PNAS
www.pnas.org

Main Manuscript for

Postdocs' Lab Engagement Predicts Trajectories of Ph.D. Students' Skill Development.

David F. Feldon[1*], Kaylee Litson[2], Soojeong Jeong[1], Jennifer Blaney[1], Jina Kang[1], Candace Miller[3], Kimberly Griffin[4], Josipa Roksa[3].

[1]Department of Instructional Technology & Learning Sciences, Utah State University, Logan, Utah 84322-2830, USA.

[2]Department of Psychology, Utah State University, Logan, Utah 84322-2810, USA.

[3]Department of Sociology, University of Virginia, Charlottesville, Virginia 22904, USA.

[4]Department of Counseling, Higher Education, and Special Education, University of Maryland, College Park, Maryland 20742, USA.

***Corresponding author:** David F. Feldon

Email: david.feldon@usu.edu

Classification

Social Sciences

Keywords

graduate training, mentorship, research skills, postdocs, doctoral education

Author Contributions

DFF conceptualized the paper, KL, SJ, JB, JK, CM, KG, and JR analyzed the data, DFF acquired funding, DFF, KL, and SJ developed the methodology, SJ, JK, and KL provided data visualizations, DFF,

KL, SJ, JB, CM, KG, and JR contributed to the writing and editing of the paper. DFF and KL contributed equally to the writing of this paper as lead authors.

This PDF file includes:

Main Text
Figures 1 to 4

Abstract

The doctoral advisor—typically the principal investigator (PI)—is often characterized as a singular or primary mentor who guides students using a cognitive apprenticeship model. Alternatively, the "cascading mentorship" model describes the members of laboratories or research groups receiving mentorship from more senior lab members and providing it to more junior members (i.e., PIs mentor postdocs, postdocs mentor senior graduate students, senior students mentor junior students, etc.). Here we show that PIs' laboratory and mentoring activities do not significantly predict students' skill development trajectories, but the engagement of postdocs and senior graduate students in laboratory interactions do. We found that the cascading mentorship model accounts best for doctoral student skill development in a longitudinal study of 336 Ph.D. students in the United States. Specifically, when postdocs and senior doctoral students actively participate in laboratory discussions, junior Ph.D. students are over 4 times as likely to have positive skill development trajectories. Thus, postdocs disproportionately enhance the doctoral training enterprise, despite typically having no formal mentorship role. These findings also illustrate both the importance and the feasibility of identifying evidence-based practices in graduate education.

Significance Statement

Conventional wisdom in the laboratory sciences suggests that Ph.D. students develop their research skills as a function of mentorship from their faculty advisors (i.e., principal investigators [PIs]). However, no prior research has identified empirically a relationship between specific PI practices and the development of research skills. Here we show that PIs' laboratory and mentoring activities do not significantly predict students' skill development trajectories, but the engagement of postdocs and senior graduate students in laboratory interactions do. These findings support the practice of "cascading mentorship" as differentially effective and identify a critical but previously unrecognized role for postdocs in the graduate training process. They also illustrate both the importance and the feasibility of identifying evidence-based practices in graduate education.

Main Text

Introduction

Developing a highly skilled scientific workforce through doctoral training is critical to the advancement of science, but faculty who supervise these students consistently articulate reliance on their own experiences as students, rather than evidence-based practices, to inform their approaches to training (1-3). The doctoral advisor—typically the principal investigator (PI)—is often characterized as a singular or primary mentor who guides students using a cognitive apprenticeship model (4). Alternatively, the "cascading mentorship" model (5) describes the members of laboratories or research groups receiving mentorship from more senior lab members and providing it to more junior members (i.e., PIs mentor postdocs, postdocs mentor senior graduate students, senior students mentor junior students, etc.). However, it is unclear how successful each model may be in fostering the development of doctoral students' research skills.

Understanding research skill development requires attention to both the growth of specific skills (e.g., experimental design, data analysis) and collective profiles that reflect consistent patterns of growth within and across skills over time. For example, discrete graduate training experiences such as teaching (6) or coauthoring with a faculty mentor (7) are associated with growth in certain research skills. Other analyses have documented differences in rates of skill development overall as a function of small differences in initial skill level (8). However, little is known about the sustained effects of programmatic features of doctoral training that predict collective skill development over the course of multiple years.

Despite the popularity of the classic, single-mentor model in characterizations of graduate training and its positive association with scholarly productivity (9), no studies to date have linked the quality of mentorship to differential learning or skill outcomes. Likewise, the cascading mentorship model has been described as a "signature pedagogy" of laboratory-based sciences (5), but no studies have tested the efficacy of that structure.

In this 4-year longitudinal study, we measured research skills of a cohort of 336 Ph.D. students in the biological sciences who began their programs of study in Fall 2014, drawn from 53 universities across the United States (see Materials and Methods for details regarding participant recruitment). Specific sub-disciplines included cellular and molecular biology, developmental biology, microbiology, and genetics. Research skills were measured annually using sole-authored writing samples (e.g., draft manuscripts, qualifying or comprehensive examinations, dissertation proposals) that proposed or reported the results of empirical studies. Each writing sample was scored on all target skills by two blind raters using a validated rubric ($0.818 \leq ICC \leq 0.969$; see SI Appendix, Table S1).

The specific research skills measured for this study were: introducing/setting the study in context (INT), appropriately integrating primary literature (LIT), establishing testable hypotheses (HYP), using appropriate controls and replication (CTR), experimental design (EXP), selecting data for analysis (SEL), data analysis (ANA), presenting results (PRE), basing conclusions on results (CON), identifying alternative explanations of findings (ALT), identifying limitations of the study (LIM), and discussing implications of the findings (IMP). These planks were selected through a review of relevant literature and iterative development of criteria (10) as well as analyses from previous studies (6-8). We acknowledge that there are other important skills that likely contribute to expertise in the biological sciences, which were not measured in this study (e.g., bench skills). Written research products that make a scientific argument are an end-goal of research. Thus, they represent authentic and ecologically valid embodiments of skill with the potential to directly impact scientific career trajectories (11, 12).

We applied latent profile transition analysis (13) (LPTA) to examine year-over-year growth in all research skills by identifying discrete patterns of performance amongst the target skills that were common across subgroups of participants within and across years. LPTA is a longitudinal, person-oriented technique for modeling change and stability in subgroup membership (i.e., latent class) across time. Following cross-sectional latent profile analyses that identified discrete participant subgroups of the sample based on constellations of twelve research skills within each year, LPTA estimated the transitions of participants between these subgroups from year to year. Based on a comparison of model fit criteria (see full description in Materials and Methods and SI Table 7), results showed three latent subgroups among our sample of doctoral students with regard to their scores in a set of research skills: *low-*, *medium-*, and *high-skill* groups within each year (Fig. 1). The collective scores within each subgroup increased their average skill mean values over time in differing amounts on a 3-point scale: the low group increased its mean by 0.096, the medium group increased by 0.283, and the high group increased by 0.522 (Fig. 2), reflecting similar patterns to those reported in (8). Students transitioned among these subgroups across time with the following transitions being considered positive: *low-* to *medium-skill*, *low-* to *high-skill*, *medium-* to *high-skill*, and *high-* to *high-skill*. From Year 1 to 2, Year 2 to 3, and Year 3 to 4, 37%, 24%, and 7% of students had positive transitions, respectively.

To capture the overall trajectories of skill growth across the four years of data collection, we examined latent growth curves (LGCs; 14) within a structural equation modeling framework that included all measured skills. We identified four distinct LGCs, with one reflecting positive linear growth over time

(n=75; 35.4% of the sample), one reflecting decreasing then increasing growth (n=39; 13.1% of the sample), and two reflecting flat or slightly decreasing trajectories (n=153; 51.5% of the sample) (Fig. 3). Participants with positive transitions at any transition point were between 2.28 (95% CI = [1.48, 3.51]) and 4.62 (95% CI = [2.93, 7.29]) times as likely to have a positive linear LGC over 4 years than any of the other three LGC trajectories, showing that positive transitions between years were substantially associated with general positive growth across all years.

Next, we tested the extent to which different features of participants' doctoral training differentially predicted their year-over-year LPTA transitions and 4-year LGC trajectories. To do this, we collected survey data from participants on an annual basis to elicit details of their interactions with faculty as mentors and information about the roles that various individuals (i.e., PI, other faculty, postdocs, senior graduate students, junior graduate students, undergraduates, lab technicians) took on in the context of their laboratory. We used data from the second year of participants' Ph.D. programs as predictors because students began permanent lab placements with a designated faculty PI at the outset of their second year (15).

Using logistic regression models, we examined predictors of positive LPTA transitions and LGC trajectories (coded as 1), contrasted with all other transition and trajectory patterns (coded as 0). First, we tested whether lab roles during year 2 predicted positive transitions. The Year 1 to Year 2 transition was excluded from analysis, because data related to permanent lab features and faculty mentor interactions were not available prior to Year 2. All lab roles for all possible lab members were included as predictors of positive transitions and trajectories. In addition, we interviewed n=82 participants (24.4% of the total sample) on an annual basis to explore contemporaneous descriptions of participants' experiences and interactions within their respective laboratories.

Results

Latent profile transition analysis results

Our results indicate that the latent profiles yielded by the latent profile analyses tended to differ from each other mostly by skill level rather than by difference in shape. Specifically, participants in Latent Class 1 scored relatively high on all of the twelve research skills, and were accordingly referred to as *high-skill students*, comprising 13%, 21%, 21%, and 9% of the sample for Years 1, 2, 3, and 4, respectively. Participants in Latent Classes 2 and 3, on the other hand, scored moderately or low on all research skills relative to the rest of the sample. We therefore referred to these latent classes as *medium-skill students*, comprising 23%, 41%, 31%, and 35%, and *low-skill students*, comprising 63%, 38%, 48%, and 56%, of the full samples for Years 1, 2, 3, and 4, respectively (see Figs. 1 and 2).

The LPTA model examined how students transitioned among skill levels from year to year. Students showed positive, negative, and no transitions among latent classes across time as shown in SI Appendix, Fig. S1. For example, 20% of students in the *high-skilled student researchers* latent class in year 1 moved to the *moderate-skilled student researchers* latent class in year 2, indicating negative movement; 23% of students in the *low-skilled student researchers* latent class in year 2 moved to the *moderate-skilled student researchers* latent class in year 3, indicating positive movement; and 67% of students in the *low-skilled student researchers* latent class in year 3 remained in the low research skills latent class in year 4. The percentages given are based on the most likely latent class at each time point for each student. Means and standard deviations for the performance-based research skills measures across time are shown in SI Table 2.

Results show that senior graduate students engaging with primary literature strongly predicted positive LPTA transitions for study participants in Years 2 to 3 (Odds ratio [OR] = 3.85; 95% CI = [1.07, 13.89]), and both postdocs and senior graduate students participating in lab discussions strongly predicted positive LPTA transitions in Years 3 to 4 (OR = 5.14; 95% CI = [1.69, 15.60]; OR = 4.50; 95% CI = [1.39, 14.58]).

Latent growth curve results
Results from the LGC trajectory logistic regression likewise showed that postdocs participating in lab discussions strongly predicted the likelihood of participants belonging to the linear positive growth LGC (OR = 4.20; 95% CI = [1.59, 11.10]). Other predictors, including demographic characteristics, presence of postdocs in lab (independent of activities), faculty interactions, and peer interactions were nonsignificant in predicting either positive LPTA transitions or LGC trajectories (SI Tables 3-5). Positive, significant predictors of positive transitions and trajectories are shown in Fig. 4.

Qualitative results
Of the interviews conducted (n=82, drawn from the larger study sample), 48.8% of interviews conducted with study participants yielded at least one characterization of postdocs as valued mentors and instructors within the laboratory context. Four themes emerged related to how students interact with and receive support from postdocs. Specifically, we found that postdocs provide hands-on instruction in the lab (n=18), give professional and academic feedback (n=17), model how an academic career may look for the graduate student (n=13), and provide personal/emotional support (n=13). Representative quotes are provided in SI Table 6.

Discussion

These findings indicate that Ph.D. students in the biological sciences are 4.50 times as likely to have positive year-to-year LPTA transitions when senior graduate students are active participants in lab discussion. They are also 5.14 times as likely to have a positive year-to-year LPTA transitions when postdocs are active participants in lab discussion. Similarly, they are 4.20 times as likely to have positive LGC growth trajectories when postdocs are active participants in lab discussion. Further, the qualitative data indicate that postdocs mentor doctoral students in myriad ways, most commonly by being present in the lab to provide on-going and hands-on instruction and professional guidance. Notably, PI activities and reported faculty mentorship measures do not predict positive year-to-year transitions or overall positive trajectory. In combination with the extensive set of variables found to have no effect on positive transitions and growth trajectories, our results suggest that active engagement in collective laboratory discussion by senior peers (i.e., senior graduate students and postdocs) better predict Ph.D. students' skill development than the mentoring or laboratory activities of faculty mentors. As such, the cascading model of mentorship is not only a descriptive norm of doctoral student support in university-based laboratory environments (5), but also a differentially beneficial practice that uniquely predicts positive research skill development.

In this context, our findings have substantial implications for both programmatic doctoral training in the biological sciences and the conceptualization of the value that postdocs contribute to the larger research enterprise. As the practice of science has shifted toward larger team enterprises and an increasing pace and volume of workload, the nature of the PI's role has shifted to one that often entails less direct contact with students (16, 17). Postdocs and others within the lab may step into the gap that is created, with unexpected dividends. In this context, our findings suggest that adoption of a cascading mentorship model which encourages active engagement of postdocs within the laboratory as mentors to Ph.D. students may be beneficial to student skill development. Accordingly, it is possible that providing training to postdocs in effective mentoring practices may further enhance the benefits to graduate students identified in this study.

Recent analyses of the postdoctoral role within the research enterprise indicate that postdocs are underpaid relative to the value they contribute to scholarly productivity (18). However, their total value within the laboratory may be substantially more than currently recognized based on their skill development contributions. Conversely, postdocs may realize value in terms of their own development from engaging in informal mentoring of graduate students. Previous research has identified benefits for

graduate students' research skill development from a combination of teaching and research activities over research as a sole focus (6). It may be that postdocs benefit similarly from their roles as mentors in the laboratory. Exploratory studies suggest that postdoctoral mentoring of student researchers may facilitate further thinking and risk-taking on postdocs' topics of research, along with the development of other skills, such as teaching and the use of scientific communication skills (19). Opportunities to develop such skills align with current recommendations for postdoctoral training (20, 21).

Materials and Methods

Participant Recruitment

Study recruitment materials instructing prospective participants to contact the research team were disseminated in two phases. First, we contacted program directors and department chairs of the 100 largest PhD programs in the biological sciences across the United States as well as public flagship universities and minority serving institutions (i.e., historically Black colleges and universities and Hispanic-serving institutions) with PhD programs in the biological science subfields of interest. All program directors and department chairs were given information about the purpose of the study and asked to share recruitment materials with incoming PhD students in fall of 2014. Next, the research team sent recruitment emails to several listservs, including those of the American Society for Cell Biology and the Center for the Integration of Research, Teaching, and Learning Network for broader dissemination. All students who responded to these emails were entering PhD programs that we contacted in the first phase of recruitment, suggesting that recruitment efforts approached saturation at the institutional level. All prospective participants who contacted the research team were then screened to ensure that they met the criteria for participation and understood the expectations for participation. Participants signed informed consent per the requirement by the Utah State University Institutional Review Board (IRB) for human subjects research under protocol #5888. To incentivize study participation, students received a $400 annual incentive. The full procedure for this study was approved by the IRB.

In total, we recruited 336 participants from 53 institutions across the United States. Of the institutions represented, 42 are classified as R1 (highest research activity), seven institutions are R2 (higher research activity), and the remaining four institutions fall in other Carnegie categories.

Data Collection

Data for the present study were obtained through web-based surveys and the collection of single-authored writing samples via email. Both survey data and writing samples were collected annually during the first four years of the doctoral program.

After removing cases with missing data on all key variables and accounting for attrition (both from the study and the doctoral program), the present study relies on a longitudinal sample of n=297 students. Most participants were female (n=183), continuing-generation (n=210), from majority racial/ethnic groups (n=240), or were domestic students (n=237). Fewer participants were male (n=114), first-generation (n=83), from under-represented racial/ethnic minority groups (n=53), or were international students (n=57). Four students did not provide data on their generation status nor racial/ethnic identity, and three did not provide data on international student status.

Measures

Background Variables. During the first year of the study, students completed a demographic questionnaire that included questions about their race/ethnicity, gender, parents' education level, and international student status.

Race. Students indicated their race/ethnicity by selecting one or more of the following: American Indian or Alaska Native; Asian or Asian American; Black or African American; Latino/a; Native Hawaiian or Other Pacific Islander; White. Students' responses were aggregated to create a measure of under-represented racial/ethnic minority (URM) status (0=majority; 1=URM) where students who selected only a White and/or Asian identity were coded as majority; all other students were coded as URM.

Gender. As a proxy for gender, students reported their sex as female or male (female=0; male=1).

First-Generation College Status. Students were asked to indicate the highest degree obtained by their parent(s), students who had no parent with a 4-year college degree were coded as first-generation (0=continuing generation; 1=first-generation).

International Student Status. Students also self-reported whether or not they are an international student (0=No; 1=Yes).

Performance-Based Research Skills. To measure research skills, students submitted a sole-authored research product each year immediately following the spring semester of their doctoral training. Expectations that 1) documents were written within the preceding 4 months and 2) were not to have been edited or contributed to by others were clearly communicated. Consequently, writing samples were typically unpublished manuscripts, which may or may not have later been published in subsequent collaboration with others. Two independent reviewers rated each document on twelve research skills according to clearly defined rubric criteria. Rubric criteria draw heavily from prior studies (6, 10). Skills included:

1. Introducing/setting the study in context (INT)
2. appropriately integrating primary literature (LIT)
3. establishing testable hypotheses (HYP)
4. using appropriate experimental controls and replication (CTR)
5. experimental design (EXP)
6. selecting data for analysis (SEL)
7. data analysis (ANA)
8. presenting results (PRE)
9. basing conclusions on results (CON)
10. identifying alternative explanations of findings (ALT)
11. identifying limitations of the study (LIM)
12. discussing the implications of the findings (IMP)

All raters reviewed criteria on a scale from 0 to 3.25. Interrater reliability as measured by intraclass correlations (ICCs; two-way random effects) was good, 0.818 to 0.969. Exact ICC values are shown in SI Table 1. Scores were averaged across raters to create a composite measure for each skill.

Lab Roles. Roles of other lab members were evaluated by asking students:

In your research experience during your PhD program so far, who participates in:

1. lab discussions to understand contemporary concepts in your field?
2. making use of the primary scientific research literature in your field (e.g., journal articles)?
3. identifying a specific question for investigation based on the research in your field?
4. formulating research hypotheses based on a specific question?
5. designing an experiment or theoretical test of hypotheses?
6. developing the “controls” in research?
7. collecting data?
8. statistically analyzing data?
9. interpreting data by relating results to the original hypothesis?
10. reformulating original research hypotheses (as appropriate)?

Students responded to each question by selecting all persons in the lab who participated in each task. Possible responses included principal investigator(s), other faculty, research scientists/postdocs, senior graduate students, junior graduate students, lab technicians, and undergraduate students. Affirmative responses were coded as 1. Empty responses, as long as the student answered other questions, were coded as 0. Items that were seen by the participant but left unanswered, or items that were not seen by the participant were considered missing data. Responses to lab roles from year 2, after students entered a permanent laboratory, were included in the present study.

Faculty Interactions. We examined the role of student-faculty interactions, relying on previously developed items (22). The *occurrence* of faculty interactions was measured using a composite variable made up of 4 items from the annual surveys, asking students whether or not they do any of the following with program faculty: engage in social conversation; discuss topics in his/her field; discuss other topics of intellectual interest; and talk about personal matters. These items showed adequate reliability, McDonald’s omega = .71. Items were added together to form a scale ranging from 0 (little to no interactions) to 4 (many types of interactions).

The *quality* of faculty interactions was computed using a 6-item composite variable where students indicated their agreement with the following items: the faculty are accessible for scholarly discussions outside of class; I feel free to call on the faculty for academic help; the faculty are aware of student problems and concerns; I can depend on the faculty to give me good academic advice; I am treated as a colleague by the faculty; the faculty sees me as a serious scholar. Students responded to each item using a 5-point scale ranging from strongly disagree to strongly agree. These items showed good reliability, McDonald's omega = .85. Again, items were added together to form a scale, ranging from 6 (low quality interactions) to 30 (high quality interactions).

Peer Interactions. We also examined the role of the student-peer interactions using two subscales (22). The *social* interaction with peers was measured using two items that asked students to indicate (yes = 1, no = 0) whether they have interacted with peers in their department in the following ways: engage in social conversation and talk about personal matters. The *academic* interaction with peers was also assessed using two items that asked students to indicate (yes = 1, no = 0) whether they have interacted with peers in the following ways: discuss topics in his/her field and discuss other topics of intellectual interest. Items were summed to create each of the two subscales. These items yielded adequate to good reliability estimates, McDonald's omega = .81 (academic) and .83 (social).

Interviews

To provide a more nuanced interpretation of the findings, we used qualitative analysis to explore contemporaneous descriptions of postdocs' interactions with graduate student participants. Specifically, we analyzed interview data from 82 participants who were recruited from the larger sample. The qualitative sample was largely representative of the quantitative survey sample; URM students and first-generation college students each represented nearly a third (29.3%) of participants, and women made up 68.3% of the sample. All 82 participants completed an hour-long, semi-structured phone interview with a member of the research team during the summer after their fourth year in the doctoral program. The interview protocol focused on students' experiences over the course of their doctoral program and included questions about experiences in the lab, along with probing questions asking specifically about interactions with PIs, faculty, postdocs and other research staff. All interviews were recorded and transcribed verbatim.

Statistical Analyses

All quantitative analyses accounted for students nested within university. Quantitative analyses were conducted in Mplus v8.1. Response variable rubric scores had 31% missing data at the first time point, up to 57% missing data at the fourth time point. A missing values analysis [$\chi^2(312) = 346.41$, $P = 0.09$] showed that the missing data met the assumption for missing completely at random (MCAR; 23). Missing data were handled more conservatively under missing at random (MAR; 24) assumptions by using a maximum likelihood estimation algorithm robust to nonnormally distributed data (MLR).

The potential heterogeneity of doctoral student skill development was evaluated using two different methods: latent profile transition analysis and latent class growth analysis. Multinomial logistic regression analyses were conducted to examine positive transitions and positive growth. To account for family-wise error due to multiple sets of logistic regression analyses, a false discovery rate was applied to results (25).

Latent Profile Transition Analysis. Latent profile transition analysis (LPTA) is an extension of latent profile analysis (LPA) to longitudinal measures. LPA is a person-oriented technique used for identifying unobserved subgroups in a sample based on the patterns of means (and variances) of observed variables within a given time point. LPTA additionally evaluates the stability and mobility of subgroup memberships over time by evaluating the probabilities of individuals transitioning from one latent class at time *t* to another latent class at time *t*+1 (13). This analysis allowed us to examine discrete *transitions* of individuals moving from one skill profile to another skill profile between consecutive time points.

Prior to conducting LPTAs, we first performed LPAs to identify distinct latent subpopulations among doctoral students who were scored on a set of 12 indicators of research skills at each assessment point (Y1-Y4). Our analysis estimated LPA solutions with one through five latent classes, and were estimated using 500 starting values. We inspected different model fit criteria (Bayesian Information

Criterion [BIC] and Entropy) across solutions for selecting the best-fitting solution (SI Table 7). BIC is recommended as the most powerful measure for evaluating competing models to determine the optimal number of latent classes, with lower values representing better model fit. Entropy is an indicator of the precision with which individuals are assigned to each latent class, with values close to 1 representing more accurate latent class assignments. Based on these measures, our results showed that the 3-class LPA model was the best fit to the data for all four assessment points. Although the values of BICs continued to decrease without actually reaching the minimum value, the plots of these values showed that the slope plateaued between three and five latent classes, indicating that the 3-class solution was better than other solutions. The 3-class model also had very high entropy (0.95, 0.94, 0.94, 0.96 for Y1, 2, 3, 4 respectively). Accordingly, we chose the 3-class LPA model as the measurement model for the subsequent LPTA models.

After evaluating the profiles of skills *within time*, we examined how students transitioned between profiles *across time*, using LPTA. Two different LPTA models were evaluated: a model without assumptions about the model structure across time and a model that assumed measurement equivalence across time. Confirmation of measurement equivalence would indicate that the skill profiles are the same across time (e.g., high skills at time 1 are the same as high skills at time 2, etc.). The model without measurement equivalence across time best fit the data (SI Table 8).

Latent Class Growth Analysis. Latent class growth analysis (LCGA) is a longitudinal, person-oriented analysis (14). The primary use of LCGA is to identify latent subgroups (or "classes") of participants based on similar latent growth curve (LGC) trajectories. Group membership is static and unchanging in LCGA because the goal of LCGA is to identify latent trajectories rather than latent transitions. This analysis is allowed us to examine skill *trajectories* across all time points.

LCGA models with one to six latent classes were evaluated. All models were evaluated using 500 starting values. Models were compared using BIC and Entropy (see SI Table 9) to determine the number of latent classes that best represented the skills trajectories data. Like the LPTA models, the BIC values continued to decrease as the number of latent classes increased. Although these values decreased, examining plots of the BIC showed that the BIC slope plateaued at the four-class solution, and this solution was chosen as the best-fitting model.

To characterize doctoral student skill development across latent classes, we examined the mean values of each skill at each time point within each latent class. Class 1 represented 26% of the sample and showed decreasing skill levels across time. Class 2 represented 25% of the sample and showed stable, low skill levels across time. Class 3 represented 13% of the sample and showed initially decreasing then increasing skill levels across time. Class 4 represented 35% of the sample and showed linear, increasing skill levels across time.

Multinomial Logistic Regression Analyses. Logistic regression analyses were used to examine whether independent predictors impacted positive transitions as well as positive skill trajectory. All variables (as shown in Tables 6-8), a total of n=78, were included as predictors of positive transitions between years 2 and 3 as well as years 3 and 4. All variables were also included as predictors of the positive latent growth curve skill trajectory. Analyses were evaluated in separate statistical models due to the relatively low sample size. Confidence intervals were adjusted post-hoc using false discovery rate (FDR) (*28*) to maintain a familywise error rate of .05. Results for all independent variables predicting positive transitions between times 2 and 3, 3 and 4, and positive skill trajectories are presented in SI Tables 6-8.

Qualitative Analyses. Data was analyzed using NVivo 12 software. Because we were particularly interested in students' experiences with postdocs and full-time research staff related to the current paper, we purposely selected transcripts from all participants who mentioned their interactions with postdoc(s) and other research staff as part of their interview. To identify these participants, we first conducted exploratory analyses using a sample of 20 transcripts to identify the language used by students to discuss relevant interactions in the lab. Based on this preliminary analysis, we used NVivo's text search query function to search for interviews where participants used phrases containing any variation of the following terms: "postdoc" OR "post doc" OR "post-doc" OR "lab manager" OR "research scientist". Notably, the query also detected instances where the identified terms were the *stem* of the words used in the interview (e.g., if students discussed "postdoctoral researchers", this would have also been captured by the query). Finally, results from the query were reviewed to remove any

interviews/excerpts that were not relevant. After closely reviewing the query results, we identified 53 participants with relevant data.

All 53 transcripts were coded in a systematic, two-phase process (26). During the first phase, we read and re-read each excerpt to identify emergent themes related to how students interact with postdoc and other full-time research staff. Next, we developed a codebook with descriptions and examples of the themes and used this codebook to analyze each transcript. Two members of the research team independently coded 28% of the transcripts (n=15) to ensure reliability and trustworthiness. Additionally, we met throughout the analysis process to participate in peer debriefing (27).

Data and Code Availability

The data generated and analyzed during the current study are available from the corresponding author on reasonable request. The statistical code used to analyze the data during the current study is available from the corresponding author on reasonable request.

Acknowledgments

The authors gratefully acknowledge the support of the National Science Foundation. This material is based upon work supported under Awards 1431234, 1431290, and 1760894. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors also thank Alok Shenoy for his data management services.

References

1. S. Delamont, O. Parry, P. Atkinson, Creating a delicate balance: the doctoral supervisor's dilemmas. *Teac. High. Educ*. **3**, 157-172 (1998).
2. A. Lee, How are doctoral students supervised? Concepts of doctoral research supervision, *Stud. High. Educ*. **33**, 267-281 (2008).
3. S. Stephens, The supervised as the supervisor. *Edu. Train.* **56**, 537-550 (2014).
4. A. E. Austin, Cognitive apprenticeship theory and its implications for doctoral education: A case example from a doctoral program in higher and adult education. *Int. J. Acad. Dev*. **14**, 173-183 (2009).
5. C. Golde, A. C. Bueschel, L. Jones, G. Walker, "Advocating apprenticeship and intellectual community: Lessons from the Carnegie Initiative on the Doctorate" In *Doctoral education and faculty of the future*, R. G. Ehrenberg, C. V. Kuh, Eds. (Cornell University Press, 2009), pp. 53-64.
6. D. F. Feldon *et al*., Graduate students' teaching experiences improve their methodological research skills. *Science* **333**, 1037-1039 (2011).
7. D. F. Feldon, K. Shukla, M. A. Maher, Faculty-student coauthorship as a means to enhance STEM graduate students' research skills. *Int. J. Res. Dev*. **7**, 178-191 (2016).
8. D. F. Feldon, M. A. Maher, J. Roksa, J. Peugh, Cumulative advantage in the skill development of STEM graduate students: a mixed-methods study. *Am. Educ. Res. J.* **53**, 132-161 (2016).
9. L. Paglis, S. Green, T. Bauer, Does adviser mentoring add value? A longitudinal study of mentoring and doctoral student outcomes. *Res. High. Educ*. **47**, 451-476 (2006).
10. B. E. Timmerman, D. C. Strickland, R. L. Johnson, J. Payne, Development of a rubric for assessing undergraduates' science inquiry and reasoning skills using scientific writing across multiple courses. *Assess. Eval. High. Educ*. **36**, 509-547 (2011).
11. D. K. Simonton, Scientific creativity as constrained stochastic behavior: The integration of product, person, and process perspectives. *Psychol. Bull*. **129**, 475-494 (2003).
12. D. F. Feldon, M. Maher, B. Timmerman, Performance-based data in the study of STEM graduate education. *Science* **329**, 282-283 (2010).

13. L. Collins, S. Lanza, *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences* (John Wiley & Sons, Hoboken, NJ, 2010).
14. D. S. Nagin, Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychol. Methods.* **4**, 139-177 (1999).
15. M. A. Maher, A. M. Wofford, J. Roksa, D. F. Feldon, Doctoral student experiences in biological sciences laboratory rotations. *Stud. Grad. Postdoc. Educ.* **10**, 69-82 (2019).
16. J. M. Gappa, A. E. Austin, A. *Trice, Rethinking faculty work: Higher education's strategic imperative*. (Jossey-Bass, San Francisco, CA, 2007).
17. B. F. Jones, S. Wuchty, B. Uzzi, Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* **322**, 1259-1262 (2008).
18. S. Kahn, D. Ginther, The impact of postdoctoral training on early careers in biomedicine. *Nature Biotechnol*. **35**, 90-94 (2017).
19. E. Dolan, D. Johnson, Toward a holistic view of undergraduate research experiences: An exploratory study of impact on graduate/postdoctoral mentors. *J. Sci. Educ. Tech*. **18**, 487-500 (2009).
20. R. Daniels, L. Beninson, *The next generation of biomedical and behavioral sciences researchers: Breaking through* (The National Academies Press, Washington, DC, 2018).
21. Committee to Review the State of Postdoctoral Experience in Scientists and Engineers. *The Postdoctoral Experience Revisited* (The National Academies Press, 2014).
22. J. C. Weidman, E. L. Stein, Socialization of doctoral students to academic norms. *Res. High. Educ.* **44**, 641-656 (2003).
23. R. J. A. Little, A test of missing completely at random for multivariate data with missing values. *J. Am. Stat. Assoc.* **83**, 1198-1202 (1988).
24. C. K. Enders, *Applied Missing Data Analysis* (Guilford Press, 2010).
25. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc*. **57**, 289-300 (1995).
26. W. H. Percy, K. Kostere, S. Kostere, Generic qualitative research in psychology. *Qual. Rep.* **20**, 76-85 (2015).
27. J. Creswell, D. Miller, Determining validity in qualitative inquiry. *Theory into Prac.* **39**, 124-130 (2000).

Figure Legends

Figure 1. High-, medium-, and low-skill latent profiles, showing estimated mean scores for each research skill within each year according to the final LPTA model.

Figure 2. Latent profiles by time, showing minimum, maximum, means, and number of participants in each profile.

Figure 3. Latent growth model results, reflecting common trajectories of skill development within each latent class.

Figure 4. Odds ratios with 95% confidence intervals displaying significant predictors of positive transitions (Year 2 to Year 3, Year 3 to Year 4) and skill trajectories (LGC), as well as contrasting nonsignificant predictors. Odds ratios with confidence intervals containing 1 (indicated by vertical dashed line) are nonsignificant.