occur during a developmental process. They are congruent with Kirkpatrick's level-based evaluation model, commonly depicted as a triangle. Substituting Kirkpatrick's first and second levels that only provide quantitative assessments of participant reactions and learning (i.e., surveys and pre-post test) with Evan's qualitative assessments (i.e., participant interviews) yields a more in-depth analysis of participant attitudinal and intellectual development. This combined approach allows for a better understanding of why development occurs or not, instead of just reporting on what took place during the development process in the first two development levels. The third level (i.e., behavior) represents an overlap between Kirkpatrick's and Evan's frameworks. Hence this level's assessment commences both quantitatively and qualitatively. For example, utilizing a quantitative assessment of how many participants in an FRD program submitted a grant proposal to a funding agency can be examined parallel to conducting interviews with those participants to delve deeper into their behavioral development to understand better how the program changed their behavior relating to research and grant funding activities. The fourth level (i.e., results) is research productivity (i.e., grant acquisition), which is quantitative. For this stage, employing rigorous analytic techniques to control for selection bias and confounding variables to best estimate a causal link between the program and its outcome is recommended, which is especially important in the absence of randomization. Figure 4 illustrates this combined qualitative/quantitative evaluation approach by integrating Kirkpatrick's and Evan's evaluation models.

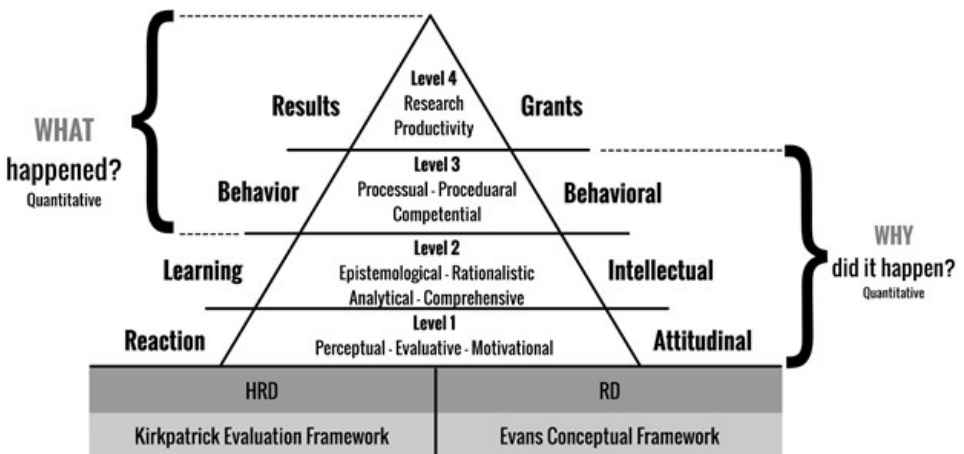


Figure 4. Kirkpatrick/Evans Combined Frameworks

[Click here for larger image](#)



3. *Program ROI (Economic Assessment)*—RD is an intensive process. It drains institutional resources, such as faculty and administrator time and effort. HEI hope RD efforts result in a positive ROI in the form of grant revenue exceeding financial input invested for its operation. The Kirkpatrick/Evans combined approach can provide a quantitative measure of FRD program effectiveness and a qualitative measure of its' implementation fidelity. Nonetheless, it neither accounts for the total cost associated with the provision of such programs nor does it compare that total cost to their total benefit (i.e., grant dollars acquired) in a formal cost-benefit analysis to ascertain program ROI. Adding a fifth step to the Kirkpatrick/Evans model, which employs a sound cost analysis (i.e., calculation of all FRD program resources) then comparing it to program outcome (i.e., total grant revenue) in a formal cost-benefit analysis, provides for a rigorous economic assessment of RD efforts at HEI. This step is significant in light of the new financial norm that HEI finds itself in, characterized by governmental divestment, mounting financial pressures, and demands for efficient public funds use.

This three-pronged micro-level evaluation approach provides education leadership with resolutions to the following questions: "What Happened?" during an FRD program, "Why did it happen?"; "How much did it cost?"; and "Was it worth it?". Levels 1-2 of the evaluation model qualitatively (e.g., participant interviews) assess participant attitudinal and intellectual development. Level 3, both quantitatively and qualitatively (e.g., institutional record, surveys, and interviews), homes in on participant attitudinal, intellectual, and behavioral development. Level 4 quantitatively assesses program results (e.g., grant acquisitions). Level 5 concludes with conducting a formal cost analysis to cost out the total cost associated with FRD program provision (e.g., salaries, fringe benefits, facilities), which can then be compared to its' outcome in a formal CBA to ascertain program ROI as illustrated by Figure 5.

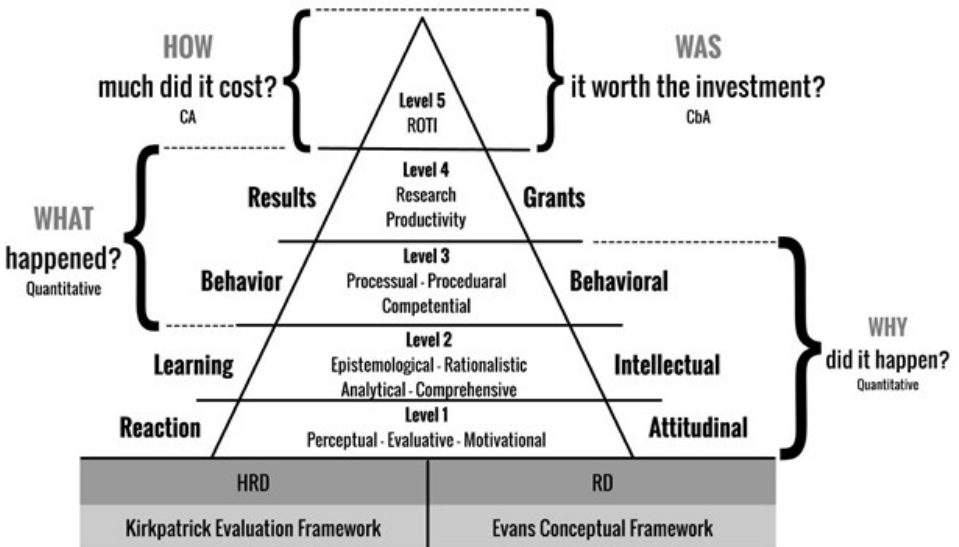


Figure 5. Micro-level Approach

[Click here for larger image](#)

### Combined Approach

*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. LM assist in understanding the FRD program theory of change. They holistically describe/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). As illustrated by Figure 6, LM are flowcharts that summarize a program's critical elements, such as 'Inputs', 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, and 'Impact', which is the ultimate change to the organization resulting from the program. (McLaughlin & Jordan, 2004).

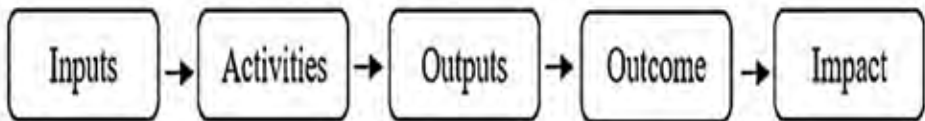
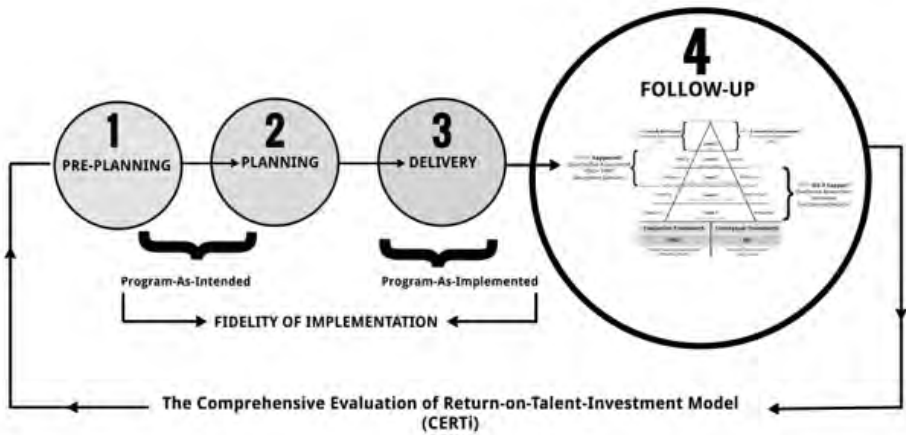


Figure 6. Logic Model

[Click here for larger image](#)

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. Understanding the FRD program theory of change is essential because it explains linkages between activities and outcomes and how and why the desired change is expected, based on past research or experiences. LM are essentially a graphic representation of change theory illustrating the linkages among resources, activities, outputs, audiences, and short-, intermediate- and long-term outcomes. Figure 7 illustrates the *CERTi* comprehensive macro-micro approach.



### CERTi - The Comprehensive Evaluation of Return-on-Talent Investment Model

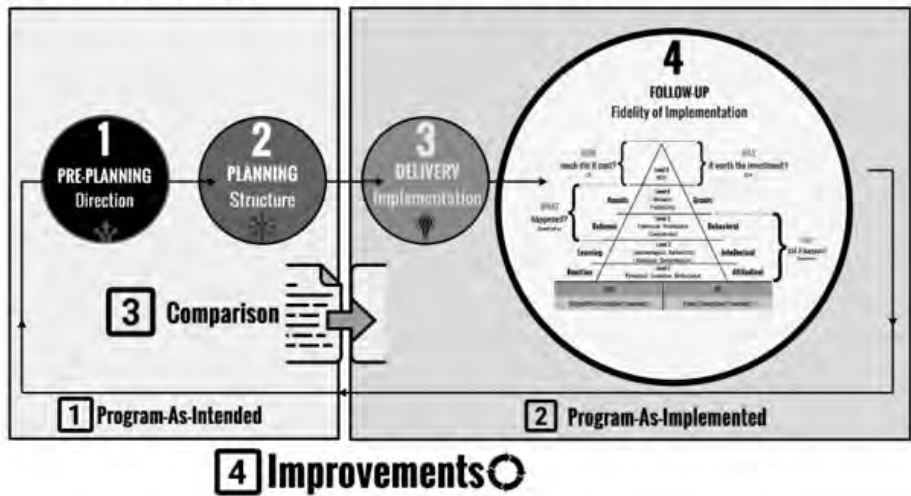


Figure 7. The CERTi Model

[Click here for larger image](#)

### CERTi Model Application

To demonstrate *CERTi's* applicability, we present a hypothetical case study of an FRD program. Facing governmental divestment in HEI and a recent decline in grant acquisitions, HEI leadership at the College of Public Health at an R1 research-intensive university implemented an FRD program to increase its faculty's grant acquisition skills. The College faces a leveling off of federal

grant funding due to a leveling of grant submissions and a decline in funded grant proposals. The program relies on senior level (i.e., Professor rank) faculty with a demonstrable record of grant acquisition to mentor a cohort of their junior level (i.e., Assistant rank) counterparts. The program was one year in length and coincided with federal agency proposal submission deadlines to culminate in grant proposal submissions to that agency.

### *Macro-Level*

A *CERTi* evaluator begins by creating a program-as-intended LM representing the program's macro-level (i.e., pre-planning and planning) stages by documenting inputs, activities, outputs, outcomes, and desired impact as envisioned by its designers. Program artifacts (e.g., a timeline of events) and interviews with its designers (i.e., leadership team) represent these data points. Next, key program indicator measures are developed, facilitating comparison between program-as-intended and program-as-implemented LM to ascertain program FOI. The evaluator documents the program's inputs (e.g., faculty time and effort, facilities, supply costs), activities (e.g., group sessions and mentor/mentee meetings), outputs (e.g., grant proposals submissions), outcomes (e.g., mentee attitudinal, intellectual, behavioral) data, and impact (e.g., ROI data) utilizing the three-pronged qualitative, quantitative, and economic micro-level approach that take place during the delivery and follow-up stages.

### *Micro-Level*

*Quantitative Assessment*—Researchers often aim to determine the effects of 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-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 of receiving treatment, conditional on baseline covariates, which opens the door for oversampling in either direction. Hence, we strongly encourage using causal estimation methodologies to estimate the causal effect of FRD programs in randomization's absence.

Randomized control trials (RCTs) are admittedly expensive to administer, consume researchers' valuable time and are often impractical to implement, explaining observational study prevalence in the educational field. However, researchers are increasingly employing statistical methods to mimic RCTs to increase their studies' rigor in the absence of randomization (Austin & Stuart, 2015). One such method increasingly used for addressing confounding and moving towards more causal estimates is using propensity scores to balance observable baseline covariates between treatment and control groups. A propensity score is the probability of treatment assignment conditional on measured baseline covariates, which allows for reducing or eliminating the confounding effects when using observational data (Rosenbaum & Rubin, 1983).

Pan and Bai (2015) outlined four steps to estimating the causal effect of programs:

1. Estimate propensity score
2. Match
3. Evaluate match quality
4. Evaluate outcomes

The first step entails estimating the likelihood of an individual data unit experiencing treatment given a set of characteristics (i.e., covariates). The second step involves matching scores of treated individual units within the data set to non-treated ones 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. The third step involves evaluating match quality (i.e., the balance of covariates). The fourth and final step entails evaluating outcomes and estimating causal effects.

The statistical literature describes four methods of using propensity scores to address selection bias: stratification, adjustment, matching, and, more recently, inverse probability of treatment weighting (IPTW). Among all these methods, both matching and IPTW have demonstrated the greatest efficiency in reducing imbalance in baseline covariates (Pirracchio, Resche-Rigon, & Chevret, 2012). Austin and Stuart (2015) observed that the "use of IPTW has increased rapidly in recent years" (p. 3664) because this method creates non-confounded pseudo-populations. In such cases, there is oversampling of treated or control groups based on specific covariates. Countering this oversampling by weighting facilitates achieving balance. Figure 8 illustrates such a situation; there is oversampling in the treated group compared to the control group. Nine out of ten subjects, in this example, are treated, which creates an imbalance. This oversampling must 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, which creates balance.

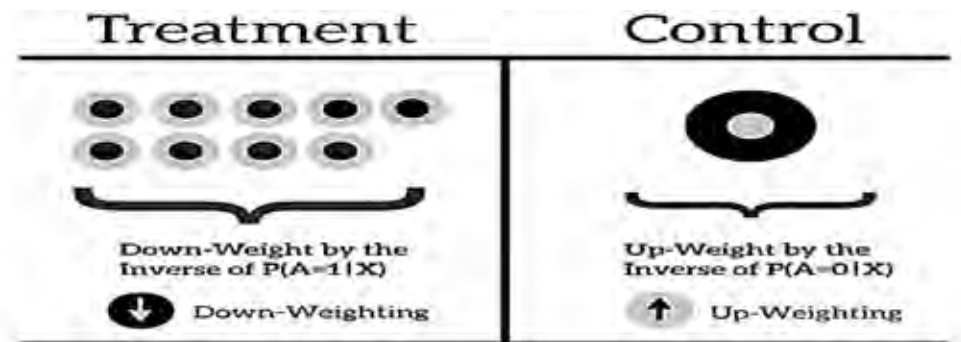


Figure 8. IPTW

[Click here for larger image](#)

Achieving this balance results in a balanced pseudo-population based on observable baseline covariates, ensuring that, on average, treated subjects do not differ systematically from their control counterparts based on those characteristics, allowing for direct comparison between the groups to estimate treatment effect as illustrated by Figure 9. Each treated subject counts as nine-tenths of a subject (i.e., down-weighted) while the control subject 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 present in the original one. In the original group, subjects had a higher probability of receiving treatment based on shared baseline covariates, while in the new one, that probability is equal. Although this does not rise to the rigor of randomization, it essentially mimics the desired characteristics of randomization present in RCTs.

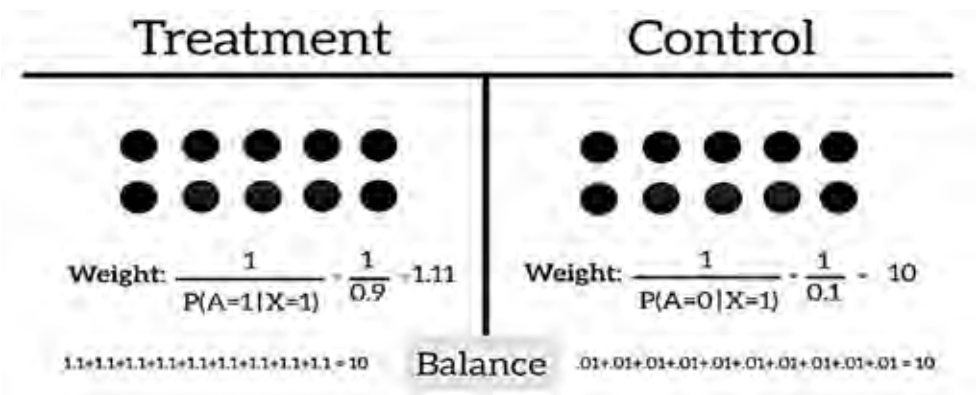


Figure 9. Balanced Pseudo Populations

[Click here for larger image](#)

In summary, for researchers aiming to use IPTW to link an FRD program to its' outcome, good practice includes 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. A *CERTi* evaluator can utilize the process shown in Figure 10 to create a pseudo-control group from non-participating faculty within the College based on shared baseline covariates (e.g., Race, Gender, Rank). Compared to FRD program participants, the outcomes of the pseudo-control group isolate treatment effect. They can accomplish this by identifying an appropriate data set, defining treatment, control, and outcome, selecting relevant covariates, estimating propensity scores to match groups, assessing match quality, and estimating program causal effect.

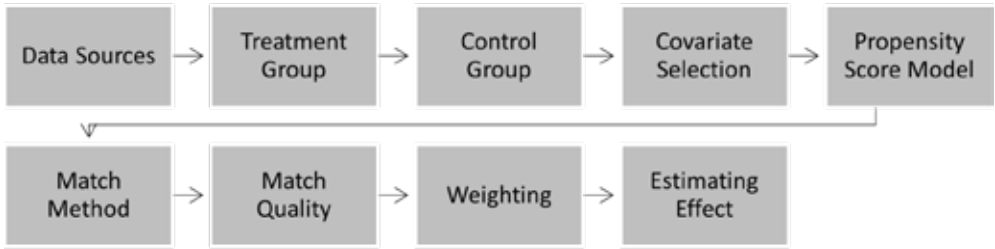


Figure 10. Estimating causal effects process

[Click here for larger image](#)

*Qualitative Assessment*—Supplementing the quantitative assessment by conducting interviews with all stakeholders (e.g., leaders, mentors, and mentees) facilitates mapping out the "missing middle" between the program's activities and its' potential impact. Qualitative data provides an in-depth appraisal of "Why" things happened, which is essential to explaining program effectiveness and understanding the theory of change undergirding the program. Reviewing program records and artifacts (e.g., program timeline, the outline of events, session handouts, communications), along with university institutional records, facilitate developing the program-as-intended LM and identifying key program indicators. Semi-structured interviews with program developers (i.e., Leadership team) provides data on program pre-planning and planning activities.

Program records and artifacts (e.g., attendance records, communications, and presentations), university institutional records, and semi-structured interviews with participants (i.e., mentees and mentors) provide data facilitating the development of the program-as-implemented LM. Interviews with mentees and their mentors provide feedback on their experiences and allow for triangulation and verification of their interactions, providing a more holistic examination. These data elucidate the minutia of the mentoring process. Mentor perceptions regarding their interactions with mentees and between-mentee comparisons add rich context to mentees' attitudinal, intellectual, and behavioral data, providing a comprehensive picture of the development process.

*Economic Assessment*— Economic evaluations combine economics, a field concerned with allocating scarce resources, with evaluations. This data-informed field helps decision-makers choose among alternative policies or decision-making programs (Levin et al., 2017). Enhancing FRD program evaluations' robustness by supplementing the combined quantitative and qualitative approaches with a sound economic evaluation based on opportunity cost is essential to achieving a comprehensive evaluation methodology. Opportunity cost is "the value of what is sacrificed by using a specific resource in one way rather than in its best alternative use" (Levin & Belfield, 2015, p. 403). The assumption among decision-makers and evaluators is that cost information is readily available from budgets and business personnel. However, 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 neglect to account for opportunity cost.

In contrast, the ingredients method of cost estimation is based on the economic principle of opportunity cost and provides more accurate cost estimations. It assumes that all the ingredients (e.g., personnel, training, facilities, equipment, materials, other outputs) associated with programs have cost implications. Operating under this assumption, the method documents all resources utilized in the program, regardless of whether each resource has a budgetary cost or not, to fully capture its actual cost. Next, it involves matching each ingredient with its respective costs. Monetizing ingredients' most common method is market prices because competition produces an equilibrium price representing the good's value. 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 sometimes local average pricing is advantageous, especially when addressing local constituents such as policymakers. The critical consideration in choosing between national and local average pricing is transparency in detailing how ingredients were valued. Shadow pricing, "societal willingness to pay for a specific impact" (Levin & McEwan, 2000, pp. 60-61), is utilized in the absence of market prices. Various methods can calculate shadow pricing. 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). The ingredients method concludes with calculating total program costs, which provides evaluators with a proper accounting of the cost of each program to conduct their economic evaluation of choice; Cost-Effectiveness Analysis (CEA), Cost-Feasibility Analysis (CFA), Cost-Utility Analysis (CUA), Cost-benefit Analysis (CBA).

CBA's are analytical tools that compare 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 program ROI, which makes it the most appropriate for calculating an FRD program's ROI. CBA evaluates all potential costs, including opportunity costs. This method produces the necessary information to gauge whether the program examined is worth university investment. It compares the program's benefit (i.e., total grant dollars) to its' total cost of provision to determine its ROI. 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).

Net Present Value (NPV) represents the discounted value of the benefit minus the costs' discounted value. Discounting is a process of determining money's present value since money is worth more today than tomorrow according to the time value of money (TVM) principle (Lokken, 1986). One method is the consumer saving options (i.e., returns sacrificed by consumers in order to consume resources now instead of saving them), and another is the average ROI 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 many methods for choosing a discount rate. "The disagreement in the literature suggests evaluators should choose an initial discount rate of 3%



to 5% as a baseline discount rate and then test for uncertainty by conducting sensitivity analyses that vary discount rates between 0% and 10% to check the robustness of the findings." Levin et al. (2017, p. 99) 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.

$$B_{PV} = \sum_{t=1}^n \frac{B_i}{(1+i)^{t-1}} \qquad C_{PV} = \sum_{t=1}^n \frac{C_i}{(1+i)^{t-1}}$$

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

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

According to Levin et al. (2017), "The NPV metric has the advantage of being the most straightforward to report and interpret" (p. 222). Programs with higher NPVs are always preferred, while programs with an NPV amount of less than zero are assumed inefficient and rejected.

Although NPV is a simple and straightforward method for ascertaining program ROI, it does come with a trade-off. The method's simplicity makes it difficult to compare 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 way of overcoming this shortcoming, as illustrated by this equation.

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

A BCR above 1 represents benefits exceeding costs, and contrarywise, a BCR lower than 1 represents costs exceeding benefits, allowing for a better ROI comparison between programs. In this case study example, the FRD program aimed to attain NIH large-scale R-level grants. A *CERTi* evaluator would then utilize program benefit data (i.e., total grant dollars) from the quantitative assessment to compare the cumulative costs resulting from applying the ingredients method to assess the program's ROI via either the NPV or BCR metrics. They can also conduct a sensitivity analysis for cost estimates to test for their robustness.

### *Combined Approach*

The data and ensuing analyses from this comprehensive (i.e., Macro-Micro) approach provide an estimate of program effectiveness, a realistic depiction of what transpired during the program's implementation, and a measure of ROI, allowing comparison between program intent and implementation in actuality to uncover incongruities. Findings resulting from LM comparison

may lead to one of these 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 Stolovitch and Keeps (2006). Any LM comparison data resulting in a negative outcome requires utilizing the macro/macro data to redesign the original program for continuous improvement. The comparison data would undergird the development of a new and improved program addressing the first's shortcomings. Figure 11 illustrates *CERTi*'s cyclical approach with LM outcomes.

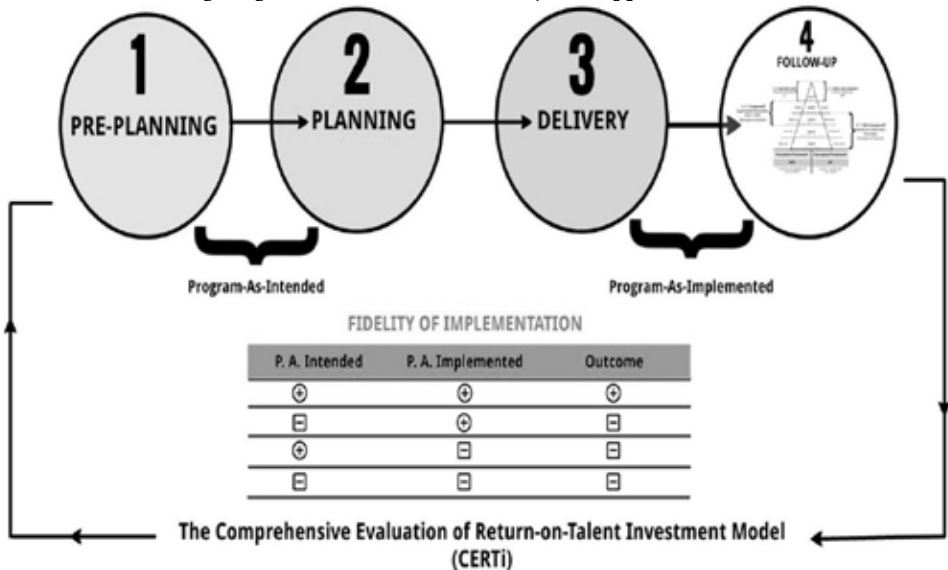


Figure 11. Cyclical approach

[Click here for larger image](#)

## Conclusion and Implications

As the world grapples with the financial implications of the COVID-19 global pandemic, HEI, who are already under fiscal strain, are sure to reduce funding for faculty PD 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. *CERTi* provides an innovative, comprehensive, and interdependent approach that combines quantitative, qualitative, and economic methodologies to advance adult PD 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.

## Authors

### **Mazen Aziz, Ph.D.**

Managing Director, Master of Human Resources Program  
Darla Moore School of Business  
Department of Management  
University of South Carolina  
Columbia, South Carolina, USA  
ORCID ID: 0000-0002-9415-235X  
[azizm@mailbox.sc.edu](mailto:azizm@mailbox.sc.edu)

### **Henry Tran, MPA, PHR, SHRM-CP, Ph.D.**

Assistant Professor, Department of Educational Leadership and Policies  
Wardlaw College of Education  
University of South Carolina  
Columbia, South Carolina, USA  
ORCID ID: 0000-0002-2229-7111  
[TRANHE@mailbox.sc.edu](mailto:TRANHE@mailbox.sc.edu)

## References

- AAAS. (2019). R&D at Colleges and Universities. Retrieved from <https://www.aaas.org/programs/r-d-budget-and-policy/rd-colleges-and-universities>
- Astin, A. W. (2012). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education*. Rowman & Littlefield Publishers.
- 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.
- Cronan, M. (2012). Grant strategies in a difficult funding climate. *Research Development & Grant Writing News*, *2*(9).
- Evans, L. (2011). What Research Administrators Need to Know about Researcher Development: Towards a New Conceptual Model. *Journal of Research Administration*, *42*(1), 15-37.
- Feldman, M. D., Steinauer, J. E., Khalili, M., Huang, L., Kahn, J. S., Lee, K. A., . . . Brown, J. S. (2012). A mentor development program for clinical translational science faculty leads to sustained, improved confidence in mentoring skills. *Clinical and translational science*, *5*(4), 362-367.

- 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., & Dewhurst, M. W. (2017). Multidisciplinary Mentoring Programs to Enhance Junior Faculty Research Grant Success. *Academic medicine: journal of the Association of American Medical Colleges*, *92*(10), 1410-1415. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5589470/pdf/nihms-845093.pdf>
- 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.
- Haras, C., Taylor, S. C., Sorcinelli, M. D., & von Hoene, L. (2017). INSTITUTIONAL COMMITMENT TO TEACHING EXCELLENCE: Assessing the Impacts.
- Jagsi, R., Griffith, K. A., Jones, R. D., Stewart, A., & Ubel, P. A. (2017). Factors associated with success of clinician-researchers receiving career development awards from the National Institutes of Health: a longitudinal cohort study. *Academic medicine: journal of the Association of American Medical Colleges*, *92*(10), 1429.
- Kena, G., Hussar, W., McFarland, J., de Brey, C., Musu-Gillette, L., Wang, X., . . . Dunlop Velez, E. (2016). *The Condition of Education 2016*. Retrieved from The Condition of Education:
- Kirkpatrick, D. (1994). *Evaluating Training Programs: The Four Levels*. 1994 San Francisco. Calif Berrett-Koehler.
- Kulage, K. M., & Larson, E. L. (2017). Intramural Pilot Funding and Internal Grant Reviews Increase Research Capacity at a School of Nursing. *Nursing Outlook*.
- Lawler, P. A., & King, K. P. (2000). Refocusing faculty development: The view from an adult learning perspective.
- 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.
- Lokken, L. (1986). The Time Value of Money Rules. *Tax L. Rev.*, 42, 1.
- Longo, D. R., Katerndahl, D. A., Turban, D. B., Griswold, K., Ge, B., Hewett, J. E., . . . Schubert, S. (2011). The research mentoring relationship in family medicine: findings from the grant generating project. *Family Medicine-Kansas City*, 43(4), 240.
- 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.
- McLaughlin, J. A., & Jordan, G. B. (2004). Using logic models. *Handbook of practical program evaluation*, 2, 7-32.
- Morrison, L. J., Lorens, E., Bandiera, G., Liles, W. C., Lee, L., Hyland, R., . . . Heathcote, E. J. (2014). Impact of a formal mentoring program on academic promotion of Department of Medicine faculty: a comparative study. *Medical teacher*, 36(7), 608-614.
- 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.
- Newgard, C. D., Morris, C. D., Smith, L., Cook, J. N., Yealy, D. M., Collins, S., . . . Kimmel, S. (2018). The first national institutes of health institutional training program in emergency care research: productivity and outcomes. *Annals of emergency medicine*, 72(6), 679-690.
- NORDP. (2019). What is Research Development? Retrieved from <https://www.nordp.org/what-is-research-development->
- 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.
- Pan, W., & Bai, H. (2015). *Propensity score analysis: Fundamentals and developments*: Guilford 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.

- 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.
- 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.
- 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.
- Stolovitch, H. D., & Keeps, E. J. (2006). *Handbook of human performance technology: Principles, practices, and potential*: John Wiley & Sons.
- Tran, H. (2020). Revolutionizing school HR strategies and practices to reflect talent centered education leadership. *Leadership and Policy in Schools*, 1-15.
- 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.
- 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.
- Villar, A., & Strong, M. (2007). Is mentoring worth the money? A benefit-cost analysis and fiveyear rate of return of a comprehensive mentoring program for beginning teachers. *ERS Spectrum*, 25(3), 1-17.
- Wingard, D. L., Garman, K. A., & Reznik, V. (2004). Facilitating faculty success: outcomes and cost benefit of the UCSD National Center of Leadership in Academic Medicine. *Academic Medicine*, 79(10), S9-S11.
- Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement. Issues & Answers. REL 2007-No. 033. *Regional Educational Laboratory Southwest (NJ1)*.