



8-2021

Addressing the Inconsistencies in Counselor Skill Development and Measurement Through Deliberate Practice and Assessment Tools

Zach Budesa

University of Tennessee, Knoxville, zbudesa@vols.utk.edu

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss



Part of the [Counselor Education Commons](#)

Recommended Citation

Budesa, Zach, "Addressing the Inconsistencies in Counselor Skill Development and Measurement Through Deliberate Practice and Assessment Tools. " PhD diss., University of Tennessee, 2021. https://trace.tennessee.edu/utk_graddiss/6479

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Zach Budesca entitled "Addressing the Inconsistencies in Counselor Skill Development and Measurement Through Deliberate Practice and Assessment Tools." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Counselor Education.

Casey Barrio Minton, Major Professor

We have read this dissertation and recommend its acceptance:

Joel Diambra, Melinda Gibbons, R. Eric Heidel, Jennifer A. Morrow

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

**Addressing the Inconsistencies in Counselor Skill Development and Measurement Through
Deliberate Practice and Assessment Tools**

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Zach Budesa
August 2021

Dedication

*To my wife, **Sam**.*

This was only possible through your support, patience, and determination.

*To my **mom**.*

Your enthusiasm and encouragement made it possible for me to succeed.

Acknowledgments

*It is impossible to consider completing a Master's or PhD without the support of my mentor, friend, and colleague, Dr. **Chad Luke**.*

*Dr. **Casey Barrio Minton**, my dissertation chair, provided support and feedback that made the completion of this dissertation an incredible learning process.*

*My first shot at a dissertation study went poorly, and this work would not have been possible with my committee, Drs. **Joel Diambra**, **Melinda Gibbons**, **Eric Heidel**, and **Jennifer Morrow**, who helped me pull success from disaster.*

*I would also like to thank Dr. **Louis Rocconi**, from whom I learned most of the statistics I used, for his excellent teaching, support, and feedback.*

*Finally, I want to thank **Leigh Bagwell**, **Kristen Carlosh**, **Charaya Upton**, and **Adam York** for their roles in helping me feel less alone through school, comps, and a dissertation during a pandemic.*

Abstract

Counselor educators and supervisors have an ethical duty to monitor client welfare and student counselor development, but the tools and processes programs put in place may miss the mark. This dissertation proposes deliberate practice and assessment improvement as avenues for counselor educators to leverage existing skills and resources for the benefit of clients and students. In the first manuscript, deliberate practice is proposed as a framework which can unify the various methods and teaching practices which have proliferated over the history of counselor education. This manuscript reviews the literature surrounding counselor skill development and deliberate practice and provides examples of how these methods can be implemented together. The second manuscript explores existing methods of evaluating students in field experiences. Data from an existing student assessment ($n = 867$) are analyzed through Interclass Correlation Coefficients, Generalizability Theory, and Item Response Theory to identify common pitfalls and problems which may affect the quality of student evaluation. Taken together, these manuscripts advance ideas about how counselor educators can improve important facets of counselor education.

Table of Contents

Introduction	1
Chapter I: Enhancing Counselor Education and Supervision through Deliberate Practice	4
Enhancing Counselor Education and Supervision through Deliberate Practice	6
Deliberate Practice	8
The Process of Deliberate Practice	9
Deliberate Practice for Clinical Effectiveness	10
Integrating Deliberate Practice into Counselor Education and Supervision	12
Interpersonal Skills Courses	15
Other Curricular Applications	16
Pre- and Post-Graduate Supervision and Field Experiences	17
Limitations of Deliberate Practice	19
Implications for Future Research	20
Conclusion	22
References	23
Chapter II: A Psychometric Case Study of Field Experience Evaluations in A Single Counselor Preparation Program	30
Abstract	31
Evaluation of Students During Field Experiences	33
Difficulties of Student Evaluations in Field Experiences	35
Purpose of the Study	37
Method	37
Program and Participants	37
Instrument	39
Procedure	40
Data	40
Assessment	45
Decision Making Utility	45
Scale Information	47
Discussion	48
Implications for Scale Development and Improvement	51
Limitations	54
Conclusion	54
References	56
Appendices	62
Appendix A	62
Appendix B	64
Appendix C	65
Appendix D	66
Appendix E	67
Appendix F	68
Appendix G	69
Conclusion	70
Vita	72

List of Tables

Table 1	64
Table 2	65
Table 3	66

List of Figures

Figure 1	67
Figure 2	68
Figure 3	69

Introduction

Reliable processes for counselor skill development and assessment are key to ensuring client welfare and counselor development. Counselor educators and counseling programs have ethical and educational requirements to facilitate and monitor students and supervisees' development of counseling skills and professional dispositions (ACA, 2014; CACREP, 2016, 2021). Not only must counselor educators attend to these factors of counselor development, they must be able to provide students and supervisees quality feedback, engage in reasonable gatekeeping practices, and only endorse individuals who meet the requirements of entrance into the profession. If counselor educators fail to uphold these expectations, they can result in harm to clients, counselors, the mental health fields, and the counseling profession.

Despite the importance of skill development and counselor assessment counselor educators find themselves needing to make important decisions with conflicting information. While ethical standards (ACA, 2014) and accreditation standards (CACREP, 2016, 2021) emphasize the importance of counselor development and assessment, the leaders of counseling's professional bodies provide little to no guidance which educators can apply in their own work. At the same time, counselor education programs exist within the world of higher education, which includes pressures to admit more students, ensure institutional goals are met, and graduate students in a timely fashion. Programs themselves have their own special interests, which, whether faculty- or funder-inspired, may influence which skills are the focus of teaching or assessment. Without clear standards, the educators and programs in the counseling profession have developed idiosyncratic ways of facilitating and measuring counseling skills and professional dispositions.

The process of teaching counseling skills continues to be a hot topic in the counseling and psychotherapy professions. One of the early methods of teaching counseling skills, microskills, continues to be the subject of debate and disagreement (Ridley et al., 2011). In order to accommodate class sizes that prevent individual attention, some educators have sought group-based methods (Paladino et al., 2011), team methods (Landis & Young, 1994), and other, less clinically focused, methods. While a plethora of methods can help counselor educators reach students who learn in a variety of ways, questions remain regarding how much any teaching or supervision method affects client outcomes (Erekson et al., 2017; Watkins, 2011).

The methods available for evaluating counselor development are similarly broad and varied. Reviews of published and program-based counselor assessments have shown that there is a wide variety of existing assessments (Kemer et al., 2017; Tate et al., 2014). The existing assessments overlap on key topics in counselor education (e.g., counseling skills, professional dispositions, multicultural skills, etc.), but the lack of psychometric vetting of many of them leads to inconsistent assessments of student ability. The use of program-created assessments of students can be useful, but when this results in inconsistency between verbal discussions and written assessments of students, it becomes difficult to justify gatekeeping and remediation decisions (Personal Communication, Barrio Minton, May 2021).

The purpose of this dissertation is to consider how counselor educators can address the inconsistencies which have resulted from a proliferation of methods for teaching and evaluating counseling skills. The first manuscript proposes Deliberate Practice (Ericsson et al., 1993) as a method to incorporate existing elements of counselor education into a more unified framework. Deliberate practice is an evidence-based method which emphasizes the importance of regular constructive feedback, solo practice, and scaffolding to develop and maintain skill. Only recently

applied in counseling and psychotherapy (Chow et al., 2015), deliberate practice shows promise as a method for facilitating counselor development. Manuscript one provides examples for the application of deliberate practice across the counselor education curriculum and in supervision.

The second manuscript is a psychometric analysis of a single program's assessment of counseling students' performance during field experiences. Using psychometric tools including Generalizability Theory (Briesch et al., 2014) and Item Response Theory (Embretson & Reise, 2000), this analysis explores the results of an internally developed, under-studied assessment form and resulting pitfalls programs may encounter. The results of this analysis are considered in the context which the program collects and uses the assessment in question. Based on this analysis, recommendations about developing and implementing assessments in counselor education field experiences are discussed.

Chapter I: Enhancing Counselor Education and Supervision through Deliberate Practice

Abstract

Professional and student counselors must develop and maintain their counseling skills in order to engage in ethical and effective counseling. Recent research suggests that graduate education and supervision has little effect on client outcomes. Investigations of expertise development have given rise to deliberate practice, a framework which structures skill development research and instruction. Deliberate practice involves individualized coaching, repetition, and solo practice in development of optimal performance. This conceptual article introduces deliberate practice as a framework for enhancing effectiveness of counselor education and supervision practices. Applications of deliberate practice to teaching and supervision and suggestions for future research are provided.

Keywords: Deliberate Practice, Skill Development, Counselor Education

Enhancing Counselor Education and Supervision through Deliberate Practice

Interpersonal skills are essential for helping clients achieve positive outcomes in counseling (Anderson et al., 2009; Wampold & Imel, 2015). The evidence-base for interpersonal skills is supported by the codification of relational skills in educational and ethical standards within counselor education. Council for Accreditation of Counseling and Related Educational Programs (CACREP; 2016) *Standards* require programs to attend to methods for developing and maintaining relationships (2.F.5.d), counselor characteristics that influence counseling (2.F.5.f), essential interviewing and counseling skills (2.F.5.g), and evidence-based counseling strategies (2.F.5.j). Following graduation, counselors have an ethical duty to maintain skills which facilitate client outcomes (American Counseling Association [ACA], 2014).

The centrality of interpersonal skills for effective counseling has led researchers and educators to develop models which help student counselors learn these skills. Skills training methods over the past five decades include Ivey's (1971) microskills focused on enactment of specific behaviors, the Triad Training Model (Pedersen, 1977) focused on developing self-other awareness and perspective taking, and the Reflecting Team Model (Landis & Young, 1994) in which student counselors provided immediate feedback to each other. Further, Paladino et al. (2011) proposed the Interactive Training Model to improve realism of role plays and integrate strengths of previous models. Other educators proposed mindfulness training (Gockel et al., 2013) and theatrical improvisation (Romanelli & Berger, 2018) to facilitate student counselors' interpersonal skill development. Although each of these methods have their supporters, the mental health professions have limited empirical evidence about which methods facilitate development of effective clinical skills.

Evidence regarding the impact of graduate education and clinical supervision on helping professionals' interpersonal skill development and ability to facilitate client outcomes remains mixed. In two studies, clinical psychology graduate students' work with clients showed consistency in client-reported wellbeing outcomes regardless of clinicians' level of education (Budge et al., 2013; Erekson et al., 2017). In other studies, clients working with more experienced clinicians reported reduced symptomology (Hill et al., 2015), were less likely to terminate prematurely (Rønnestad et al., 2019), and experienced improvement in fewer sessions (Erekson et al., 2017) compared to clients working with less experienced clinicians. In longitudinal studies of supervision, participation in supervision did not lead to improved client outcomes (Rousmaniere et al., 2016). Watkins (2011) argued that the empirical evidence did not answer the question regarding whether supervision impacted client outcomes. In all, this evidence suggests that education and supervision may need a shift to greater focus on enhancing counselor skills and, in turn, improving client outcomes.

In recent years, researchers in clinical and counseling psychology advocated the use of deliberate practice to improve practitioner effectiveness. Deliberate practice is based on the science of expertise development (Ericsson, 2019). Research on deliberate practice spans a variety of fields and shows that much of what is considered expert performance is based on effort and practice, not on innate talent or biological factors (Ericsson et al., 1993). Deliberate practice offers an evidenced-based teaching and learning method from which counselor educators and supervisors can frame skill development. To aid counselor educators in including deliberate practice in their work, this article includes explorations of deliberate practice and applications to the unique context of counselor education and supervision.

Deliberate Practice

Deliberate practice is a framework for engaging in focused, intense practice in a spirit of continuous improvement that develops and maintains optimal performance (Ericsson et al., 1993). Deliberate practice requires the involvement of a more expert individual who facilitates focused repetitive solo practice (Ericsson, 2019; Ericsson et al., 1993). The more experienced teacher helps to design practice plans; facilitate goal setting, feedback, and repetition; assess performance; and sequence tasks with increasing difficulty (Ericsson, 2019).

Deliberate practice was first researched in music where researchers sought to understand what differentiates the top performers from more average performers. Ericsson et al. (1993) used a variety of novel methods to illustrate how world-class violinists and pianists developed and maintained a high level of skill. Deliberate practice research now spans many fields, including athletic coaching (Coughlan et al., 2014; Ericsson, 2019), medical education (Barrett-Naylor et al., 2020; Ericsson, 2008), and competitive games like SCRABBLE and chess (Ericsson et al., 1993; Tuffiash et al., 2007). More recently, deliberate practice has been applied to counseling and psychotherapy (Chow et al., 2015; Goldberg, Rousmaniere et al., 2016).

The same framework first established with musicians holds true for expert performers across a wide variety of fields. Because deliberate practice crosses academic and professional fields, some foundation literature refers to coaches, performers, and trainees and considers topics like expert, superior, or elite performance. These are related to counselors, counselor educators, supervisors, supervisees, students, and teachers within a counselor education paradigm where the focus is on highly effective counseling skills. For clarity, we will use “teachers” and “students” unless discussing clinical supervision.

The Process of Deliberate Practice

Through its history, the key components of deliberate practice have been refined, and Ericsson (2019) has explicated the three major tasks: individualized design of effective practice, active responses to a task, and individualized assessment. In the first task, teachers and students develop practice tasks by identifying the most important area for improvement. Teachers provide clear and direct guidance about more effective ways to perform (Ericsson, 2019). Active responses to a task necessitate that teachers and students set an unambiguous goal, students engage in repetitive solo practice, and teachers provide students with immediate formative feedback (Ericsson et al., 1993). In the final task, teachers assess performance and decide when students should transition to the next task, which is set just beyond their current abilities (Ericsson, 2019). To balance the challenge of training, teachers organize tasks in increasing difficulty. Teachers' role in deliberate practice is to support students' practice, but this role cannot overshadow the importance of students' individual efforts.

Enhancing performance through deliberate practice also requires solo practice, environmental support, intrinsic motivation, and unique practice strategies. In interdisciplinary studies of expert performance, superior performers differentiated themselves from average performers through solo practice, with elite performers practicing around twice as much as more average performers (Chow et al., 2015; Ericsson, 2019; Ericsson et al., 1993; Tuffiash et al., 2007). Solo practice can include reviewing recordings of performance (Chow et al., 2015; Tuffiash et al., 2007), repeated practice of discrete movements (Coughlan et al., 2014; Ericsson, 1993), and repeating and refining affective and verbal responses (Barrett-Naylor et al., 2020).

Although rewarding, deliberate practice is not enjoyable or easy (Coughlan et al., 2014; Ericsson, 2016; Ericsson et al., 1993), and students need environmental support and effective

coping strategies to benefit (Goldberg, Babins-Wagner, et al., 2016). Because deliberate practice is not intrinsically enjoyable, students must maintain deliberate focus on their practice task in order to benefit (Coughlan et al., 2014; Ericsson et al., 1993).

Finally, experts practice in different ways than non-experts. By working with experienced teachers, students can benefit from practice strategies such as focusing on addressing limitations rather than reinforcing strengths, investigating unique practice strategies, and repeating discrete skills until attaining desired performance (Coughlan et al., 2014; Ericsson et al., 1993; Tuffiash et al., 2007). Focused solo practice in combination with supervised work can enable performers to develop expertise in a wide variety of fields.

Deliberate Practice for Clinical Effectiveness

Although still in its early days and somewhat piecemeal in approach, deliberate practice holds promise for improving enhancing counselor skills and, in turn, impacting client outcomes. In an adaptation of the original deliberate practice study (Ericsson et al., 1993), Chow et al. (2015) showed that clients experienced the greatest reduction in symptoms and increase in wellbeing when working with clinicians who spent more than double the amount of time in solo practice than clinicians with average client outcomes. Clinicians most often engaged solo practice of reviewing difficult cases and reflecting on past and future sessions. In another study, providing a supportive environment in which clinicians engaged in deliberate practice resulted in a greater reduction in client symptoms over time (Goldberg, Babins-Wagner, et al., 2016). In samples of graduate psychology students, modeling effective methods of responding to clients and providing opportunity for repetitive role plays with immediate feedback resulted in improvement in interpersonal skills assessed by experienced raters (Anderson et al., 2019). Graduate students in psychology who received deliberate practice coaching focused on specific

skills also developed greater self-efficacy (Hill et al., 2019) and became more accepting of their own mistakes (Rosén, 2020), while graduate counseling students improved case conceptualization skills (Lipp, 2019). Finally, psychology graduate students and psychiatric nurses demonstrated improved interpersonal skills and interventions following engagement in workshops emphasizing repetitive practice and immediate feedback (Barrett-Naylor et al., 2020; Westra et al., 2020).

Along with positive results from initial research, educators and supervisors developed methods to apply deliberate practice to aid clinical supervisors and mental health clinicians. Combining deliberate practice and Feedback-Informed-Treatment (Prescott et al., 2017), the Expertise-Development Model (EDM; Rousmaniere et al., 2017) is a supervision and consultation model which incorporates client feedback into goal setting and skill development. Other recent deliberate practice applications included a self-help guide for clinicians wanting to integrate deliberate practice into their work (Rousmaniere, 2016) and a guide to intrapersonal skill development (Rousmaniere, 2019). More recent work includes a guide on the science and application of deliberate practice for clinicians (S. D. Miller et al., 2020) and an organization which certifies clinical deliberate practice trainers (Rousmaniere & Vaz, n.d.), which includes new self-directed training for multicultural skills. These efforts are gaining steam in psychology while raising questions about how deliberate practice may help professional counselors (Clements-Hickman & Reese, 2020).

Many components of deliberate practice will be familiar to counselor educators and are found throughout the teaching and supervision literature. Chow et al. (2015) found that the most effective practice activities were reviewing difficult cases alone, attending workshops focused on specific approaches to counseling, and reflecting on past and future sessions. Along with reading

and learning through self-study and supervision, these activities are common within counselor education. Counselor educators will also recognize the importance of articulating learning outcomes clearly and providing regular feedback throughout the developmental process (Best Practices in Clinical Supervision Task Force, 2011; CACREP, 2016; Wood et al., 2016). The professional identity literature highlights clearly the role of more experienced mentors in counselors' developmental process (Gibson et al., 2010), and CACREP (2016) standards are designed to ensure appropriate educator experience and preparation. Each of these components represents a key aspect of deliberate practice (Ericsson, 2019). Deliberate practice does not replace previous approaches; instead, it provides a framework that counselor educators can use to direct the tools already available within the field more effectively.

Integrating Deliberate Practice into Counselor Education and Supervision

Deliberate practice is a flexible framework that counselor educators can use individually, in triads, or in larger groups with student and professional counselors at all stages of development. Across contexts, counselor educators' use of deliberate practice remains the same: a) assess the quality of counselors' performance; b) aid counselors in finding more effective ways to connect with clients; c) structure practice activities that include an explicit goal, skill repetition, and immediate feedback; and d) monitor practice and scaffold from simpler to more complex tasks (Ericsson, 2019). Counselor educators can return to this basic framework in order to develop classroom, supervision, or solo practice activities. Because this framework can be used with many individuals and situations, the activities present here and elsewhere should be seen as building blocks of practice design. Not only can the activities be adapted to individual goals, but they can also be repeated throughout counselor development.

Solo practice is integral to high quality skill development (Chow et al., 2015; Ericsson, 2019), although it may be the least used method of practice within counselor education. Solo practice involves active engagement with repetitive practice on a specific skill. Chow et al. (2015) investigated practice activities including reflecting on previous sessions, reviewing recordings of challenging sessions and case records, planning for future sessions, and watching experienced clinicians' sessions. Solo practice methods can involve use of audio and video recordings, transcripts, or role plays. Solo practice could even use non-counseling media prompts to practice responding to strong emotions or difficult situations (Rousmaniere, 2019). Counselor educators can use brief practice during class or supervision sessions to assess and correct skills and provide direction for future homework. Developing homework tailored to student counselors' developmental ability offers an opportunity to engage in solo practice that is challenging but rewarding.

Counselor educators hoping to integrate deliberate practice into their teaching and supervision may find practice diaries and recordings useful. In studies of deliberate practice, participants used practice diaries to record their practice, work, school, leisure, and daily routine activities in 15-minute increments (Ericsson et al., 1993). Student counselors could be encouraged or required to keep a diary in which they track time, length, focus, and experiences related to solo practice (Rousmaniere, 2016). By expecting practice time to be tracked, counselor educators may encourage students and supervisees to increase their practice time. Counselor educators and supervisors might also assign a minimum amount of solo practice time each week (one to three hours) to be aligned with assignments and recorded in a practice diary or journal.

Recordings are a common tool for counselor educators to assess student performance and are used throughout counselor education curriculum. Because practice sessions and counseling

sessions are often recorded for assignments, they can provide opportunities to identify and practice different responses. Counselor educators may instruct students to review their recordings alone or with a teacher, pause after each response, and either identify a better response or explain why the response they made was optimal. Students may also “think aloud” as they review recordings to make their cognitive process clearer. As student counselors refine their skills, repetition allows them to develop more self-efficacy in using different interpersonal skills. Recordings can also be used to facilitate feedback to student counselors, and teachers and supervisors can explore alternative responses with students.

In order to manage time and include deliberate practice for multiple student counselors, counselor educators may vary group sizes and activity timing. Multiple models of skill instruction already include triadic, small group, and large group practice (Paladino et al., 2011), and deliberate practice can be incorporated as needed. Instructors can pair or group student counselors at the beginning of class and provide brief case studies or prompts focused on a specific skill for the class. Student counselors can practice this single skill and giving each other feedback. Instructors then have the opportunity to conduct a brief assessment of performance and provide feedback, modeling, or shaping of responses. Large group or triadic settings allow for role play and immediate feedback, with supervisor or teacher-led rehearsal.

From introductory skills courses through field placements and supervision, the deliberate practice framework can provide effective practice for student and professional counselors. In the remainder of this section, we illustrate specific applications of deliberate practice within counselor education.

Interpersonal Skills Courses

Introductory skills courses are often one of the first courses new students encounter, and they focus on helping students develop important interview and counseling skills while beginning to integrate them with personal characteristics (CACREP, 2016). An advanced goal in these courses is for students to develop specific relational skills and be able to enact them in appropriate time and context. Deliberate practice can fit easily into skills courses as students focus on the basic skills counseling researchers have established as the smallest microskills. Although microskills have been subject to criticism (Ridley et al., 2011), microskills define clear goals for performance, and counselor educators can quickly assess student performance and provide feedback. Role play and mock counseling sessions, whether in pairs, triads, or larger groups provide an excellent opportunity for skills practice.

To improve skill selection and use, counselor educators use role play activities in which student counselors focus on using a specific skill. In these role plays, one student acts as a counselor while another acts as a client. When the “counselor” stumbles on a skill, the pair can replay the moment and practice the skill until the “counselor” feels more comfortable with said skill. Teacher may coach the “counselor” on the use of a particular skill, providing the teacher with an opportunity to engage in assessment and help define practice goals and activities with students. Other coaching methods might include directing the “client” to repeat moments while the “counselor” uses different skills. Conducting practice in this manner will focus on one or two microskills or responses at first. As student counselors develop greater skills levels, teachers can increase complexity by including more skills or responses to practice or extending the time of practice.

Other Curricular Applications

Deliberate practice can be applied to conceptual and clinical skills across the counselor education curriculum. Lipp (2019) focused on helping students develop case conceptualization skills through three fifteen-minute coaching sessions. These coaching sessions included 1) a review of recent homework (either from class or from previous coaching sessions) and a discussion of personal strengths and limitations, 2) a discussion about deliberate practice and how to use it, and 3) an individualized practice exercise designed to push students into the zone of proximal development. Each session concluded with formative feedback and a homework assignment. These assignments scaled in difficulty as sessions progressed and included expectations that students focus on their internal experiences and struggles as they completed the work. Participants who received deliberate practice coaching showed improvement in case conceptualization skills when compared with a control group. Deliberate practice for conceptualization skills may be readily applied within context of counseling theories, developmental theories, multicultural and social justice conceptualization, and diagnostic decision-making.

Additional opportunity for applying deliberate practice to more advanced counseling skills may include coaching students in assessment use, suicide assessment, broaching cultural differences, interprofessional communication, or treatment planning. For example, the Deliberate Practice Institute (Rousmaniere & Vaz, n.d.) hosts a Multicultural Deliberate Practice training which guides counselors' practice broaching and addressing cross-cultural topics with clients. Counselor educators might assign these modules, ultimately viewing students' responses to simulated content and providing directive feedback for practice.

Pre- and Post-Graduate Supervision and Field Experiences

Counseling field experiences and supervision provide a context where deliberate practice can be adapted with little modification. Supervision is required in counseling programs, field placements (CACREP, 2016) and post-graduate licensure (Kaplan & Kraus, 2018). Field experience instructors and supervisors who utilize deliberate practice (Ericsson, 2019) will assess supervisees' strengths and limitations, identify goals for practice, provide supervisees with assistance in better connecting with clients, develop clear goals, provide immediate feedback, and expect repetitive practice. Counselor educators and supervisors already engage in constant assessment of supervisees to identify when they reached a desired level of proficiency and are ready to practice more complex tasks. Each of these tasks contributes to the overall framework of deliberate practice and is necessary for the development of skilled performance (Ericsson, 2019). Students and supervisees benefit most from deliberate practice when supervisors emphasize the importance of repetitive solo practice and maintaining motivation. By providing structure and entrusting supervisees with the responsibility for practice, supervisors may aid supervisees in developing a consistent self-supervision practice.

Practicum and internship field experiences are often the first exposure to clients and students for student counselors. While these students are more advanced, they may need to review basic counseling skills and begin to work on advanced skills. The interpersonal skills practice activities identified above can still be used, and more advanced students may focus on developing more complex skills and improving self-efficacy. Practicum students may be tasked with completing assessments or setting goals in their work with clients, and deliberate practice-minded teachers would help them identify areas for improvement and assign repetitive practice.

As new counselors progress from to internship and beyond graduation, they must develop advanced skills, caseload management, and complex conceptualizations of clients. To continue challenging new counselors, supervisors can use deliberate practice focused on theory-based conceptualization and intervention skills. Theory-based interventions, such as Motivational Interviewing (W. R. Miller & Rollnick, 2013), are more complex than relational skills and can be an excellent target for deliberate practice. In much the same way, counselor educators can identify areas for improvement and facilitate student counselors' practice of these skills. When considering theory-based interventions, counselor educators may begin by having the student counselors focus on rolling with a single statement of resistance. As with other applications of deliberate practice, activities should be organized from simple (the single statement) to more complex (session-length applications of theory) to facilitate development.

One example of integrating deliberate practice into field experience curriculum or new counselors is through ASIST workshops (Lang et al., 2013). As described by Shannonhouse et al. (2018), ASIST is a multiday suicide intervention training which includes lectures, discussions, and role plays. While taking part in an ASIST workshop, participants practice, receive feedback, and learn better ways of intervening with persons with thoughts of suicide. Much like Westra et al.'s (2020) deliberate practice-based workshop, participants show an improvement in skills, knowledge, and competence following these brief interventions (Shannonhouse et al., 2018). In much the same way, student counselors can use repetitive practice and receive immediate feedback about risk assessment skills, they may develop stronger skills when faced with the reality of assessing clients' level of risk.

Limitations of Deliberate Practice

Although performers across fields experience show improvement through deliberate practice, it has drawbacks and limitations applicable to counselor education. First and foremost, research in counseling and psychotherapy has been limited. This body of research is growing but needs continued exploration by clinicians and educators. Second, counselor education programs often operate with limited time, funding, and content-saturated curricula. In order to accommodate small course sizes in field experience, some programs offer large didactic courses for other core areas. Under these conditions, devoting time to individualized assessment of students and personalized practice plans can be time consuming for counselor educators. Although supervisors may be better positioned to focus on specific skills with supervisees, supervisees may need assistance with a broad number of concerns or cases in each session, reducing time available for focused development of a singular skill. Performers report that deliberate practice activities hold limited intrinsic enjoyment (Chow et al., 2015; Ericsson et al., 1993), which requires educators and student counselors to develop and maintain motivation to continue the high level of solo practice that contributes to optimal performance. Together, these may present logistical challenges to implementing deliberate practice with fidelity.

Perhaps most importantly, the expectation that students focus efforts and feedback on weaker areas may seem antithetical to the strengths-based foundations of the counseling profession. Some counselor educators may be concerned that the emphasis on “weakness” or “limitations” goes against one of counseling’s core philosophies. Similarly, some counselor educators may be reluctant to consider themselves experts who hold the answers and should model “best” responses to a variety of client situations. Rather, we encourage counselor educators to consider deliberate practice from a growth mindset in which they integrate scientific

evidence of expert performance, an acceptance of working with limitations or growth edges as necessary for improvement, and capacity for optimal counselor development. Just as counselors co-construct goals with clients, counselor educators can co-construct personalized targets and areas of emphasis with students.

Implications for Future Research

Despite a broader evidence base that goes back three decades, deliberate practice is still an innovative practice within counselor education. The first investigation of deliberate practice in counseling and psychotherapy was published in 2015 (see Chow et al., 2015). Since that time, deliberate practice research in the mental health professions accounts for a small number of publications. Most of these studies focused on professional mental health clinicians or psychology doctoral students, not student counselors. In order to better understand the value deliberate practice may hold, counselor educators must consider how deliberate practice can inform work with student counselors, prioritize the skills important in counselor education, and operationalize what constitutes skilled performance in counseling.

The first question is whether deliberate practice is an effective method for improving the abilities of counselors and, in turn, producing better client outcomes. Conclusions from previous research are mixed, with some authors finding skill improvement resulting from deliberate practice (Anderson et al., 2019), while others finding little to no effect on interpersonal skills (Hill et al., 2019). One method which may appeal to counselor education researchers is developing a specific protocol for using deliberate practice in interpersonal skill development. This protocol will allow researchers to compare the use of deliberate practice with a control group who does not use the protocol. Research along these lines could use Anderson et al.'s (2009) Interpersonal Facilitative Skills assessments, Lambie et al.'s (2018) Counselor

Competency Scale, or client outcomes measures to assess counselor performance. Using the deliberate practice protocol as an intervention in a multiple baseline single case design with client outcomes as the outcome variable may also provide evidence that deliberate practice can improve student counselors' client outcomes. Finally, by assessing supervisees' solo practice through practice diaries or assessments designed to measure engagement in and motivation for deliberate practice, researchers may evaluate the quality of solo practice associated with deliberate practice and improved performance.

The optimal focus for deliberate and solo practice is another question that should be addressed within counselor education and supervision. Within clinical education broadly, there remains debate about what skills clinicians should possess to facilitate client outcomes. Microskills, like those measured by the Counselor Competency Scale-Revised (Lambie et al., 2018), are common but may miss nuance of broader relational skills such as instilling hope and positive expectations and navigating rupture and repair (Ridley et al., 2011). Clements-Hickman and Reese (2020) argued that research into therapist effects and client change has not provided clear evidence of a specific skill set which can be enhanced through practice. Some efforts, including the Facilitative Interpersonal Skills (Anderson et al., 2009) and Evidence-Based Relationships (Parrow et al., 2019) have attempted to operationalize these skills, but lack necessary validation. Clarifying these skills represents an important step toward agreement about what makes counselors effective and developing effective methods to practice these skills.

Important to deliberate practice is the reality that experts practice, perform, and self-assess differently. There is no current agreement about what comprises expertise or superior performance in counseling. Previous researchers used clients' outcomes (e.g., Chow et al., 2015; Hill et al., 2015; Rønnestad et al., 2019) and measures of interpersonal skill (Anderson et al.,

2009) to measure clinician performance. Researchers can draw on previous research regarding expertise development in counseling and statistical modeling to facilitate exploration of skilled clinicians. Once high performing counselors are identified, researchers can assess their practice efforts via practice diaries (Chow et al., 2015; Ericsson et al., 1993), “think aloud” techniques, or other novel methods to disentangle the strategies they use (Ericsson & Smith, 1991). By establishing a clearer picture of outcomes and process of high-quality skill development in counseling, counselor educators can focus more on how to help student counselors develop these skills more effectively.

Conclusion

Based on the science of expertise development, deliberate practice differentiates top performers from more average performers in a variety of fields (Ericsson, 2019; Ericsson et al., 1993; Tuffiash et al., 2007). Recently, psychologists began investigating if deliberate practice holds the same promise for mental health professionals. Deliberate practice resembles a variety of current practices in counselor education and can provide a framework for facilitating greater skill development among student counselors. With further research into methods, targets, and outcomes of deliberate practice, deliberate practice holds promise for helping counselors improve the quality of care they provide to clients.

References

- American Counseling Association. (2014). *ACA code of ethics*. Author.
- Anderson, T., McClintock, A. S., Himawan, L., Song, X., & Patterson, C. L. (2015). A prospective study of therapist facilitative interpersonal skills as a predictor of treatment outcome. *Journal of Consulting and Clinical Psychology, 84*(1), 57–66.
<https://doi.org/10/f75xp6>
- Anderson, T., Perlman, M. R., McCarrick, S. M., & McClintock, A. S. (2019). Modeling therapist responses with structured practice enhances facilitative interpersonal skills. *Journal of Clinical Psychology, 76*(4), 659–675. <https://doi.org/10/gg2z57>
- Barrett-Naylor, R., Malins, S., Levene, J., Biswas, S., Mays, C., & Main, G. (2020). Brief training in psychological assessment and interventions skills for cancer care staff: A mixed methods evaluation of deliberate practice techniques. *Psycho-Oncology, 1*–8.
<https://doi.org/10/gg2z6c>
- Best Practices in Clinical Supervision Task Force. (2011). *Best practices in clinical supervision*. Author.
- Budge, S. L., Owen, J. J., Kopta, S. M., Minami, T., Hanson, M. R., & Hirsch, G. (2013). Differences among trainees in client outcomes associated with the phase model of change. *Psychotherapy, 50*(2), 150–157. <https://doi.org/10/f42vk6>
- Chow, D. L., Miller, S. D., Seidel, J. A., Kane, R. T., Thornton, J. A., & Andrews, W. P. (2015). The role of deliberate practice in the development of highly effective psychotherapists. *Psychotherapy, 52*(3), 337–345. <https://doi.org/10/dc79>

- Clements-Hickman, A. L., & Reese, R. J. (2020). Improving therapists' effectiveness: Can deliberate practice help? *Professional Psychology: Research and Practice*.
<https://doi.org/10/gg2z69>
- Coughlan, E. K., Williams, A. M., McRobert, A. P., & Ford, P. R. (2014). How experts practice: A novel test of deliberate practice theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*(2), 449. <https://doi.org/10/gg2z7j>
- Council for Accreditation of Counseling and Related Educational Programs [CACREP]. (2016). *2016 CACREP Standards*. <http://www.cacrep.org/wp-content/uploads/2017/08/2016-Standards-with-citations.pdf>
- Erekson, D. M., Janis, R., Bailey, R. J., Cattani, K., & Pedersen, T. R. (2017). A longitudinal investigation of the impact of psychotherapist training: Does training improve client outcomes? *Journal of Counseling Psychology*, *64*(5), 514–524. <https://doi.org/10/gcgsgnb>
- Ericsson, K. A. (2008). Deliberate practice and acquisition of expert performance: A general overview. *Academic Emergency Medicine*, *15*(11), 988–994. <https://doi.org/10/d2v2n7>
- Ericsson, K. A. (2016). Summing up hours of any type of practice versus identifying optimal practice activities: Commentary on Macnamara, Moreau, & Hambrick (2016). *Perspectives on Psychological Science*, *11*(3), 351–354. <https://doi.org/10/gf7rft>
- Ericsson, K. A. (2019). Towards a science of the acquisition of expert performance in sports: Clarifying the differences between deliberate practice and other types of practice. *Journal of Sports Sciences*, *38*(2), 159–176. <https://doi.org/10/gg2z7x>
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406.
<https://doi.org/10/dfs2vp>

- Ericsson, K. A., & Smith, J. (1991). Prospects and limits in the empirical study of expertise: An introduction. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (pp. 1–38). Cambridge University Press.
- Gibson, D. M., Dollarhide, C. T., & Moss, J. M. (2010). Professional identity development: A grounded theory of transformational tasks of new counselors. *Counselor Education and Supervision, 50*(1), 21–38. <https://doi.org/10/fx5zx7>
- Gockel, A., Burton, D., James, S., & Bryer, E. (2013). Introducing mindfulness as a self-care and clinical training strategy for beginning social work students. *Mindfulness, 4*(4), 343–353. <https://doi.org/10/dsnh>
- Goldberg, S. B., Babins-Wagner, R., Rousmaniere, T. G., Berzins, S., Hoyt, W. T., Whipple, J. L., Miller, S. D., & Wampold, B. E. (2016). Creating a climate for therapist improvement: A case study of an agency focused on outcomes and deliberate practice. *Psychotherapy, 53*(3), 367–375. <https://doi.org/10/f85zgz>
- Goldberg, S. B., Rousmaniere, T. G., Miller, S. D., Whipple, J., Nielsen, S. L., Hoyt, W. T., & Wampold, B. E. (2016). Do psychotherapists improve with time and experience? A longitudinal analysis of outcomes in a clinical setting. *Journal of Counseling Psychology, 63*(1), 1–11. <https://doi.org/10/f77p7d>
- Hill, C. E., Baumann, E., Shafran, N., Gupta, S., Morrison, A., Rojas, A. E. P., Spangler, P. T., Griffin, S., Pappa, L., & Gelso, C. J. (2015). Is training effective? A study of counseling psychology doctoral trainees in a psychodynamic/interpersonal training clinic. *Journal of Counseling Psychology, 62*(2), 184–201. <https://doi.org/10/gfxcqh>
- Hill, C. E., Kivlighan, D. M., Rousmaniere, T. G., Kivlighan, D. M., Gerstenblith, J. A., & Hillman, J. W. (2019). Deliberate practice for the skill of immediacy: A multiple case

- study of doctoral student therapists and clients. *Psychotherapy*, 1–11.
<https://doi.org/10/gf7862>
- Ivey, A., E. (1971). *Microcounseling: Innovations in interviewing training*. Charles C Thomas.
- Kaplan, D. M., & Kraus, K. L. (2018). Building blocks to portability: Culmination of the 20/20 initiative. *Journal of Counseling & Development*, 96(2), 223–228.
<https://doi.org/10/gg4d7k>
- Lambie, G. W., Mullen, P. R., Swank, J. M., & Blount, A. (2018). The counseling competencies scale: Validation and refinement. *Measurement and Evaluation in Counseling and Development*, 51(1), 1–15. <https://doi.org/10/dkrj>
- Landis, L. L., & Young, M. E. (1994). The reflecting team in counselor education. *Counselor Education and Supervision*, 33(3), 210–218. <https://doi.org/10/fx52dt>
- Lang, W. A., Ramsay, R. F., Tanney, B. L., Kinzel, T., Turley, B., & Tierney, R. J. (2013). *ASIST trainer manual* (11th ed.). LivingWorks Education.
- Lipp, S. L. (2019). *The effects of case conceptualization training and deliberate practice coaching on counselor competence* [Doctoral dissertation]. Florida Atlantic University.
- Miller, S. D., Hubble, M. A., & Chow, D. L. (2020). *Better results: Using deliberate practice to improve therapeutic effectiveness*. American Psychological Association.
- Miller, W. R., & Rollnick, S. (2013). *Motivational interviewing: Helping people change* (3rd ed). Guilford Press.
- Owen, J., Drinane, J. M., Kivlighan, M., Miller, S. D., Kopta, M., & Imel, Z. (2019). Are high-performing therapists both effective and consistent? A test of therapist expertise. *Journal of Consulting and Clinical Psychology*, 87(12), 1149–1156. <https://doi.org/10/gg2z94>

- Paladino, D. A., Minton, C. A. B., & Kern, C. W. (2011). Interactive training model: Enhancing beginning counseling student development. *Counselor Education and Supervision, 50*(3), 189–206. <https://doi.org/10/fxq22w>
- Parrow, K., Sommers-Flanagan, J., Cova, J. S., & Lungu, H. (2019). Evidence-based relationship factors: A new focus for mental health counseling research, practice, and training. *Journal of Mental Health Counseling, 41*(4), 327–342. <https://doi.org/10/gg22b8>
- Pedersen, P. B. (1977). The triad model of cross-cultural counselor training. *The Personnel and Guidance Journal, 56*(2), 94–100. <https://doi.org/10/fzpf5h>
- Prescott, D. S., Maeschalck, C. L., & Miller, S. D. (Eds.). (2017). *Feedback-informed treatment in clinical practice: Reaching for excellence* (1st ed.). American Psychological Association. <https://doi.org/10.1037/0000039-000>
- Shannonhouse, L. R., Elston, N., Lin, Y.-W. D., Mize, M. C., Rumsey, A., Rice, R., Wanna, R., & Porter, M. J. (2018). Suicide intervention training for counselor trainees: A quasi-experimental study on skill retention. *Counselor Education and Supervision, 57*(3), 194–210. <https://doi.org/10/ghkngg>
- Romanelli, A., & Berger, R. (2018). The ninja therapist: Theater improvisation tools for the (daring) clinician. *The Arts in Psychotherapy, 60*, 26–31. <https://doi.org/10/gg22bh>
- Rønnestad, M. H., Nissen-Lie, H. A., Oddli, H. W., Benum, K., Ekroll, V. B., Gullestad, S. E., Haavind, H., Reichelt, S., Råbu, M., Stänicke, E., von der Lippe, A. L., & Halvorsen, M. S. (2019). Expanding the conceptualization of outcome and clinical effectiveness. *Journal of Contemporary Psychotherapy, 49*(2), 87–97. <https://doi.org/10/gg22bm>

- Rosén, E. (2020). *Learning how to learn: Psychology trainees' self-compassion while implementing deliberate practice, with FIT at a psychotherapy training clinic* [Unpublished master's thesis]. Umeå Universitet.
- Rousmaniere, T. G. (2016). *Deliberate practice for psychotherapists: A guide to improving clinical effectiveness* (1st ed.). Routledge. <https://doi.org/10.4324/9781315472256>
- Rousmaniere, T. G. (2019). *Mastering the inner skills of psychotherapy: A deliberate practice handbook* (1st ed.). Gold Lantern Books.
- Rousmaniere, T. G., Goodyear, R. K., Miller, S. D., & Wampold, B. E. (Eds.). (2017). *The cycle of excellence: Using deliberate practice to improve supervision and training*. Wiley.
- Rousmaniere, T. G., Swift, J. K., Babins-Wagner, R., Whipple, J. L., & Berzins, S. (2016). Supervisor variance in psychotherapy outcome in routine practice. *Psychotherapy Research, 26*(2), 196–205. <https://doi.org/10/gfxkgh>
- Rousmaniere, T. G., & Vaz, A. (n.d.). *What is deliberate practice?* Deliberate Practice Institute. Retrieved June 12, 2020, from <https://www.dpfortherapists.com/about-1>
- Tuffiash, M., Roring, R. W., & Ericsson, K. A. (2007). Expert performance in SCRABBLE: Implications for the study of the structure and acquisition of complex skills. *Journal of Experimental Psychology: Applied, 13*(3), 124–134. <https://doi.org/10/bh56ct>
- Watkins, C. E. (2011). Does psychotherapy supervision contribute to patient outcomes? Considering thirty years of research. *The Clinical Supervisor, 30*(2), 235–256. <https://doi.org/10/cff44j>
- Westra, H. A., Norouzian, N., Poulin, L., Hara, K., Coyne, A., Olson, D., & Antony, M. M. (2020). Testing a deliberate practice workshop for developing appropriate responsivity to resistance markers. *Psychotherapy, 36*.

Wood, S., Baltrinic, E., Barrio Minton, C., Cox, J., Kleist, D., Lingertat-Putnam, C., Merino, C., Osterlund, L., Perjessy, C., Rodriguez, A., & Sheely, A. (2016). *Best practices in teaching in counselor education report*. Association for Counselor Education and Supervision.

**Chapter II: A Psychometric Case Study of Field Experience Evaluations in A Single
Counselor Preparation Program**

Abstract

Counseling field experiences provide authentic opportunities to assess student growth and counselor development, but the complexities of counseling environments and counseling skills make effective evaluation a difficult task. In this manuscript, data collected through a CACREP-accredited counselor education program were analyzed using a variety of psychometric techniques to demonstrate common pitfalls and problems with informally developed assessment forms. These analyses show that the assessment lacks key types of reliability, is subject to excessive measurement error, and only provides informative results for a narrow range of student abilities. These findings inform a discussion of methods for addressing difficulties inherent to evaluating counselor development.

Keywords: Student evaluations, Counselor development, Assessment development

A Psychometric Case Study of Field Experience Evaluations in A Single Counselor Preparation Program

The need for high quality assessment and evaluation¹ permeates the counseling profession. For counselors, assessment provides the foundation for effective interventions (Watson & Flamez, 2015), and an entire section is dedicated to assessment in the profession's *Code of Ethics* (ACA, 2014). Counselor educators must engage in frequent evaluation of students, provide effective feedback, and identify problems of professional competence to safeguard client welfare and facilitate student counselor development (ACA, 2014; CACREP, 2016). As focus in the profession shifted from content delivery to student learning outcomes (SLOs) and competency-based standards, the clinical and educational importance of evaluating student learning became entwined (Akos et al., 2019; Barrio Minton & Gibson, 2012). Current accreditation standards include evaluation of individual students and programs at multiple points using multiple methods (CACREP, 2016). Initial drafts of the 2024 CACREP standards show increased emphasis on formative and summative evaluations of student performance; clear identification of performance benchmarks; and systematic methods to collect, use, and aggregate assessment data (CACREP, 2021). This manuscript will explore the intricacies of student evaluation during counseling field experiences, use a variety of psychometric analysis to demonstrate potential problems of uninvestigated program-developed assessments, and provide recommendations about ways to improve the quality of student evaluations.

¹ Throughout this manuscript, the terms “assessment” and “evaluation” may be used somewhat interchangeably. “Assessment” will refer to the process of assessing clients in counseling, as well as the forms developed and used by counselor educators. “Evaluation” will refer to the process or application of the assessment, either in specific or general circumstances.

Evaluation of Students During Field Experiences

Counseling field experiences receive particular attention in counseling programs and accreditation standards. An opportunity for “the application of theory and the development of counseling skills under supervision” (CACREP, 2015, p. 14), practicum and internship experiences make up a minimum of 700 clock hours of direct and indirect service covering a minimum of two semesters. These semesters of field experience serve as one of the best opportunities for authentic assessment (Barrio Minton et al., 2016), where multiple supervisors can observe student counselors’ professional behaviors. To best capture student skills during practica and internships, students should be assessed on their use of foundational interpersonal skills, advanced counseling skills, and professional dispositions.

Foundational interpersonal skills are the basic units of the therapeutic conversation and often play a central role in evaluation of student counselors. Also called essential skills, basic skills, or microskills, these include reflections (of content, feeling, and meaning), questions, minimal encouragers, paraphrasing, and summaries. Few tasks in counseling mirror mechanical, psychomotor skills in other healthcare fields (Akos et al., 2019), and microskills come closest. Microskills are an excellent target for evaluating skill development and make up around half of one of the primary examples of counseling skill assessments (Lambie et al., 2018). Through recordings and live observation, supervisors can readily observe these skills.

As students begin to develop confidence in their use of foundational interpersonal skills, counselor educators can focus on more advanced counseling skills. Field experiences provide opportunities to practice more advanced, integrative skills (Akos et al., 2019) in authentic context including diagnostic processes, theoretical conceptualizations, and multicultural competence (Barrio Minton et al., 2016). Assessments such as the Counselor Competency Scale-

Revised (CCS-R; Lambie et al., 2018) include some advanced skills (i.e., Confrontation [item 1.H], Focus of Counseling [item 1.J], Multicultural Competence [item 2.F]), but might not assess other key skills like crisis assessment and intervention (CACREP Standard 2.F.3.g, 2015), diagnosis (Standard 5.B.2.i), treatment planning (Standard 2.F.5.h), and ethical decision making (ACA, 2014, Section I.1.b). That these skills fall outside of the purview of common skill assessments is reasonable due to the complexity of issues, but this complicates evaluating students' skill development.

The last, and perhaps most important, domain on which students are assessed is professional dispositions. Professional dispositions are the “commitments, characteristics, values, beliefs, interpersonal functioning, and behaviors that influence the counselor's professional growth and interactions with clients and colleagues” (CACREP, 2015, p. 43). Counselor educators evaluate professional dispositions in admissions (Garner et al., 2016; Swank & Smith-Adcock, 2014), at multiple points within the program (CACREP, 2016), and in endorsement for licensure (ACA, 2014). Researchers have developed variety of professional disposition assessments (Garner et al., 2016; Miller et al., 2020; Tate et al., 2014), but supervisors continue to rely on subjective standards (Sabella et al., 2020). Recent psychometric efforts have sought to add rigor to the evaluation of professional dispositions, but lack agreement in factor structure (Garner et al., 2016; Miller et al., 2020). Although they differ across authors and programs, professional dispositions often overlap with counseling skills, assessment and conceptualization skills, professional behavior, self-awareness and self-reflection, supervision behaviors, and multicultural competence, making them more difficult to assess (Kemer et al., 2017). Translating professional dispositions to concrete student learning outcomes or competencies is a difficult, program-specific undertaking, for which CACREP (2015, 2021) provides limited guidance.

Difficulties of Student Evaluations in Field Experiences

Evaluating student development through counselor education programs is a difficult task, which is made more difficult by student non-disclosure, site supervisors with limited training in supervision, lack of agreement on expectations, and limited psychometric choices. On one hand, faculty members and students interact through classes, program activities, and advising, while coursework and assignments provide frequent evaluation opportunities. On the other hand, student counselors may withhold important information from supervisors because they feel uncomfortable or rushed during supervision, disagree with their supervisors, or are experiencing intrapersonal difficulties (Cook et al., 2019). Counselor educators rely on limited information about students' dispositions (Sabella et al., 2019), and may only provide feedback or remediation when severe situations occur (Sabella et al., 2020).

Supervision is a complex task, which can be harmed by pressures inherent to mental health or school settings. Even when counseling programs offer training, site supervisors may be uncomfortable with the role of supervisor (Bjornestad et al., 2014). Although accomplished in their professional roles, site supervisors often have limited preparation to engage in effective supervision and evaluate student skills and dispositions (Wambu & Myers, 2019), instead focusing on administrative supervision and defaulting to subjective and informal methods of evaluating students (Sabella et al., 2020). The effect of informal evaluations can be compounded by the different expectations between programs and sites alongside the varying professions from which some supervisors hail.

In pursuit of effective evaluations of counseling students, researchers have developed and validated tools to assess microskills (e.g., Lambie et al., 2018), interpersonal and process skills (e.g., Anderson et al., 2018), and counseling-related interpersonal skills (e.g., Hamlet & Burnes,

2013). Christensen et al. (2018) proposed finding consensus on professional dispositions, and in the process proposed yet another version. Likewise, Miller et al. (2020) reviewed a variety of conceptualizations in the professional disposition literature, before putting forth their own psychometrically-based disposition assessment. Despite the existence of valid tools used with similar populations, educators and programs have developed their own tools to address the unique combination of institutional and personnel factors with which programs wrestle. In an analysis of practicum and internship student assessments, Kemer et al. (2017) identified 27 separate assessments with 1,034 total items which assessed counseling skills, assessment and conceptualization skills, professional behaviors, self-awareness and reflection skills, engagement in supervision, and multicultural skills.

As standards, expectations, and research have proliferated, researchers and programs' attempts to ensure students meet expectations while remaining true to their training philosophy have multiplied as well. The psychometric quality of these assessments has not kept pace. In a review of existing assessments of counselor performance, Tate et al. (2014) found that around two-thirds of the 41 instruments they reviewed were self-report only, and 13 of 26 domain-specific assessments focused on multicultural competence, while few accounted for social desirability. Few of these instruments used direct observation of behaviors, and clients' ratings of counselors were all but missing. In this same review, eight of sixteen expert-based instruments reported inter-rater reliability, and there was no consistency in reports of reliability or validity. Reviews of more recent literature suggests that these psychometric limitations have not improved (La Guardia, 2021; Barrio Minton & Hightower, 2020), thus reducing the utility of assessments from educational and measurement perspectives (Barrio Minton et al., 2016; Flake & Fried, 2020).

Purpose of the Study

Despite the ubiquity of student evaluation challenges, the counseling literature has struggled to fill these gaps. As students move through field experiences, programs cannot ensure quality evaluation and may not be meeting their ethical responsibilities for accurate formative and summative evaluation of counselor development. The purpose of the present study is to analyze data obtained through one CACREP-accredited counseling program's field experience assessment to highlight potential limitations and opportunities for addressing challenges.

Through an analysis of assessment data, the following research questions were addressed:

- How reliable is the assessment in evaluating students across field experiences?
- How useful is the assessment for making decisions about students' ability?
- Does the assessment provide useful information about counseling students across a range of abilities?

Method

This study involved a psychometric analysis of archival data collected during the 2017-2018, 2018-2019, and 2019-2020 academic years using a single assessment form for students completing practicum and internship in clinical mental health and school settings.

Program and Participants

The dataset analyzed includes 867 individual evaluations of practicum and internship students by site and doctoral supervisors. Anonymized participant information has been used for analysis as indicated below.

Program

The program was a mid-sized CACREP-accredited counseling program in the southeastern United States. The program had three accredited program areas with Clinical

Mental Health Counseling and School Counseling master's programs and a Counselor Education doctoral program. The program had eight core faculty members with a strong counselor professional identity who taught the vast majority of courses. The program was housed in a very high research activity (R1) brick and mortar public institution, with most classes taught in-person and the vast majority of students attending the program full-time. Students were grouped into cohorts, which typically enrolled in Practicum in the spring semester of their first year, Internship 1 in the fall semester of their second year, and Internship 2 in the spring semester of their second year. Other course enrollment possibilities included Internship 3 for clinical mental health counseling students in the summer semester of their second year, Internship 3 and Internship 4 in the fall and spring semesters of dual clinical mental health and school counseling students third year, and doctoral practicum in the fall semester of doctoral students' first year. The final semester of data corresponded with beginning of the COVID-19 pandemic, meaning that most courses and field experiences were shifted to telehealth and teleconferencing.

The program used a practicum and internship model in which students were placed in clinical or school settings with site supervisors employed by the settings who provided weekly supervision in accordance with CACREP (2016) standards. This supervision was supplemented on campus by weekly individual supervision with doctoral student supervisors for practicum students and weekly rotations of individual and triadic supervision with doctoral student supervisors for internship students. Each student also attended group supervision with a faculty member on a weekly basis.

Students

The dataset includes midterm and final evaluations of 108 individual students across practicum and internship courses. These students represent each accredited track within the

counseling program, including Clinical Mental Health Counseling ($n = 59$), School Counseling ($n = 32$), Dual Clinical Mental Health and School Counseling ($n = 8$), and Counselor Education Doctoral practicum students ($n = 16$). Due to the nature of these evaluations, demographic data is not available for these students, but university-provided program data for the years in questions provides some description of students. The students were 83.15% women and 16.85% men; the university did not provide a non-binary option. Students had average age of 28.00 ($SD = 7.45$, $Med = 25.43$) at the beginning of their program. The majority of students were white (92.13%), with Black or African American (5.62%), Hispanic (1.12%), and students of unidentified race/ethnicity (1.12%) making up the remaining student population; no students identified as American Indian or Alaska Native, Asian or Pacific Islander during the enrollment period.

Supervisors

One hundred and thirty-six individual supervisors evaluated an average of six students each. These supervisors consisted of site supervisors ($n = 111$) and doctoral student supervisors ($n = 35$). Site supervisors were split between clinical mental health ($n = 48$) and school ($n = 63$) settings. Due to the nature and method of data collection, no demographic information is available for site or doctoral student supervisors. Based on university-provided program data for doctoral student supervisors shows that 69.23% of these students were women and 30.77% men, with an average age of 31.34 ($SD = 7.13$, $Med = 29.74$) at the beginning of their program. The majority of students were white (76.92%), with Black or African American (19.23%), and Hispanic (3.85%) students comprising the student population.

Instrument

The assessment under investigation was a program-developed field experience rating form consisting of a body of 27 questions which contained three nominal subscales: Counseling

Skills (9 questions; Cronbach's $\alpha = .87$), Professional Dispositions (11 questions; $\alpha = .94$), and Conceptualization and Planning (7 questions; $\alpha = .89$), and two setting-specific scales corresponding to clinical mental health (3 questions) or school (5 questions) settings. Each item was rated on a 3-point Likert-type scale which contained options for "Does not Meet Expectations" (1), "Meets Expectations" (2), and "Exceeds Expectations" (3), and an option which allowed supervisors to indicate they had not observed a skill. The assessment included items which identified students, supervisors, setting, time of assessment (midterm or final), supervisor role (site or doctoral student), students' course (practicum or internship), and the counseling services provided. The assessment allowed for free text entry of students' strengths, limitations, and additional comments. Free text entry, setting-specific scales, and counseling services were excluded from analysis.

Procedure

Data collection was conducted during the normal course of counseling field experience evaluation procedures. For each field experience (i.e., Practicum, Internship 1, Internship 2) students were under supervision of both a doctoral student supervisor and a site supervisor. Each supervisor assessed each student twice per semester: once around midterm, and once near the conclusion of the semester. Thus, each master's student had at least 12 assessments using the same set of questions by the conclusion of the second semester of internship. Data used in this analysis were collected between September 2017 and December 2020 via Qualtrics and QuestionPro. IRB indicated that use of de-identified archival data was exempt from review.

Data Analysis

A program administrator provided the primary researcher an initial dataset that had been retrieved from program records with names of students, sites, and supervisors redacted and

replaced with a ID to match cases longitudinally; additionally, all text responses were deleted from the dataset. Data were visually inspected and cleaned according to best practices for data processing (Morrow & Skolits, 2017). This included addressing missing data and “Not Observed” responses. Not Observed responses were relabeled as missing, because the resulting increase in total missing data was negligible (an increase from 9.05% missing data to 10.90%). Inspection of relationships between missing and non-missing variables showed that the missing data were Missing At Random (MAR), which indicated that missing data were dependent on other observed data (Rubin, 1976). Due to MAR data, multiple imputation is possible where complete cases are needed (Dong & Peng, 2013), and was utilized for item response theory analysis. Following data preparation, each research question was analyzed using R statistical computing software (R Core Team, 2020). For each analysis, a sensitivity analysis was conducted to compare results across possible divisions of the dataset, which produced results much like those presented below.

Two separate Interclass Correlation coefficients (ICC) were calculated to assess test-retest reliability and interrater reliability (Koo & Li, 2016). ICC estimates and their 95% confidence intervals were calculated using the *irrNA* package (Brueckl & Heuer, 2021) to account for missing data. Test-retest reliability ICC for each subscale was calculated using a mean-rating ($k = 9.08$), absolute-agreement, 2-way mixed-effects model. Interrater reliability was calculated using a mean-rating ($k = 7.32$), consistency, 2-way mixed-effects model.

Generalizability theory (G-theory) was used to assess utility of the assessment for estimating student ability. G-theory (Cronbach et al., 1972) is an extension of Classical Test Theory which disentangles different sources of variation in order to determine how much an object of measurement (i.e., students) contribute to assessment scores in comparison to other

sources of variation (i.e., items, courses, supervisors, error, etc.). Using the *gtheory* package (Moore, 2016), a Generalizability Study (G study) was used to identify sources of variance in scores, and a Decision Study (D study) was used to assess whether the assessment was effective for making comparisons within the sample and absolute assessments. The G study conducted used a 4-facet crossed design using the following model:

$$\sigma^2(X_{p,c,i,f,sc}) = \sigma_p^2 + \sigma_c^2 + \sigma_i^2 + \sigma_f^2 + \sigma_{(s:c)}^2 + \sigma_{p,c}^2 + \sigma_{p,i}^2 + \sigma_{p,f}^2 + \sigma_{p,(s:c)}^2 + \sigma_{i,c}^2 + \sigma_{i,f}^2 + \sigma_{i,(s:c)}^2 + \sigma_{c,f}^2 + \sigma_{p,c,i} + \sigma_{p,c,f}^2 + \sigma_{c,f,i} + \sigma_{f,(s:c)}^2 + \sigma_{p,c,i,f,s:c,e}^2$$

where $\sigma^2(X_{p,c,i,f,sc})$ is the total variance across the entire measure X , σ_p^2 is the variance across individual students, σ_c^2 is the variance across course enrollment, σ_i^2 is the variance across scale items, σ_f^2 refers to the variance across forms (midterm or final), and $\sigma_{(s:c)}^2$ refers to variance across individual supervisors nested within each semester. The use of a nested facet was necessary because individuals were not all assessed by all supervisors. The remaining variance figures with subscripts $p, c; p, i; p, f; p, (s:c); i, c; i, f; i, (s:c); f, (s:c); c, f; p, c, i; p, c, f; c, f, i$ refer to the interactions of the main facets. This variance was due to the interactions between 1) the person (student) and the course, 2) the person and the scale item, 3) the person and the form, 4) the person and the nested effect of supervisor within course, 5) the item and the course, and 6) the item and the form, 7) the item and the nested effect of supervisor within course, 8) the form and the nested effect of supervisor within course, 9) the course and the form, 10) the person, course, and form, and 11) the course, form, and item. Finally, $\sigma_{p,c,i,f,s:c,e}^2$ refers to the variance due to undifferentiated error that remains and cannot be otherwise disentangled.

The Item Response Theory (IRT) framework was used to investigate how much information the assessment provided across a range of abilities. IRT (Embretson & Reise, 2000) is a family of probability models which use item-level data to estimate underlying trait/ability

levels. In an IRT framework, fitting probability models to the available data allows for estimation of item discrimination values (α), which indicate the ability of the item to differentiate between different levels of ability, and item difficulty values (β), which measure the latent trait levels associated with a 50/50 probability that a respondent will choose the correct answer. Item difficulty is measured on the same scale as the estimated latent trait (θ), with a mean of 0 and standard deviation of 1. Although there are a variety of IRT models for a given set of circumstances, the model used was a graded-response model. A graded-response model is a polytomous probability model which can be used scales with ordinal responses ranging from low to high. Under a graded-response model, item difficulty values are thresholds for movement from a lower categorical response to a higher category. Using the *mirt* package (Chalmers, 2012), a graded-response model was fit to each of the three subscales. This package enabled imputation of missing data through latent ability models, which can be used when data is MAR.

In this analysis, IRT was used to fit a unidimensional graded-response model to each of the three subscales. Before fitting the graded-response models, confirmatory factor analysis with robust standard estimators was used to verify unidimensionality of each scale, a key assumption of unidimensional IRT. Using the *lavaan* package (Rosseel, 2012), a unidimensional structure was analyzed, and model fit was assessed through the root mean square error of approximation (RMSEA), the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the standardized root mean square residual (SRMR). Acceptable model fit is indicated by fit statistics in the following ranges: $RMSEA \leq .08$, $CFI \geq .90$, $TLI \geq .90$, and $SRMR < .08$. Confirmatory Factor Analysis results showed that a single factor solution was acceptable for the Counseling Skills subscale ($RMSEA = .073$, $CFI = .934$, $TLI = .912$, $SRMR = .043$), the Professional Dispositions

subscale (RMSEA = .071, CFI = .939, TLI = .923, SRMR = .0443), and the Conceptualization and Planning subscale (RMSEA = .069, CFI = .956, TLI = .933, SRMR = .035).

After fitting a graded response model to the available data, model fit, item fit, and person fit must be assessed. Demonstrating that the graded response model has good fit with the data shows that the selected model is appropriate for the data and that the generated parameters have been estimated with a greater degree of reliability. Model fit was assessed through the model fit index M2, an IRT-specific fit statistic, which provides an RMSEA estimation, as well as CFI, TLI, and SRMR, using the same criteria as before. The Counseling Skills subscale (RMSEA = .050, CFI = .989, TLI = .984, SRMR = .042), Professional Dispositions subscale (RMSEA = .043, CFI = .993, TLI = .990, SRMR = .041), and the Conceptualization and Planning subscale (RMSEA = .081, CFI = .984, TLI = .969, SRMR = .034) all show good model fit. Item fit was evaluated through the $S-X^2$ index, which also provides a RMSEA value. For each item, RMSEA values were less than .06, and inspection of empirical plots showed acceptable fit. Finally, person fit was established through inspection of the Zh index. Person misfit is demonstrated through Zh values of less than -2. For each models, less than 3% of cases had values of $Zh < -2$, which shows acceptable fit.

Having adequate sample size for each analysis is important to ensure that parameter estimation is reliable, and error is reduced as much as possible. Sample size determination is difficult in G-theory analyses but following the loose guideline of 20 persons and 2 occasions per facet (Briesch et al., 2014), the sample size ($n = 108$) and evaluation occasions ($k = 9.08$) exceeded the minimum needed for four facets. Defining an acceptable sample size is difficult for IRT analyses, as some models struggle to estimate reliable parameters with 350 cases (Embretson & Reise, 2000). Simulated data suggest that a graded-response model can provide

reliable parameter estimates at samples of at least 500 cases (Reise & Yu, 1990), which the current dataset exceeded. In order to account for each of these sample size requirements, each analysis uses all of the available cases and includes each of the 27 items across the three main subscales.

Results

Assessment Reliability

Using an ICC calculation is a common method of evaluating measure reliability. ICC values less than 0.5 suggest poor reliability, values between 0.5 and 0.75 suggest moderate reliability, values between 0.75 and 0.9 suggest good reliability, and values greater than 0.90 indicate excellent reliability (Koo & Li, 2016). The test-retest reliability was poor for the Counseling Skills subscale ($ICC = 0.162$, 95% CI [-.017, .335]), the Professional Dispositions subscale ($ICC = 0.159$, 95% CI [-.020, .331]), and for the Conceptualization and Planning subscale ($ICC = 0.155$, 95% CI [-.023, .328]). The interrater reliability was poor for the Counseling Skills subscale ($ICC = 0.422$, 95% CI [.261, .559]), the Professional Dispositions subscale ($ICC = 0.426$, 95% CI [.265, .562]), and for the Conceptualization and Planning subscale ($ICC = 0.432$, 95% CI [.273, .567]). These results suggest that the assessment shows poor test-retest and interrater reliability.

Decision Making Utility

The first analysis within G-theory, a G study, can help disaggregate sources of variance using information about the facets contributing to variance in scores (Briesch et al., 2014). For each individual analysis, the overall model resulted in overfitting and produced variance close to zero. Overfitting shows that the model fits too closely to a small number of datapoints and may inadequately fit aspects of the data. In each case, the removed variance terms were .00005 or

less, which was then included in the residual term. For the Counseling Skills subscale analysis, the variance terms $\sigma_{p,c}^2$, $\sigma_{i,f}^2$, and $\sigma_{i,c}^2$ were removed. For the Professional Dispositions Subscale, the variance terms $\sigma_{p,c}^2$, $\sigma_{i,c}^2$, $\sigma_{c,f,i}$, and $\sigma_{c,i,f}^2$ were removed. For the Conceptualization and Planning Subscale, the variance terms $\sigma_{p,c}^2$, $\sigma_{p,f}^2$, $\sigma_{i,f}^2$, and $\sigma_{c,f,i}$ were removed.

The results of the G study suggested that assessment scores varied due to factors other than student development. The variance attributable to course enrollment (whether students are enrolled in practicum, internship 1, internship 2, etc.) ranged from 0.9% of explained variance (Professional Disposition) to 3.3% of explained variance (Conceptualization and Planning). Students themselves explained very little variance, ranging from 0.8% of explained variance (Counseling Skills) to 1.9% (Professional Development). Overall, the facet which explained the most variance was the facet for supervisors nested in courses. In each scale, this facet accounted for at least 22% of the explained variance. Finally, the G study showed that unexplained variance accounted for the greatest amount of variance, with a minimum of 37.1% of variance remaining unexplained. The amount of unexplained variance indicates that it is impossible to identify the source of nearly half of the variance despite the number of facets available for analysis. Table 1 in Appendix B provides full variance break down for each subscale.

Following the G study analyses, results were used to conduct a decision (D) study, which provides information about the usefulness of a measure (Briesch et al., 2014). D studies provide a variety of statistics to evaluate measures, but we focused on the generalizability coefficient (ρ), similar to Cronbach's alpha, and the dependability coefficient (ϕ), which evaluates the reliability of the measure for absolute decision making. For each subscale, ρ is .04 or less, which was poor or unacceptable reliability (Briesch et al., 2014). The values of ϕ for each subscale were less than .022, which suggested that were not useful for making absolute decisions about the skill level of

the students being assessed (Briesch et al., 2014). Table 2 in Appendix C provides detailed information regarding D study analyses.

Scale Information

A graded-response model was fit to data from each of the three subscales, which produced the item discrimination and threshold values found in Table 3 in Appendix D. Of note, all of the questions showed good discrimination ($\alpha > 1$), which means that each question could differentiate between individuals with high levels of the latent trait and individuals with low levels of the latent trait (as measured by the entire subscale). In the case of a three-point scale, two threshold ratings are estimated: β_1 estimates the threshold between “Does not meet expectations” and “Meets expectations,” while β_2 estimates the threshold between “Meets expectations” and “Exceeds expectations.” These thresholds values can be “tipping points,” or the latent trait score needed to have an equal probabilities of receiving the higher rating and the lower rating. The β_1 threshold values for 23 of 27 items were around -3 standard deviations (range: -4.12 to -2.47), indicating that individuals who possessed an estimated latent ability score around three standard deviations below the mean had a 50/50 probability of receiving ratings of “Meets expectations” than “Does not meet expectations.” The β_2 threshold values for these same items show that a latent trait score of around 0 (range: -0.51 to 0.57) showed that individuals of average latent ability had a 50% chance of receiving ratings of “Exceeds expectations” and a 50% chance of receiving a rating of either “Does not meet expectations” or “Meets expectations”

Figure 1 in Appendix E provides a graphical depiction of these values, with the probability ($P(\theta)$) of a response on the y-axis, the latent trait (θ) on the x-axis, and each line representing a response option. The remaining 4 items (“Nonverbal Behaviors”, “Promotes Confidence in the Counseling Profession”, “Advocates for the Profession”, “Self-Evaluation”) have β_1 values of

around 0 (range: -0.11 to 0.63), and no β_2 value. This means that an individual of around average ability has a 50% chance of receiving ratings of “Meets expectations” or “Does not meet expectations,” and the probability of receiving an “Exceeds expectations” is low. Figure 2 in Appendix F depicts these items graphically.

The use of a graded-response model allows for graphical depiction of information provided by the scales. Information is a function of the parameters estimated by the model and provides a sense of the amount of psychometric information provided by items and a scale across a range of ability levels. Figure 3 in Appendix G shows test information curves for each of the three subscales. All three show a similar pattern, where each scale provides the most information about respondents with an average level of ability, and a second peak occurs around -3.5 or -4 standard deviations below. Beyond those peaks, the scale provides very little information about respondents' ability. As the threshold parameters would suggest, these scales provide little to no information about students whose ability is greater than average.

Discussion

Ensuring quality evaluations of students in counseling field experiences is an ethical responsibility for all counselor educators and supervisors (ACA, 2014; CACREP, 2016). The purpose of the current study was to analyze one program-created counseling field experience assessment in attempts to investigate the psychometric properties of the assessment, identify potential problems related to evaluating counseling students using the assessment, and provide suggestions to improve the quality of field experience evaluations of student counselors.

The test-retest and interrater reliability and G study results showed that there was little consistency between evaluations of individual students at multiple time points, in part, because the individual students and the time periods over which they are assessed contributed little to

assessment scores. Similarly, supervisors explained the greatest percentage of explained variance, but had little consistency between them due to the large amount of unexplained error variance. The G study showed that in each of the three subscales, error variance (variance which cannot be otherwise decomposed) was the source of around half of the variance present in each measure, while supervisors were the next greatest source of variance. These results suggest that much of the variation in assessment scores is the result of error and individual supervisors' idiosyncratic evaluation methods, rather than differences between students' skill levels or growth across time. These results were echoed by the D study, which showed that the scales' generalizability (analogous to Cronbach's alpha in utility) and dependability coefficients were very poor ($\leq .035$ and $\leq .019$, respectively), which indicated that they provided little useful measurement of students' performance. Finally, a graded-response model showed that, across most items, individuals with an average ability score ($\theta = 0$) had a greater probability of receiving ratings of "Exceeds expectations," while individuals did not receive "Does not meet expectations" ratings until they dropped below an ability score at least 3.5 standard deviations below average. Rather than detecting students' ability levels, the frequent use of "Exceeds expectations" may be indicative of supervisors' generosity and desire to rate students highly. Moreover, scales provided information about performance in two narrow peaks around these ability levels but could not detect performance across a range of abilities.

These results have several possible interpretations. The poor test-retest reliability may be indicative of students demonstrating extensive growth in skills between evaluations, as counselor development is a developmental process during which educators and supervisors hope to see student growth over time. In combination with the G-theory results, though, it is possible that poor test-retest reliability is not due to developmental factors, as students' progression through

field experience courses accounts for little variation in scores, and supervisors may naturally adjust developmental expectations for students as they progress their programs (Sabella et al., 2020). The limited use of the “Does not meet expectations” response might indicate that supervisors are unwilling to rate a student below expectations without exceptional problems. It might also suggest that the program and supervisors do an excellent job in identifying potential problems and engaging in effective remediation and gatekeeping, meaning that the evaluations are only reflective of the students who successfully complete remediation or do not need it. This, in turn, might could mean that the estimated “average” latent trait level for the group of students evaluated is at least “exceed[ing] expectations.” In these cases, it is not that the evaluation is unable to detect poor or concerning student performance, but that it does not adequately capture the variation and nuance of differing levels of counselor skill both above and below average. Overall, these hypotheses cannot be confirmed by the available data.

The problems highlighted by this analysis are not unique to evaluations of counseling field experiences. A need for quality evaluations has led programs to develop internal standards and assessments to meet accreditation and programmatic standards. Assessments may cover student outcomes (Haberstroh et al., 2014), professional dispositions (Spurgeon et al., 2012), or SLOs tied to specific CACREP standards (Barrio Minton et al., 2016). Each new assessment adds to the large existing body of assessments (Kemer et al., 2017; Tate et al., 2014), but these may fall into the common problems of psychological and educational measurement (Flake & Fried, 2020).

In the field, supervisors’ evaluations of students are often limited because they cannot always observe student counselors’ performance (Miller et al., 2020), they rely on subjective evaluation (Sabella et al., 2020), and they need greater training in supervision (Bjornestad et al.,

2014; Wambu & Myers, 2019). These limitations may result in supervisors conflating professional dispositions, skills, and personality as they assess counseling students (Landon et al., 2021). As counselor education moves toward a competency-based curriculum (Akos et al., 2019), the reliance on the authentic assessment opportunities and need for more effective assessments will increase. Although there are reliable and valid assessments for student counselors, they may still be limited in the quality or range of assessment.

As counselor educators and programs move toward more rigorous student evaluations, use of psychometric and research processes can improve the quality of evaluation. As CACREP has developed more detailed standards (2021), it will become easier to define clear learning outcomes and benchmarks for performance. Using these student learning outcomes, counselor educators can use guides (e.g., Lambie et al., 2017) to develop and establish clear, valid constructs on which to build assessments. By refocusing from self-report to expert assessment (Tate et al., 2014) or demonstration of skills (Anderson et al., 2018), programs are better able to assess the quality of care clients will receive, rather than students' self-efficacy or knowledge of desirable responses. Consideration of client experiences as a focus of evaluation through feedback-informed-treatment can further contribute to ensuring effective student evaluations (Yates et al., 2016). Finally, this assessment and others include opportunities for open ended feedback, which may provide a clearer description of students' skill levels, which should be incorporated in student evaluation.

Implications for Scale Development and Improvement

In order to improve the quality of data obtained from existing assessments, there are several implications from the above analysis, including the use of response options, training supervisors, and clarifying expectations. When considering potential item response options, a

three-point Likert-type scale reduces the level of detail available. One of the limitations of the analyzed evaluations is the narrow range of ability which they measure. This can be a side effect of the limited response options and limited use of the lowest response option. By moving from a 3- to a 5-point scale, respondents will have greater options for rating a wider range of student abilities. Although use of a 3-point scales have been used in counseling field experience assessments (e.g., Haberstroh et al., 2014), scales which use 5-points or more have greater psychometric support (Lambie et al., 2018; Miller et al., 2020).

When students are assessed based on specific expectations using an idiosyncratic assessment, it is important that raters are properly trained on the use of the assessment. Not only is it necessary to train supervisors in counseling programs' evaluation procedures, CACREP accreditation standards require that program faculty prepare site supervisors with "knowledge of the program's expectations, requirement, and evaluation procedures..." (2015, p. 15). Although there are common accepted expectations for student counselors as they develop counseling skills and attitudes (CACREP, 2016, 2021), results suggest that supervisors have their own ideas about when students meet these expectations. Training can take many forms, depending on program and supervisors' needs, but consistency and accuracy in evaluation are key. Programs may consider research-like rater trainings such as those used by Garner et al. (2016) and Timmerman et al. (2011), which included an overview of the purpose and method of assessment, an assessment guide, and practice scoring exemplar case studies which illustrate students who are below, at, or exceeding expectations until desired interrater reliability is achieved. Once supervisors are trained, occasional random spot-checking for problems in interrater reliability and regular retraining can reduce drift in supervisors' evaluations (Castorr et al., 1990).

Finally, the quality of data received through evaluations could be enhanced by detailed expectations within behavioral anchors for each item. A behavioral anchor provides information to the respondent about what each point on a scale may look like. These may be provided at the extreme ends of the scale (i.e., a description of the worst and best possible performance), or at each point on the scale. Effective behavioral anchors can be as simple as the extent to which respondents' agree with items (see Miller et al., 2020), or be as complicated as providing a clear example of counselor behaviors (see Anderson et al., 2018; Uhlin, 2011) and expectations about the frequency of skill usage (see Lambie et al., 2018).

A final option, and an extension of behavioral anchors, is to write clear developmental expectations for skills and behaviors for at each stage of development. By delineating student performance at each stage of development, faculty and supervisors can ensure client welfare and counselor development. Kemer et al. (2017) suggest the use of separate forms for different courses in order to improve the clarity of assessments. Counseling faculty have different expectations for students at different stages of development, and well-developed assessments can reflect these developmentally-oriented expectations.

Overall, decisions about evaluating student counselors during field experiences is a complex issue. Programs must consider the internal and external expectations unique to their contexts, buy-in of site supervisors and other community stakeholders, and the particular set of psychometric skills available to them. A proliferation of idiosyncratic and formal assessments has also muddied the existing literature about evaluating students, resulting in less agreement. A possible solution would be for individual programs to focus on implementing existing assessments which are reliable and validated for evaluating student counselors, while the leaders of the counseling profession focus the collective education and research talent on developing

consensus about what constitutes competency with regard to counseling skills, assessment skills, professional dispositions, self-awareness and reflection skills, and multicultural skills. Without collective action, counseling evaluations may remain mired in the collective disagreement which exists.

Limitations

The present study was a psychometric analysis of a single program's field experience assessments. As such, the results are not generalizable to all counselor education programs, student counselors, or field experience evaluations. Rather, results are presented as examples to highlight difficulties involved with evaluating student counselors and facilitate discussion regarding opportunities or responding to these challenges. Due to the nature of data collection and the methods involved, no demographic analysis was possible, which further limits interpretation of results. Because the assessment was internally developed and unvalidated, no conclusions can be drawn about the participants' counseling skill or professional behaviors. The supervisors included in the dataset represent a wide range of potential experience levels, which may account for some of the inconsistency between supervisors, but there are not data which allow for an investigation of this possibility. Finally, the unique population, location, and field experience opportunities might influence the way in which evaluations were completed, and possible colliders cannot be analyzed.

Conclusion

A signature aspect of counseling programs is the requirement that student counselors complete at least 700 hours of field experience over at least two semesters during which they move from trainee to supervised clinician. While students complete these field experiences, it is incumbent on counselor education programs to engage in effective evaluation that safeguards

client welfare, monitors student development, and meets ethical and legal responsibilities regarding clear communication to students regarding performance and remediation. Despite several validated assessments for student counselors, options for evaluating field experiences are limited, and programs may create their own to address internal and external expectations. Through an analysis of one program-created assessment, I have illustrated the perils of psychometrically bereft scale development. The assessment has limited reliability, is influenced by the individual supervisor who completes it, and cannot provide informative scores for many students. Rather than a unique situation, limited attention to effective evaluation of counselor development during field experiences means that poor assessments may be ubiquitous in counselor education. I hope that a clear demonstration of the pitfalls in assessment development can lead to greater attention toward measuring development and evaluating counselors before, during, and after their field experiences. As in counseling, effective measurement is a necessary first step to remedying problems, monitoring growth, and contributing to client wellness.

References

- ACA. (2014). *Code of ethics*. American Counseling Association.
- Akos, P., Wasik, S. Z., McDonald, A., Soler, M., & Lys, D. (2019). The challenge and opportunity of competency-based counselor education. *Counselor Education and Supervision, 58*(2), 98–111. <https://doi.org/10/gj2hw5>
- Anderson, T., Patterson, C. L., McClintock, A. S., Song, X., & The Psychotherapy and Interpersonal Lab Team. (2018). *Facilitative interpersonal skills task and rating manual*. Ohio University.
- Barrio Minton, C. A., & Gibson, D. M. (2012). Evaluating student learning outcomes in counselor education: Recommendations and process considerations. *Counseling Outcome Research and Evaluation, 3*(2), 73–91. <https://doi.org/10/gg2z6d>
- Barrio Minton, C. A., Gibson, D. M., & Wachter Morris, C. A. (2016). *Evaluating student learning outcomes in counselor education*. American Counseling Association.
- Barrio Minton, C. A., & Hightower, J. M. (2020). Counselor education and supervision: 2018 annual review. *Counselor Education and Supervision, 59*(1), 2–15. <https://doi.org/10/gj2hxw>
- Bjornestad, A., Johnson, V., Hittner, J., & Paulson, K. (2014). Preparing site supervisors of counselor education students. *Counselor Education and Supervision, 53*(4), 242–253. <https://doi.org/10/gg2z6n>
- Briesch, A. M., Swaminathan, H., Welsh, M., & Chafouleas, S. M. (2014). Generalizability theory: A practical guide to study design, implementation, and interpretation. *Journal of School Psychology, 52*(1), 13–35. <https://doi.org/10/ghmsn5>

- Brueckl, M., & Heuer, F. (2021). *IrrNA: Coefficients of interrater reliability – generalized for randomly incomplete datasets* (0.2.2) [Computer software]. <https://CRAN.R-project.org/package=irrNA>
- CACREP. (2016). *2016 CACREP Standards*. Council for Accreditation of Counseling and Related Educational Programs.
- CACREP. (2021). *CACREP 2024 Standards draft 1*.
- Castorr, A. H., Thompson, K. O., Ryan, J. W., Phillips, C. Y., Prescott, P. A., & Soeken, K. L. (1990). The process of rater training for observational instruments: Implications for interrater reliability. *Research in Nursing & Health, 13*(5), 311–318. <https://doi.org/10/fchn66>
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29. <https://doi.org/10/gckfrg>
- Christensen, J. K., Dickerman, C. A., & Dorn-Medeiros, C. (2018). Building a consensus of the professional dispositions of counseling students. *The Journal of Counselor Preparation and Supervision, 11*(1), 27.
- Cook, R. M., Welfare, L. E., & Sharma, J. (2019). Exploring supervisees' in-session experiences of utilizing intentional nondisclosure. *The Clinical Supervisor, 38*(2), 202–221. <https://doi.org/10/gj2hxq>
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. Wiley.
- Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus, 2*(1), 222. <https://doi.org/10/gcp325>

- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. L. Erlbaum Associates.
- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456–465. <https://doi.org/10/ghnbdg>
- Garner, C. M., Freeman, B. J., & Lee, L. (2016). *Assessment of student dispositions: The development and psychometric properties of the professional disposition competence assessment (PDCA)*. VISTAS Online. 14.
- Haberstroh, S., Duffey, T., Marble, E., & Ivers, N. N. (2014). Assessing student-learning outcomes within a counselor education program: Philosophy, policy, and praxis. *Counseling Outcome Research and Evaluation*, 5(1), 28–38. <https://doi.org/10/fgzs>
- Hamlet, H. S., & Burnes, T. (2013). Professional school counseling internship: Developmental assessment of counseling skills (CIDACS). *Counseling Outcome Research and Evaluation*, 4(1), 55–71. <https://doi.org/10/gj2hxf>
- Kemer, G., Eustice, K., & Luby, M. (2017). An examination of master’s level practicum and internship evaluation forms. *Counseling Outcome Research and Evaluation*, 8(2), 105–120. <https://doi.org/10/gj2hx2>
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. <https://doi.org/10/b84r>
- La Guardia, A. C. (2021). Counselor education and supervision: 2019 annual review. *Counselor Education and Supervision*, 60(1), 2–21. <https://doi.org/10/gj2hvx>

- Lambie, G. W., Blount, A. J., & Mullen, P. R. (2017). Establishing content-oriented evidence for psychological assessments. *Measurement and Evaluation in Counseling and Development, 50*(4), 210–216. <https://doi.org/10/gj2hxr>
- Lambie, G. W., Mullen, P. R., Swank, J. M., & Blount, A. (2018). The counseling competencies scale: Validation and refinement. *Measurement and Evaluation in Counseling and Development, 51*(1), 1–15. <https://doi.org/10/dkrj>
- Landon, T. J., Sabella, S. A., McKnight-Lizotte, M., & Bernacchio, C. (2021). Rehabilitation counselor supervisors' perceptions of counselor professional dispositions for rural service delivery. *Rehabilitation Counseling Bulletin, 64*(3), 145–157. <https://doi.org/10/gj2hxd>
- Miller, S. M., Larwin, K. H., Kautzman-East, M., Williams, J. L., Evans, W. J., Williams, D. D., Abramski, A. L., & Miller, K. L. (2020). A proposed definition and structure of counselor dispositions. *Measurement and Evaluation in Counseling and Development, 53*(2), 117–130. <https://doi.org/10/d48v>
- Moore, C. T. (2016). *gtheory: Apply generalizability theory with r* (0.1.2) [Computer software]. <https://CRAN.R-project.org/package=gtheory>
- Morrow, J. A., & Skolits, G. (2017). *Twelve steps of quantitative data cleaning: Strategies for dealing with dirty evaluation data* [Paper]. American Evaluation Association Conference, Washington, DC.
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Reise, S. P., & Yu, J. (1990). Parameter recovery in the graded response model using multilog. *Journal of Educational Measurement, 27*(2), 133–144. <https://doi.org/10/d57sx3>

- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10/f3r4v8>
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. <https://doi.org/10/fhqxxb>
- Sabella, S. A., Bernacchio, C. P., Boland, E. A., & Schultz, J. C. (2019). The conceptualization and assessment of professional dispositions in rehabilitation counselor education. *Rehabilitation Research, Policy, and Education*, 33(3), 198–211. <https://doi.org/10/ghtzmz>
- Sabella, S. A., Landon, T. J., McKnight-Lizotte, M., & Bernacchio, C. P. (2020). How do supervisors assess and develop professional dispositions among counselors in vocational rehabilitation agencies? A qualitative inquiry. *The Clinical Supervisor*, 39(1), 106–127. <https://doi.org/10/gj2hxx>
- Spurgeon, S. L., Gibbons, M. M., & Cochran, J. L. (2012). Creating personal dispositions for a professional counseling program. *Counseling and Values*, 57(1), 96–108. <https://doi.org/10/dqwm>
- Swank, J. M., & Smith-Adcock, S. (2014). Gatekeeping during admissions: A survey of counselor education programs. *Counselor Education and Supervision*, 53(1), 47–61. <https://doi.org/10/gj2hxn>
- Tate, K. A., Bloom, M. L., Tassara, M. H., & Caperton, W. (2014). Counselor competence, performance assessment, and program evaluation: Using psychometric instruments. *Measurement and Evaluation in Counseling and Development*, 47(4), 291–306. <https://doi.org/10/gj2hxx>

- Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a 'universal' rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education*, 36(5), 509–547. <https://doi.org/10/b87q67>
- Uhlin, B. D. (2011). *Therapist in-session rated facilitative interpersonal skills (FIS-IS) in the psychotherapy process* [Doctoral dissertation, Ohio University]. <https://etd.ohiolink.edu/>
- Wambu, G. W., & Myers, C. E. (2019). School counselor site supervisors' perceptions of preparedness and training needs. *Journal of School Counseling*, 17, 30.
- Watson, J. C., & Flamez, B. (2015). *Counseling assessment and evaluation: Fundamentals of applied practice*. Sage Publications.
- Yates, C. M., Holmes, C. M., Smith, J. C. C., & Nielson, T. (2016). The benefits of implementing a feedback informed treatment system within counselor education curriculum. *The Professional Counselor*, 6(1), 22–32. <https://doi.org/10/gj2hw6>

Appendices

Appendix A

UT Counseling Site & Doctoral Supervisor Evaluations

1. Which evaluation are you completing?
 - a. Midterm
 - b. Final
2. Semester
 - a. Fall
 - b. Spring
 - c. Summer
3. Which describes your role?
 - a. Site Supervisor
 - b. Doctoral Student Supervisor
4. What describes your student?
 - a. Master's practicum
 - b. Master's intern
 - c. Doctoral practicum
 - d. Doctoral intern
5. What services does your student provide? [Mark all that apply]
 - a. Individual counseling
 - b. Group counseling
 - c. Family/couple counseling
 - d. Psychoeducation, group presentations, or guidance
 - e. Direct consultation
 - f.

For the following items [6-41], rate student performance on the following scale:

0 – Does Not Meet Expectations for an individual at this level of development

1 - Meets Expectations for an individual at this level of development

2 - Exceeds Expectations an individual at this level of development

N/A - No opportunity to observe this skill

6. Establishes emotional connections and builds rapport with clients/students
7. Conveys empathy, unconditional positive regard, and genuineness
8. Demonstrates active listening skills relevant to content, feeling, and meaning (e.g., paraphrasing, reflection of content, reflection of feeling, reflection of meaning, reflection of themes, summarization)
9. Uses appropriate questioning skills (e.g., open-ended, probing, therapeutically relevant)
10. Uses effective nonverbal communication (e.g., eye contact, head nods, voice, tone, posture; external behavior consistent with internal affect; appropriate timing)
11. Applies effective and developmentally appropriate interventions for groups and systems (e.g., classroom guidance, psychoeducation, process groups)
12. Collaborates with stakeholders in clients'/students' lives (e.g., family, school, workplace, other helpers)
13. Demonstrates ability to assess for imminent danger and intervene effectively (e.g., crisis, suicide, abuse/neglect)
14. Demonstrates intentional use of theory-based interventions and techniques
15. Demonstrates sensitivity to ethical and legal issues

16. Conducts self in an ethical manner promoting confidence in the counseling profession
17. Advocates for clients'/students' interests and rights in a manner that respects and empowers
18. Advocates for the counseling profession as appropriate
19. Manages time and tasks in an appropriate, dependable manner (e.g., arrives on time, meets deadlines, completes reporting requirements)
20. Engages fully in the supervision and consultation process (e.g., initiates requests for feedback/discussion, is open to feedback, applies recommendations)
21. Demonstrates commitment (i.e., investment, counselor identity, advocacy, civic engagement, collaboration, interpersonal competence)
22. Demonstrates openness (i.e., open to ideas, learning, and change; to giving and receiving feedback; to growth, and to others; interpersonal communication; understanding of micro/macro perspective)
23. Demonstrates respect (i.e., perceives and honors diversity, self-care skills, wellness)
24. Demonstrates integrity (i.e., personal responsibility, maturity, honesty, courage, congruence)
25. Demonstrates self-awareness (i.e., integrity, humility, self-reflection, exploration, place in history)
26. Attends to developmental levels and stages when working with diverse clients/students
27. Attends to issues of culture when working with clients/students
28. Collaborates with the client/student or stakeholders to establish and facilitate movement toward measurable therapeutic goals
29. Develops and applies effective counseling treatment plans consistent with individual and/or group counseling
30. Evaluates effectiveness of counseling and counseling programs as appropriate to setting
31. Demonstrates case conceptualization skills that include attention to theoretical framework, culture, and development
32. Integrates understanding of systemic interactions into conceptualization and planning
33. Navigates diverse service delivery models and programs within CMHC settings
34. Uses formal and informal assessment strategies to guide diagnosis and conceptualization
35. Demonstrates effective case management skills, including those related to record keeping, third party reimbursement, and interfacing with other care providers
36. Demonstrates skills for developing and managing a comprehensive, developmental school counseling program that aligns with the TN model
37. Promotes academic development
38. Promotes career and college readiness
39. Promotes social-emotional development
40. Engages in opportunities to be an advocate for all students (e.g., students with disability, special needs, gifted)
41. Demonstrates accountability by using data and sharing results

Appendix B
Table 1

<i>Variance Components</i>			
Scale	Variance Component	Variance	% Explained Variance
<i>Counseling Skills</i>			
	Student (σ_p^2)	.002	0.8
	Course (σ_c^2)	.008	3.0
	Item (σ_i^2)	.014	5.1
	Form (σ_f^2)	.012	4.4
	Supervisor*Course ($\sigma_{(s:c)}^2$)	.059	22.0
	$\sigma_{p,(s:c)}^2$.010	4.1
	$\sigma_{f,(s:c)}^2$.013	4.9
	Other*	.036	13.3
	Residual ($\sigma_{p,c,i,f,s:c,e}^2$)	.114	42.4
<i>Professional Dispositions</i>			
	Student (σ_p^2)	.005	1.9
	Course (σ_c^2)	.002	0.9
	Item (σ_i^2)	.009	3.5
	Form (σ_f^2)	.007	2.8
	Supervisor*Course ($\sigma_{(s:c)}^2$)	.060	22.9
	$\sigma_{i,f}^2$.018	7.0
	$\sigma_{i,(s:c)}^2$.015	5.9
	$\sigma_{f,(s:c)}^2$.022	8.4
	$\sigma_{p,c,i}$.011	4.0
	Other*	.011	4.0
	Residual ($\sigma_{p,c,i,f,s:c,e}^2$)	.102	38.7
<i>Conceptualization and Planning</i>			
	Student (σ_p^2)	.004	1.4
	Course (σ_c^2)	.008	3.3
	Item (σ_i^2)	.001	0.2
	Form (σ_f^2)	.013	5.2
	Supervisor*Course ($\sigma_{(s:c)}^2$)	.071	28.0
	$\sigma_{p,(s:c)}^2$.011	4.4
	$\sigma_{i,(s:c)}^2$.016	6.3
	$\sigma_{f,(s:c)}^2$.020	8.0
	Other*	.016	6.1
	Residual ($\sigma_{p,c,i,f,s:c,e}^2$)	.095	37.1

*Interaction variance components less than 2% have been condensed into this category.

Appendix C
Table 2

D Study Coefficients

Scale	Generalizability Coefficient (ρ)	Dependability Coefficient (ϕ)
<i>Counseling Skills</i>	.014	.008
<i>Professional Dispositions</i>	.035	.019
<i>Conceptualization and Planning</i>	.030	.014

Appendix D
Table 3

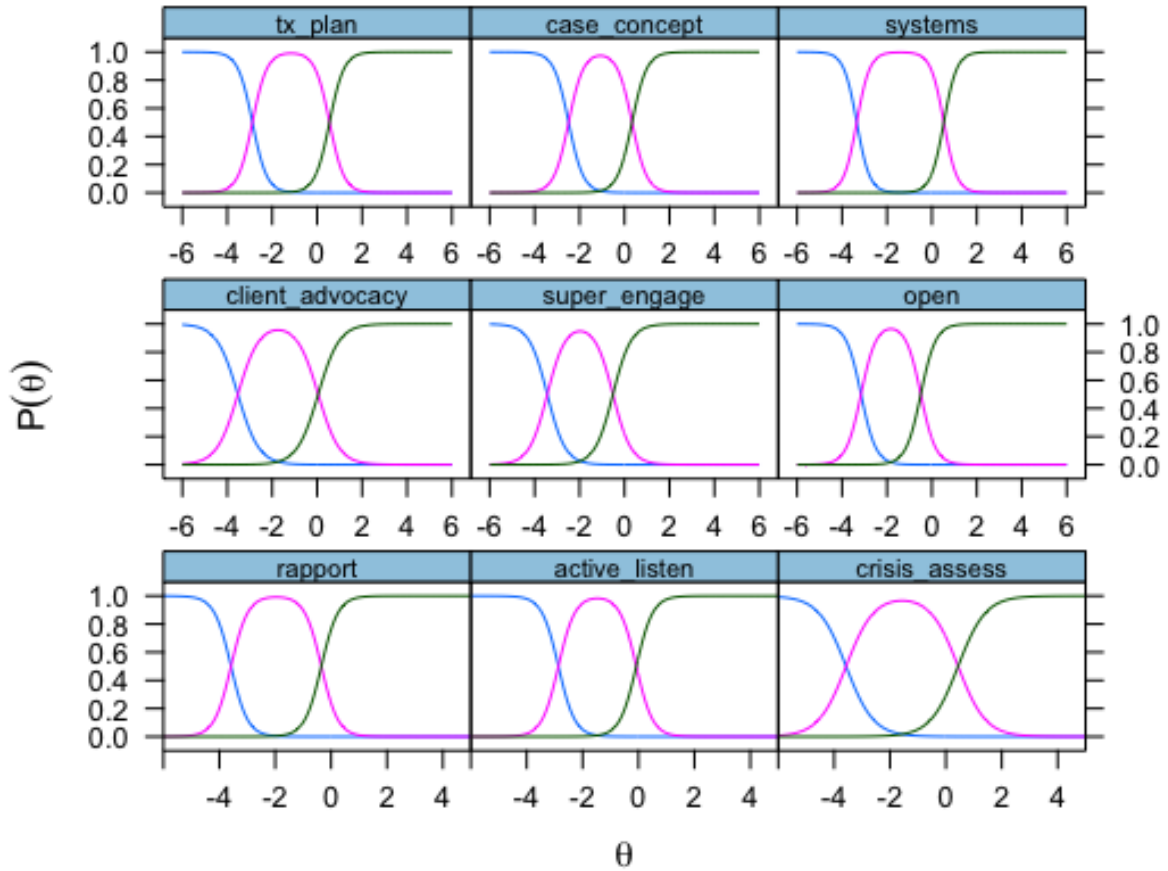
Graded-Response Model Item Parameters

Scale	Item	α	β_1	β_2
<i>Counseling Skills</i>				
	Rapport	3.34	-3.58	-0.35
	Empathy, Unconditional Positive Regard, & Genuineness	3.08	-3.65	-0.42
	Active Listening	3.39	-2.87	-0.08
	Questions	3.36	-2.872	0.30
	Nonverbal Behaviors	2.91	0.08	
	Evaluates Effectiveness of Interventions	2.11	-3.66	0.45
	Collaborates with Stakeholders	1.41	-4.30	0.50
	Crisis Assessment	2.03	-3.57	0.44
	Use of Theory	2.15	-2.54	0.49
<i>Professional Dispositions</i>				
	Ethical and Legal Sensitivity	2.59	-3.37	0.09
	Promotes Confidence in the Counseling Profession	3.41	-0.10	
	Advocates for Clients	2.13	-3.53	0.05
	Advocates for the Profession	3.35	0.54	
	Time Management	1.64	-3.12	0.12
	Engagement in Supervision	2.48	-3.45	-0.51
	Commitment	2.91	-3.45	-0.04
	Openness	3.01	-3.15	-0.49
	Respect	3.83	-3.13	-0.21
	Integrity	4.40	-3.03	-0.27
	Self-Awareness	3.21	-2.76	-0.15
<i>Conceptualization and Planning</i>				
	Developmental Awareness	3.45	-3.35	0.43
	Cultural Awareness	3.33	-3.15	0.46
	Collaborates with Clients	2.44	-2.92	0.39
	Treatment Planning	3.10	-2.88	0.55
	Self-Evaluation	3.47	0.63	
	Case Conceptualization	3.05	-2.49	0.34
	Integration of Systemic Understanding	3.46	-3.35	0.53

Appendix E

Item Trace Graphs

Does not meet expectations — (blue line)
Meets expectations — (magenta line)
Exceeds expectations — (green line)



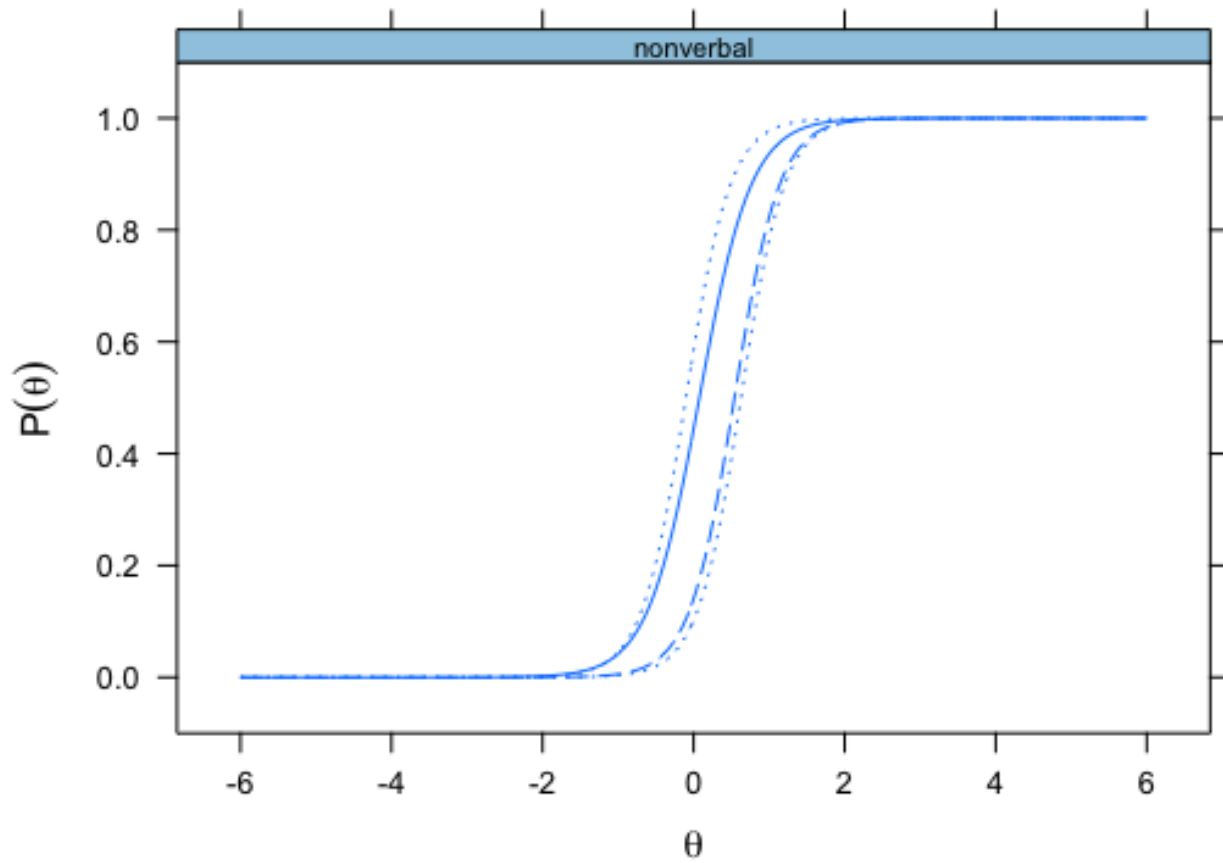
Item Trace Graphs

Figure 1

Appendix F

Item trace lines

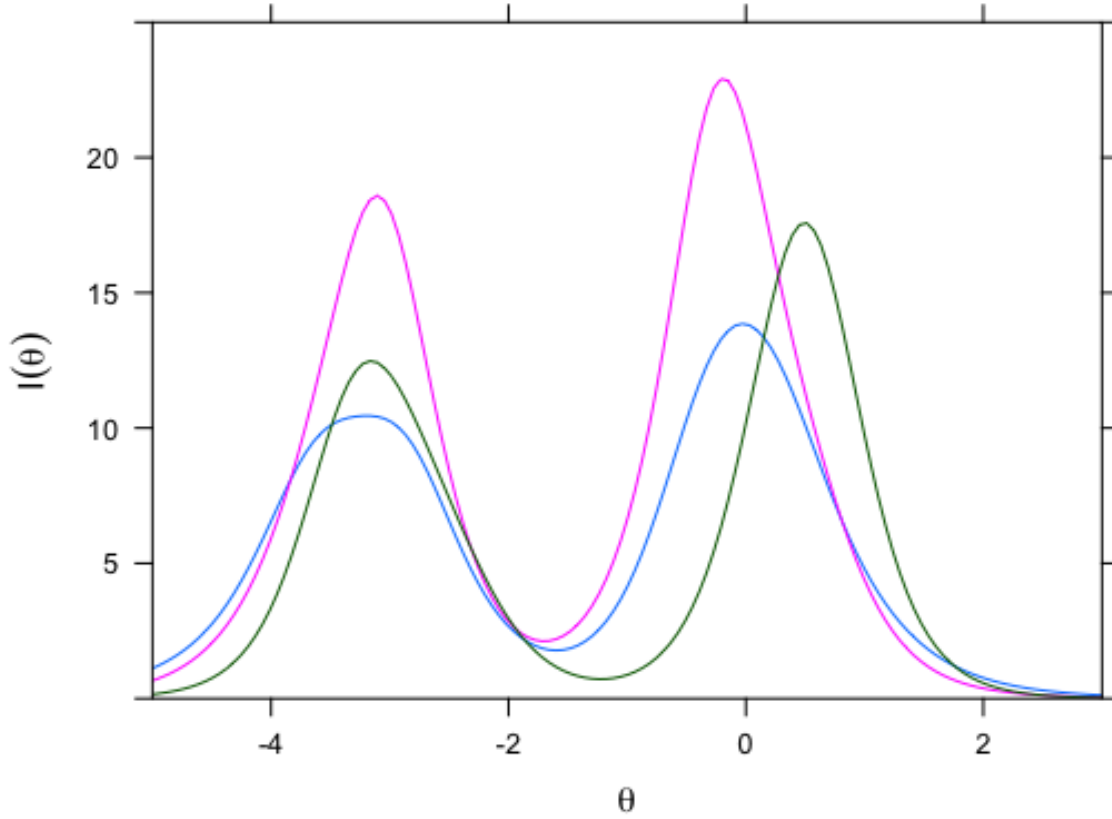
Nonverbal Behaviors ———
Promotes Confidence in the Counseling Profession
Advocates for the Profession - - -
Self-Evaluation



Item Trace Line Graphs for Poorly Performing Items
Figure 2

Appendix G

Counseling Skills —
Professional Dispositions —
Conceptualization and Planning —



Test Information Graph for Subscales
Figure 3

Conclusion

Inconsistencies in the ways we prepare and assess counseling students can jeopardize the development of high quality counselors, make gatekeeping and remediation difficult, and bring harm students' current and future clients. There is no magic bullet for these issues, but there are existing methods which can help improve the quality and consistency of instruction and assessment across the counseling profession. This dissertation proposed two possible solutions to the problem of skill development and assessment: deliberate practice and assessment development. I intend for the manuscripts presented above to add to the conversation by identifying existing tools available to counselor educators to improve the quality of education and assessment already occurring.

Deliberate practice may hold particular promise for counselor education because it can enhance the effectiveness of what counselors are already doing. Rather than an entirely new intervention, deliberate practice integrates the feedback, scaffolding, and supervision processes which exist into a framework which highlights the role that solo practice has in skill development. Counselor educators and counseling students can benefit from deliberate practice by understanding how these processes work together and implementing practice, homework, or rehearsal in coursework and supervision. Deliberate practice also hopes to help learners develop clear mental models of high quality skill, which, in turn, will improve their own practice as they advance through their careers.

When assessing student counselors' field experiences, counselor educators may rely on rubric development skills, but neglect important psychometric attributes which can allow them to trust assessment results. The results can be assessments which provide little useful evidence for monitoring student development or demonstrating program outcomes. As psychometric skill

exists within both counselor education programs and in the programs with which counselor education is adjacent, integrating these practices into student assessment can improve programs' abilities to collect useful data about students. As programs improve the quality of data collection about students' performance as counselors, they can reintegrate this information into program planning and decision making, further improving the quality of both program and program graduates.

Practically speaking, change in education, health, and research fields is a slow process. Funders, participants, administrators, and gatekeepers must be convinced, and it requires cultural and political will to make large changes a reality. The counseling profession, like the other mental health professions, faces problems which directly impact the care of clients and the development of students. Above, I propose that integrating deliberate practice into counselor education and focusing on improving assessment methods through the application of psychometrics as methods of addressing existing limitations. These methods align with the goals and existing processes of counselor education and may be amenable to the many stakeholders in counselor education. Rather than broad changes to the counselor profession, small changes by programs and individuals may help bring counseling into greater alignment with its goals and philosophy.

Vita

Zach Budesá grew up in Mt. Juliet, Tennessee. Zach Budesá complete a Bachelor of Science degree in Psychology and a Master of Arts in Clinical Mental Health Counseling at Tennessee Technological University. Following working as a counselor for 2 years, he chose to pursue a Doctor of Philosophy degree in Counselor Education at the University of Tennessee, Knoxville with a focus on quantitative methods. His research interests include skill development and measurement. Following graduation, he will begin as a postdoctoral research fellow with the University of Tennessee Health Science Center.