

RESEARCH ARTICLE

Developing evidence-based resources for evaluating postgraduate trainees in the biomedical sciences

Jacqueline E. McLaughlin^{1*}, Rebekah L. Layton², Paul B. Watkins³, Robert A. Nicholas^{2,4}, Kim L. R. Brouwer³

1 Division of Practice Advancement and Clinical Education, Center for Innovative Pharmacy Education and Research, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **2** Office of Graduate Education, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **3** Division of Pharmacotherapy and Experimental Therapeutics, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, **4** Department of Pharmacology, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America

* Jacqui_mclaughlin@unc.edu



OPEN ACCESS

Citation: McLaughlin JE, Layton RL, Watkins PB, Nicholas RA, Brouwer KLR (2022) Developing evidence-based resources for evaluating postgraduate trainees in the biomedical sciences. PLoS ONE 17(12): e0278297. <https://doi.org/10.1371/journal.pone.0278297>

Editor: Sina Safayi, Graduate Career Consulting LLC, UNITED STATES

Received: March 30, 2022

Accepted: November 14, 2022

Published: December 13, 2022

Copyright: © 2022 McLaughlin et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data collection instruments and survey data can be accessed in Open ICPSR: <https://www.openicpsr.org/openicpsr/project/174721/version/V1/view>.

Funding: JM, RL, PW, RN, and KLRB received funding for this research from the same National Institute of General Medical Sciences of the National Institutes of Health under Award Number T32GM086330-10S1. The content is solely the responsibility of the authors and does not necessarily represent the official views of the

Abstract

Postgraduate trainees elevate the academic strength of institutions by conducting research, promoting innovation, securing grant funding, training undergraduate students, and building alliances. Rigorous and systematic program evaluation can help ensure that postgraduate training programs are achieving the program's intended outcomes. The purpose of this project was to develop evidence-based evaluation tools that could be shared across federally funded biomedical training programs to enhance program evaluation capacity. This manuscript describes the evidence-based process used to determine program evaluation needs of these programs at a research-intensive university. Using a multi-phased sequential exploratory mixed methods approach, data were collected from trainees, employers, leaders, and program directors. Data analyses included document analysis of program plans, inductive coding of focus groups and interviews, and descriptive analysis of surveys. Two overarching categories—Trainee Skills and Program Characteristics—were identified including six themes each. Program directors prioritized communication, social and behavioral skills, and collaboration as the trainee skills that they needed the most help evaluating. Furthermore, program directors prioritized the following program characteristics as those that they needed the most help evaluating: training environment, trainee outcomes, and opportunities offered. Surveys, interview scripts, and related resources for the categories and themes were developed and curated on a publicly available website for program directors to use in their program evaluations.

Introduction

Postgraduate trainees play a critical role in the success of academic institutions [1, 2]. Under the guidance of mentors who help develop their intellectual independence, postgraduate trainees

National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. <https://www.nigms.nih.gov/>.

Competing interests: The authors have declared that no competing interests exist.

elevate the academic strength of institutions by conducting research, promoting innovation, training and mentoring undergraduate students, securing grant funding, and building alliances [1–3]. In the United States, biomedical sciences doctoral students typically complete their coursework and research in 4–6 years, whereas postdoctoral trainees (colloquially known as “post-docs”) typically complete 1–5 years of activities that promote scholarly autonomy, disciplinary specialization, and entrepreneurial skills [3–6].

Recent research highlights the complexity of postgraduate career and professional development and raises concerns about the extent to which training programs align with the needs of trainees, employers, and society [1, 2, 7–9]. Career outcomes can vary widely, with biomedical trainees transitioning into various academic and non-academic careers [10, 11]. Furthermore, recent research from the NIH Broadening Experience in Scientific Training (BEST) consortium suggests that participating in professional development, such as career panels, skill-building workshops, job search workshops, site visits, and internships, neither reduces efficiency (i.e., no delayed time to degree) nor productivity (i.e., no fewer total or first author publications) [12]. Despite initial attempts to gain stakeholder perspectives [13], more research is needed to better develop and align postgraduate training with trainee goals, employer needs, entrepreneurship, and economic workforce predictions.

Nationally, this reality has been recognized across funding agencies in numerous ways. Federally funded postgraduate training programs (e.g., T32 and R25 programs) are intended to further develop and enhance scientific training and help trainees become more competitive for the job market. While goals typically vary from program to program, well-designed curricula should increase trainee competitiveness for peer-reviewed research funding, strengthen trainee publication records, and foster institutional environments conducive to success in the biomedical and health sciences [14]. Ensuring that these programs achieve their intended outcomes is crucial for promoting the success of postgraduate trainees and their universities.

As a requirement for funding, National Institute of General Medicine Science (NIGMS)-funded training programs must conduct ongoing evaluations to monitor the success of the training and mentoring activities, with an expectation that findings are disseminated to the broader community [15]. In 2020, NIGMS published a Notice of Special Interest regarding availability of funds for Administrative Supplements to Enhance Program Evaluation Capacity at institutions with NIGMS training programs [16]. If programs are not identifying training goals when creating their programming plans, it is unlikely they will be able to meet the needs of the defined outcomes, and without measuring defined outcomes, it becomes impossible to quantify the effectiveness of the proposed programming. Therefore, the emphasis of evidence-based program evaluation can support the creation of more effective programs—as well as the dissemination of program aspects that are successful and could be replicated elsewhere.

The University of North Carolina (UNC) has a rich history of NIGMS training programs across multiple schools and departments, currently including two T32 programs for postdoctoral trainees, ten T32 programs for predoctoral trainees, and one R25 for predoctoral trainees. These programs are bolstered by a strong culture of collaboration and coordination, such as the Biological and Biomedical Sciences Program (BBSP), which serves as a single admission portal and first-year program for 15 UNC graduate programs and enables students to join predoctoral T32 programs once they become eligible [17]. The breadth of T32/R25 training programs and related institutional infrastructure makes UNC ideal for systematically evaluating these programs. However, uniform processes were not in place for evaluating trainees, mentors, and program effectiveness at the institution, and suitable evaluation tools were not readily available when UNC received the NIGMS Administrative Supplement. The purpose of this project was to develop evidence-based evaluation tools that could be shared across these T32/R25 training programs to enhance program evaluation capacity. This manuscript

describes the evaluation process and embedded mixed methods study utilized to understand the evaluation needs of UNC T32/R25 programs and develop related evaluation support for program directors.

Materials and methods

As seen in Fig 1, the project team followed the evaluation framework described by the Centers for Disease Control and Prevention (CDC) Evaluation Working Group [18]. While the framework steps are interdependent and typically implemented linearly, some steps (e.g., engage stakeholders) were revisited repeatedly. Portions of our process that involved systematic data collection and analysis followed a multi-phased sequential exploratory mixed methods design (Fig 2).

At the start of this project, (1) **an evaluation team was convened**, comprised of several key stakeholders: three T32 program directors [KLRB, RAN, PBW], two education specialists (1 from the UNC Eshelman School of Pharmacy [JEM] and one from the UNC School of Medicine Office of Graduate Education [RLL]), and one T32 program administrator. Team members were chosen based on the expertise needed to achieve the goals of the project. The evaluation team met approximately every two weeks to discuss progress and critical decisions. In addition, an evaluation consultant was engaged periodically throughout the process. These individuals were instrumental for providing feedback about goals, needs, process, and impact.

To better understand and (2) **describe the programs** included in this project, the evaluation team used individual T32 program plans to develop a logic model for each program.

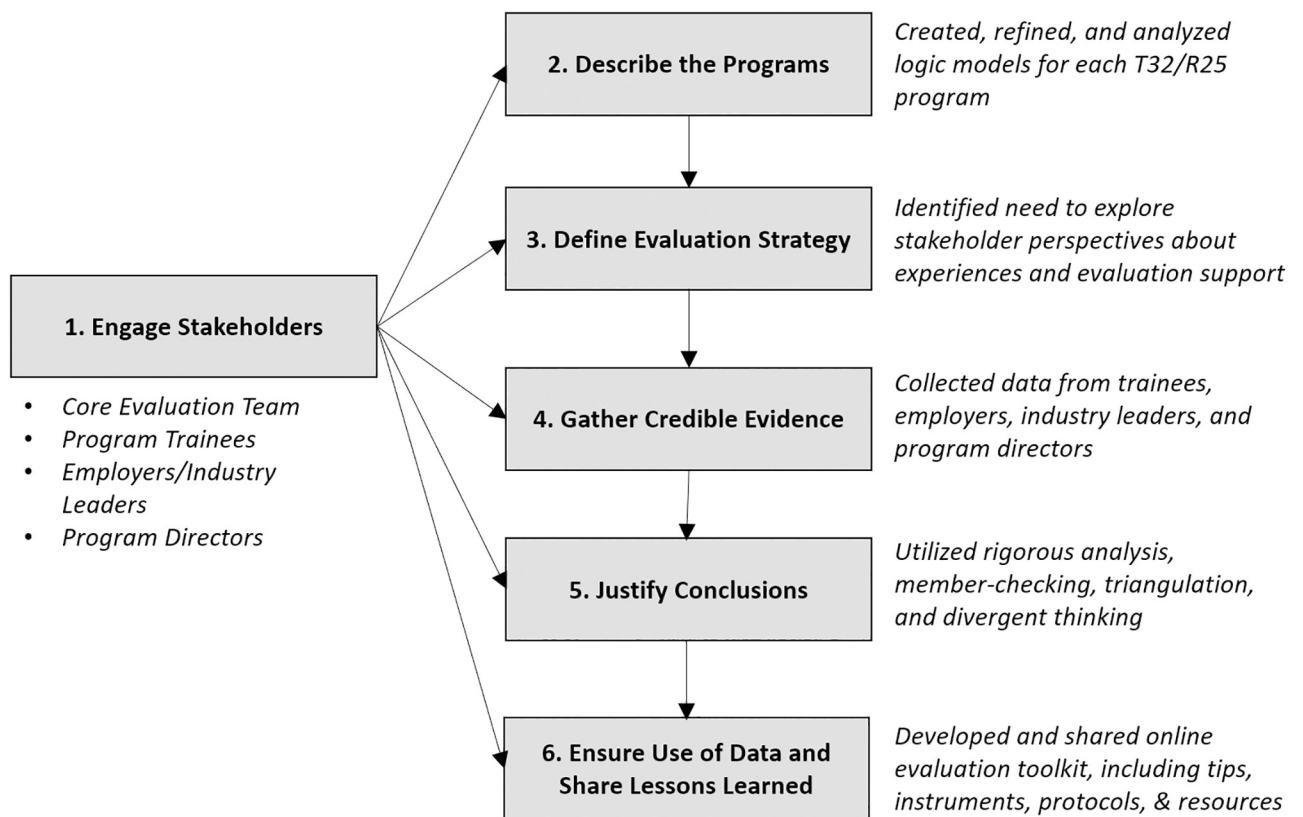


Fig 1. Evaluation process for determining evaluation needs for T32/R25 training programs.

<https://doi.org/10.1371/journal.pone.0278297.g001>

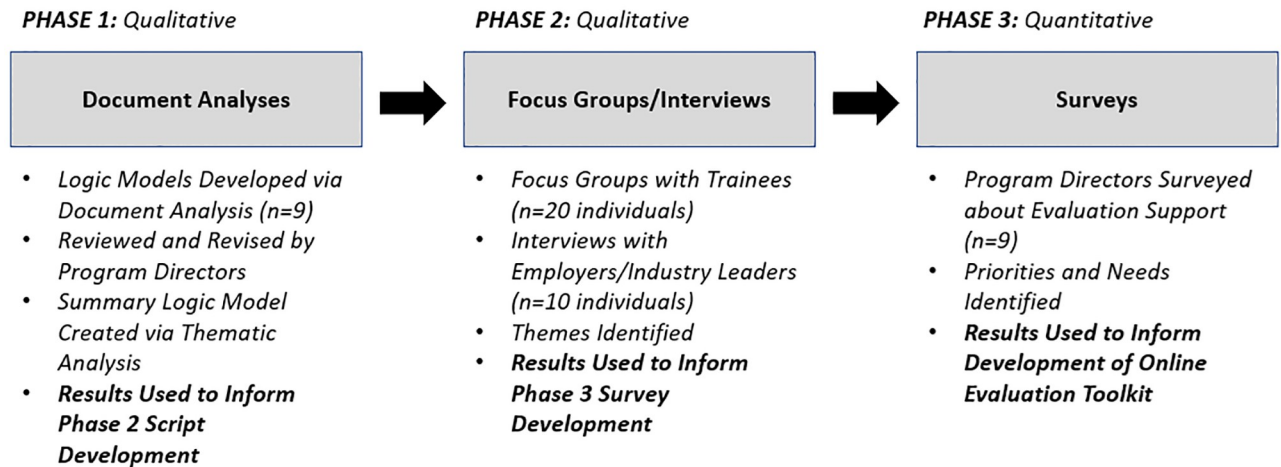


Fig 2. Multiphase sequential mixed methods design for determining evaluation needs of T32/R25 training programs.

<https://doi.org/10.1371/journal.pone.0278297.g002>

Logic models represent shared relationships among program resources, activities, outputs, and outcomes [19, 20]. Graphically, these models detail the process by which we believe the program meets its long-term outcomes (i.e., program theory): *if we have [inputs], we can do [activities], which will result in [outputs] that impact [outcomes]*. Nine logic models were developed by a single researcher (JEM) utilizing the T32 program plan provided by each program director. The draft logic model was provided to the program director for review and revised based on feedback. Once all nine logic models were complete, an aggregate logic model was created, reviewed, and revised by the evaluation team to reflect common elements across all programs.

With a better understanding of the training programs, the evaluation team was able to identify issues that were relevant to training program plans, articulate questions and items that remained unclear, and (3) **define an evaluation strategy** focused on relevant users, uses, and needs. Of note, the evaluation team wanted more insight into the evaluation needs of training programs from various stakeholder perspectives. As such, an evaluation strategy was developed that included data collection from trainees, program directors, employers, and industry leaders.

During the next phase, (4) **credible evaluation data were gathered** via trainee focus groups to better understand their experiences in these programs—an important issue that was not sufficiently uncovered in the document analysis. Purposive sampling was utilized to identify current training program students and postdocs, and participants were recruited via email. Four 1-hour focus groups were conducted in Fall 2020 with 20 graduate students and postdoctoral trainees from multiple training programs. Example script questions included: *What aspect of your program do you feel all doctoral/postdoctoral trainees in your program gain value from participating in?* and *What challenges have you experienced as part of your training?*

Next, interviews were conducted with select employers of training program alumni and industry leaders, as these stakeholders are well-positioned to provide insight into needs and gaps in the field. Purposive sampling was utilized, and participants included Dean-level academic leaders, former federal program officers, leadership of professional organizations, and employers from Clinical Research Organizations (CRO), pharmaceutical industry, biotechnology companies, and regulatory agencies. Participants were recruited via email and 1-hour interviews, which were conducted in December 2021 and January 2022 (n = 10). Example script questions included: *What are appropriate program metrics that should be included in*

a program evaluation, from your perspective? and *What changes do you see in your industry in the next 5–10 years that we should be considering?* All focus group and interview sessions were conducted via Zoom with at least one interviewer and one note-taker.

Sessions were recorded using field notes that were inductively coded independently for themes by two researchers. This qualitative analysis approach enabled the researchers to identify semantic themes, provide a description of what existed in the data, and allowed for the data to be organized into interpretative patterns (i.e., categories, themes). Saturation of themes was achieved, providing a rich description of the entire data set, and themes were pervasive across participants. Achieving data saturation indicated that the sample size was sufficient for the intended purpose (i.e., additional interviews were not necessary) and that findings were relevant regardless of participant role (e.g., Dean, CRO). Categories and related themes were refined, and discrepancies were resolved via consensus-building discussions [21, 22].

Using the resulting categories and themes from the focus groups and interviews, program directors were surveyed about their evaluation priorities and needs. The survey was administered to the 11 NIGMS Program Directors with one reminder email during Spring 2022. The survey introduction described the purpose of the study and how the themes and categories were developed. For each category, participants were asked to *rate how important you think each of the [themes] are for evaluating your T32/R25 training program* (1-Not At All Important to 5-Extremely Important), and to *rank order the [themes] that you would most like help with evaluating in your T32/R25 training program* (1-Least Help to 6-Most Help). An open text field invited additional items that participants wanted help evaluating. Data were analyzed using descriptive statistics (mean \pm standard deviation (SD)).

Taken together, the logic models, focus groups, interviews, and surveys provided rich data and relevant insight into the evaluation needs of the training programs. This multiphase process enabled us to **(5) justify conclusions** through member-checking, triangulation, divergent thinking, and disclosure of limitations. For example, in April 2021 the members of the evaluation team met with the program directors to review and discuss findings from the survey with the hopes of further understanding their evaluation needs (i.e., member-checking). This discussion and feedback, in turn, enabled the evaluation team to identify and develop a new online evaluation toolkit comprised of educational resources and tips about evaluation, protocols and instruments for data collection, key findings from this work, related summary information about core training skills, and contact information for evaluation experts. In December 2021, the core evaluation team met with program directors via zoom to share and discuss the online evaluation toolkit. This positioned us to **(6) ensure use of evaluation data and share lessons learned**.

Ethical considerations, consent

This project was submitted to the UNC Institutional Review Board (#18–3140) and was determined to be non-human subjects research according to 45 CFR 46.104. Due to this classification, formal consent was not required; however, study communications included information about participation via email invitations and verbal reminders, such as prior to beginning each focus group or interview, with reminders that participation was voluntary, and that all data would be de-identified prior to sharing.

Results

Fig 3 depicts an abbreviated version of the aggregate logic model (*NOTE: the full logic model can be found online at <https://tarheels.live/t32programevaluation/logic-models/>*). Of note, the logic model highlights common elements across all included T32/R25 programs, as well as an

PROBLEM STATEMENT	PROCESS			OUTCOMES
	Inputs: <i>what we invest</i>	Activities: <i>what we do</i>	Outputs: <i>what we get</i>	
Workforce gaps must be filled with exceptionally skilled and diverse scientists and clinician-scientists who are specifically prepared to advance health and medicine	Personnel	Recruiting, Selecting, and Matching	Trained Postgraduates (e.g., doctoral degree, certificate)	Short/Long Term: <i>what we achieve</i>
	Institutional Support	Mentoring/Advising		
	Facilities	Training (e.g., didactic, experiential, career development)	Research Products (e.g., abstracts, presentations, grants, publications)	Improved research quality
	Recruitment & Selection	Training Assessment	Trained Faculty Mentors & Advisors	Increased diversity
	Curriculum	Faculty Selection & Training		Increased capacity
	Time	Program Management		Enhanced collaboration
		Program Retreats		Improved health treatments/cures
		Program Review		Empowered leaders
		Evaluation		
Assumptions: <i>beliefs we have about the program, how we think the program will work</i> <ul style="list-style-type: none"> UNC and research environments (e.g., collaborations, excellence/quality, facilities) will attract candidates and help enable success of program Faculty/mentors have expertise, motivation, and ability to participate as designed The curriculum will be adopted, used in the way intended, and effective BBSP leads to efficiency and synergism in admissions and training, while greatly expanding the diversity of research topics available to students 		Contextual Factors: <i>conditions that influence implementation and outcomes</i> <ul style="list-style-type: none"> Economic factors can influence program design and success UNC programs play a critical role in recruiting, supporting, tracking trainees and in training program faculty UNC provides core research technologies and related support Faculty often collaborate/share resources) across programs, depts, and centers 		

Fig 3. Abbreviated logic model depicting summary inputs, activities, outputs, and outcomes of UNC T32/R25 training programs.

<https://doi.org/10.1371/journal.pone.0278297.g003>

overarching goal of addressing workforce gaps with *exceptionally skilled and diverse scientists and clinician-scientists who are specifically prepared to advance health and medicine* (Fig 3).

Common inputs included personnel, institutional support and infrastructure, and time. Common activities addressed needs for trainees (e.g., recruitment, selection, training, assessment), faculty (e.g., selection, training), and the program (e.g., management, retreats, evaluation). Common outputs consisted of trained personnel (e.g., trainees, faculty mentors) and research products (e.g., publications, grants) leading to various outcomes, such as improved research quality, increased diversity, increased capacity, enhanced collaboration, and ultimately, advances in healthcare and the next generation of leaders. From the logic models, focus group and interview questions were drafted to learn more about the training and assessment activities.

In the focus groups and interviews, two overarching categories were identified related to needs in postdoctoral training: trainee skills and program characteristics (Table 1). Trainee skills included themes of communication (e.g., written, oral, scientific, interpersonal), social and behavioral skills (e.g., adaptability, resilience), collaboration, career awareness/development (e.g., business acumen, job market readiness), discipline-specific skills (e.g., technical skills), and productivity. Program characteristics included themes of training environment (e.g., work-life balance, quality of mentoring), trainee outcomes (e.g., job placement), opportunities offered (e.g., funding for professional development, mentor selection/matching), trainee productivity (e.g., number and types of presentations, publications, grants), program transparency (e.g., clarity regarding selection of trainees and mentors, mentor selection/matching), and assessment culture (e.g., interviews/focus groups, surveys).

In addition, focus group and interview participants provided various suggestions and ideas regarding assessment strategies that could help elucidate key aspects of their program. This compiled list included: measurement instruments (e.g., surveys, focus groups, interviews/exit interviews); evaluation focus (e.g., applicant characteristics, trainee evaluations, mentor evaluations, post-program career progression, program efficacy); and program metrics (e.g., number of participants, graduates, research products, job placement, trainee perceptions, and employer perceptions).

Table 1. Results from focus groups, interviews, and program director surveys, including categories, themes, and the support needed for evaluation.

Category	Theme	Examples	Example Quote	Importance to PD* (n = 9) Mean ± SD	Support Needed by PD** (n = 9) Mean ± SD
Trainee Skills	Communication Skills	Written, oral, conflict resolution, client-facing, receiving feedback	<i>Effective communication skills really make a difference in the long-term career trajectory. These skills can make or break a career. (Employer/Leader)</i>	4.43 ± 0.49	4.23 ± 0.75
	Social & Behavioral Skills	Adaptability, initiative, negotiation, resilience, well-being	<i>Employers are asking for intangible personal soft skills needed to be successful in companies, like resilience, communication, persistence. (Leader)</i>	3.71 ± 0.88	2.83 ± 1.77
	Collaboration Skills	Interpersonal, teamwork, community building	<i>Trainees need to be experts in their own disciplines, but [also be] boundary crossers and able to work across disciplines. (Employer/Leader)</i>	4.29 ± 0.45	2.67 ± 1.25
	Career Awareness Skills	Business acumen, career exploration, career preparation, market readiness	<i>We need exposure outside of academia earlier in the program. (Trainee)</i>	4.14 ± 0.64	2.50 ± 1.12
	Discipline-Specific Skills	Analytical thinking, knowledge expertise, problem solving, technical skills	<i>As more aspire to careers outside of academia, competencies become more important. . .are trainees gaining competencies that they need for THEIR career choice? (Employer/Leader)</i>	4.43 ± 0.73	1.77 ± 1.11
	Productivity Skills	Completing projects, meeting deadlines, preparing presentations	<i>Practice is the most important thing when it comes to science. (Trainee)</i>	4.43 ± 0.73	1.00 ± 1.83
	Program Characteristics	Training Environment	Work-life balance, quality of mentoring, peer community institutional support, satisfaction	<i>Mentors don't always understand that they can change the course of a student's path. (Trainee)</i>	4.22 ± 0.63
Trainee Outcomes		Job placement, number of job offers, met professional goals, fellowships, alumni involvement	<i>Are they satisfied with the careers once they go into them? Related to that is underemployment. . .are they utilizing their skills and whether they feel they want more related to career satisfaction? (Employer/Leader)</i>	4.44 ± 0.68	2.89 ± 1.03
Opportunities Offered		Funding for training/travel, mentor selection and matching, learning contexts (e.g., lab, networking)	<i>Being more well-funded means less stress we have about things that aren't [related to our research]. (Trainee)</i>	4.22 ± 0.63	2.56 ± 1.17
Trainee Productivity		Number/types of presentations, publications, awards	<i>Each program should have core metrics related to the goals of the training grants, for example pubs & presentations & conferences, workshops & local presentations. (Trainee)</i>	4.22 ± 0.79	1.89 ± 1.85
Program Transparency		Clarity about how trainees and mentors are selected and matched, program successes	<i>I didn't understand the importance of funding and how it works. . .how much time and money would be provided for travel? Which PIs get money and which students receive funding? (Trainee)</i>	3.75 ± 0.66	1.78 ± 1.03
Assessment Culture		Focus groups, surveys, stakeholder engagement (e.g., T32 Director, Alumni)	<i>Focusing more on getting the perceptions of students/trainees, alumni, and employers. . .find key gaps or what you wish you would've learned in our program. (Employer/Leader)</i>	3.67 ± 0.66	1.78 ± 1.62

PD = program director; SD = standard deviation

* Program directors were asked to rate how important you think each of the [training skills or program characteristics] are for evaluating your T32/R25 training program, from (1-Not At All Important to 5-Extremely Important)

**Program directors were asked to rank order the [training skills or program characteristics] that you would most like help with evaluating in your T32/R25 training program, from (1-Least Help to 6-Most Help)

<https://doi.org/10.1371/journal.pone.0278297.t001>

When asked about the importance of the trainee skills identified in the focus groups and interviews (1-Not At All Important to 5-Extremely Important), program directors (n = 9, 81.81% response rate) rated communication as most important (4.43 ± 0.49), followed by discipline-specific skills (4.43 ± 0.73), and productivity skills (4.43 ± 0.73) (Table 1). Program directors ranked (1- Least Help to 6-Most Help) communication as the item they needed the most support evaluating (4.23 ± 0.75), followed by social and behavioral skills (2.83 ± 1.77), and collaboration (2.67 ± 1.25). Communication was clearly prioritized by program directors in terms of importance and support needed.

When asked about the importance of the program characteristics identified in the focus groups and interviews, program directors rated trainee outcomes (4.44 ± 0.68), training environment (4.22 ± 0.63), opportunities offered (4.22 ± 0.63), and trainee productivity (4.22 ± 0.79) as most important (Table 1). Program directors ranked training environment (4.11 ± 1.29), trainee outcomes (2.89 ± 1.03) and opportunities offered (2.56 ± 1.17) as items they needed the most support evaluating. Training environment, trainee outcomes, and opportunities offered were clearly prioritized by program directors in terms of importance and support needed.

Survey results and subsequent discussion with the surveyed program directors were used to identify the need for data collection instruments aimed at supporting evaluation of communication, social and behavioral skills, and collaboration along with training environment, trainee outcomes, and opportunities offered. In some cases, program directors also articulated a need for scholarly resources related to skills considered challenging to address or develop (e.g., adaptability). As such, the core evaluation team developed a *core survey* for T32/R25 program evaluation that included multiple survey items related to each theme, along with 11 short (i.e., 5–10 items) *modular surveys* that each addressed a specific theme of interest to the program directors (e.g., communication, adaptability). Most modular surveys also were accompanied by a one-page evidence-based summary that provided background information, tips for trainees, tips for mentors, and resources for more information.

To promote accessibility, an online evaluation toolkit was created (<https://tarheels.live/t32programevaluation/>). The website included background information about evaluation, tips for logic models, downloadable surveys and related evidence-based summary resources, interview scripts and protocols, additional resources, and contact information for the core evaluation team.

Discussion

Program evaluation is a critical process for determining the extent to which federally funded postgraduate training programs are achieving their intended outcomes, such as elevating outcomes in biomedical sciences training, making trainees more competitive for the job market, and meeting the research needs of the nation. Although NIGMS training grant proposals must include an evaluation plan, funded programs may benefit from support in designing and operationalizing their evaluations (e.g., identifying shared components across programs and supporting perceived areas of need). Using a common evaluation framework [18], we demonstrated a strategy for creating synergy and collaboration across training programs at a single institution. Specifically, the systematic process described in this manuscript helped our interdisciplinary team identify common program evaluation priorities and develop related, evidence-based evaluation tools for use across the institution. We believe this work will advance evaluation efforts and ensure that postgraduate training programs in the biomedical sciences are well-positioned to collect and utilize data to inform decision-making.

Results of this study align with a growing body of research that highlights the importance of postgraduate career skills that extend beyond traditional, discipline-specific skills. In the

pharmaceutical sciences, for example, researchers have identified communication, collaboration, adaptability, flexibility and curiosity, among others, as key skills [23, 24]. Similarly, Sinche and colleagues identified written and oral communication skills, project management and creativity/innovative thinking as highly-rated transferrable skills by alumni [25]. Recently, some training programs also have focused on the development of skills related to resiliency, mental health, and wellness [26]; the ability to lead and function in diverse teams [27–29]; and communication outside disciplinary channels and audiences [30, 31]. Surprisingly, there appeared to be a gap between program director perspectives regarding the importance of Social and Behavior Skills and their need for support to evaluate these skills. As a growing body of research highlights the importance of social and behavioral skills, program directors may be simultaneously recognizing the importance of evaluating these skills but struggling to understand exactly how they impact trainee development.

Targeted career-development programming and professional-development training can assist in career placement [32] without hindering productivity or efficiency [12]. In addition to NIGMS funding requirements for career outcomes reporting, recent efforts to increase career outcomes reporting have enabled more rigorous evaluation of such outcomes [4, 10, 33–36]. A recent initiative to create more transparent access to evidence-based research on graduate and postdoctoral training was built through dozens of stakeholder insights across national organizations, indicating a true need for better program evaluation and dissemination thereof [37]. Our results indicate the need to focus on crucial program characteristics, namely training environment and opportunities offered, on trainee outcomes. Furthermore, trainee outcomes need to be more broadly defined in the research, including expanded factors beyond basic skill development, time-to-degree, publications, first job placements, and alumni job satisfaction. Although assessment culture and program transparency were identified as themes during the focus groups and interviews, they were not highly rated by program directors, which may be an important gap to explore in future research.

While the findings in this study may be transferable to other institutions, programs should give careful consideration to trainee skills and program characteristics most relevant to their program outcomes and institutional values. It is not our intent to suggest that the findings of our evaluation are the only important aspects of postgraduate training programs—only that these themes emerged from our data, collected from a select group of stakeholders at a specific point in time. Further, our analysis prioritized overarching themes at the potential expense of more specific aspects of training that may be essential for these programs. Additional aspects of postgraduate training known to affect trainee outcomes, such as equity, diversity, and inclusion (EDI), and entrepreneurship also should be considered for inclusion in program evaluations.

Several evaluation strategies utilized in this process are worth noting. Although often overlooked as an evaluation tool, logic models are critical for articulating how various elements of a program are believed to relate to one another [38]. While logic models typically include common core elements (e.g., resources, activities, outputs), they are a flexible tool that can vary by content, visual representation, and stakeholder beliefs about how a program works [39]. The logic models developed in this study enabled us to better understand individual programs, identify commonalities across programs, focus data collection on relevant activities and stakeholders, and determine next steps. However, creating a consensus logic model (Fig 3) required us to prioritize commonalities over program specifics and nuances, potentially overshadowing distinctive features of the programs included. As such, future research could explore strategies for developing more balanced consensus models and should focus specifically on relationships between various elements (e.g., activities and outputs) [40].

The engagement of the various stakeholders was also a strength of this study. The importance of stakeholders is often acknowledged, yet researchers rarely attend to stakeholder needs,

concerns, priorities, and perspectives in the evaluation design [41, 42]. In our study, program directors offered input into program evaluation needs, provided feedback on key findings (i. e., member-checking), helped focus instrument development on relevant constructs, and determined next steps. Employers and industry leaders provided specific insights into contemporary needs of the workforce, while trainees provided rich descriptions of their experiences in the training programs and institution. Although existing program evaluations addressed some identified constructs and needs, we believe that the input from these stakeholders will help synergize program evaluations broadly while providing opportunities for customizing measures of interest for each program. Of the recognized stakeholder groups in evaluation [43], our study included people who have direct responsibility for the programs (e.g., program directors) and people who are the intended beneficiaries of the programs (e.g., trainees, employers). In subsequent work, program alumni also could provide useful insight.

This study has several limitations. First, this evaluation was conducted at a single, research-intensive institution with a scientifically diverse graduate student and postdoctoral population. While this design limits generalizability, the needs and experiences of these participants are likely reflected at similar types of institutions. Second, the participants interviewed were volunteers, which may have introduced self-selection bias, and did not include program alumni or every type of employer (e.g., management consultant). However, the mixed methods design of this study enabled us to collect data from various stakeholders, utilize results to inform subsequent data collection and decision-making, and demonstrate triangulation across data sources. Third, the focus was broad (across programs) at the expense of exploring specific design elements of the training programs in depth [44]. Fourth, our sample of employers was relatively small and professionally diverse. Fifth, this research focused on design and implementation, leaving the impact of the developed toolkit unknown. Subsequent evaluation efforts will include an outcomes evaluation that explores website engagement (e.g., number of times accessed, number of times the materials are downloaded) and effectiveness of the evaluation tools.

In conclusion, this work contributes to a clear need concerning program evaluation in federally funded biomedical sciences training programs and provides a model for enhancing program evaluation capacity. Institutions must invest resources in evaluation to ensure programs are designed, implemented, and enhanced effectively. Based on our experience, strategic partnerships between program directors and evaluation experts can promote critical evaluation discourse, data collection, and decision-making. We recommend that other institutions offer similar evaluation expertise and encourage collaborations that promote evaluation capacity. At our university, staff and faculty with evaluation expertise are available to provide ongoing evaluation support for these programs, individually and collectively. The results of this work informed improvements at UNC and are offered as a resource to help others support their postgraduate training programs.

Acknowledgments

The authors would like to acknowledge Dr. Thomas A. Angelo for his consultation during the evaluation and Kirsten Leysieffer for her administrative support throughout this project. We would also like to thank all of our participants for their time and willingness to engage in focus groups, interviews, and surveys.

Author Contributions

Conceptualization: Jacqueline E. McLaughlin, Rebekah L. Layton, Paul B. Watkins, Robert A. Nicholas, Kim L. R. Brouwer.

Data curation: Jacqueline E. McLaughlin.

Formal analysis: Jacqueline E. McLaughlin, Rebekah L. Layton.

Funding acquisition: Jacqueline E. McLaughlin, Rebekah L. Layton, Paul B. Watkins, Robert A. Nicholas, Kim L. R. Brouwer.

Investigation: Jacqueline E. McLaughlin.

Methodology: Jacqueline E. McLaughlin, Rebekah L. Layton.

Project administration: Kim L. R. Brouwer.

Resources: Kim L. R. Brouwer.

Supervision: Paul B. Watkins, Robert A. Nicholas, Kim L. R. Brouwer.

Validation: Jacqueline E. McLaughlin.

Visualization: Jacqueline E. McLaughlin.

Writing – original draft: Jacqueline E. McLaughlin, Rebekah L. Layton.

Writing – review & editing: Jacqueline E. McLaughlin, Rebekah L. Layton, Paul B. Watkins, Robert A. Nicholas, Kim L. R. Brouwer.

References

1. Nowell L, Dhingra S, Kenny N, Jacobsen M, Pexman P. Professional learning and development framework for postdoctoral scholars. *Studies in Graduate and Postdoctoral Education*. 2021 Jul 13.
2. National Academies of Sciences Engineering & Medicine. *Graduate STEM Education for the 21st Century*. 2018.
3. National Academy of Sciences (US), National Academy of Engineering (US), Institute of Medicine (US), Committee on Science, Engineering, and Public Policy. *Enhancing the Postdoctoral Experience for Scientists and Engineers: A Guide for Postdoctoral Scholars, Advisers, Institutions, Funding Organizations, and Disciplinary Societies*. Washington (DC): National Academies Press (US); 2000.
4. Mathur A, Cano A, Kohl M, Muthunayake NS, Vaidyanathan P, Wood ME, et al. Visualization of gender, race, citizenship and academic performance in association with career outcomes of 15-year biomedical doctoral alumni at a public research university. *PloS one*. 2018 May 17; 13(5):e0197473. <https://doi.org/10.1371/journal.pone.0197473> PMID: 29771987
5. Silva EA, Des Jarlais C, Lindstaedt B, Rotman E, Watkins ES. Tracking career outcomes for postdoctoral scholars: a call to action. *PLOS Bio*; 2016: 14(5);e1002458. <https://doi.org/10.1371/journal.pbio.1002458> PMID: 27152650
6. Xu H, Gilliam RS, Peddada SD, Buchold GM, Collins TR. Visualizing detailed postdoctoral employment trends using a new career outcome taxonomy. *Nature biotechnology*. 2018 Feb; 36(2):197–202. <https://doi.org/10.1038/nbt.4059> PMID: 29334368
7. Best Places to Work Postdocs 2013. *The Scientist*. <https://www.the-scientist.com/features/best-places-to-work-postdocs-2013-39557>. Accessed on 25 May 2022.
8. Faupel-Badger JM, Raue K, Nelson DE, Tsakraklides S. Alumni perspectives on career preparation during a postdoctoral training program: A qualitative study. *CBE—Life Sciences Education*. 2015 Mar 2; 14(1):ar1. <https://doi.org/10.1187/cbe.14-06-0102> PMID: 25673353
9. McConnell SC, Westerman EL, Pierre JF, Heckler EJ, Schwartz NB. United States National Postdoc Survey results and the interaction of gender, career choice and mentor impact. *Elife*. 2018; 7:e40189. Published 2018 Dec 18. <https://doi.org/10.7554/eLife.40189> PMID: 30561332
10. Stayart CA, Brandt PD, Brown AM, Dahl T, Layton RL, Petrie KA, et al. Applying inter-rater reliability to improve consistency in classifying PhD career outcomes. *F1000Research*. 2020; 9. <https://doi.org/10.12688/f1000research.21046.2> PMID: 32089837
11. National Institutes of Health. *Biomedical Research Workforce Working Group Report*. Bethesda, MD: National Institutes of Health; 2012.
12. Brandt PD, Sturzenegger Varvayanis S, Baas T, Bolgioni AF, Alder J, Petrie KA, et al. A cross-institutional analysis of the effects of broadening trainee professional development on research productivity. *PLoS Bio*. 2021 Jul 15; 19(7):e3000956. <https://doi.org/10.1371/journal.pbio.3000956> PMID: 34264929

13. Ramadoss D, Bolgioni AF, Layton RL, Alder J, Lundsteen N, Stayart CA, et al. Using stakeholder insights to enhance engagement in PhD professional development. *PLoS one*. 2022 Jan 27; 17(1): e0262191. <https://doi.org/10.1371/journal.pone.0262191> PMID: 35085283
14. Department of Health and Human Services. T32 Training Program for Institutions That Promote Diversity. <https://grants.nih.gov/grants/guide/rfa-files/RFA-HL-22-001.html> Accessed on 23 Feb 2022.
15. Department of Health and Human Services. National Institute of General Medical Sciences Predoctoral Institutional Research Training Grant. <https://grants.nih.gov/grants/guide/pa-files/PAR-20-213.html> Accessed on 23 Feb 2022.
16. Department of Health and Human Services. Notice of Special Interest (NOSI): Administrative Supplements to Enhance Program Evaluation Capacity. <https://grants.nih.gov/grants/guide/notice-files/NOT-GM-21-024.html> Accessed on 23 Feb 2022.
17. Biological and Biomedical Sciences Program. <https://bbbsp.unc.edu/> Accessed on 22 Feb 2022.
18. Centers for Disease Control and Prevention. Framework for program evaluation in public health. *MMWR* 1999;48(No.RR-11):1–42. <https://www.cdc.gov/eval/materials/frameworksummary.pdf> Accessed on 13 Jan 2022.
19. Ladd S, Jerniga J, Watkins N, Farris R, Minta B, Brown S. Evaluation Guide: Developing and Using a Logic Model. CDC Division for Heart Disease and Stroke Prevention. Accessed June 25, 2020 from https://www.cdc.gov/dhdsp/docs/logic_model.pdf
20. Lawton B, Brandon PR, Cicchinelli L, Kekahio W. Logic Models: A Tool for Designing and Monitoring Program Evaluations. REL 2014–007. Regional Educational Laboratory Pacific. 2014 Feb.
21. Braun V, Clark V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006; 3:77–101.
22. Ponterotto JG. Brief note on the origins, evolution, and meaning of qualitative research concept “thick description.” *Qual Rep*. 2006; 11(3):538–549.
23. McLaughlin JE, Minshew LM, Gonzalez D, Lamb K, Klus NJ, Aubé J, et al. Can they imagine the future? A qualitative study exploring the skills employers seek in pharmaceutical sciences doctoral graduates. *Plos one*. 2019 Sep 9; 14(9):e0222422. <https://doi.org/10.1371/journal.pone.0222422> PMID: 31498853
24. Wu-Pong S, Gobburu J, O’Barr S, Shah K, Huber J, Weiner D. The future of the pharmaceutical sciences and graduate education: Recommendations from the AACP Graduate Education Special Interest Group. *Am J Pharm Edu*. 2013; 77(4): Article S2. <https://doi.org/10.5688/ajpe774S2> PMID: 23716757
25. Sinche M, Layton RL, Brandt PD, O’Connell AB, Hall JD, Freeman AM, et al An evidence-based evaluation of transferrable skills and job satisfaction for science PhDs. *PLoS one*. 2017 Sep 20; 12(9): e0185023. <https://doi.org/10.1371/journal.pone.0185023> PMID: 28931079
26. Milgram, S. NIH OITE Becoming a Resilient Scientist Series. https://www.training.nih.gov/nih_becoming_a_resilient_scientist_series Accessed 23 Feb 2022.
27. Center for the Improvement of Mentored Research Experiences in Research (CIMER). <https://cimerproject.org/> Accessed 23 Feb 2022.
28. Center for the Integration of Research, Teaching and Learning (CIRTL program). https://www.cirtl.net/about/core_ideas Accessed 23 Feb 2022.
29. Leadership and Management in Action Plan. <https://sites.wustl.edu/lmap/apply/> Accessed 23 Feb 2022.
30. ComSciCon: The Communicating Science workshop for graduate students. <https://comscicon.com/> Accessed 23 Feb 2022.
31. Science Communication Lab: Capturing the Journey and Wonder of Science. <https://www.sciencecommunicationlab.org/mission> Accessed 23 Feb 2022.
32. Collins TRL, Hoff K, Starback M, Layton RL. Creating and sustaining collaborative multi-institutional industry site visit programs: a toolkit. *F1000Research* 2020; 9:1317. <https://doi.org/10.12688/f1000research.26598.2> PMID: 33335720
33. Blank R, Daniels RJ, Gilliland G, Gutmann A, Hawgood S, Hrabowski FA, et al. A new data effort to inform career choices in biomedicine. *Sci*. 2017 Dec 15; 358(6369):1388–9. <https://doi.org/10.1126/science.aar4638> PMID: 29242335
34. Burghart S, Collins TRL, Hao Y, MacDonald JE, Ramadoss D, Robinson-Hamm J, et al. Institutional Graduate Career Outcomes Database 2020. Open Science Framework Project. <https://doi.org/10.17605/OSF.IO/ZP9RY>
35. Collins T, Ramadoss D, Layton RL, MacDonald J, Wheeler R, Bankston A, et al. Making Strides in Doctoral-Level Career Outcomes Reporting: Surveying the Landscape of Classification and Visualization Methodologies and Creating a Crosswalk Tool. *bioRxiv*. 2021. <https://doi.org/10.1101/2021.07.12.451657>

36. Lenzi RN, Korn SJ, Wallace M, Desmond NL, Labosky PA. The NIH "BEST" programs: Institutional programs, the program evaluation, and early data. *FASEB J.* 2020 Mar; 34(3):3570–82. <https://doi.org/10.1096/fj.201902064> PMID: 31960495
37. Professional Development hub (PD|Hub). pd|hub Collections: Building a national resource for advancing PhD education. <https://www.pdhub.org/> Accessed on 22 Feb 2022.
38. Cooksy LJ, Gill P, Kelly PA. The program logic model as an integrative framework for a multimethod evaluation. *Evaluation and program planning.* 2001 May 1; 24(2):119–28.
39. Greene JC. The role of theory in qualitative program evaluation. *Theory and concepts in qualitative research: Perspectives from the field.* 1993:24–45.
40. Rogers PJ. Causal models in program theory evaluation. *New directions for evaluation.* 2000 Sep; 2000(87):47–55.
41. Bryson JM, Patton MQ, Bowman RA. Working with evaluation stakeholders: A rationale, step-wise approach and toolkit. *Evaluation and program planning.* 2011 Feb 1; 34(1):1–2. <https://doi.org/10.1016/j.evalprogplan.2010.07.001> PMID: 20674980
42. Fleischer D. Evaluation use: A survey of US American Evaluation Association members. Unpublished Masters Thesis, Claremont Graduate University. 2007.
43. Greene JC. Stakeholders. In Friedman AL, Miles S. *Stakeholders: Theory and practice.* OUP Oxford; 2006 Apr 13., pp. 397–398.
44. Minshew LM, Olsen AA, McLaughlin JE. Cognitive Apprenticeship in STEM Graduate Education: A Qualitative Review of the Literature. *AERA Open.* 2021 Oct; 7:23328584211052044.