THE COMPETENCY PIPELINE: EXAMINING THE ASSOCIATION OF DOCTORAL

TRAINING WITH EARLY CAREER OUTCOMES

Andrea Ortiz, M.A.

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

December 2023

APPROVED:

Jennifer L. Callahan, Committee Chair
C. Edward Watkins, Jr., Committee Member
Ateka Contractor, Committee Member
Donald Dougherty, Chair of the Department of
    Psychology
Albert Bimper, Executive Dean of the College
    of Liberal Arts and Social Sciences
Victor Prybutok, Dean of the Toulouse Graduate
    School

Ortiz, Andrea. *The Competency Pipeline: Examining the Association of Doctoral Training with Early Career Outcomes*. Doctor of Philosophy (Clinical Psychology), December 2023, 77 pp., 1 table, 1 appendix, references, 52 titles.

Participants from earlier nationwide studies on predictors of internship match were contacted 7-10 years after obtaining their doctoral degree to gather additional data concerning their attained early career competencies and benchmarks (e.g., scores on the national licensing exam). In this sample ($N = 190$), licensure exam scores were significantly positively associated with scores obtained on the Graduate Record Examination (GRE), family of origin socioeconomic status, and student debt load. However, obtained licensure scores were not found to be significantly associated with any pre-doctoral training variables (e.g., intervention/ assessment hours, number of integrated reports, number of publications, rank order of matched internship site) or most post-doctoral early career activities. Weak positive associations between licensures scores and engaging in teaching / supervision / consultation services were observed. Few self-reported early career competencies were found to be weakly positively associated with scores on the national licensing exam (i.e., integrity, seeking supervision, scientific mindedness, evidence-based practice). Significantly inverse associations were found between national licensing exam scores and self-reported competencies in the areas of management and systems change. Findings are discussed and implications for the national licensing exam considered.

TABLE OF CONTENTS

CHAPTER 1

REVIEW OF LITERATURE

Callahan, Smotherman, Dziurzynski, Love, Kilmer, Flores Niemann, and Ruggero (2018) contend that exposure to diverse student groups in higher education programs is key to fostering multicultural encounters that can lead to multicultural competence and the development of workforce leadership skills. Additionally, they observed that members of diverse groups (i.e. minoritized race/ethnicity, disability, gender identification groups) comprise only 15% or less of the student population of clinical training programs. As a result, doctoral students tend to experience few multicultural encounters within their training cohorts, which in turn creates risk of polarized and hostile study environments that may hamper recruitment as well as retention of diverse individuals.

Perhaps reflective of those issues, non-Hispanic white psychologists are disproportionately overrepresented in the workforce. Callahan and colleagues' (2010, 2014) quantitative analyses examined the number of diverse students in American Psychological Association's (APA) accredited doctoral psychology programs to determine the following: (1) is the pipeline diverse and representative of the national population, and (2) are there any significant differences on undergraduate grade point average (GPA), scores on the Graduate Record Examination (GRE), or other academic qualification indicators that may explain the disparate numbers of white versus diverse individuals entering the professional psychology workforce? With respect to their first question, the authors' analysis of the annual reports of all APA accredited doctoral programs nationwide from 2005 to 2015 revealed that African American and Black, and Hispanic and Latino/a students were grossly underrepresented in APA doctoral programs. Furthermore, the data revealed non-Hispanic white students in APA

programs represented 83% of the student body population (despite representing only 63% of the 2010 census population for the United States; US), while African Americans and Blacks represented only 5.3% of the student population (despite comprising 13.3% of the US population). Similar to African American and Black students and Hispanic and Latino/a students were underrepresented and comprised only 10.75% of the student population, despite representing 17.8% of the US population. Related to their second question, the researchers found no statistically significant differences on indicators of academic qualification (i.e., grade point average, GPA, or scores on the Graduate Record Examination) between diverse students and non-Hispanic white students.

Callahan and colleagues' (2018) study prompted three responsive commentaries, each highlighting a different issue. As summarized more fully below, Luebbe and Ogbaselase (2018) provided compelling data to demonstrate that underrepresentation of marginalized individuals begins even before the point of applying to doctoral programs. Borrego (2018) identified recruitment and ongoing mentorship as key to diversifying the profession, while Lund (2018) cautioned that variables after admission are also important considerations in creating diversity-affirming training environments. Each of these commentaries merits a brief summary to provide context for the current study.

First, Luebbe and Ogbaselase's review of relevant literature provided complementary data to Callahan, Smotherman, Dziurzynski, Love, Kilmer, Niemann, and Ruggero's (2018) examination of diversity within the training-to-workforce pipeline for professional psychology. Luebbe and Ogbaselase's data suggested that individuals from most minority groups were also underrepresented in psychology at the undergraduate level as well, leading the authors to question how early in the training pipeline the loss of representation of minorities begins.

Though they identified means of improving representation such as early exposure to psychology and making psychology more accessible to ethnic minorities, financial hardship and looming accumulated debt still posed significant barriers. As such, faculty are encouraged to provide a sense of supportiveness coupled with micro-affirmations (i.e. listening, inclusion, and camaraderie) within racial/ethnic congruent as well as incongruent mentoring relationships.

Second, Borrego's 2018 commentary provides an overview of the ongoing dialogue about diversity outcomes in doctoral student populations. The author asserts that graduate programs must identify sustainable systems of training curriculum evaluation in order to increase diversity and equity within the field. The author's review of Callahan and colleagues' (2018) study on the training-to-workforce pipeline underscores the role of improved recruitment and retention of diverse students as critical to efforts at increasing representation of diverse individuals in the professional psychology workforce. The authors posited such changes are most viable when faculty are willing to examine their own commitment to diversity in conjunction with evaluating student competency. Specifically, Borrego (2018) identified mentoring and increased recruitment of racial and ethnic minorities as effective ways to diversify the training-to-workforce pipeline. By gaining a thorough command of intersectionality (according to the APA's revised multicultural training guidelines), programs may successfully implement meaningful change (APA, 2017). Borrego asserts that equity and inclusion can only occur in the face of unified advocacy and social justice.

Third, Lund (2018) provided a commentary on Callahan et al.'s (2018) research, with particular attention to representation of psychology trainees with disabilities. Lund underscored the observation that students with disabilities were not only underrepresented, but also susceptible to higher attrition rates, regardless of their academic qualifications. Lund advocated

that the data were indicative of a need for training programs and faculty to implement disability-affirmative support networks in order to create a more collegial environment for all learners. Overall, Lund asserts these efforts could greatly improve trainee recruitment and retention outcomes.

Lund's observation concerning the importance of affirming training environments is underscored in a forthcoming article by Gregus, Stevens, Seivert, Tucker, and Callahan (2020). Gregus and colleagues used the Multicultural Environmental Inventory-Revised (MEI-R) to assess students' perceptions of program climate and how well multicultural issues were addressed. The MEI-R is a four-subscale, 27-item self-report measure that assesses students' perceptions of how well they believe their clinical program addresses important multicultural issues, and their perceptions of their program climates (Pope-Davis, Liu, Nevitt & Toporek, 2000). Participants were comprised of 397 students from clinical psychology graduate programs across the United States. In a mixed methods design study, students were asked to complete the MEI-R and were asked the following open-ended questions to determine the strengths and weakness of their respective programs: "What do you think your program is doing well with respect to diversity in training? Please give specific examples of how this is done" and, "What do you think your program could do differently to enhance diversity in training? Please give specific examples of how this could be done." Several trends were found among different demographic groups.

Students who identified as members of marginalized groups rated their program as having less favorable multicultural training, quality curriculum, supervision and program climate and comfort. Students with disabilities reported lower climate comfort subscale scores. For example, African American and Black participants reported lower total scores, $F(4, 385) =$

10.94, $p < .001$, with regard to curriculum and supervision and lower climate and comfort subscale scores $F(4, 385) = 5.18$, $p < .001$, and lower honesty in recruitment scores $F(4, 385) = 3.02$, $p = .02$, when compared to other racial/ethnic groups. There was a medium effect $F(2, 378) = 3.67$, $p = .03$ of sexual orientation on MEI-R scores, with gay and lesbian students reporting higher scores on honesty in recruitment, $F(2, 378) = 5.71$, $p = .004$, and higher total scores than other demographic groups. Students from programs with diversity committees rated their programs more favorably than programs without such committees. In the qualitative analysis, participants indicated program strengths included opportunities to work with diverse individuals, whereas program weaknesses included lack of multiculturalism in curriculum and lack of recruitment and retention of underrepresented students and faculty. The authors recommend programs use the MEI-R in order to identify which students are not feeling supported, as it helps elucidate students' perspectives of their respective programs. Implementation of diversity committees, integration of multicultural learning across curriculum, diversity training, deep discussions of multicultural issues and recruitment of diverse program members are recommended based on these results.

Importantly, the challenges faced by individuals from marginalized populations do not end with training. A recent review of the nationally required licensure examination summarizes longstanding concerns that are encountered following training as individuals seek to emerge into the profession. In their review of the current state of the national psychology licensure examination, Callahan, Bell, Davila, Johnson, and Strauman (2020) outlined several strong criticisms of forthcoming changes in the examination. The Examination for Professional Practice in Psychology (EPPP) is administered by the Association of State and Provincial Psychology Boards (ASPPB) and required by state and provincial licensing boards across the United States,

its territories, and Canada. The current EPPP measures foundational knowledge, while the intention of the EPPP (Part 2 – Skills) is to evaluate functional competencies.

As noted by Callahan and colleagues, the general approach of using use competency-based evaluations as a prerequisite to licensure is in keeping with contemporary expectations of health care professions. These exams generally require the trainee to apply knowledge, skills, values, attitudes, and behaviors under real world conditions. In contrast to other healthcare fields involvement of stakeholders in the development and validation of the exam, Callahan and colleagues contend that the Part 2 – Skills exam has not involved a broad base of stakeholders. The authors also assert that the Part 2 – Skills exam has not yet been sufficiently validated, creating risk of negative implications for the general public by limiting the number of licensed professionals who come from underrepresented groups. The authors point out that the Spanish version of the EPPP, the S-EPPP, was implemented before it was validated, which led to a severe shortage of licensed professionals in Puerto Rico. They express concern that the validation process of the Part 2 – Skills exam is similarly insufficiently developed, particularly with respect to addressing risks of bias (Callahan et al., 2021).

Returning to the issue of the doctoral training years, the EPPP-2 has also been criticized for the proposed timing of the test within the doctorate training timeline. With the new exam, students would take the first part of the test, during their training years within their doctoral program, around the same time most students are preparing for compressive exams and dissertation proposal defenses. The addition of a high stakes exam at this crucial juncture adds a significant barrier that may undermine student success. Studying for that exam may also necessitate, by reality or perception, the purchase of expensive test preparation materials. Relatedly, the EPPP-2 exam will ultimately cost substantially more than the standard exam;

disproportionately disadvantaging students who have a high student debt load. Other concerns raised about the future of the EPPP-2 include lack of theoretical grounding, peer review, or evidence of external or incremental validity. Given these shortcomings, the undue burden on students of lower SES, in particular, seems inappropriate. Callahan and colleagues conclude that Licensure jurisdictions not adopt the EPPP-2 until information on intended and unintended referents are made available. Essentially, they contend that the EPPP-2 should demonstrate that it reliably measures competence, not socioeconomic status, gender, or disability, prior to adoption.

Collectively, across Callahan and colleagues' (2018) study, the three associated commentaries (Borrego, 2018; Luebbe & Ogbaselase, 2018; Lund, 2018), and the recent Callahan and colleagues (2020) review of the Examination for the Professional Practice of Psychology, it appears that diversity is salient to (1) the evaluation of competency among marginalized individuals across the entire training to workforce pipeline, and (2) the experience of training and supervision as a doctoral student. In the following sections the literature associated with each of those areas is reviewed to provide a foundation for the current study.

1.1   Evaluation of Competency

In a recent comprehensive review of empirical studies that examined foundational components of graduate admissions, research and training among doctoral health services psychology programs, Callahan and Watkins' (2018a) noted that most programs' admissions processes place significant emphasis on quantitative scores on the GRE. Such an emphasis is potentially problematic for several reasons, but the primary concern that is salient to the current study is that the GRE has been consistently linked to poorer admission outcomes for underrepresented racial/ethnic minorities. Given that the field of psychology both underserves and underrepresents racial/ethnic minorities, this creates a strong likelihood for little to no

increase of diversification in the future. Graduate admissions faculty were specifically encouraged by the authors to consider more holistic measures when evaluating applicants. The authors posited that interviews were fairly accurate indicators of early graduate career success, and therefore training programs were encouraged to focus on how interviews are conducted, and leveraged, in the admissions process.

Callahan and Watkins (2018a) also posited multiculturally-focused curriculum and culture-centric interventions which help trainees explore and develop their cultural identities along with self-care as essential to the graduate training experience. They noted that students also tended to rate programs that provided these experiences more favorably overall. Programs which had higher success ratings placed stronger emphasis on multicultural interventions with clients and principle-based feedback. Variables such as trainee commitment to a research lab, strong relationships with mentors, and an overall collegial research environment were also found to be essential components in the research training process (Callahan & Watkins, 2018a).

More research is needed to determine which factors contribute to student perseverance. Specifically, future research must examine the experiences of students who identify as members of diverse groups and may need additional support. Overall, Callahan and Watkins' (2018a) research suggest that underrepresentation in the training to workforce pipeline was notable within each category of diversity examined, and that tools such as the GRE negatively contribute to lower admission of marginalized students (African American and Black and/or Hispanic and Latino/a applicants, in particular). The authors assert that holistic reviews of admissions materials, paired with interviews, and full disclosure about the graduate training process (e.g., costs, time commitments, expectations, etc.) are essential factors to improving graduate psychology admissions, curriculum, and research training.

Roberts and Ostreko (2018) reviewed Callahan and colleagues' (2018) empirical analysis of psychology doctoral trainee population data, and identified several potential barriers to admission of diverse applicants for doctoral study in professional psychology. Specifically, the authors assert that since the GRE has a longstanding, well-documented history of bias toward marginalized groups, less emphasis (if any) should be placed on these scores. Furthermore, the authors suggest programs remove public postings of GRE scores from their admissions sites/materials because they may adversely affect minority-group applicants' decisions to apply at all. Roberts and Ostreko strongly recommend graduate programs remain aware of their commitment to inclusion and diversity and demonstrate that commitment by breaking with the incongruent practice of posting GRE scores publicly. To create a uniform admissions process founded in holistic review, programs are encouraged to develop rubrics and metrics of what faculty believe to be essential to graduate training and early career success. Qualities such as perseverance, leadership qualities, professional experiences, creativity, critical thinking, problem solving, and how one responds to adversity were among those mentioned. The authors noted such holistic reviews require more time and attention from faculty and staff, but yield more in-depth insights into applicant readiness for doctoral training.

Aligned with Roberts and Ostreko's (2018) conceptualization of GRE scores as only one piece of data that must be appreciated as part of a larger context, Collins, Callahan and Klonoff (2007) outlined a developmental model of the emerging professional training sequence and identified three necessary steps thought to scaffold competency development during the doctoral program training years. Those included: trainee characteristics, doctoral program opportunities, and the pre-doctoral internship year—each of which was conceptualized as building upon one another, culminating in the competency of the emerging professional.

Trainee characteristics were categorized as encompassing the personal aspects of a students' educational experiences and their ability to impact training outcomes. Goals and expectations were also an essential characteristic, though described as likely changing over time. The review of the second step, doctoral programs, elicited recommendations that doctoral programs be held accountable for and publicly disclose their match rates. The internship year, which serves as a capstone to graduate coursework, was measured based on its ability to provide students with entry-level competency. The writers concluded that the role of emerging professional was attained by meeting benchmarks assigned by faculty assessments of trainees' emerging competency. The authors concluded by advocating for empirical study aimed specifically at assessing the determinants of competency development. Given that trainees' individual characteristics and levels of competency vary, more research is needed to effectively and uniformly evaluate graduate students' competency.

In a comprehensive review of the literature, Callahan and Watkins (2018b), examined supervision and competency within health services psychology doctoral training programs. The current literature suggests the majority of professional growth occurs during graduate training; therefore, quality training is important for the future of the field. Competency guidelines are in place in order to protect the public and should be implemented using an evidence-based approach. Similarly, these competencies are of special concern to doctoral trainees in the process of applying to internship. The literature suggests that faculty should be mindful of informing trainees of the internship application process, providing emotional support, and prompt completion of recommendation letters, as these behaviors reduce anxiety about internship (Callahan & Watkins, 2018b).

Burn out is common for doctoral students, therefore focusing on self-care is also an

important focal point in supervision. Self-care improves client outcomes and should be included in the foundation of health service professional training. Individuals who matched for internship differed from those who did not on measures of peer-reviewed publications/presentations, practicum training experiences, number of integrated reports and previous experience within varied clinical settings. The data suggest students from the Northeast region of the country had the lowest match rate followed by the West, Midwest and South. Continued use of evidence-based multicultural training, conference and workshop attendance, self-study presentations, and support for publications are all essential to professional development and successful internship matches.

Subsequent work by the same authors reported that GRE scores were not significantly or meaningfully associated with a major doctoral training outcome: internship match. If neither GRE scores nor accrued clinical hours are good proxies for trainee competency, what measures might be more useful? Only within the past few years has empirical research on professional-wide competencies begun to emerge. In their 2017 quantitative analysis, Price, Callahan and Cox, compared 270 competency ratings across different training levels, training programs, and supervisors. Using Rasch analysis, the reliability for trainees (.98) and competency items (.92) were both good. The authors suggest using standardized practicum evaluation forms across programs, trainees, and specialties and as a means of providing structured supervisory feedback to supervisees. Price, Callahan, and Cox (2017) found that students developed the most competency during the transition from pre-practicum to internal practicum placement. The authors suggest the foundational competency factor of the benchmarks be used prior to field experience, and functional competency be rated during field experience. Peer supervisors and supervisors rated students significantly similarly ($r = .84$; *large*; $p = .04$) unless the peers were at

different competency levels. Since peer supervision dyads are rarely set up to facilitate matching peers of similar competency levels, the authors caution against routine use of 360° evaluations as summative assessments of student competency.

Callahan, Neumann, Cox, and Ruggero's (2017) quantitative analysis examined the archival training records of 38 trainees spanning a 10-year period in order to evaluate the psychometric properties of the Supervisor Trainee Quarterly Evaluation (STQE) (2017). The STQE is a psychometric measure of trainees' competency that also serves to predict the likelihood of trainees meeting essential benchmarks prior to internship. Factor analysis reveal significant ($p = < .01 - .001$) loading on general competency with two second-level factors of functional competency and foundational competency. Both functional (coefficient alpha = .91) and foundational (coefficient alpha = .81) consistency demonstrated acceptable internal consistency. Total scores on the STQE significantly positively correlated with number of practicum intervention hours ($p = .001$) and supervision hours ($p = .03$).There was a significant correlation between the STQE total score and remediation and/or termination ($r_{pb} = .35, p = .04$). Notably, the STQE did not correlate with number of assessment hours or internship match outcome prompting the authors to question the validity of clinical hours as indirect indicators of competency at the point of applying for internship. The authors assert the STQE could be used at a program level for self-study and as a means of identifying trainees with competency problems.

Gonsalvez's (2015) and colleagues examined supervisor ratings of clinical psychology trainees on the Clinical Psychology Practicum Competencies Rating Scale (CΨPRS) ($N = 204$). The primary objective of the study was to classify the rating scale items into sub clusters, clusters and super clusters which were used to create a hierarchical statistical technique to determine the relative proximities of the relationship between the items. The data were divvied

into two sets, each with 8 or 9 domains designed to assess trainees' knowledge, skills and attitudes about concepts such as ethics, personal capacities, scientist practitioner competencies, and response to supervision. A main analysis of the ANOVA X Hierarchical Placement Cluster technique yielded results relatively stable across Data Sets I and II, and across the split-case analyses.

In a forthcoming update to his programmatic line of inquiry into competency assessment Gonsalvez, along with Deane, Terry, and Gooi (online first), attempted to infer the composite structure of competency as a construct. In this forthcoming work, items on the Clinical Psychology Practicum Competencies Rating Scale were presented to supervisors in a fixed order to capture their evaluation of 195 trainees. Subsequently, items were presented in a random order for use in evaluating 353 trainees. Results revealed that the hierarchical structure of competency varied to some extent depending upon whether items were presented in fixed versus random order. However, as pointed out in a forthcoming invited commentary by Borden and McIllvried (online first), the generalizability and implications are unclear. Callahan and Braggs (online first) also caution against over-interpreting Gonsalvez and colleagues recent findings noting multiple concerns about the statistical analyses performed and conclusions drawn. Beyond psychometric concerns though, Callahan and Braggs' commentary underscores the need for the field of professional psychology to uniformly implement evidence-based competency assessments. However, in order to bring this endeavor to fruition, the authors identified several necessary milestones that still need to be achieved. Perhaps the foremost need is for the field to adopt a concrete definition of competency (from both a theoretical and a functional lens). Based on the state of the science at this time, Callahan and Braggs recommend an intricate combination of clinical and supervision hours, in combination with other professional milestones, be used as

external validity markers. As a result, the authors encourage training programs to consider these milestones according to a specific timeline in an applied standardized fashion in terms of how they align with nascent competency assessments. The ability of a standardized measure of supervisor ratings to identify trainees with problems of professional competency (TPPC) may be a significant advancement, but such evaluations inherently rely on expert supervisor judgments, which may be biased.

1.2   Role of the Supervisor in Trainee Competency Development and Evaluation

Gonsalvez and Freestone (2007) were among the first to identify risk of bias. In this analysis of archival data, the researchers identified several trends in the assessment of clinical psychology trainees in practicum placement. The researchers analyzed supervisors' end-of-placement report, covering 11 dimensions of clinical skills across 291 field placements (encompassing 131 students), over a 12-year period. Overall, supervisors rated students more favorably than would be suggested by their academic grades ($p < .05$), suggesting the possibility of a leniency bias. Supervisors' level of experience did not affect the tendency towards leniency. The validity of these ratings was further questioned as supervisors' ratings at one site were not significantly associated with their ratings in a different setting. There was also little predictive power in the ratings from the beginning of training and ratings at the end of training.

Gonsalvez and Freestone (2007) posit that because supervisors have a role in being supportive and nurturing towards supervisees, they may be less likely to provide an objective judgement about trainee competency. Furthermore, relying upon trainee self-reports may significantly skew objective assessment of performance as well. The authors recommend using recorded sessions to facilitate evaluation by more than one supervisor in order to rely less on subjective reports of competency.

Unfortunately, there is emerging evidence that bias may be tied to individual differences or, possibly, educational background advantages. Although no studies were found with the specific discipline of psychology, educational background advantages have been demonstrated among medical students. More specifically, in a retrospective study, Craig, Gordon, Clarke and Oldmeadow (2009) compared 914 summative assessment results from medical school students at the University of Sydney Medical Program over a 3-year period from 1999 through 2002. The distribution of prior degrees among admitted students fell across 5 categories: biomedical science (54%), other biological science (7%), physical science (6%), health professional (20%), or non-science (12%).

Students from non-science backgrounds, who were notably more demographically diverse than students in the other prior degree groups, were significantly (almost 3 times) more likely to fail the second-year barrier assessment. At the point of the third-year assessment, these students were no longer at elevated risk, suggesting that knowledge gaps tied to their prior degree area had closed effectively. In contrast, students from biomedical backgrounds were twice as likely to fail the third-year assessment, suggesting that knowledge gained as a result of their prior degree may not have afforded any lasting advantage in medical school. Those from health service backgrounds appeared to perform the most consistently, evidencing less risk of failing exams throughout their education ($OR = .39$; 95% CI $= 0.12, 1.28$; $p < 0.05$).

The authors concluded that efforts to demographically diversify the medical profession by admitting students from a broader range of degree backgrounds was resulting in uneven preparation for success in the early years of training. Underscoring the ultimate success of these students as they progressed further in their training, the authors noted that the use of summative assessments in early years of the program may disadvantage students from non-science

backgrounds. They noted that single best answer exam formats in the early years of training, in particular, advantage those with prior biomedical science degrees. As a result, Craig and colleagues (2009) advocate for additional input and support to aide students from non-science degree backgrounds in developing their knowledge base so that they are not disadvantaged early in training.

Within health services psychology, there is some evidence of negative supervisory biases that are tied to individual differences. In a 2001 qualitative analysis, researchers Chung, Marshal, and Gordon conducted a study to evaluate whether racial or/ and gender bias towards supervisees occurred between cross-cultural supervisors and minoritized groups (i.e. racial/ethnic people, women). They recruited 76 counseling professionals ($N = 106$) from diverse backgrounds, degree levels, work settings, and geographical locations with experience supervising trainees to anonymously complete the Supervisee Evaluation Scale (SES). Among the participants, 61% were female, and with 12% self-identifying as African American and Black, 8% as Asian, 1% as Native American, and 78% as white, and 1% as other. The SES asked supervisors to rate their supervisees on 12 items: coverage of information, style of writing, understanding of the client's current functioning, understanding of how the client's problems evolved, case conceptualization, theoretical foundation of conceptualization, awareness of dynamics in the counseling relationship, counselor impairment due to counter-transference, treatment provided, treatment plan, cultural sensitivity, and overall evaluation. The data suggested that female trainees with male supervisors were more likely to receive negative competency evaluations. Conversely, the data did not reflect racial bias within racially dissimilar supervisor-trainee dyads.

Training directors appear to have some awareness that trainee demographics may interact with supervisor competency in some instances. In a qualitative study, researchers Shen Miller,

Forrest, and Elman (2009), coded the semi-structured interviews of 14 training directors from counseling psychology doctoral programs for a grounded theory comparative analysis focused on student remediation. The 14 training directors described 47 separate cases of trainee remediation. Transcripts were coded for: definitional clarity, conceptual sophistication about race and ethnicity, differential integration of approaches to diversity, remediation, and strong emotions associated with race/ethnicity in trainee remediation.

Findings revealed a range of diversity considerations that impacted remediation plans of students with professional competence problems. Four training directors cited gender norms and assumptions as impacting male trainees' ability to reveal their vulnerable emotions within training. Another hypothesized that professional norms are based on female development and therefore enforce the gender imbalance of the field. One respondent reported feeling discomfort discussing a trainee's eating disorder in supervision due to the fear the topic may be viewed as sexual harassment in the context of the supervisor/supervisee gender mismatch. Of particular note, the authors observed that training directors who cited literature in their answers were more likely to do so in reference to gender considerations, rather than race/ethnicity (Shen Miller, Forrest, & Elman, 2009).

In fact, they observed that training directors consistently reported difficulty reconciling issues of professional competence with cultural differences. Training directors cited fear of litigation as a reason for not addressing diversity factors with trainees exhibiting problems of professional competence. More specifically, white, non-Hispanic training directors reported discomfort in making culturally specific recommendations due to not sharing the same cultural background. A common way in which these training directors attempted to reconcile their own discomfort was to focus on faculty competence and communication about multicultural

competence with trainees. Given such discomfort, the study authors asserted that lower numbers

of ethnic minorities in the field may limit opportunities to discuss race and ethnicity (Shen

Miller, Forrest, & Elman, 2009). They suggested using racial and ethnic identity development

models within the context of supervision as a means of improving conversations around these

issues with trainees identified as having problems of professional competence. Implementation

of such models allow supervisors and supervisees the opportunity to engage in culturally

appropriate dialogues about the potential impact of race and ethnicity on the supervision dyad.

Inadequate supervision competency among contemporary supervisors may, in part,

reflect poor historical attention to training in supervision competencies themselves. Scott,

Ingram, Vitanza, and Smith's (2000) quantitative survey of training directors from APA

accredited psychology programs ($n = 123$) and pre-doctoral internship sites ($n = 209$), examined

the extent to which students were trained in conducting supervision. Half of all academic

programs surveyed offered supervision training as an elective or a practicum experience in

supervision. In 90% of these programs training was provided by core faculty. Reasons cited by

programs for not offering courses or practicum in supervision were full student course loads,

budget constraints, and the belief that this training should occur during internship. Notably,

counseling programs emphasized didactic instruction in supervision significantly more than

clinical programs and were more likely to have proficiency standards for supervision χ^2 ($2, N =$

$120) = 13.90, p < .001.$

Of the 209 internship sites surveyed, 39% offered a didactic seminar in supervision and

about half of all programs either offered or required supervision experience. On average, among

interns who obtained supervision experience, interns supervised 2.45 trainees over 7 months.

There were some differences reported as a function of the internship setting. More specifically,

university counseling centers provided significantly more opportunities for didactic seminars and practice in supervision than other program types, χ^2 (2, $N = 121$) = 27.24, $p < .001$. These programs were also more likely to use formal measures of evaluating supervision proficiency.

Perhaps reflecting diffusion of responsibility, academic programs were more likely to report that supervision should be taught during internship, while internship programs indicated that this training should occur prior to internship. Based on their obtained data, Scott and colleagues (2009) concluded that supervision should be taught and practiced in academic programs and implemented with a standardized evaluation method. The authors further suggested that if costs were prohibitive for programs, they consider partnering with para-professional or related helping professional training programs to create supervision practicum experiences. Further evidence suggesting poor supervisory competency around the role of trainee diversity and multiculturalism in supervision is suggested by a few quantitative studies in health service psychology.

Even before Shen Miller, Forrest, and Elman (2009) documented training directors qualitatively conceptualizing trainee diversity and competency evaluations by supervisors as intersecting meaningfully, Gatmon and colleagues (2001) sought to understand the impact of supervisee and supervisor match on a few individual diversity variables. Their study, surveyed 289 pre-doctoral psychology interns from APA accredited internship sites to examine the relationship between supervisor-supervisee congruence on variables of ethnicity, gender, and sexual orientation, working alliance and satisfaction with supervision. Overall, participants reported a low frequency of multicultural discussion within supervision. Supervisee/supervisor demographic match or non-match did not significantly impact supervision satisfaction or working alliance ratings. However, supervision satisfaction and working alliance were

significantly positively correlated ($r = .59$, $p < .01$) with quality, frequency, depth, and feelings of safety in discussion about issues of ethnicity, gender, and sexual orientation. The authors consequently suggested that supervisors routinely initiate discussions of diversity issues, as this improves working alliance. They also recommended using training in multicultural supervision as a framework for addressing and discussing multicultural issues within the context of supervision. The authors emphasize the importance of creating a safe supervision environment with a supervisor that frequently initiates in-depth, quality conversations about multicultural issues.

In a small study examining the relationship between supervisor/supervisee racial match discussion of multicultural issues within supervision, Jernigan, Green, Helms, Perez-Gualdron, and Henze (2009) surveyed 15 graduate students from a large urban university on the east coast. Six respondents, who identified as People of Color, provided data for content analysis of their responses to open ended questions. Using Helms' model of racial identity, which consists of 6 levels of racial identity development, content analysis indicated a regressive supervision relationship when the supervisee was at least one level of racial identity development above the supervisor (Helms, 1995). Supervisees of Color reported bringing up racial and cultural topics more often than their supervisors of Color, and reported feeling burdened by the responsibility of having to educate their supervisors. This disparity led supervisees to feel unsupported, confused, frustrated, and discouraged.

In contrast, supervision relationships in which the supervisor had a more advanced level of racial identity development fostered supervision environments in which the supervisee could explore, and reflect on personal experiences, beliefs and values related to racial and cultural perspectives. In these relationships the dyad could focus on the ways in which racial and cultural

dynamics were present in the working alliance. One participant noted that her supervisor used assigned readings to introduce cultural topics. The authors conclude that though supervisors and supervisees may both be People of Color, comfort in discussing multicultural issues are dependent on the distribution of racial identity development in the dyad. Supervisors and supervisees should be careful not to assume that either is an expert on their race. Within training, supervisees should be in an environment in which they are able to process racial and cultural information with multiple professionals and mentors.

The literature offers some guidance for incorporating cultural supervision practices and addressing value conflicts that may emerge between supervisees and their supervisors. With respect to cultural supervision practices, Williams and Raney's (2020) qualitative analysis of relational cultural theory (RCT) outlines a viable framework for enhanced clinical supervision for trainees and postdoctoral residents alike. RCT is both a therapeutic approach and a theory. It places great emphasis on the connections and disconnections in a client's life, and how they interact upon the client's ability to function within roles and relationships (2020). With respect to this study, the authors made several parallels to RCT and competency evaluation.

Though RCT has traditionally been applied to therapist-client dyads, the authors suggest it is an effective means of building the supervisor-supervisee relationship. RCT is unique for several reasons. Specifically, the emphasis on mutual empathy and relational authenticity between both parties. When applied consistently, this framework serves to create myriad opportunities for mutual learning and respect. Both trainees and early career professionals are tasked with meeting professional development milestones which are then subject to supervisor evaluations. The authors' illustrations of RCT in practice elucidated the anxiety-reducing potential of this method of supervision. Not only does RCT decolonize the training process, it

also fosters growth, self-confidence, and improved perceptions of clinical abilities (Williams & Raney, 2020). Overall, Williams and Raney outline a new framework for managing critical incidents both between therapists and clients, as well as supervisors and supervisees.

With respect to addressing value conflicts in the supervision dyad, Dunn, Callahan, Farnsworth, and Watkins' (2017) framework proposal modified an existing model of understanding value differences between clients and clinicians for use in the supervisory relationship. They posit that value differences are likely to occur in the relationship between supervisors and supervisees, making a framework for discussing these differences necessary. In their framework, the supervisor is conceptualized as having the onus of responsibility to (1) detect, (2) articulate and (3) respond to value conflicts. The authors emphasize the importance of power dynamics within these discussions and encourage supervisors to proactively discuss value conflicts with empathy.

Value conflicts are further categorized, with recommendations made for each category of conflict. For instance, the authors postulate that preemptive conflict exists prior to the working alliance and occur when the value itself is the source of conflict. In these cases, Dunn and colleagues suggest that supervisees be assigned a different supervisor. When the conflict in values is adjacent to the client's presenting problem, outcome tracking, focused supervision and avoidance of overinterpretation is recommended (Dunn et al., 2017). When both supervisor and supervisee share the same values but disagree on implementing them, an operational values conflict exists. In this case, the authors suggest clients be presented with all recommendations to guide the process based on the client's preferences and values (2017). Unarticulated value conflicts should be monitored by focused supervision, outcome tracking, and education to determine whether the relationship should terminate, continue, or further focus discussion on the

role of value conflict in a supervisory working alliance.

1.3   Trainees with Problems of Professional Competence

Emphasis on competency as a training goal has developed slowly across the past 30 years. To best appreciate the arc of empirical studies in that area over time it may be useful to consider studies based on the order in which they appeared in the literature. The first quantitative study located, by Mearns and Allen (1991), surveyed 29 faculty members and 73 graduate students' experiential and affective experiences of dealing with impaired peers. Graduate students were asked to complete a 3-part, 40-item self-report survey assessing their program climate, experiences with impaired peers, and their responses to known impairments and improprieties (i.e. ethical lapses). Faculty were asked to predict graduate students' responses, as well as to rate their perception of their peers' level of response to impairment. Approximately 95% of graduate student respondents ($n = 69$) reported being aware of a peer's impaired functioning, while 49% ($n = 36$) indicated an awareness of impropriety. A Pearson's Rho, $r(44) = .53$, $p < .0001$, revealed a moderate positive correlation between faculty and students' impairment rankings, but revealed only a small positive correlation with respect to impropriety, $r(44) = .29$, $p < .05$.

Mearns and Allen's (1991) research suggests faculty tend to variably under- and over-estimate students' emotional and behavioral responses. Though faculty predicted that students would report worry, disloyalty to perpetrators, and apprehension when faced with responding to peers' impairment, students actually reported feeling angry, conflicted and frustrated. Similarly, students reported feeling angry, conflicted, and justified when responding to peers' impropriety (i. e. ethical infractions, substance abuse, academic dishonesty), while faculty predicted students would feel angry, apprehensive, and disloyal to their peers. Importantly, faculty generally

overestimated students' willingness and ability to take action despite negative affective responses to impairment and impropriety.

In contrast, faculty and students shared similar perspectives about responsibility. Students tended to feel responsible for addressing impairment and impropriety with peers, though they reported faculty did not appear to put forth enough effort to do so themselves. Faculty and students both agreed that faculty were significantly more obligated to confront issues of impairment and impropriety with graduate students, however, students perceived faculty as being less involved than they should be. In contrast, faculty rated their colleagues as significantly more involved in screening efforts than students, despite both groups agreeing it is primarily faculty's responsibility to vet potential future clinicians.

Although dated, the data in Mearns and Allen's study suggest graduate students' tendency to respond to the impairment and impropriety of peers may increase if they believe they can make a difference. Training programs and faculty must remain aware of how they can respond to student concerns through direct actions while empowering students to do the same. Settings such as focus groups, experiential ethics courses, and involving students in evaluation processes (via inclusion of formal reports) are integral parts of the training process. Additionally, Mearns and Allen (1991) reported faculty can provide better training and support to student trainees by using a coping versus mastery modelling of problem solving. Direct observation opportunities and other experiential activities are integral components of clinical training; therefore, faculty are better able to serve students through effective modelling in the midst of ethical dilemmas.

In a quantitative study that appeared fairly early, Yourman and Farber (1996) surveyed doctoral students ($N = 93$) enrolled in clinical psychology programs in the New York

Metropolitan area. Participants were asked about the extent to which they do not disclose information during supervision to examine if there were any predictive variables for nondisclosure (see Table 1). Nondisclosure was defined in that study as omitting information or suppressing comments/feelings about the events of sessions, clients, and/or supervisor, or ignoring instructions. Using a multiple regression analysis, a significant positive relationship was found between supervisee satisfaction and nondisclosure during supervision ($F(11, 74) = 12.57$, $p < .001$). Further, the discussion of countertransference was significantly positively correlated with lower rates of nondisclosure ($F(11, 74) = 8.69$, $p < .005$) and supervisee satisfaction ($F(10, 75) = 11.00$, $p < .001$). Additionally, frequent discussion about countertransference during supervision, predicted less non-disclosure among supervisees and overall satisfaction with supervision.

Table 1

Selected Variables' Relationship to Supervisee Nondisclosure and Supervisee Satisfaction

Variable	Non-disclosure[a] $F (11, 74)$	Direction	Satisfaction[b] $F (10, 75)$	Direction
Supervisee satisfaction[b]	12.57**	+		
Discuss countertransference[b]	8.69*	+	11.00**	+
Supervisee gender[c]	3.01	+	0.69	-
Supervisee age	1.02	+	0.38	-
Year in program	0.96	+	0.51	+
Video/audio taping of therapy sessions[b]	0.80	+	1.13	+
Supervisor gender[d]	0.31	+	2.06	+
Minority status	0.23	+	0.84	+
Supervisor-supervisee gender match	0.14	-	2.41	-
Supervisor treats supervision like therapy[b]	0.08	+	8.45*	-
Orientation match	0.05	-	1.82	+

Source: Yourman & Farber (1996). Note. A multiple regression model was used where all variables were entered simultaneously. [a] As represented by total scores. The higher the total score the less supervisee nondisclosure and distortion is being reported. [b] Scored on a 7-point Likert type scale where 1 = never, 4 = sometimes, and 7 = always. [c] Female trainees had higher total scores and were less frequently satisfied. [d] Trainees with female supervisors had higher total scores and were more frequently satisfied. * $p \leq .005$, ** $p \leq .001$.

Overall, the findings by Yourman and Farber (1996) suggest supervisees, in general, are honest and forthcoming within the context of supervision and generally consider supervision to be a satisfying experience. Yet, some level of withholding information about perceived clinical errors was common and nearly half of the participants reported a high frequency of telling their supervisor what they wanted to hear. A number of specific suggestions emerged from the data. For example, supervision relationships that mimicked psychotherapy were associated with less satisfaction with supervision suggesting supervisors may do well to avoid creating or otherwise fostering such dynamics. Additionally, Yourman and Farber suggest that discussing countertransference within the context of supervision may reduce instances of non-disclosure among trainees and improve their satisfaction with supervision.

Gaubatz and Vera (2002) conducted the first known quantitative study in the closely related field of counseling education that examined the prevalence of "deficient" students (p. 296). An 11-item survey was mailed to faculty members in counseling training programs across the United States. Of the 118 faculty respondents, which represented 85% of programs approached for study participation, there were a wide range of program-dependent responses to the research questions. Some programs reported graduating 1% of students who were unqualified and other programs reporting up to 75% of graduates who were unqualified to practice. Programs that routinely engaged in formal review processes to evaluate trainees were associated with better gatekeeping and more likely use of remediation efforts. A significant inverse correlation was found between formal review procedures and reports of graduating unqualified professionals ($r = -.44$, $p < .0001$), yielding a medium effect size.

This relationship was stronger for non-accredited programs. Faculty at accredited universities reported more formal review procedures than non-accredited universities. The size of

the university did not play a role in remediation rates. The authors of this study suggest that formal gatekeeping procedures should be included in professional training in order to graduate well-qualified health service professionals and to prevent graduating unqualified professionals. Within their sample, 19% of the variance in "deficient" students being advanced in the program anyway was tied to whether or not the program used formalized gatekeeping evaluations. Some limitations to gatekeeping procedures identified in this study were concerns about receiving poor teaching evaluations (expressed primarily by untenured or adjunct faculty), frequent faculty turnover due to a high number of adjunct staff and fear of civil lawsuits.

Vacha-Haase, Davenport, and Kerewsky (2004) examined the gatekeeping practices of APA-accredited doctoral programs across the United States ($n = 103$), training directors were asked to share their insights about their programs' practices regarding guidance and intervention for problematic or impaired students (Vacha-Haase, Davenport, & Kerewsky, 2004). Participants were surveyed on program demographics, policies, and procedures for problematic student intervention. The survey, which consisted of questions compiled from a variety of surveys and feedback from faculty, graduate students, and training directors. Though the survey yielded a 37% response rate, it was the largest of that kind, at that time.

The study revealed the average program had an average of 48 graduate students ($SD = 28.06$), with the majority being housed in the psychology department (65%). Training director reports indicated that inadequate clinical skills (65%), defensiveness in supervision (52%), and deficient interpersonal skills (42%) were the most frequently observed problematic behaviors for impaired graduate students. Furthermore, the training directors reported using increased supervision (49%), leave of absence (47%), repetition of practicum (39%), and extra classwork or practicum experience (39%) as remediation in lieu of termination. The authors also noted

training directors citing personal counseling (60%) as the most likely recommendation for impaired graduate students, despite lacking empirical evidence of its efficacy. These findings suggest over half of graduate programs terminated at least 1 ($M = 1.11$, $SD = 5.3$) graduate student per 3-year period, resulting in 114 terminations across the programs surveyed. Of the terminated graduate students, inadequate clinical skills ($n = 22$) was the most frequently cited cause. Overall, these findings suggest training directors must work together to help their programs to reach a consensus on the terminology used to describe and define problematic student behaviors, in order to ascertain prevalence and identify viable remediation measures, including termination.

In an exploratory mixed methods study by Oliver, Bernstein, Anderson, Blashfield, and Roberts (2004), researchers surveyed doctoral students ($N = 28$) who were enrolled in clinical psychology programs throughout the United States. Participants were asked to share their perspectives on their programs' management of student impairments that cause significant interferences with graduate students' professional functioning. For the purpose of this study, student impairment was defined as, "graduate students who experience difficulties associated with interpersonal problems, substance abuse, Axis II disorders, and so forth and are consequently unable to meet the expectations of their programs." Participants were provided with a list of open-response questions, which yielded 50 pages of feedback notes.

Using phenomenological coding, the study revealed that the average participants' program had 50.6 students ($SD = 44.57$), though the range was very wide (25 – 225). All participants were from scientist-practitioner programs with membership in the Council of University Directors of Clinical Psychology (CUDCP) professional organization. Within this sample, themes revealed that 53% of students were unsure whether their programs had policies

pertaining to trainee impairment. While 33% reported knowing such policies existed, 4% characterized them as vague. Additionally, participants estimated that roughly 12% of their peers were impaired. At least within this sample, graduate students appear to be sensitive to their peers' real or perceived impairment. As a result, the authors suggested that graduate programs establish formal standards and procedures to address student impairment, provide supportive remediation, and clearly define student impairment in an ethical and empathic manner.

In a mixed methods exploratory pilot study, Rosenberg, Getzelman, Arcinue, and Oren, (2005) distributed a total of 325 surveys to current students and reported a 40% return rate. Respondents ($n = 129$) were asked answer 25 questions ranking how a problematic peer affected them emotionally, clinically, and/or academically. Results revealed that 109, or roughly 85% of participant-trainees believed a problematic peer had significantly affected their learning capabilities and emotional wellbeing. Similar to other studies of problematic students (Mearns & Allen, 1991), the mean number of students who were identified as problematic peers by any one respondent was 3.32 ($SD = 3.15$). Of these, 55% of participants reported becoming angry at faculty for not directly addressing the problem with the problematic peer. The authors identified five specific areas of concern: trainee awareness of problematic peers, trainee perspectives of who was responsible for addressing problematic peers, learning the types of problems encountered, learning which actions students took to address their concerns, how problematic peers affected respondents' overall functioning, and best practices for negotiating concerns about peers. These results not only underscore areas for future research but also highlight the impact a problematic trainee has within their department and on their peers (e.g., evidencing problematic symptoms, impaired personal functioning, and lagging professional competency and knowledge acquisition).

The year 2012 saw a brief explosion of publications. El-Ghoroury, Galper, Sawaqdeh, and Bufka (2012) examined barriers to graduate students' coping strategies among a geographically diverse sample of 387 psychology graduate students. Within their sample, over 70% of participants reported stressors that negatively impacted their functioning. Common stressors in this sample included: academic responsibilities, finances, anxiety, and poor work/school/life balance. Students reported coping with support from friends, family, and classmates. Among participants, barriers to help-seeking included lack of time as well as financial constraints. Notably, using both univariate and multivariate analyses, the researchers found that trainee ethnicity/race was significantly associated with stressors and coping strategies, as well as barriers to wellness. More specifically, minority students were more likely (than white students) to report experiences of discrimination among their sources of stress. Although minority students were significantly more likely to report spiritual practices as a facilitative coping strategy, they reported significantly more barriers to engaging in wellness activities (e.g., cost, lack of time, confidentiality concerns).

In their qualitative analysis, researchers Shen Miller, Forrest and Burt (2012) conducted phone interviews with 22 faculty and training directors from counseling psychology programs across the United States. The researchers explored the relationship between intersectional diversity and trainees with problems of professional competence. Specifically, they asked participants "to what extent have issues of diversity played a role in a student being identified and/or placed on remediation?" In this study, "diversity" was presented as including race, ethnicity, social class, gender, sexual orientation, nationality and religion. The researchers then codified responses and categorized them according to grounded theory methodologies.

Findings in this study suggest that faculty believed discussing students with competency

issues would invite uncomfortable discussion about diversity among staff. Participants reported avoiding topics of students' competence problems due to fear of other faculty reactions, or being perceived as using the trainee's issues to address existing program conflicts. Similarly, participants described tense faculty discussions of students' skill levels and diversity factors. Participants also felt these discussions often reflected the variable approaches to diversity while underscoring the differing standards for multicultural competence among faculty.

In a 2012 study of 321 student affiliate members of the American Psychological Association, Shen Miller, Grus, Van Sickle, Schwartz-Mette, Elman, Jacobs and Kaslow assessed the experiences of health services psychology trainees with problems of professional competence (TPPC). Shen Miller and colleagues' quantitative study was perhaps seminal in yielding statistically-informed recommendations for remediation in training. More than half of participants reported experiencing peers with problems of professional competence, while a small number of participants ($n = 19$) self-identified as a TPPC. While those with professional competence problems are only a small subset of trainees, their presence in professional programs are a source of complex problems.

Students who reported taking action after identifying a TPPC (i.e., consulting with faculty, a peer, or the individual with problems), believed this action only had a moderate effect and reported they would not take action in a future similar situation. Among actions taken, students ranked peer consultation as most helpful, talking with faculty second, and speaking to the individual directly as least helpful. Trainees were most likely to take action about a peer's problems with professional competence when the impairment was related to alcohol or substance use. Shen Miller et al.'s (2012) findings suggest that ongoing training throughout the graduate program was a best practice for supporting trainees with problems of professional competence.

Additionally, they suggest that trainees are provided with opportunities to provide feedback on the impact of peer impairment on their training environment. Adapting this practice would allow trainees and supervisors to improve their working alliance.

To explore perceptions of problem students within their academic program, Veilleux, January, VanderVeen, Reddy, and Klonoff (2012) collected survey data from a nationwide sample of 570 clinical psychology graduate students. Participants were asked to identify if they believed a peer was impaired and if they believed this peer had been identified by faculty. They were then asked if the impairment was an issue of "clinical impairment" or lack of characteristics required for the profession, or the result of "lack of clinical competence" in which issues are more likely to be resolved with additional training. The impact of impaired students on their peers was also assessed.

Participants identified individuals unsuited for the profession as lacking skills and traits such as critical thinking, ethical behavior, good communication and self-awareness. These individuals were seen as less sympathetic than those who were impaired due to circumstance. Participants who rated impaired students as having character trait deficits, also indicated a belief that these students would not improve with training. The presence of students who were impaired while remaining in graduate programs, was rated as contributing to a loss of faith in the program's faculty, reduction in perceived value of a doctoral degree, reduced ability to learn and elicitation of negative emotions. The authors recommend improving formal gatekeeping policies, informing students of these policies, and consistent procedural adherence to remediation of impaired students to ensure the integrity of the profession.

Mehr, Ladany, and Caskie's quantitative study of 201 health services psychology doctoral students assessed the relationship between trainee anxiety, trainee perception of

supervisory working alliance, counseling self-efficacy, and willingness to disclose in supervision (2015). They found that a higher-level of counseling self-efficacy and perceived supervisory working alliance significantly predicted less anxiety in supervision (ß = --.52, *p* < .001). The perception of a stronger working alliance was also found to be significantly positively associated with trainees' willingness to disclose within supervision (ß = .46, *p* < .001). These findings suggest a significant relationship between the degree of trainee willingness to disclose in supervision, their perception of the supervisory alliance, and anxiety level. This information may be useful in guiding disclosure during supervision.

Humphries, Crino, and Wilson's 2017 quantitative analysis of personality traits, psychological symptomology and clinical competence, surveyed 59 students from postgraduate degree programs in clinical or forensic psychology enrolled at an Australian university. In this sample, students generally rated themselves as having a level of anxiety, depression, extraversion, and agreeableness that was within the normal range. Of those sampled, 27% of students responded with at least one scale on a questionnaire in the problematic range. Forty percent of them endorsed symptomology in the problematic range on more than one questionnaire, with half reporting experiencing difficulties for an extended period of time. A clinical level of depression reduced the ability to acquire clinical competence, whereas conscientiousness positively predicted clinical competence; explaining one-third of variance in competency among participants.

There were a small number of participants who rated themselves as having significant psychological distress, but who had not been identified by faculty. The presence of this unseen distress in psychology programs highlights the need to address competency issues with students immediately after signs become present. These outcomes suggest improving trainees' self-

reflection, self-awareness, and self-care should be targeted within training programs.

Although not an empirical study, a very recent publication provides a snapshot view of contemporary thinking. Several illustrative examples, and a review of the literature, to highlight issues related to supervisor and trainee self-disclosure are provided by Boyle and Kenny (2020). The authors posited that as supervisors self-disclosed supervisees viewed their supervisors more positively and their supervisory working alliance improved. Furthermore, this improved supervisory working alliance was viewed as enhancing development and learning. As described more fully by the authors, supervisor self-disclosure can also be used to teach, build rapport and improve supervisee disclosure, particularly when a supervisee is struggling or having an emotional reaction to their client. In contrast, the vignettes provided by Boyle and Kenny (2020) also illustrate that if supervisor self-disclosure is dismissive, rather than normalizing of trainees' experiences, both learning and the working alliance are harmed. Self-disclosures that shifted the attention to the supervisor instead of the learning goals of the supervisee also tended to be more harmful than helpful. Finally, low levels of supervisor self-disclosure were linked to a poorer attachment relationship within supervision.

Boyle and Kenny (2020) suggest that trainees be encouraged to frequently discuss the supervision relationship and training goals with their supervisors in order to improve the working alliance and clarify the relationship. They also suggest validating vulnerability and handling uncomfortable feelings with sensitivity in order to combat shame within the supervisory relationship. This may be particularly important for supervisees who have a low tolerance for vulnerability. Such trainees may also struggle with supervisor self-disclosure, thus reducing the effectiveness of this behavior on normalizing the experience of the supervisee (Boyle & Kenny, 2020).

1.4   Current Study

The current study sought to provide follow-up to the internship match studies described in Callahan, Collins, and Klonoff (2010) and Callahan, Hogan, Klonoff, and Collins (2014). Across those studies, advanced doctoral (Ph.D.) students in clinical psychology who were enrolled in one of the member programs of the nationwide Council of University Directors of Clinical Psychology provided information about their training and experiences. As described more fully below, the gathered data was found to be predictive of success in attaining important required benchmarks (i.e., securing a competitive internship) for completion of the doctoral degree.

Callahan, Collins, and Klonoff's 2010 quantitative study of 330 applicants to the Association of Psychology Postdoctoral and Internship Centers (APPIC) match system sought to better understand factors that influence match outcome. In the sample, 96% were matched to an APA accredited internship. Applicants on average submitted 14.47 applications and were invited to an average of 7.81 interviews. The authors collected a large number of data points on each applicant surveyed including: debt load, program type, demographic data, academic and professional experience, interview offers were the primary determinant of match success. The number of applications submitted ($r = .41, p < .001$) had the largest correlation with interview offers. Other positive correlations with interview offers were significant but small, including peer reviewed publications ($r = .24, p < .001$), assessment and intervention hours ($r = .14, p = .014$) and completion of the dissertation proposal ($r_{pb} = .12, p = .042$). Furthermore, geographical restriction did not significantly impact match outcome, though these applicants tended to submit fewer applications, resulting in fewer interviews ($p = .003$). There was no significant difference

between applicants from Ph.D. programs and Psy.D. programs on match success, although Psy.D. students submitted significantly more applications ($p = .005$).

The following factors were found to not significantly impact the possibility of matching to internship: program type, dissertation proposal completion, number of peer reviewed publications, number of intervention assessment hours, number of integrated reports, number of supervision hours and number of applications. Similarly, there was no correlation between match success and demographic information including: sexual orientation, racial/ethnic status, disability status, citizenship status, student debt load or dependents. Though this study aimed to determine the specific factors that lead to a successful internship match, no such factors were identified. Based upon these findings, the authors suggest that students forgo purchasing any materials that may claim to increase internship match chances. They also propose that the burden is upon APPIC and doctorate programs to make the process more straightforward for students, more efficient for sites, and less anxiety-inducing for all involved.

The authors made several recommendations for reducing these obstructions, though some are now fairly dated. An example of that is their recommendation for implementing a lottery system for unmatched applicants, rather than the now defunct clearinghouse (Callahan, Collins, & Klonoff, 2009). The authors argued that the clearinghouse process was sufficiently chaotic that it was functionally already acting as a lottery for students but with the impression of it being based on applicant merits. The authors suggest this lottery could occur at the same time as the match so students would be told if they were matched, placed into the lottery, and the site of that lottery all at the same time. The goal of their recommendations was to reduce the emotional distress felt by students who were un-matched in the first round.

Callahan, Hogan, Klonoff and Collins' subsequent (2014) quantitative analysis surveyed

601 clinical psychology doctorate students, over a three-year period, to determine personality characteristics associated with internship match outcomes. Moderate, but significant (($r = .40$, $p < .001$) correlations were again found between number of applications submitted and interview offers. Weaker significant correlations were also found to impact interview offers including: intervention and assessment hours ($r = .23$, $p < .001$), total number of publications ($r = .18$, $p < .001$), and number of supervision hours ($r = .16$, $p = < .001$). On a trait level of the big 5 personality domains, no significant correlations were found in regard to interview offers. The trait, Assertiveness, neared significance, with a positive trend towards improving internship offers ($r_{pb} = .09$, $p = .05$). Within the Big 5 traits several facets of personality demonstrated small associations. Artistic Interests ($r_{pb} = .11$; $p = .02$), Self-Efficacy ($r_{pb} = .11$, $p = .02$), Friendliness ($r_{pb} = .10$, $p = .02$), and Agreeableness ($r = .09$, $p = .045$) all significantly correlated with interview offers.

Using a hierarchical logistic regression, Callahan et al. (2014) demonstrated that demographic variables such as marital status, geographic restriction, and type of degree, only accounted for .05% of the variance in match outcome. The personality facets that demonstrated significance (Trust, Cooperation, Sympathy, Orderliness, Artistic Interest, Self-Efficacy and Friendliness), collectively accounted for 7.35% of variance, while 28.8% of variance was explained by the number of interviews offered. The authors found that those offered 7 or more interviews have a 2.5 times greater likelihood of being matched than those offered fewer interviews. Overall, Callahan et al. (2014) recommended that faculty mentor students to present themselves as friendly, assertive, self-efficient during interviews. They are also encouraged to remember that personality traits describe little variance in the match process, therefore applicants should be themselves in the interview process.

Supplemental analyses from these larger 2010 and 2014 match studies were subsequently published in by Hogan and colleagues (2014) and Callahan and colleagues (2014). In Callahan, Swift, Hogan, Tompkins, Connor, and Klonoff (2014), supplemental analyses provided an empirical examination of applicant differences according to the internship setting where they ultimately matched. A subsequent hierarchical analysis focused on changes in subjective well-being among internship applicants as a function of their match outcome (Hogan, Callahan, Tompkins, Swift, Connor, & Klonoff, 2014). Hogan's examination of the phenomenological experiences of unmatched trainees revealed that unmatched trainees experience significant amounts of shame, discomfort, reduced moral, and self-isolation. Moreover, they found that unmatched applicants' subjective well-being (SWB) ($M = 10.96$, $SD = 2.91$) was significantly less than those who secured an internship ($M = 13.67$, $SD = 2.42$, $t[436] = -7.02$, $p < .001$, $d = 1.01$). By examining these variables, clinical training programs may provide greater support to assist trainees with identifying areas of improvement with regard to competencies that may facilitate improved match outcomes.

1.5   Research Questions and Hypotheses

The current study builds upon the wealth of data gathered across the two larger internship match studies (Callahan et al., 2010; Callahan et al., 2014) and the corresponding supplemental analyses (Callahan et al., 2014; Hogan et al., 2014) from those data. The aim of the current study was to determine whether doctoral training data gathered from students was ultimately predictive of subsequent early career competencies (e.g., clinical skills), professional activities (e.g., involvement with research), and/or attainment of benchmarks (e.g., licensure). Based on the literature reviewed and the results of previous research (Callahan et al, 2014; Callahan et al., 2018; Hogan et al, 2014), the following research questions and hypotheses were proposed:

Research Question 1: Is there evidence of bias in EPPP scores?

H^0: Scores on the professional licensing exam (EPPP) will not evidence significant group differences related to race/ethnicity or economic privilege (as measured by SES and accrued student debt).

H^1: Underrepresented minorities (African American and Black; Hispanic and Latino/a individuals) will obtain significantly lower scores on the EPPP than white, non-Hispanic individuals.

H^2: Socio-economic status growing up (SES) will systematically vary with scores on the EPPP, with higher EPPP scores associated with backgrounds of greater privilege (i.e., higher SES) and lower accrued student debt.

Research Question 2: Are scores on the EPPP meaningfully related to pre-internship preparatory training?

H^0: Scores on the EPPP will not significantly correlate with the number of accrued pre-internship hours in assessment/intervention services, supervision hours, completed integrated reports, scientific presentations, or peer-reviewed publications.

H^1: Scores on the EPPP will be significantly positively correlated (single-tail test) with the number of accrued pre-internship hours in assessment/intervention services.

H^2: Scores on the EPPP will be significantly positively correlated (single-tail test) with the number of accrued pre-internship supervision hours.

H^3: Scores on the EPPP will be significantly positively correlated (single-tail test) with the number of integrated reports that were completed prior to applying for internship.

H^4: Scores on the EPPP will be significantly positively correlated (single-tail test) with the number of scientific presentations accomplished prior to applying for internship.

H^5: Scores on the EPPP will be significantly positively correlated (single-tail test) with the number of peer review publications at the time of applying for internship.

H^6: Scores on the EPPP will be predicted by the number of accrued pre-internship hours in assessment/intervention services, supervision hours, completed integrated reports, scientific presentations, and/or peer-reviewed publications. To test this hypothesis, forward stepwise regression will be used to examine the contribution of each variable to the explained variance in EPPP scores.

Research Question 3: Are scores on the EPPP meaningfully related to internship training variables?

H^0: Scores on the EPPP will not significantly correlate with internship match outcomes, including the dichotomously considered outcome of matched versus not match for

internship (using a point bi-serial correlation) and the rank order associated with the site for which individuals were matched.

H^1: Scores on the EPPP will significantly correlate (point bi-serial) with internship match outcomes, with higher EPPP scores associated with being successfully matched for internship.

H^2: Scores on the EPPP will significantly inversely correlate with the rank order associated with the site for which individuals were matched so that higher EPPP scores are associated with sites ranked highly (e.g., top ranked site is coded as "1").

Research Question 4: Are scores on the EPPP meaningfully related to early career activities?

H^0: Scores on the EPPP will not be significantly associated with the nature of early career activities, as measured by each of the following dichotomous (yes/no) variables:

- Conducting/assisting with original research

- Engaging in direct clinical service

- Engaging in evidence-based practice

- Engaging in supervision of clinical work

- Teaching evidence based practice

- Engaging in consultation, in-services, and/or peer-supervision

- Advising others to use evidence-based practices

- Engaging in dissemination activities

- Helping others to engage in evidence-based practice?

- Teaching/mentoring of undergraduates or graduate students, postdoctoral fellows, or allied health professionals

- Teaching others how to engage in scientific research, how to consider the scientific bases of clinical psychology, or how to engage in evidence-based practice

- Engaging in volunteer or paid service to the profession to help advocate for clinical psychology research, scientifically-oriented training, and/or evidence-based practice

H^1: Scores on the EPPP will be significantly associated (point bi-serial correlation) with conducting/assisting with original research in the early career phase.

H^2: Scores on the EPPP will be significantly associated (point bi-serial correlation) with engaging in direct clinical service in the early career phase.

H^3: Scores on the EPPP will be significantly associated (point bi-serial correlation) with engaging in evidence-based practice in the early career phase.

H^4: Scores on the EPPP will be significantly associated (point bi-serial correlation) with engaging in supervision of clinical work in the early career phase.

H^5: Scores on the EPPP will be significantly associated (point bi-serial correlation) with teaching evidence based practice in the early career phase.

H^6: Scores on the EPPP will be significantly associated (point bi-serial correlation) with engaging in consultation, in-services, and/or peer-supervision in the early career phase.

H^7: Scores on the EPPP will be significantly associated (point bi-serial correlation) with advising others to use evidence-based practices in the early career phase.

H^8: Scores on the EPPP will be significantly associated (point bi-serial correlation) with engaging in dissemination activities in the early career phase.

H^9: Scores on the EPPP will be significantly associated (point bi-serial correlation) with helping others to engage in evidence-based practice in the early career phase.

H^{10}: Scores on the EPPP will be significantly associated (point bi-serial correlation) with teaching/mentoring of undergraduates or graduate students, postdoctoral fellows, or allied health professionals in the early career phase.

H^{11}: Scores on the EPPP will be significantly associated (point bi-serial correlation) with teaching others how to engage in scientific research, how to consider the scientific bases of clinical psychology, or how to engage in evidence-based practice in the early career phase.

H^{12}: Scores on the EPPP will be significantly associated (point bi-serial correlation) with engaging in volunteer or paid service to the profession to help advocate for clinical psychology research, scientifically-oriented training, and/or evidence-based practice in the early career phase.

Research Question 5: Are scores on the EPPP meaningfully related to self-appraisals of competency at the time of applying for internship?

H^0: Scores on the EPPP will not significantly associate with either pre-doctoral or early career self-appraisals of professional competencies.

H^1: Scores on the EPPP will be significantly positively correlated (single-tail test) with self-appraisal scores for global competency at the time of applying for internship.

H^2: Scores on the EPPP will be significantly positively correlated (single-tail test) with self-appraisal scores for clinical competency at the time of applying for internship.

H^3: Scores on the EPPP will be significantly positively correlated (single-tail test) with self-appraisal scores for research competency at the time of applying for internship.

H^4: Scores on the EPPP will be significantly positively correlated (single-tail test) with self-appraisal scores for foundational knowledge competency at the time of applying for internship.

H^5: Scores on the EPPP will be predicted by self-appraisal scores at the time of applying for internship. To test this hypothesis, only those competencies found to significantly correlated with EPPP scores when testing Hypotheses 1- 4 of this research question will be entered into a forward regression predicting EPPP scores.

Research Question 6: Are scores on the EPPP meaningfully related to early career self-appraisals of profession-wide competencies?

H^0: Scores on the EPPP will not significantly associate with early career self-appraisals of professional competencies.

H^1: Scores on the EPPP will significantly positively associate with early career self-appraisal scores for professional wide competencies (PWC). The following PWC will be significantly correlated with EPPP scores. In each area listed the raw item scores range from 1-100. Where more than one item exists, the average score is computed and that average score correlated (one-tailed) with EPPP scores.

- Professional values and attitudes (composite score across 5 items)

- Individual and cultural diversity (composite score across 4 items)

- Ethical and legal standards and policy (composite score across 3 items)

- Reflective practice / self-care (composite score across 4 items)

- Relationships (composite score across 3 items)

- Scientific knowledge and methods (composite score across 3 items)

- Research / evaluation (composite score across 2 items)

- Evidence-based practice (1 item)

- Assessment (composite score across 6 items)

- Intervention (composite score across 4 items)

- Consultation (composite score across 4 items)

- Teaching (composite score across 2 items)

- Supervision (composite score across 4 items)

- Interdisciplinary systems (composite score across 4 items)

- Management – administration (composite score across 4 items)

- Advocacy (composite score across 2 items)

H^2: Scores on the EPPP will predict early career self-appraisal scores for professional wide competencies (PWC). To examine this hypothesis, EPPP scores will be regressed onto each of the above PWCs that are found to significantly correlated with EPPP scores.

CHAPTER 2

METHOD

2.1   Participants

Across the larger studies from which doctoral training data were collected (Callahan et al.

2010; Callahan et al. 2014), 977 individuals were eligible for this follow-up study. All eligible

participants were individuals who completed at least a majority of their doctoral training in

psychology by 2011, with a mean age of 29 at the time of providing doctoral training data. Of

those eligible participants, 80% were female. In terms of race/ethnicity, the eligible participants

were 80% non-Hispanic white, 7% Asian/Pacific Islander, 4% African American and Black, 4%

Hispanic and Latino/a, 4% Biracial/Multiracial, 2% Other. A total of 190 individuals were

located (via unchanged email information or found via a search of state licensing boards

nationwide) for this follow-up study. The obtained sample was approximately similar to the

eligible pool of participants, with 78.4% female and 83.2% non-Hispanic white, 6.3%

Biracial/Multiracial, 5.3% Asian/Pacific Islander, 3.7% Hispanic and Latino/a; 1.1% African

American and Black.

2.2   Measures

Archival data was captured via an online survey, as described in Callahan et al., 2010;

Callahan et al., 2014). New data was also gathered via an online survey to capture early career

activities, competencies, and EPPP score information.

2.3   Early Career Alumni Competency Form

Hatcher and colleagues' (2013) workgoup pilot tested a rating form, based on the

Benchmarks article, intended to serve as a more concise, user-friendly assessment tool. Based on

the results of that pilot test, they (1) constructed three domains of competency (i.e.,

44

Professionalism, Relational, and Science) as a consolidation of the seven foundational competencies from the Benchmarks document, (2) created three clusters of functional competences out of the original eight domains (i.e., Application, Education, and Systems), (3) moved on previously identified foundational competency (i.e., Interdisciplinary Systems) into the functional category of Systems, and (4) moved one previously identified functional competency (i.e., Research Evaluation) into the foundational competency area of Science. Separate rating forms were then created for each developmental stage. The current study drew from the version created for those emerging into independent practice. Although Hatcher and colleagues encouraged psychometric evaluations of their forms, no published studies were identified in the published literature. In total, the Early Career Alumni Competency Form is comprised of 55 separate items, with four categories of behavioral indicators presented in a scoring codebook: (1) Does Not Meet Expectations, (2) Pre-internship Level, (3) Internship Level, and (4) Readiness for Entry to Practice. A dimensional, Visual Analog Scale (VAS), rating between 1 and 100 allows raters to slide a bar to the area on the line that they feel accurately reflects the competence level of the individual being rated. No numbers were visually presented. For the current study, participants were asked to self-evaluate and provide ratings of their competencies on the Early Career Alumni Competency Form.

2.4   Procedures

With Institutional Review Board permission (Appendix A), an email was sent to the contact email provided during participation in one of the earlier longitudinal internship match studies. For those without a functional email address, a search of each state licensure database across the United States was conducted in an attempt to locate current contact information. A consent notice was provided prior to participation in this survey-based study. Completion of

study measures subsequent to the notice constituted informed consent. Participants were asked to complete one survey, which took approximately 60 minutes, via an online survey hosted by Qualtrics. Data was automatically coded upon clicking the "submit" button and sent directly to the researchers without identification of the sender. All data obtained from this study has been and will continue to be kept confidential and stored securely.

Participants did not receive any compensation and no personal benefits were expected in exchange for participation. More generally, the information resulting from this study may aid in understanding the relationship between the variables that are salient to the internship match, the developmental unfolding of competencies in students, and the continued development of competencies in early career. There were no known risks associated with this project that are greater than those ordinarily encountered in daily life. All participants and their data were treated in accordance with ethical standards.

CHAPTER 3

RESULTS

As described earlier, the dataset for this investigation was constructed by compiling data gathered in two earlier internship match studies (Callahan et al., 2010; Callahan et al., 2014), but had not been previously combined or analyzed in the manner presented in the current study. Newly gathered data for the current study was matched to the archival doctoral training data gathered in the original surveys via examination and linking of unique identifiers. Identifiers were checked by two independent researchers to ensure accuracy of data linkages and then verified by the supervising faculty mentor. Data were then inspected and cleaned and assumptions checked before implementing the data analytic plan. An alpha level of .05 was used to determine significant relationships for all statistical analysis. Across analyses, the number of participants varies due to missing data. No data was imputed or transformed. An independent statistical consultant validated all analyses.

3.1   Research Question 1

The first research question sought to determine if there is evidence of bias on the EPPP exam based on demographic variables. Based on the literature, it was hypothesized that underrepresented minorities would obtain lower scores than their majority counterparts on the EPPP. The difference between minority students and non-Hispanic, white students on the EPPP was measured using the Welch t statistic to compare means. Students of Asian descent were excluded from these tests as they are not an underrepresented minority in the training to workforce pipeline (Callahan et al., 2018). Although there was no statistically significant difference between the EPPP scores of underrepresented ethnic minorities ($n = 8$, $M = 646$, $SD = 56.05$), and non-Hispanic, white students ($n = 92$, $M = 656.17$, $SD = 82.14$), $t(98) = 1.377$,

p=.244, this analysis is likely to have been underpowered. Importantly, Hedges' g (appropriate for use with unequal sample sizes), was .126 and consistent with a small effect size.

Family of origin socio-economic status growing up was also hypothesized to systematically vary with scores on the EPPP, with higher SES relating to higher EPPP scores. The socio-economic status growing up variable included $N = 78$ with the following frequencies, Lower class $n = 4$, Lower Middle-Class $n = 12$, Middle Class $n = 34$, Upper Middle Class $n = 23$, Upper Class $n = 5$. Missing data was deleted pairwise, leaving 45 pairs of data with both data on the EPPP and SES status growing up. A significant positive moderate correlation was found between socio-economic status growing up and scores on the EPPP $r(121) = .391, p = .004$.

Finally, it was predicted that higher EPPP scores would be associated with lower accrued student debt. Of the 77 respondents to this item, 25 of them said they had $0 accrued student loan debt. The mean student loan debt load was $36,233 and the range was between $0 and $150,000. The mode of this data was $0 with 25 respondents reporting no student loan debt. Of those who reported accrued student loan debt, 44 also reported scores on the EPPP. Missing data was deleted pairwise. There was a significant negative moderate correlation between the amount of education debt accrued and scores on the EPPP, $r(119)= -.371, p = .007$, meaning as debt increased, EPPP scores decreased.

3.2   Research Question 2

The second research question sought to determine if there were any meaningful relationships between scores on the EPPP and pre-internship preparatory training in the archival data that was collected from participants at the time of their application to internship match. Hypothesis 1 was that there would be a significant positive relationship between the number of pre-internship accrued hours and scores on the EPPP. Of the 74 respondents, internship hours

ranged from 305 to 3960, with a mean of 1040.16 and a standard deviation of 521.57. Scores on the EPPP and pre-internship accrued hours did not correlate significantly $r(115) = -.165$, $p = .145$. Hypothesis 2 was that there would be a significant, positive correlation between hours of pre-internship supervision hours and scores on the EPPP. Of the 74 respondents, supervision hours ranged from 140-900, $M = 444.86$ hours, $SD = 163.09$. There was no statistically significant relationship between these variables, $r(115) = -.227$, $p = .02$.

The third hypothesis was that there would be a significant positive correlation between the number of reports written and scores on the EPPP. Of the 75 respondents, there was a $M = 26.93$ and a $SD = 29.147$ with a range between 1 and 141 integrated reports written. There was no statistically significant relationship found between the number of integrated reports written and scores on the EPPP $r(114) = -.032$, $p = .420$.

Hypothesis four was that there would be a significant positive relationship between the number scientific presentations presented and scores on the EPPP. Of the 76 respondents, there was a mean of 10.55 presentations, $SD = 6.047$ and a range between 0 and 25. Contrary to the prediction, there was no statistically significant relationship between the number of scientific presentations and scores on the EPPP, $r(188) = -.101$, $p = .258$. Hypothesis five was that there would be a significant positive correlation between the number of peer reviewed publications and scores on the EPPP. Of the participants, $N = 74$, there was a mean of 2.65 publications with a standard deviation of 2.34. No statistically significant relationship was found between the number of peer reviewed publications and scores on the EPPP, $r(116) = .019$, $p = .452$.

A forward stepwise multiple regression analysis also found none of the following pre-internship variables, individually or collectively, to be predictive of EPPP scores: hours in assessment/intervention services, supervision hours, completed integrated reports, scientific

presentations and peer-reviewed publications. Notably, EPPP scores were found to be moderately positively correlated with scores on the Verbal section of the GRE, $r(145) = .311$, $p = .037$, and the quantitative section of the GRE, $r(144) = .405$, $p = .008$.

3.3   Research Question 3

The third research question examined whether those who matched for internship and/or those who matched with their preferred site would yield a higher score on the EPPP than those who did not match for internship, or matched with a lower choice on their internship ranking. A single-tailed point bi-serial correlation revealed no statistically significant relationship between matching for internship (yes/no) and EPPP scores $r_{pb}(121) = .133$, $p = .192$, though this sample contained a low number of unmatched participants. Of the 74 participants who provided their match rank order data, the mean rank of their site match was 2.49 ($SD = 2.36$). No statistically significant relationship between the rank of applicants matched site and EPPP scores was identified, $r(115) = .181$, $p = .122$.

3.4   Research Question 4

The aim of this research question was to elucidate the association of EPPP scores with early career activities. For questions in this hypothesis participants were asked to answer Yes or No. Yes answers were coded as 1 and No answers were coded as 0. For some questions participants were asked "if Yes, Then…" meaning not every question was responded to equally. Means in this data set are reflective of this coding process with means closer to 1 indicating that more individuals responded Yes, while Means closer to 0 indicate a higher response of No. For each of the below, a point biserial correlation was calculated to assess the strength of relationship between EPPP scores and activity engagement (yes/no). For ease of presentation, each analysis is presented separately below:

- The relationship between conducting or assisting with original research early in a professional's career and scores on the EPPP was not statistically significant, $r_{pb}(260) = .082, p = .215$ ($n = 149, M = .74, SD = .441$).

- Similarly, the relationship between engaging in direct clinical services and scores on the EPPP was not statistically significant $r_{pb}(259) = -.053, p = .305$ ($n = 148, M = .97, SD = .181$).

- There was no statistically significant relationship between engaging in early career evidence-based practice and scores on the EPPP $r_{pb}(254) = .057, p = .293$ ($n = 143, M = .99, SD = .084$).

- There was a weak positive statistically significant relationship between engaging in supervision of clinical work and scores on the EPPP $r_{pb}(259) = .291, p = .002$ ($n = 148, M = .89, SD = .320$).

- There was no statistically significant relationship between scores on the EPPP and teaching others evidence-based practice during one's early career, $r_{pb}(242) = -.002, p = .493$ ($n = 131, M = .94, SD = .240$).

- Engagement and consultation in-service training or peer supervision in early career ($n = 148, M = .91, SD = .294$) had a weak positive statistically significant relationship with scores on the EPPP $r_{pb}(259) = .198, p = .028$.

- Advising others in early career to use evidence-based practices has no statistically significant relationship with scores on the EPPP, $r_{pb}(244) = -.081, p = .231,$ ($n = 113, M = .96, SD = .191$).

- Engaging in early career dissemination services had no statistically significant relationship with scores on the EPPP $r_{pb}(259) = .144, p = .084$ ($N = 148, M = .61, SD = .488$).

- There was no statistically significant relationship between helping others engage in evidence-based practices and scores on the EPPP $r_{pb}(202) = .095, p = .244$ ($n = 91, M = .95, SD = .229$).

- Early career teaching or mentoring undergraduates, graduates, or postdoctoral or allied health professionals and scores on the EPPP did not have a statistically significant relationship, $r_{pb}(259) = -.041, p = .346$ ($n = 148, M = .89, SD = .312$).

- There was a weak statistically significant positive relationship between teaching others how to engage in the scientific basis of psychological practice or evidence-based practice and scores on the EPPP, $r_{pb}(243) = .225, p = .019$ ($n = 132, M = .93, SD = .253$).

- There was no statistically significant relationship between volunteer or paid service to advocate for clinical psychology research or evidence based practice and scores on the EPPP $r_{pb}(259) = .025$, $p = .405$, ($n = 148$, $M = .49$, $SD = .502$).

3.5 Research Question 5

A single-tailed Pearson product moment correlation was conducted to determine if there were any meaningful relationships between subjective self-appraisals of competency (on a scale of 1-10, with 10 being best/highest) at the time of applying to internship with participant's subsequently earned scores on the EPPP. At the time of applying for internship, participants were asked to rate themselves on four areas of perceived competency. No statistically significant relationship was identified between scores on the EPPP and self-appraisal of foundational knowledge competency $r(143) = -.052$, $p = .389$, self-appraisals of clinical competency $r(143)=-.182$, p=.159, self-appraisals of research competency $r(143) = .089$, $p = .313$, or self-appraisals of global competency $r(143) = -.075$, $p = .341$,

3.6 Research Question 6

The goal of research question 6 was to determine whether earned EPPP scores were associated with early career competency scores on the domains assessed via the Early Career Alumni Competency Form. It was hypothesized that there would be a significant relationship between scores on the EPPP and scores on self-appraisal scores for each of the 16 separate areas, each of which is presented below.

The Professional Values and Attitudes scale is a measure of behavior and comportment that reflect the values and attitude of psychology, and includes the subscales Integrity, Deportment, Accountability, Concern for the Welfare of Others and Professional Identity. Of these subscales, there was a significant weak positive correlation between scores on the EPPP and scores on the Integrity Subscale $r(215) = .261$, $p = .004$. The Integrity subscale describes the

ability to monitor and independently resolve situations that challenge professional values and integrity and the finding indicates that as scores on the Integrity subscale increased, scores on the EPPP also increased.

The Individual and Cultural Diversity scale describes the level of awareness, sensitivity and skills in working professionally with diverse individuals, groups and communities who represent various cultural and personal backgrounds and characteristics defined broadly and consistent with APA policy. There was no statistically significant relationship between scores on the EPPP and self-appraisal scores on this scale.

The Ethical Legal Standards and Policy scale describes the ability to apply ethical concepts and awareness of legal issues regarding professional activities with individuals, groups, and organizations. There were no statistically significant relationships between scores on this scale and scores on the EPPP.

The Reflective Practice, Self-Assessment and Self Care scale is a measure of personal and professional self-awareness and reflection, awareness of competences and appropriate use of self-care. This scale includes the subscales Reflective Practice, Self-Assessment, Self-Care, and Participation in Supervision Process. Of these subscales, there was a significant positive weak relationship between scores on the EPPP and Participation in the Supervision Process $r(204) =$ .179, $p = .043$. The Participation in Supervision Process subscale describes the ability to seek supervision when personal problems may interfere with professional activities and the ability to seek supervision when working outside of professional competency.

The Relationships scale measures the ability to relate effectively and meaningfully with individuals, groups and/or communities. There was no statistically significant relationship between this scale and scores on the EPPP.

The Scientific Knowledge and Methods scale describes the level of understanding of researchers, research methodology, techniques of data collection and analysis, biological basis of behavior, cognitive-affective bases of behavior and development across the lifespan. This scale includes subscales for Scientific Mindedness, Scientific Foundation of Psychology, and Scientific Foundation of Professional Practice. Of these subscales, there was a significant positive weak relationship between scores on the Scientific Mindedness and scores on the EPPP $r(213) = .186, p = .03$.

The Research and Evaluation scale is a measure of research generation contributing to a professional knowledge base or evaluating the effectiveness of various professional activities. No statistically significant relationship was found between this scale and scores on the EPPP.

The Evidence Based Practice scale measures the ability to integrate research and clinical expertise in the context of patient factors. There is a significant weak positive relationship between this scale and scores on the EPPP $r(211) = .206, p = .02$. This means as the ability to independently apply knowledge of evidence-based practice, empirical bases of assessment, intervention and other psychological applications, clinical expertise and client preferences increases, scores on the EPPP also increase.

The Assessment scale describes the ability to assess and diagnose problems and capabilities associated with individuals, groups and organizations. There was no statistically significant relationship between this scale and scores on the EPPP.

The Intervention scale describes the ability to design interventions for individuals, groups and/or organizations that alleviate suffering and promote health and well-being. No statistically significant relationship between this scale and scores on the EPPP.

The Consultation scale is a measure of the ability to provide expert guidance or

professional assistance in response to a client's needs or goals. There was not a statistically significant relationship between scores on this scale and scores on the EPPP.

The Teaching scale measures the ability to provide instruction, disseminate knowledge and evaluate acquisition of knowledge and skills in professional psychology. Scores on the Teaching scale and scores on the EPPP did not statistically significantly correlate.

The Supervision scale describes the ability to train and monitor the professional functioning of others. There was no statistically significant relationship between this scale and scores on the EPPP.

The Interdisciplinary Systems scale is a measure of knowledge of key issues and concepts in related disciplines. No statistically significant relationship was found between scores on this scale and scores on the EPPP.

The Management-Administration scale describes the ability to manage the direct delivery of services and the administration of organization, programs or agencies. This scale includes subscales: Appraisal of Management and Leadership, Management, Administration, and Leadership. A significant weak negative correlation was found between scores on the Management subscale and scores on the EPPP $r(178) = -.231, p = .03$. This means as the ability to participate in management of direct delivery of professional services, including scheduling, billing, and maintenance or records increase, scores on the EPPP decreased.

The Advocacy scale describes actions targeting the impact of social, political, economic or cultural factors to promote change at the individual, institutional and/or systems level. This scale includes the subscales: Empowerment and Systems Change. There is a significant weak negative correlation between scores on the Systems Change subscale and EPPP scores $r(189)=-.268, p = .009$, indicating that as the ability to promote change including the ability to engage

with groups of differing viewpoints and develop alliances within those groups increases, scores on the EPPP significantly decreased.

Finally, a forward stepwise multiple regression analysis was used to determine which professional wide competencies significantly predicted scores on the EPPP. The results of the regression indicated that two subscales predictor variables explained a significant percentage of the variance (R^2 = .260, $F(2,48)$ = 8.43, $p < .001$.) If was found that scores on the Integrity subscale (item 1.A) significantly predicted scores on the EPPP in a positive direction B = .376, p = .005, while scores on the Systems Change subscale (Item 16.B) significantly predicted scores on the EPPP in a negative direction Systems Change (B = -.417, p = .002.).

Participants predicted EPPP score is equal to 437.757-1.631 (Systems Change) + 3.634 (Integrity), where Systems change and Integrity are both continuous variables. Participants' EPPP scores increased 3.634 points for each increase in points on the Integrity scale on the PWC rating. The score decreased on the EPPP by 1.631 for every one-point increase on the Systems Change PWC score.

CHAPTER 4

DISCUSSION

Overall, scores on the EPPP were significantly correlated with very few variables in this study. There was a positive relationship between socio-economic status growing up and EPPP scores and a negative relationship between student debt load and EPPP scores. There was no relationship between EPPP scores and number of pre-internship intervention, assessment and supervision hours, number of integrated reports, number of peer reviewed publications, number of presentations. Similarly, there was no relationship between EPPP scores, match ranking, self-assessment of competency and many early career activities.

There was a positive relationship between EPPP scores and several early career activities including engaging in supervision of clinical work, engaging in consultation, in-service or peer supervision and teaching others how to engage in the scientific basis of psychology or evidence-based practice. Of the early career competency scores there was a significant relationship between scores on the EPPP and measures of Evidence Based Practice, Scientific Mindedness, Participation in Supervision and Integrity. Scores on the Systems Change subscale and the Management subscale of the PWC negatively impacted scores on the EPPP.

4.1   Implications of the Study

There were no significant findings linking the EPPP to racial/ethnic demographic variables. The EPPP is one of the final barriers in the training-to-workforce pipeline, and previous research has indicated that ethnic minorities are less likely to enter the field at the undergraduate level and are more likely to experience attrition at multiple points along the pipeline prior to reaching the point of taking the EPPP. In this study, 91% of respondents were non-Hispanic whites and 7.1% were underrepresented ethnic minorities. The 2019 census bureau

reports that the US population is made up of 60.3% non-Hispanic whites, 18.5% Hispanic and Latino/as, and 13.4% African American and Black individuals. Consistent with previous findings, non-Hispanic whites are overrepresented in psychology doctorate programs compared to the overall population, and Hispanic and Latino/a and African American and Black individuals are underrepresented in these programs. Previous research suggests this difference is due to the financial burden of higher education, not a difference in academic achievement.

In this study, there was not a significant relationship between racial/ethnic demographic variables and socioeconomic status, though a small effect size was demonstrated. Future research with a more diverse sample is strongly needed to adequately power similar analysis. There was, however, a significant relationship between scores on the EPPP, student debt load, and early life socioeconomic status. This means those who had a higher student debt load and lower SES growing up had significantly lower scores on the EPPP than those with a lower debt load and a higher SES growing up. While these individuals were able to complete the program and meet academic milestones such as matching for internship, completing comprehensive exams, accumulating practicum hours, writing integrated diagnostic reports, and successfully graduating with a doctoral degree, they were less likely to enter the workforce as psychologists due to obstacles created by the licensure exam. Previous work in this area warns of the increase in this barrier with the adoption of the EPPP-2, an even more expensive licensure exam that has not been externally validated.

The use of the EPPP-2 would mean the current EPPP would occur at the same time as the comprehensive examinations, creating an even higher student workload and cost. This increased workload and cost associated with study materials also reduces a student's ability to work for pay during their program. The evidence in this study is consistent with the EPPP creating a

financial burden for students who have otherwise demonstrated competency in their abilities to practice and responsibly contribute to the field of psychology.

This suggests that the EPPP is better described as a reliable measure of socioeconomic status, rather than a measure of clinical or scientific competency. The exam keeps individuals out of the field of psychology, further reducing their ability to pay down the debt load and increase their future socio-economic status. In Callahan's 2018 study on the training-to-workforce pipeline, it was found that several minority groups, including students with disabilities and African American and Black students, rated their programs as having poorer honesty in their recruitment practices than other demographic groups. In this study, Callahan explored the possibility that lack of honesty in recruitment may have led students to take on large debt loads without the possibility for gainful employment, despite equal academic and training outcomes.

Many possible factors may contribute to this finding, including the rise (and fall) of for-profit training programs, shifting licensure requirements in the healthcare industry for the diagnosis and treatment of mental illness, and predatory recruitment techniques that have targeted individuals who did not qualify for the limited spots in no-cost Ph.D. programs. The EPPP creates a final but significant barrier to individuals from lower SES backgrounds who have taken on high student debt loads in the training-to-workforce pipeline. This final barrier means that there is a group of individuals with all of the necessary training and academic credentials to practice psychology who are consequently unable to attain a meaningful career in the field. As the need for mental healthcare workers increases in the US, the barriers to becoming a mental health care worker are increasing with the introduction of the EPPP-2.

Inevitably, there will be a supply and demand crisis in the near future, as the current professionals retire and a smaller, less diverse workforce replaces them, and demand for services

increases. Lack of access to mental health services will inevitably have detrimental effects on individuals, communities, and our country as a whole. Clinical Psychology is a client-facing field that centers the needs and safety of clients at all other levels of training and professional practice, and should not have barriers at the licensure exam level, particularly when this exam is unrelated to a trainee's quality or quantity of academic performance. In this author's opinion, the introduction of more barriers through the unvalidated EPPP-2 is irresponsible on the part of the American Psychological Association, and the evidence in the current study should serve as a caution for a potentially vicious cycle in the future.

It could be argued, however, that the EPPP is a better indicator for quality professionals than other factors such as quantity of training, match outcomes, and academic success. If this is found to be the case in future research, the lack of relationship between these factors and scores on the exam is alarming. If, for example, more assessment/intervention hours actually have no effect on the quality of the professionals practicing in the field, a reduction in academic requirements may be warranted. Reducing the number of hours required in unpaid work may leave more time for students to engage in gainful employment during their program, thereby reducing the student loan debt load and increasing their abilities to build post-graduate professional networks of gainful employment. This may create more work-life balance and professional development opportunities, as well as improve the overall well-being of students in training programs. If the ASPPB asserts that the EPPP and the EPPP-2 are appropriate measures for gatekeeping in the field of psychology, and that these measures are unrelated to training outcomes, would then a model involving fewer client-facing hours and more direct study for this examination be a more appropriate use of a student's time and resources? This assertion is unsubstantiated given the history of unvalidated, rushed-to-market licensure exams; the Spanish

version of the EPPP, in particular, which led to a massive shortage of mental health professionals on the U. S. Commonwealth island of Puerto Rico.

A predominantly Spanish-speaking population was not able to access care from practitioners who were also Spanish-speaking due to an unvalidated licensure exam. While this incident is the most egregious, it is also illustrative of the positive feedback loop that can occur in minority groups. As the number of Spanish-speaking psychologists decreases, so does the ability of trainees to find Spanish-speaking advisors, which reduces the ability of training programs to recruit Spanish-speaking students. A lack of student and faculty diversity in training programs reduces the overall ability of all trainees to demonstrate cultural competency with their Spanish-speaking clients. Spanish speaking clients without access to culturally competent psychologists means an entire demographic group slowly loses access to mental health services. Losing access to mental health services in a demographic group creates a larger effect in that community, one which has historically been weary of utilizing mental health services due to cultural incongruency. A void in access may either lead to an overall reduction in care or may lead to a rise in services provided by less-qualified (i. e. culturally competent, language barriers, understanding of diversity and inclusion) and less-regulated professionals.

4.2   Multicultural Competency

According to Callahan, Smotherman, Dziurzynski, Love, Kilmer, Flores-Niemann, and Ruggero (2018) multicultural training can lead to multicultural competency in the workforce. There are not very many non-white students despite there being no difference in GRE scores or GPA between these groups upon matriculation. Therefore, this disparity creates a lower opportunity for multicultural competency across a diverse array of clinical trainees.

The results of this study indicated there was no effect on the profession-wide competencies related to diversity and scores on the EPPP. Conversely, results indicated that the more time spent with people in an out group (systems change description) the more likely one would perform poorly on the EPPP. They also point to a positive feedback loop (or vicious cycle) where a program will have few culturally and ethnically diverse students which leads to a more difficult time recruiting more culturally and ethnically diverse students which in turn leads to a workforce with poor multicultural competencies. All of which ultimately will become the norm by which the field is standardized. The standards of the field are what gatekeeping measures, such as the EPPP, are meant to uphold. Based on these results, this author questions what standards the APA and other gatekeepers actually endeavor to uphold.

4.3   Obstacles to Minorities

Debt loads and financial hardship is cited by Luebbe and Ogbaselase (2018), who's research suggest debt loads may have a significant relationship with EPPP scores. Essentially, this means that those with more debt tend to perform worse on the EPPP and are therefore less likely to have a financially stable future that would allow them to pay off that debt. Their data also suggested there is a significant relationship between early life socioeconomic status and scores on the EPPP. Essentially, those who had a lower SES growing up, had lower scores on the EPPP than those who came from higher SES backgrounds.

Debt load and financial hardship, according to Luebbe and Ogbaselase (2018), keep minorities from entering the field at the undergraduate level, and those who do choose to enter are significantly hindered at the level of the licensure exam, even though all other factors are the same, including the quality of minority candidates for the match process, the quantity of their training and their ratings on subjective measures of profession-wide competencies. In this study,

there was no significant relationship between SES and ethnicity, but there was a significant relationship between SES and EPPP scores and debt load and EPPP scores.

This provides a major obstacle to diversifying the field of psychology. The APA says it wants to implement "meaningful change" through better training and understanding of intersectionality and multicultural training. However, the evidence from this study suggests that these intentions are not reflected by the Association of State and Provincial Psychology Boards (ASPPB) in their work on the EPPP; there is a significant negative relationship between measures or multicultural competency and scores on the EPPP and no relationship between early career competency measures of diversity and scores on the EPPP. If the APA wants to implement meaningful change, they will need ASPPB to share that vision in their gatekeeping role with the national licensure exams.

The Examination for Professional Practice in Psychology (EPPP) is administered by the ASPPB and required by state and provincial licensing boards across the United States, its territories, and Canada. If the EPPP significantly reduces the likelihood that members of minority groups will become psychologists, then there will be a shortage of those individuals in the field. If there is a shortage of minorities in the field, there will be less culturally competent care for clients, fewer culturally competent supervisors, and less culturally competent training available. This means it is harder to retain and recruit diverse students into training programs. This also means the field of psychology training and research and the client-facing psychology practitioners will become more and more homogenous. Homogeneity of practitioners will ultimately lead to less access to care for individuals in minority groups seeking mental health care services.

Lack of access to mental health care services in historically underserved and

underrepresented minority groups can lead to elevated levels of anxiety and depression, and major health disparities resulting in long-term mental health crises (*Stress in America*, October 2020). For instance, in Puerto Rico, an unvalidated Spanish version of the EPPP led to a severe shortage in licensed professionals on the U. S. Commonwealth island, thereby resulting in fewer Spanish-speaking clinicians who were able to provide services. Following major natural disasters like the earthquakes and hurricanes of 2017, Spanish-speaking professionals and their respective communities were severely impacted, and were doubly impacted by the lack of mental health services. GRE scores and EPPP scores are significantly correlated in Callahan and Watkins' (2018a) study. Their findings illustrated that Quantitative GRE scores are emphasized in graduate admissions in psychology. In this study Quantitative GRE scores were significantly correlated with scores on the EPPP. Scores on the GRE lead to poor admissions outcomes for those who are of racial and ethnic minorities. This is one of the stops on the training-to-career pipeline that reduces the number of minorities in the field of psychology.

4.4   Supervision

Within the current study, there was a significant relationship between scores on the EPPP and engagement in supervision. Scott, Ingram, Vitanza, and Smith's (2000) assert that only half of all programs even offered this as a course, despite its importance in the licensure exam. The authors found that programs are not prioritizing training in this area due to high course loads, budget constraints and the belief that this training would occur during internship. The current study suggests supervision should be emphasized during all levels of training and consistently through all programs. Furthermore, the current study suggests doctoral programs place more emphasis on monitoring supervision relationships, with a system of mutual assessment of both the supervisor and supervisee. These outcomes can be used to develop and improve training,

increase trust, and reduce unfair power dynamics leading to student attrition and poor early career outcomes. Such assessments also create a system of accountability between training programs and their stakeholders to improve employability and increase student's confidence in their abilities to successfully enter the workforce. Lastly, mutual assessment will help reduce and/or prevent potentially abusive dynamics between supervisors and supervisees, while encouraging training programs to structure their curricula to provide equitable, inclusive, and culturally competent training by faculty and staff who embody such principles.

4.5   Matching to Internship

Matching to internship did not have a significant relationship with EPPP scores. Matching to internship did not have a relationship with several training factors, program type, dissertation proposal completion, number of peer reviewed publications, number of intervention assessment hours, number of integrated reports, number of supervision hours and number of applications (Callahan et al., 2010; Callahan, et al., 2014). The lack of relationship between quantity of training and internship match and internship match and scores on the EPPP necessitates an examination of the focus on quantity of training factors in psychology programs. Requirements for practicum experience hours, writing of integrated reports and supervision hours can increase the financial burden on students who are less able to work for pay and increases the amount of student loan debt. If these factors do not have an impact on EPPP scores, or internship match, they may be creating an undue burden on students of lower socio-economic status and increasing the likelihood of attrition of these students. Furthermore, it is worth noting that many of these requirements do not provide clear-cut guidelines or avenues of support to students who are parenting/lactating, non-traditional, identify English as their second language, and/or student with disabilities to address their often unseen and/or undisclosed barriers to

successful completion of these requirements. The onus lies on training programs, faculty, ad staff to integrate these measures into their curriculum to create a more diverse workforce training pipeline.

4.6   Limitations and Future Directions

This study was not without limitations. The authors of the Early Career Alumni Competency Form did not include psychometric information or normative references. Additionally, a large number of analyses were performed, without making any statistical correction (e.g., Bonferroni correction) for possible alpha inflation. While it is possible that some small effects were spurious, this analytic decision allowed for maximum detection of associations between EPPP scores and other variables. Yet, very few statistically significant associations were found, which strongly implicates the EPPP as not meaningfully related to competency or training and education in doctoral health service psychology.

Data relied on an archival cohort, meaning those who responded are not in the context of any events currently affecting students and early career psychologists. On March 8, 2019, outside of the selection time period for this current study, the APA announced that ten accredited psychology training programs closed without warning (APA, 2019). Trainees impacted by these closures were given 90-days to transfer or forfeit their graduate credits. This event left psychology trainees, many of whom identified as racial and/or ethnic minorities, Veterans, non-traditional adult learners, and/or people with disabilities all over the country without viable programs to attend, leaving many without options to continue towards successfully entering the workforce. Within the trainee-to-workforce pipeline, this event not only significantly affected the students in those programs, but also reduced the total number of programs for future diverse students and likely reduced trust in the current training climate. Future research should (1)

66

identify the effects of these closures on individuals directly involved and those who were stopped in this pipeline at the undergraduate level, and (2) explore the effects of multicultural competency and access to care in regions that lost training programs.

The programs that closed were professional training programs, most of which had students with large debt loads. In the current research, there is a significant relationship between large debt loads and lower scores on the EPPP. As a result of campus closures that disproportionately burdened students with high debt loads, few students with heavy debt loads may remain in the pipeline at this time. A lack of opportunity to complete psychology training through the use of student loans is predicted to reduce the overall number of professionals entering the field. Future research that includes data from 2019 onward, is predicted to find less multicultural competency and a diverse representation of clinicians in the profession overall. Ultimately, a professional replacement rate that is lower than the professional retirement rate is also likely to cause a crisis in access to care in the United States overall.

The second event likely to affect the training and early career outcomes that was not captured in the current study is the COVID-19 global pandemic. The pandemic is believed to have a significant effect on the mental health of the United States population (*Stress in America, October 2020*). This impact created an increased need for psychological care services throughout the United States. The pandemic interrupted the typical training flow for many due to local regulations, as many programs were forced to create telehealth-based training within a fully remote learning environment mid-year. Future research would predict a larger impact of trainees (especially pregnant, lactating, and female) with dependents living at home during this time, as remote work became the norm for many parents and their school-aged children. Future research should explore the relationship between decreased access to culturally competent care providers

and an increased need for these providers. Understanding if there was any impact on the method of training provided, using virtual and telehealth services in the natural experiment created by the pandemic will be vital to the future of psychology.

Future research should also examine the impact of community outcomes for areas that are affected by a reduction in access to care. A community impact survey should be conducted to determine the effects of lower access to care in areas that have fewer psychologists and/or a homogenous pool of psychologists due to unvalidated licensure exams. This survey could potentially explore the relationship between access to care and community outcomes such as local crime rates, the need for social services for vulnerable populations such as children, disabled individuals and the elderly, rates of truancy, and the impact on other healthcare fields such as primary care physicians and emergency care.

The correlation between quantitative GRE scores and EPPP scores is among the strongest relationships within this study, and is significant at an alpha of .01. Notably, only 34 participants who completed the GRE also completed the EPPP, possibly due to inconsistent requirements of the GRE scores to enter clinical psychology training programs or due to lack of score recall at the time of applying for internship. This may be an artifact of selection bias, meaning those who choose to take the GRE, may have done so because they believed they would do well on it. Nevertheless, these findings are consistent with previous research finding a significant relationship between scores on GRE and the EPPP Sharpless and Barber (2013).

4.7   Conclusions

Clinical doctoral clinical psychology training programs should be examined to better understand how these programs are serving underrepresented and historically marginalized trainees. The presence of a representative sample of trainees from these groups improves the

multicultural competency of all trainees, and improves access to care for individuals seeking services from these marginalized groups who may share congruent identities and statuses. Previous research suggests the training-to-workforce pipeline stops trainees of marginalized groups at multiple points along the way, if they decide to enter the field at all. The current study focused on one of the final points in the pipeline, the EPPP licensure exam. Scores on this exam were found to be correlated with several expected factors such as scientific-mindedness, engagement in supervision and knowledge of evidence-based practice. However, scores were also significantly correlated with other, unrelated factors, such as early life SES, student debt load, and scores on the GRE.

Training factors such as quantity of hours, number of integrated reports, peer reviewed publications, and presentations had no effect on EPPP scores. Similarly, trainee quality as evidenced by match ranking, self-reported competency and several early career practices had no discernible relationship with the EPPP. The EPPP creates an undue financial burden and correlates most strongly with income, rather than a candidate's quality or quantity of training. These factors should be carefully considered as the implementation of the EPPP-2 occurs during what is set to be a time of higher demand for practitioners. This study should contribute to the existing literature predicting an inevitable supply-demand imbalance. Future research should review the effects of this imbalance on communities, especially those historically underserved and underrepresented in the field of psychology.

APPENDIX

INFORMED CONSENT NOTICE

University of North Texas Institutional Review Board

Informed Consent Notice

Before agreeing to participate in this research study, it is important that you read and understand the following explanation of the purpose, benefits and risks of the study and how it will be conducted.

Title of Study: *The Competency Pipeline: Pre-Internship to Early Career*

Supervising Investigator: Dr. Jennifer Callahan, PhD ABPP, University of North Texas (UNT) Department of Psychology.

Purpose of the Study: You are being asked to complete a follow-up survey of the internship match survey you completed between 2009-2011. This study aims to see how people have developed early career competencies.

Study Procedures: This survey will require approximately 30-45 minutes. For purposes of maintaining data integrity, please do not submit data in pieces. Rather, proceed to the study measures by clicking on the link below to see what information you will be asked for and, if necessary, you may subsequently return to this page and follow the link to submit data in a single session. It is recommended that you have your CV and your other scores on the national licensing exam (EPPP), if you have taken it, available to you while completing responses.

Foreseeable Risks: There are no known risks associated with this project that are greater than those ordinarily encountered in daily life. No personal benefits are anticipated in exchange for participation. The information resulting from this study may aid in understanding the developmental unfolding of competencies in emerging and early career psychologists.

Benefits to the Subjects or Others: This study is not expected to be of any direct benefit to you. The information resulting from this study may aid in understanding the developmental unfolding of competencies in emerging and early career psychologists.

Compensation for Participants: There is no compensation for participating in this study.

Procedures for Maintaining Confidentiality of Research Records: Please note that this survey is confidential. Any data you submit will be automatically coded upon clicking the "submit" button" and sent directly to the researchers without identification of the sender. The data obtained from this study will be kept confidential, and the responses will not be linked to any of the participants. The data will only be stored electronically, and protected on a controlled access server. Confidentiality will be maintained to the degree possible given the technology and practices used by the online survey company. Your participation in this study involves risks to confidentiality similar to a person's everyday use of the internet.

Office of Research Integrity & Compliance   APPROVED BY THE UNT IRB
University of North Texas   5/18/2018
Last Updated: July 11, 2011

Page **1** of **2**

71

Questions about the Study: If you have any questions about the study, you may contact Dr. Callahan at jennifer.callahan@unt.edu or 940-369-8229.

Review for the Protection of Participants: This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The UNT IRB can be contacted at (940) 565-4643 with any questions regarding the rights of research subjects.

Research Participants' Rights:

Your participation in the survey confirms that you have read all of the above and that you agree to all of the following:

- *Dr. Jennifer Callahan* has explained the study to you and you have had an opportunity to contact him/her with any questions about the study. You have been informed of the possible benefits and the potential risks of the study.
- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study personnel may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You understand you may print a copy of this form for your records.

Directions: To complete this survey, all you need to do is select among the choices in the following form and SUBMIT the form when you are done by pressing the SUBMIT button at the end of the survey. IT IS VERY IMPORTANT TO SUBMIT THE PAGE; otherwise no data will be collected. Please be sure you complete the form entirely by responding to every question.

Thank you for your participation in this study.

Office of Research Integrity & Compliance   APPROVED BY THE UNT IRB
University of North Texas   5/18/2018
Last Updated: July 11, 2011

Page **2** of **2**

72

REFERENCES

American Psychological Association. (2017). *Multicultural Guidelines: An Ecological Approach to Context, Identity, and Intersectionality*. Retrieved from: http://www.apa.org/about/policy/multicultural-guidelines.pdf

American Psychological Association. (2019, March). *In wake of Argosy closure, APA seeks support for students*. https://www.apa.org/news/apa/2019/argosy-closure

American Psychological Association (2020, October). *Stress in America 2020: A national mental health crisis*..https://www.apa.org/news/press/releases/stress/2020/report-october

Borden, K. A., & McIllvried, E. J. (online first). Continuing the call for better assessment of competencies: Comment on anatomy of psychology practitioner competence. *Clinical Psychology: Science and Practice*. doi: 10.1111/cpsp.12369

Borrego, Jr., J. (2018). It takes a village for meaningful and sustainable change in diversifying psychology. *Training and Education in Professional Psychology, 12*(4), 271-300. doi: 10.1037/tep0000224

Boyle, S. L., & Kenny, T. E. (2020). To disclose or not to disclose: Examining supervisor actions related to self-disclosure in supervision. *Journal of Psychotherapy Integration*, *30*(1), 36-43. doi: 10.1037/int0000181

Callahan, J. L., Collins, F. L., & Klonoff, E. A. (2010). An examination of applicant characteristics of successfully matched interns: Is the glass half full or half empty and leaking miserably? *Journal of Clinical Psychology, 66*(1), 1-16. doi: 10.1002/jclp.20664

Callahan, J. L., Hogan, L. R., Klonoff, E. A., & Collins, F. L., Jr. (2014). Predicting match outcomes: Science, practice, and personality. *Training and Education in Professional Psychology, 8,* 68-82. doi: 10.1037/tep0000030

Callahan, J. L., Swift, J. K., Hogan, L. R., Tompkins, K. A., Connor, D. R., & Klonoff, E. A. (2014). Seeking the "perfect" match: An empirical examination of applicant differences according to internship setting. *Psychotherapy Bulletin, 49*(1)*,* 13-18.

Callahan, J. L., Neumann, C. S., Cox, R. J., & Ruggero, C. J. (2017). The Supervisor Trainee Quarterly Evaluation (STQE): Psychometric support for use as a measure of competency. *Training and Education in Professional Psychology*, *11*(3), 117-127. doi: 10.1037/tep0000140

Callahan, J. L., Smotherman, J. M., Dziurzynski, K. E., Love, P. K., Kilmer, E. D., Flores Niemann, Y., and Ruggero, C. J. (2018). Diversity in the professional psychology training-to-workforce pipeline: Results from doctoral psychology student population data. *Training and Education in Professional Psychology*, *12*(4), 273-285. doi: 10.1037/tep0000203

Callahan, J. L. & Watkins, Jr., C. E. (2018a). The science of training I: Admissions,

curriculum, and research training. *Training and Education in Professional Psychology, 12*(4), 219-230. doi: 10.1037/tep0000205

Callahan, J. L., & Watkins, C. E. (2018b). The science of training III: Supervision, competency, and internship training. *Training and Education in Professional Psychology*, *12*(4), 245-261. doi: 10.1037/tep0000208

Callahan, J. L., Bell, D. J., Davila, J., Johnson, S. L., Strauman, T. J., & Yee, C. M. (2020). The enhanced examination for professional practice in psychology: A viable approach? *American Psychologist*, *75*(1), 52-65. doi: 10.1037/amp0000586

Callahan, J. L., Bell, D. J., Davila, J., Johnson, S. L., Strauman, T. J., & Yee, C. M. (2021). Inviting ASPPB to address systemic bias and racism: Reply to Turner et al. (2021). *American Psychologist, 76,* 167-168. doi: 10.1037/amp0000801

Callahan, J. L. & Braggs, P. (2020). Anatomy of psychology practitioner competency: A commentary. Manuscript submitted for publication. *Clinical Psychology Science and Practice.*

Chung, Y. B., Marshall, J. A., & Gordon, L. L. (2001). Racial and gender biases in supervisory evaluation and feedback. *The Clinical Supervisor, 20*(1), 99-111. doi: 10.1300/J001v20n01_08

Collins, Jr., F. L., Callahan, J. L., & Klonoff, E. A. (2007). A scientist-practitioner perspective of the internship match imbalance: The stairway to competence. *Training and Education in Professional Psychology*, *1*(4), 267-275. doi: 10.1037/1931-3918.1.4.267

Craig, P., Gordon, J., Clarke, R., & Oldmeadow, W. (2009). Prior degree and student assessment performance: How can evidence guide decisions on assessment policy? *Assessment & Evaluation in Higher Education*, *34*(5), 537-551. doi: 10.1080/02602930802187308

Dunn, R., Callahan, J. L., Farnsworth, J. K., & Watkins, C. E. (2017). A proposed framework for addressing supervisee-supervisor value conflict. *The Clinical Supervisor*, *36*(2), 203-222. doi: **Error! Hyperlink reference not valid.**10.1080/07325223.2016.1246395

El-Ghoroury, N. H., Galper, D. I., Sawaqdeh, A. & Bufka, L. A. (2012). Stress, coping, and barriers to wellness among psychology graduate students. *Training and Education in Professional Psychology, 6*(2), 122-134. doi:10.1037/a0028768

Forrest, L., Elman, N. S., Huprich, S. K., Veilleux, J. C., Jacobs, S. C., & Kaslow, N. J. (2013). Training directors' perceptions of faculty behaviors when dealing with trainee competence problems: A mixed method pilot study. *Training and Education in Professional Psychology*, *7*(1), 23-32. doi: 10.1037/a0032068

Gatmon, D., Jackson, D., Koshkarian, L., Martos-Perry, N., Molina, A., Patel, N., & Rodolfa, E. (2001). Exploring ethnic, gender, and sexual orientation variables in supervision: Do they really matter? *Journal of Multicultural Counseling and Development*, *29*, 102-113. doi: 10.1002/j.2161-1912.2001.tb00508.x

Gaubatz, M. & Vera, E. (2002). Do formalized gatekeeping procedures increase programs' follow-up with deficient trainees? *Counselor Education and Supervision, 41,* 294-305. doi: 10.1002/j.1556-6978.2002.tb01292.x

Gonsalvez, C. J., Deane, F. P., Terry, J., & Gooi, C. H. (online first). Anatomy of psychology practitioner competence: Clustering approach reveals the hierarchical organization of competencies. *Clinical Psychology: Science and Practice.* doi: 10.1111/cpsp.12347

Gonsalvez, C. J., & Freestone, J. (2007). Field supervisors' assessments of trainee performance: Are they reliable and valid? *Australian Psychologist*, *42*(1), 23-32. doi: 10.1080/00050060600827615

Gonsalvez, C. J., Deane, F. P., Knight, R. Shires, A., Nicholson Perry, K. Allan, C, & Blokas, V. The hierarchical clustering of clinical psychology practicum competencies: A multisite study of supervisor ratings. *Clinical Psychology Science and Practice, 22*(4), 390-403. doi:10.1111/cpsp.12123

Gregus, S. J., Stevens, K. T., Seivert, N. P., Tucker, R. P., & Callahan, J. L. (2020). Student perceptions of multicultural training and program climate in clinical psychology doctoral programs. *Training and Education in Professional Psychology, 14,* 293-307. doi: 10.1037/tep0000289

Hatcher, R. L., Fouad, N. A., Grus, C. L., Campbell, L. F., McCutcheon, S. R., & Leahy, K. L. (2013). Competency benchmarks: Practical steps toward a culture of competence. *Training and Education in Professional Psychology*, *7*, 84-91. doi:10.1037/a0029401

Helms, J. E. (1995). An update of Helms's white and people of color racial identity models. In J. G. Ponterotto, J. M. Casas, L. A. Suzuki, & C. M. Alexander (Eds.), *Handbook of multicultural counseling* (pp. 181–198). Thousand Oaks, CA: Sage.

Hogan, L. R., Callahan, J. L., Tompkins, K. A., Swift, J. K., Connor, D. R., & Klonoff, E. A. (2014). Subjective well-being among internship applicants. *Psychotherapy Bulletin, 49*(2)*,* 13-16.

Humphreys, L., Crino, R., & Wilson, I. (2017). Psychological functioning predicts competence development for postgraduate students of professional psychology. *Training and Education in Professional Psychology*, *11*(1), 49-56. doi: 10.1037/tep0000139

Jernigan, M. M., Green, C. E., Helms, J. E., Perez-Gualdron, L., & Henze, K. (2010). An examination of people of color supervision dyads: Racial identity matters as much as race. *Training and Education in Professional Psychology*, *4*(1), 62-73. doi: 10.1037/a0018110

Luebbe, A. M. & Ogbaselase, F. A. (2018). GREs, constriction of the educational pipeline for students of color at the point of entry to doctoral work in psychology: Commentary on Callahan and colleagues. *Training and Education in Professional Psychology, 12*(4), 291-294. doi: 10.1037/tep0000211

Lund, E. M. (2018). Disability as diversity in professional psychology: A continued call for research and action. *Training and Education in Professional Psychology, 12*(4), 295-296. doi: 10.1037/tep0000218

Mearns, J. & Allen, J. G. (1991). Graduate students' experiences in dealing with impaired peers, compared with faculty predictions: An exploratory study. *Ethics and Behavior, 1*(3), 191-202.

Mehr, K. E., Ladany, N., & Caskie, G. I. L. (2015). Factors influencing trainee willingness to disclose in supervision. *Training and Education in Professional Psychology*, *9*(1), 44-51. doi: 10.1037/tep0000028

Oliver, M. N., Bernstein, J. H., Anderson, K. G., Blashfield, R. K., & Roberts, M. C. (2004). An exploratory examination of student attitudes toward "impaired" peers in clinical psychology training programs. *Professional Psychology: Research and Practice, 35*(2), 141-147. doi: 10.1037/0735-7028.35.2.141

Parker, P. (2015). The historical role of women in higher education. *Administrative Issues Journal: Connecting Education, Practice, and Research*, *5*(1), 3-14. doi: 10.5929/2015.5.1.1

Price, S. D., Callahan, J. L., & Cox, R. J. (2017). Psychometric investigation of competency benchmarks. *Training and Education in Professional Psychology*, *11*(3), 128-139. doi: 10.1037/tep0000133

Pope-Davis, D. B., Liu, W. M., Nevitt, J., & Toporek, R. L. (2000). The development and initial validation of the Multicultural Environmental Inventory: A preliminary investigation. *Cultural Diversity and Ethnic Minority Psychology, 6,* 57– 64. http://dx.doi.org/10.1037/1099-9809.6 .1.57

Roberts, M. C. & Ostreko, A. (2018). GREs, public posting, and holistic admissions for diversity in professional psychology: Commentary on Callahan et al. *Training and Education in Professional Psychology, 12*(4), 286-290. doi: 10.1037/tep0000219

Rosenberg, J. I., Getzelman, M. A., Arcinue, F. A., & Oren, C. Z. (2005). An exploratory look at students' experiences of problematic peers in academic professional psychology programs. *Professional Psychology: Research and Practice, 36*(6), 665-673. doi: 10.1037/0735-7028.36.6.665

Scott, K. J., Ingram, K. M., Vitanza, S. A., & Smith, N. G. (2000). Training in supervision: A survey of current practices. *The Counseling Psychologist*, *28*(3), 403-422. doi: 10.1177/0011000000283007

Shen-Miller, D. S., Forrest, L., & Burt, M. (2012). Contextual influences on faculty diversity conceptualizations when working with trainee competence problems. *The Counseling Psychologist*, *40*(8), 1181-1219. doi: 10.1177/0011000011431832

Shen Miller, D. S., Forrest, L., & Elman, N. S. (2009). Training directors' conceptualizations of the intersections of diversity and trainee competence problems: A preliminary analysis. *The Counseling Psychologist*, *37*(4), 482-518. doi: 10.1177/0011000008316656

Shen-Miller, D. S., Grus, C. L., Van Sickle, K. S., Schwartz-Mette, R., Cage, E. A., Elman, N. S., Jacobs, S. C., & Kaslow, N. J. (2011). Trainees' experiences with peers having competence problems: A national survey. *Training and Education in Professional Psychology*, *5*(2), 112-121. doi: 10.1037/a0023824

Vacha-Haase, T., Davenport, D. S., & Kerewsky, S. D. (2004). Problematic students: Gatekeeping practices of academic professional psychology programs. *Professional Psychology: Research and Practice*, *35*(2), 115-122. doi: 10.1037/0735-7028.35.2.115.

Veilleux, J. C., January, A. M., VanderVeen, J. W., Reddy, L. F., & Klonoff, E. A. (2012). Differentiating amongst characteristics associated with problems of professional competence: Perceptions of graduate student peers. *Training and Education in Professional Psychology*, *6*(2), 113-121. doi: 10.1037/a0028337

Williams, T. R. & Raney, S. (2020). Relational cultural supervision enhances the professional development of postdoctoral residents of color in health service psychology. *Journal of Psychotherapy Integration, 30*(1), 140-146. doi: 10.1037/int0000169

Yourman, D. B., & Farber, B. A. (1996). Nondisclosure and distortion in psychotherapy supervision. *Psychotherapy: Theory, Research, Practice, Training*, *33*(4), 567-575. doi: 10.1037/0033-3204.33.4.567