Assessment of Dispositions in Program Admissions: The Professional Disposition Competence Assessment—Revised Admission (PDCA-RA)

Curtis Garner
Gonzaga University

Brenda Freeman
University of Nevada, Reno

Roger Stewart
Boise State University

Ken Coll
University of Nevada, Reno

Assessment of Dispositions in Program Admissions: The Professional Disposition Competence Assessment—Revised Admission (PDCA-RA)

Curtis Garner, Brenda Freeman, Roger Stewart, Ken Coll

Tools to assess the dispositions of counselor education applicants at the point of program admission are important as mechanisms to screen entrance into the profession. The authors developed the Professional Disposition Competence Assessment—Revised Admission (PDCA-RA) as a screening tool for dispositional assessment in admissions interviews. In this study, 70 participants engaged in a video-based training protocol designed to increase the interrater reliability of the PDCA-RA. An intraclass correlations coefficient was calculated as an index of interrater reliability. Cronbach's alpha coefficients were calculated for internal consistency, and Fleiss' kappa, free-marginal kappa, and percent of agreement were calculated for absolute agreement. Calculations were made for pretest and posttest scores. Results of the study suggest that the PDCA-RA demonstrates "good" reliability in terms of interrater reliability and "excellent" reliability in terms of internal consistency. The video-based training improved interrater reliability.

Keywords: **dispositions, counselor education, interrater reliability, counseling admissions, PDCA-RA**

Beyond ethical codes and standardized education requirements, one criterion understood to be a demarcation of a profession is that it controls entry into its occupation (Miller, 2006). The stature of any profession is heavily influenced by the collective quality, preparation, and professional fit of those who are allowed to enter the profession. In the profession of counseling, counselor preparation programs, practicing counselors, field site supervisors, and state licensure boards share the overarching charge to screen for the profession (Freeman et al., 2016), but counselor educators alone bear the responsibility of initial screening of potential new entrants into the profession. The funnel of individuals seeking entrance into the profession begins with admission to graduate programs. This responsibility is a solemn one because post-admission gatekeeping can lead to high-stakes legal disputes (Dugger & Francis, 2014; Hutchens et al., 2013; McAdams et al., 2007).

Similar to other graduate programs, criteria for entrance into counselor preparation programs generally incorporate academic and career factors, but unlike many other graduate programs, the dispositions (traits and characteristics) of applicants are also critical factors for identifying appropriate candidates for the profession (Hernández et al., 2010). The use of admissions interviews is a common method for observing dispositions (Swank & Smith-Adcock, 2014). Characteristics such as interpersonal skills, warmth, emotional stability, and self-awareness are examples of traits deemed important to many counseling academic programs (Crawford & Gilroy, 2013; McCaughan & Hill, 2015), though counselor educators lack agreement about which dispositions should be screened at admission (Bryant et al., 2013).

Curtis Garner, EdD, NCC, NCSC, LCPC, is a professor and department chair at Gonzaga University. Brenda Freeman, PhD, is a professor at the University of Nevada, Reno. Roger Stewart, PhD, is a professor at Boise State University. Ken Coll, PhD, is the Dean of the School of Education at the University of Nevada, Reno. Correspondence may be addressed to Curtis Garner, 502 East Boone Ave., Spokane, WA 99258-0102, garnerc@gonzaga.edu.

Once applicants have been accepted into a counselor education program, if problematic dispositional issues arise the American Counseling Association (ACA) ethical codes require remediation (ACA, 2014), which is sometimes followed by suspension or dismissal. Therefore, *gatekeeping*, defined as the process of deterring program graduation of those lacking sufficient knowledge or skills (Koerin & Miller, 1995), begins at the point of program screening and admission (Kerl & Eichler, 2005). Bryant et al. (2013) emphasized that effective screening of applicants prior to formal admission into the academic program may greatly reduce the need to address problematic student behaviors after admission.

In addition to conducting admissions screening as a form of gatekeeping, the courts are more likely to support universities in admissions-related legal disputes if screening policies, standards for admission, and admission procedures are clear and fair (Cole, 1991). Legal research also underscores the importance of programs communicating clearly with students about the expected dispositions and other criteria from admission through exit (McCaughan & Hill, 2015). Reliable admissions tools designed to assess dispositions represent one method of showing fidelity in implementing policies (Hutchens et al., 2013). Despite the research support for sound structures to scaffold the admissions process, assessments with published psychometric properties measuring dispositions in admissions interviews are scarce (Hernández et al., 2010).

Jonsson and Svingby (2007) noted that a number of forms of reliability and validity are important in establishing the psychometric properties of admissions tools, but when multiple raters are involved, such as in the admissions process, interrater reliability for rubrics is particularly essential. Specific training in the tool is critical to improving interrater reliability (Jonsson & Svingby, 2007). Video training protocols to increase interrater reliability are becoming more important in professional dispositional research (Kopenhaver Haidet et al., 2009; Rosen et al., 2008). The use of video technology to train raters to capture behavioral observations has two advantages: the opportunity for admissions personnel to practice admissions interview ratings prior to real-time observations, and the relative ease of using modern, sophisticated recording equipment (Kopenhaver Haidet et al., 2009).

Admissions Processes and Criteria

Overwhelmingly, admission criteria and procedures for counselor education programs have focused upon undergraduate grade point average (GPA); standardized test scores, such as the Graduate Record Examination (GRE) or the Miller Analogies Test (MAT); a personal interview; and some form of personal statement (Bryant et al., 2013). Such procedures have been shown to be reasonably predictive of academic success, but less so for counselor development (Smaby et al., 2005). Some programs have utilized Carkhuff's Rating Scale (Carkhuff, 1969) or Truax's Relationship Questionnaire (Truax & Carkhuff, 1967) to measure applicants' ability to communicate the conditions of empathy, genuineness, and respect effectively (Hernández et al., 2010; Swank & Smith-Adcock, 2014). Carkhuff's Rating Scale and Truax's Relationship Questionnaire have been found to exhibit good interrater reliability and, when correlated with one another, have been found to exhibit considerable overlap (Engram & Vandergoot, 1978).

Dispositional Assessment

Following the gatekeeping dispute in *Ward v. Wilbanks* (2010), in which a graduate student in counselor education refused to work with a gay client, and the ensuing litigation upon that student's dismissal from their program, the need for a reliable method for evaluating counseling student dispositions has become increasingly apparent. This high-profile legal case also highlighted the need to monitor and document student dispositions (Dugger & Francis, 2014; McAdams et al., 2007). Correspondingly, in 2009 the Council for Accreditation of Counseling and Related Educational Programs (CACREP) released standards that made monitoring student dispositions a mandatory aspect of program evaluation. In the

2016 CACREP standards the expectation for the assessment of counselor-in-training dispositions was expanded to include the monitoring of dispositions at multiple points over the duration of time students are enrolled in a counselor education program. The accreditation expectations for screening at the point of admission are found in Section I.L., where the standards delineate the expectation that counseling programs consider dispositions (CACREP, 2015). Dispositions for consideration include relationship skills and cultural sensitivity.

As the need for dispositional appraisal has become increasingly imperative in the counselor education profession, there have been various efforts to design specific approaches to assess student dispositions (Frame & Stevens-Smith, 1995; Kerl et al., 2002; Lumadue & Duffey, 1999; McAdams et al., 2007; Redekop & Wlazelek, 2012; Williams et al., 2014). One early approach was the utilization of standardized personality tests (Demos & Zuwaylif, 1966; Utley Buensuceso, 2008). However, the use of personality tests fell into disfavor because of the potential for conflicts with the Americans with Disabilities Act (U.S. Department of Justice, 2010) as well as for their inherent deficit orientation. Consequently, the use of standardized tests has been generally replaced by rating scales and rubrics (Panadero & Jonsson, 2013).

One reason that rubrics were considered superior to rating scales was their transparency (Panadero & Jonsson, 2013). Transparency empowers students by equipping them with an understanding of expectations for performance prior to their creating a product or performing a skill. Rubrics also have greater potential to align with learning outcomes and they provide useful direct feedback to students (Alexander & Praeger, 2009; Panadero & Jonsson, 2013).

Examples of dispositional assessments for counselors include the Counselor Characteristics Inventory (Pope, 1996), an inventory that assesses personality characteristics of effective counselors. Also, Spurgeon et al. (2012) described a process that includes a Likert-style assessment of dispositional traits. In addition, Swank et al. (2012) developed the Counseling Competencies Scale (CCS), a tool for measuring counselor competence. Frame and Stevens-Smith (1995) developed a 5-point Personal Characteristics Evaluation Form, and finally, Lumadue and Duffey (1999) published a Professional Performance Fitness Evaluation to evaluate specific behaviors of pre-professional counselors. Few studies of the reliability and validity of the tools were found in published research, especially related to admissions. However, some do have limited published psychometric research and in some cases norms (Flynn & Hays, 2015; Pope, 1996; Swank et al., 2012; Taub et al., 2011).

One example of a dispositional tool for counselor education with published psychometrics is the Counselor Personality Assessment (CPA) developed by Halinski (2010). The CPA is a 28-item scale reporting a Cronbach's alpha reliability score of .82. Another tool, the CCS (Swank et al. 2012), is a 32-item rubric for measuring counseling skills, professional conduct, and professional dispositions in practicum. Cronbach's alpha for the CCS was reported at .93, and interrater reliability was reported at .57. Criterion validity was established by correlating the CCS score with the semester grade and was reported as moderate. The available psychometric data for the CPA and CCS represent exceptions. In general, lack of psychometric information may result in limited confidence in available assessment tools for appraising counselor student dispositions.

Interrater Reliability

Interrater reliability, essentially the extent to which the raters assign the same scores when observing the same behaviors (McHugh, 2012), is critical for fairness to applicants in counseling admissions interviews. Gwet (2014) stated, "If the inter-rater reliability is high, then raters can be

used interchangeably without the researcher having to worry about the categorization being affected by a significant rater factor. Interchangeability of raters is what justifies the importance of inter-rater reliability" (p. 4). Consistency ensures that the data collected are realistic for practical use. When interrater reliability is poor, interviews conducted by overly critical raters (*hawks*) naturally lead to negative bias against applicants when compared within the same applicant pool with the scores from interviews rated by less critical raters (*doves*). Epstein and Synhorst (2008) discussed interrater reliability as an approximation in which different people rate the same behavior in the same way. Thus, interrater reliability can also be understood as rater consensus.

Purpose of the Present Study

Effectively screening and selecting new entrants is one of the hallmarks that distinguishes a profession. Unfortunately, there is a dearth of available literature on assessment tools for rating admissions interviews. Further, lack of information on the reliability of the tools that exist represents a significant deficiency in professional literature (Johnson & Campbell, 2002). The Professional Disposition Competence Assessment—Revised Admission (PDCA-RA; Freeman & Garner, 2017; Garner et al., 2016) is a global rubric designed to assess applicant dispositions in brief graduate program interviews. The PDCA-RA includes a video training protocol developed to facilitate consistency across raters in scoring admissions interviews on dispositional domains.

The purpose of the study was to examine the internal stability and the interrater reliability of the PDCA-RA. The rationale for the study was that no similar rubrics assessing dispositions at admissions using training videos were found in published research, suggesting a gap in the literature. Interrater reliability was the key focus of this study because of the importance of interrater reliability for rubrics utilized in situations with multiple raters, a typical scenario for counselor education admissions processes.

Method

Sample

Raters for the study included 70 counselor educators, counseling doctoral students, adjunct faculty, and site supervisors. Counselor educators, doctoral students, and adjunct faculty at two universities were asked to participate in trainings on the new admissions screening tool. Site supervisors providing supervision for practicum and internship students at the two universities were offered training in the PDCA-RA as a component of continued professional development to maintain their supervision status. Training in both instances was free and included professional development credits. Informed consent for participation was obtained from all participants in accordance with ACA ethical codes (ACA, 2014) and IRB oversight at both universities. All participants in the study fully completed the PDCA-RA video-based training. The mean age of the raters was 43.9 (*SD* = 11.4, range 24–72). Sixty-four percent identified as female and 36% identified as male. Mean average years of experience indicated as a faculty or field supervisor was 12.2 (*SD* = 9.7, range 1–50). Ninety-three percent identified as White/Caucasian, 6% as Latino/a, and 1% as other ethnicity.

The counselor educators (27% of the sample) were primarily from two CACREP-accredited counseling programs in the Western United States. Participating universities included one private university and a state research university, both with CACREP-accredited programs. Counselor education doctoral students and adjunct faculty participants comprised 7% of the sample. The doctoral students participated in the training because they were involved as raters of master's-level

counselor education applicants in the admissions process at one institution. The remaining 66% of the participants were field site supervisors. Because field site supervisors were involved in gatekeeping, attending training in dispositional assessment was natural to their role as internship site supervisors.

Measure: PDCA-RA

The PDCA-RA was developed on the basis of the Professional Disposition Competence Assessment (PDCA; Garner et al., 2016). The PDCA, a dispositional gatekeeping tool, was revised to the Professional Disposition Competence Assessment-Revised (PDCA-R) after several rounds of use and with feedback from expert panels (Freeman & Garner, 2017). Advice from legal counsel was also reflected in the revision of the PDCA to the PDCA-R. The PDCA-R was originally used for both gatekeeping and admissions purposes, but it was determined that the PDCA-R was best used for gatekeeping, not for admissions screening, because the tool implied that the rater had prior knowledge of the student. Because this is often not the case in admissions screening, the PDCA-RA was developed.

The PDCA, PDCA-R, and PDCA-RA were conceptualized and developed through a comprehensive review of the literature, several rounds of field testing, and adjustments from expert faculty panels at two institutions. In addition to counseling literature on impairment and expert panel feedback, the Five-Factor Model, often referred to as the "Big Five" (Costa & McCrae, 1992), influenced three of the nine dispositional items. The Five-Factor Model consists of five personality traits consistently associated with positive mental health, academic success, and healthy habits and attitudes across the life span: Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Openness. The PDCA-RA dispositions are identical to the PDCA-R, with the exception of the disposition of Ethics. Ethics was removed from the PDCA-RA because the description assumed knowledge of professional ethical standards, a doubtful expectation for program applicants with no prior training in counseling. The behavioral descriptions in the PDCA-RA were narrowed so they described only those behaviors that can be observed in admissions interviews with no prior knowledge of the applicants. In addition, the rubric item descriptions were shortened to align with the practical context of brief (20- to 30-minute) admissions interviews in which there may be limited time for in-depth assessment.

If dispositions are thought of as traits, as per the definition of dispositions in the CACREP glossary (CACREP, 2015), the PDCA-RA is not technically directly measuring dispositions. Based upon advice from legal counsel, as well as the practicality of assessing applicants during short admissions interviews, the PDCA-RA assessed *behaviors* associated with dispositions and not the actual dispositions. Behaviors identified for each disposition can be observed during a short admissions interview, whereas personality traits would require a more in-depth assessment approach, one that counselor educators fear might be found legally problematic (Freeman et al., 2019; Schuermann et al., 2018).

The nine dispositions assessed in terms of observable behaviors via the rubric are Conscientiousness, Coping and Self-Care, Openness, Cooperativeness, Moral Reasoning, Interpersonal Skills, Cultural Sensitivity, Self-Awareness, and Emotional Stability. Each disposition in the PDCA-RA is rated on a scale of three levels—*developing*, *meets expectation*, and *above expectation*. The PDCA-RA is described in more detail in a manual that includes the tools as well as three suggested admissions questions for each of the nine dispositions (Freeman & Garner, 2017). The measure of internal consistency for faculty ratings of the original PDCA rubric was a Cronbach's alpha estimated at .94 (Garner et al., 2016). Cronbach's alpha for self-ratings was .82, and Cronbach's alpha for peer ratings was .89. The straightforward modifications from the original PDCA to the PDCA-RA were minimal and unlikely to significantly affect these measures of internal consistency.

Procedure

A video-based training protocol was developed for the purpose of training faculty in counselor education programs, doctoral students, site supervisors, and other admissions raters to use the PDCA-RA to assess the dispositions of graduate program applicants (Freeman & Garner, 2017; Garner et al., 2016). The video was presented to participants by a trainer. The trainer also greeted participants, obtained informed consent, passed out PDCA-RA forms when prompted by the training video, and collected completed PDCA-RA forms for later analysis. Training in the use of the PDCA-RA was important not only as a mechanism to establish interrater reliability but also as a means of informing adjustments to the tool during its initial iterative development process. Development of the video-based training protocol progressed through several stages. At first, the original 90-minute training consisted of a faculty team of seven working together as a group to read and discuss each disposition, followed by each faculty viewing an admissions interview video and rating the applicant independently. Faculty then discussed their ratings, leading to subtle adjustments in the rubric item descriptions. Additional benefits to the training were an increase in faculty self-awareness of *dove* and *hawk* tendencies when rating admissions applicants and self-awareness associated with interview bias. With continued training and feedback, the original training protocol was significantly improved.

To complete the next step in the creation of the video-based training protocol, counseling student volunteers were offered a minimal incentive to come to the film studio, and after signing waivers to allow the film clips to be used, the student volunteers were asked to respond to various admissions interview questions. The faculty filming the students instructed them to "give a strong answer" or "give a weak answer." The researchers treated all responses as unscripted role plays. The questions asked by the interviewer for each disposition were those found in the PDCA-RA materials (Freeman & Garner, 2017). Finally, the authors and developers of the training video reviewed over 100 film clips, removed those in which the acting interfered with the purpose of the video, and rated the remaining clips using the PDCA-RA, resulting in ratings of *1*, *3*, or *5*. These numerical ratings corresponded to descriptive ratings of *developing*, *meets expectation*, and *above expectation*, respectively. Clips in which the researchers found the rating to be difficult were removed from consideration. In selecting the final 18 clips (two for each of nine dispositions), the researchers considered diversity in age, ethnicity, gender, and disability of the student volunteers. The goal was to create video clips of student volunteers with diverse characteristics.

The result was a video-based training protocol that could still be completed by trainees in 90 to 120 minutes. The video training protocol began with an introduction to the PDCA-RA, followed by prompts to rate the video-recorded vignettes using the PDCA-RA prior to receiving training. This initial rating of the vignettes was considered the pretest condition. Training on the application of the PDCA-RA to the vignettes was next. Training included revealing ideal scores as determined by the authors, the reasoning behind the scoring, and opportunities to discuss scoring among participants. Following the training on the PDCA-RA, participants were, once again, given the PDCA-RA rubric along with a new set of video-recorded vignettes. This was considered the posttest condition. Participants were asked to rate the new vignettes using the PDCA-RA.

The video-based training protocol, designed for use in small groups, allowed for group discussion of ratings after each participant completed the PDCA-RA independently. This was indicated by a written message on the video reading, "Pause video for discussion." The training tape ended with a narrator discussion of how to use the PDCA-RA in actual admissions interviews, including comments on cultural sensitivity in admissions interviews.

The video-based training protocol was used as the means of training participants in dispositional assessment. The purpose of the trainings was to increase consistency of admissions raters in evaluating the admissions interviews of applicants to a master's-level counselor education program. Typically, participants completed the video training in small groups consisting of approximately six to 10 people. In addition to viewing the training video, participants also took part in group discussion and established a consensus of opinion on group ratings of video clips. Coming to a consensus on ratings, which also included feedback on rubric items and video clips, was an important aspect of the training.

Statistical Analysis

The PDCA-RA scores from the counselor education faculty, adjunct faculty, doctoral students, and site supervisors' ratings of the vignettes before training were used as the pretest or baseline interrater reliability. The PDCA-RA scores after participants were trained in the tool were used as the posttest. The intraclass correlation coefficient (ICC) was calculated as a measure of interrater reliability. Interrater reliability correlations quantify rater subjectivity (Herman et al., 1992). The ICC was calculated for pretest and posttest scores. Cronbach's alpha coefficients were calculated for internal consistency, and Fleiss' kappa (κ) was calculated for absolute agreement. In addition, Fleiss' free-marginal kappa (κ_{free}) and percent overall agreement were calculated. Calculations were made for both the pretest and posttest ratings, and a *t*-test was conducted, using SPSS, to determine whether training improved interrater reliability.

Results

The ICC estimates and associated 95% confidence intervals were calculated using SPSS statistical package version 23 and based on an individual rating, absolute agreement, 2-way random-effects model. ICC single measures for absolute agreement were calculated for the pretest administration of the PDCA-RA at .53 (95% CI [0.333–0.807]). The ICC single measures for absolute agreement were calculated for the posttest administration of the PDCA-RA at .76 (95% CI [0.582–0.920]). Cronbach's alpha was calculated at .99 for both pretest and posttest scores. Pretest and posttest ICCs were compared using a *t*-test with an a priori significance level set at .05. The test was significant (p < .05), suggesting that there was a difference between the pretest and posttest reliability, with reliability improving from the "moderate" range to the "good" range (Koo & Li, 2016) with training.

Using Excel, kappa (κ) was calculated as a measure of overall agreement for pretest and posttest scores. This particular kappa was extended by Fleiss (1971) and accommodates multiple raters like those rating the PDCA-RA. Assumptions underpinning Fleiss' kappa include categorical data (i.e., nominal or ordinal) with mutually exclusive categories, symmetrical cross-tabulations, and independence of raters. Data in this study met all assumptions. Data was ordinal with three mutually exclusive response categories for each dispositional area assessed, which resulted in all cross-tabulations being symmetrical. Although raters were trained in a collaborative setting where discussions about ratings were fostered, when the actual ratings of study participants occurred, raters did not discuss their ratings with others and were thus independent of one another. Pretest scores for the nine rubric items reflected a κ of .33, fair agreement according to Landis and Koch (1977). After training, posttest scores on the nine items reflected a κ of .55, moderate agreement according to Landis and Koch.

As an additional analysis, percent overall agreement and κ_{free} was calculated. κ_{free} is appropriate when raters do not know how many cases should be distributed into each category. In addition, κ_{free} is resistant to influence by prevalence and bias (Randolph, 2005). The percent of overall agreement is the measure of agreement between raters and historically has also been used to calculate interrater

reliability (McHugh, 2012). Table 1 illustrates that the κ_{free} for the pretest was .36 while the percent of overall agreement was 57.6%. The posttest for the κ_{free} was .56 and the percent of overall agreement was 70.4%. After examining the change in pretest to posttest calculations from both the κ_{free} and the percent of overall agreement, both offer additional support for and provide evidence that training improved the agreement of dispositional ratings on the PDCA-RA.

Table 1

Pre and Post Statistics: Percent Overall Agreement and Free-Marginal Fleiss' Kappa

Time of Rating	Percent Overall Agreement	Free-Marginal Kappa	95% CI for Free-Marginal Kappa
Before Training: Pre	57.6	.36	[.23, .49]
After Training: Post	70.4	.56	[.31, .80]

The overall κ, κ_{free} and percent of agreement results were promising, but a comparison of the percent of correct responses (the response intended by the research team) by disposition showed that the ratings of correct responses decreased by more than 2% from pre- to posttesting for three dispositions (Openness, Cooperativeness, and Moral Reasoning). Because this was an unexpected finding, the research team analyzed the ratings for incorrect responses and learned that the raters appeared to be better able to discern the difference between a rating of *1* (*developing*) and *3* (*meets expectation*) than between *3* and *5* (*above expectation*). As a post-hoc analysis, we calculated the percent of agreement with the correct score, collapsing the *3* and *5* ratings. The percent of correct responses with dichotomous categories of *1* and a collapsed category for *3* and *5* are shown in Table 2. As is evident in Table 2, using the collapsed category, the percent of correct responses for eight of the nine dispositions improved from pretest to posttest. The percent of correct responses for one disposition, Cooperativeness, decreased by more than 2% from pretest to posttest.

Table 2

Pre and Post Percent of Correct Responses by Disposition

Disposition	Pre Percent Overall Agreement 1, 3, 5 Ratings	Post Percent Overall Agreement 1, 3, 5 Ratings	Pre Percent Overall Agreement 1, 3 & 5 Collapsed	Post Percent Overall Agreement 1, 3 & 5 Collapsed
1. Conscientiousness	62.0	97.1	77.1	98.6
2. Coping & Self-Care	59.9	94.4	22.9	97.1
3. Openness	51.0	49.4	94.3	100.0
4. Cooperativeness	47.3	39.0	94.3	87.1
5. Moral Reasoning	84.1	68.8	91.4	98.6
6. Interpersonal Skills	48.0	94.4	98.6	97.1
7. Cultural Sensitivity	69.3	94.4	100.0	100.0
8. Self-Awareness	40.7	40.0	54.3	64.3
9. Emotional Stability	56.3	56.5	67.1	95.7

Discussion

The results of the study suggest that the PDCA-RA has potential as a reliable instrument for assessing counseling applicants at the point of program admission. The PDCA-RA demonstrated strong reliability from the standpoint of internal consistency. The interrater reliability, as measured by the ICC, moved from the "moderate" to the "good" range with the application of the standardized training protocol.

The results of the study also provide evidence that counselor educators, supervisors, and doctoral students can improve their agreement on ratings of student dispositions with adequate and appropriate training. Multiple statistical techniques for measuring agreement, including the ICC, κ, κ$_{free,}$ and percent agreement measured under pre-training and post-training conditions demonstrated overall improvement in rater agreement with training. The observed post-training improvement in interrater reliability corroborates the literature, underscoring the necessity of training protocols as the pathway to improved interrater reliability (Jonsson & Svingby, 2007).

The results from the second analysis conducted through collapsing the *meets expectation* and *above expectation* categories suggest that the PDCA-RA has higher reliability as a tool to screen out inappropriate candidates than to distinguish excellence within the pool of acceptable candidates. For programs seeking to eliminate problematic applicants, the PDCA-RA could prove reliable. However, for academic programs with large numbers of applicants with an objective to accept a small group of students from a large group of acceptable candidates, the PDCA-RA may be less reliable from an interrater reliability perspective. The PDCA-RA item descriptions for *above expectation* need further consideration.

The percent of correct responses after training with collapsed categories was over 87% for seven of the nine dispositions. The results suggest that the PDCA-RA or the PDCA-RA training protocol needs revision on two dispositions, Cooperativeness and Self-Awareness. The decrease in correct responses to Cooperativeness may be due to a posttest interview with a higher level of difficulty than the pretest interview. The posttest percent was 87%, suggesting that overall the rubric descriptions functioned as acceptable with this sample of raters, though not excellent. The percent of correct ratings for Self-Awareness increased from pre- to posttesting, but only to 64% agreement. One explanation could be that the Self-Awareness rubric descriptions are behavioral (as recommended by legal counsel), yet Self-Awareness as a trait is difficult to describe in behavioral terms. This could leave raters confused about the difference between their intuitive sense of the self-awareness of the applicant and the narrow behavioral descriptions on the rubric. An alternative explanation is that there is a lack of agreement in the profession on the extent of self-awareness expected from students entering the academic program, leading some raters to find the applicant's level of self-awareness acceptable, while others found the level unacceptable. In either case, the training protocol for the PDCA-RA and perhaps the rubric description need improvement. The 100% posttest agreement on the dichotomous categories for Openness and Cultural Awareness were encouraging, given the critical importance of these two dispositions (Freeman et al., 2019).

Interrater reliability is of paramount importance for the responsible use of rubrics. To improve the interrater reliability of the PDCA-RA, three issues may need to be addressed. First, the training protocol may need to be lengthened to encompass three rather than two opportunities to rate video clips. Second, structuring the discussion between raters with questions focusing attention on the gaps in ratings could be beneficial. Third, because alternate forms of the videos are being used in the training (different actors with different responses to the same question), a comparison of the

complexity of the video clips should be conducted. It may be desirable to revise the training protocol to utilize less complex responses for Part 1 training, followed by equivalent complex interviews for Part 2 training, and more complex interview responses for Part 3. More complex responses, meaning the responses are partially descriptive of two categories on the rubric, are realistic to actual admissions interviews in the field.

In conducting trainings for the PDCA-RA, a potentially interesting observation was that raters appeared predisposed to using their own subjective experience to rate the video interviews instead of applying the item descriptions in the rubric. Often the trainers observed that the disposition title, such as Self-Awareness, triggered an automatic response of high rater confidence in their ability to rate self-awareness without carefully reading the rubric descriptions. The tendency of raters to believe they are "right" rather than applying a rubric description is a potential barrier for any dispositional measure.

Implications of the Study

The implications of this study relate primarily to counselor education programs. As evident from the review of literature, careful admissions processes are critical to prevent or diminish the number of gatekeeping and remediation situations that occur in academic programs after admission. In addition to the importance of fair admissions procedures from a legal perspective, the effort required of applicants to engage in the application process justifies the importance of developing fair processes in which acceptance or denial decisions are not based solely upon the subjectivity of faculty.

For those academic programs utilizing admissions interviews, one important implication of the study is that the results suggest that without training, raters will have high variability in their ratings of admissions applicants, as illustrated by the variability of the pretest scores in this study. Structuring the rating of admissions interviews by using an assessment is one method of mitigating the variability of faculty ratings of applicants. A holistic (global) rubric such as the PDCA-RA is unlikely to ever garner the almost perfect interrater reliability associated with analytic rubrics, but the PDCA-RA is available as one practical, field-tested tool with promising reliability to help facilitate transparent and fair admissions interview rating processes.

Limitations and Future Research

In light of the lack of an established list of professional dispositions, the PDCA-RA's utility may be limited, as the selected dispositions may not align with the values of all counselor education programs. A second limiting factor is that the sample included both field site supervisors and faculty, and all participants were from the rural Western United States. The reliability of the tool is limited by the demographics of the sample. Another limitation was that the study's pretest and posttest video clips, although similar, were different from one another. The initial decision to use different pretest and posttest video clips was based on an attempt to reduce the influence of testing as a threat to internal validity. However, this also introduced the possibility that either of the sets of video clips was inherently easier or more difficult to rate than the other. Further research would include randomly juxtaposing pretest and posttest video clips, or perhaps using the same video clips pre- and posttest to eliminate the possibility that differences in pretest and posttest video clips were responsible for the improvements in score reliability rather than the intended independent variable, the training. Another potential limitation to the results is that it is possible that some of the graduate students who were filmed in the vignettes may have been known by six of the faculty members from one of the institutions. The impact of this possibility was reduced by the use of multiple student actors, but prior knowledge of the student could have influenced raters' scores.

A final issue for consideration is the decision to use site supervisors as raters for the research. Site supervisors more commonly utilize the PDCA-R rather than the PDCA-RA, the version specific to admissions screening. The PDCA-R is used by supervisors to monitor and to communicate with counselor educators and counseling program clinical personnel. Further, at least one of the counselor education programs utilizes site supervisors for the admissions process. The training protocol for both versions of the PDCA is the same, and with site supervisors routinely participating in the training, the decision was made to include site supervisors as raters. It is possible, however, that site supervisors may differ in their abilities to respond to the training protocol when compared to counselor educators, adjunct faculty, and doctoral students.

A possibility for future research is to measure the extent to which the improvement in reliability can be maintained over time. At this point, little is known about whether and how often educators and site supervisors would need training updates to function optimally as raters of student dispositions. Accordingly, rating reliability could be observed at intervals of 3 months, 6 months, or 1 year after training to monitor decay.

Future research is also needed to determine the extent to which the length of the training protocol influences interrater reliability. In addition, cultural and gender bias in the use of the PDCA-RA should be studied, as one criticism of rubrics is the potential for cultural bias.

As a tool for consistently rating counselor education program applicants, the PDCA-RA demonstrates potential, though more research needs to be conducted to increase the interrater reliability. Training improved the interrater reliability results but not to the extent that excellent interrater reliability was achieved. Adjusting the training protocol may be fruitful as a mechanism to improve interrater reliability.

Conclusion

There is a need for reliable admissions tools to assess dispositional behaviors of counseling program applicants. Interrater reliability is an important form of reliability in situations such as admissions interviews in which there are often multiple raters involved in the process. The importance of interrater reliability is founded in the critical premises of fairness and transparency to applicants, though legal protection of counselor education programs is also enhanced by using clear, standardized processes. Dispositional assessment is in its infancy, especially when applied to counselor education in general and to program admissions in particular. How exactly to define dispositions as well as how exactly the role of the counselor will serve as a means of selection and gatekeeping for the profession is yet to be determined. Yet counselor educators perceive both an ethical and professional responsibility for monitoring counseling student dispositions as a means for safeguarding the integrity of the profession (Freeman et al., 2019; Schuermann et al., 2018). The continued development of the PDCA-R and the PDCA-RA, as well as the associated training materials, represents initial steps toward standardizing and improving dispositional appraisal. The video-based training and the exploration of the training as a means of improving rater consistency will potentially increase the ability of counselor educators to consistently assess and monitor developing counseling students. Consistent dispositional ratings can also contribute to the development of a common language for discussing student progress. The current research represents a promising effort to continually improve the dispositions assessment process for counselor educators, counseling programs, and the counseling profession.

References

Alexander, C. R., & Praeger, S. (2009, June). *Smoke gets in your eyes: Using rubrics as a tool for building justice into assessment practices*. Paper presented at the Annual Conference of the Australian Teacher Education Association (ATEA). Australian Teacher Education Association. http://files.eric.ed.gov/fulltext/ED524 704.pdf

American Counseling Association. (2014). *ACA code of ethics*.

Bryant, J. K., Druyos, M., & Strabavy, D. (2013). Gatekeeping in counselor education programs: An examination of current trends. In *Ideas and research you can use: VISTAS 2013*. American Counseling Association. https://www.counseling.org/docs/default-source/vistas/gatekeeping-in-counselor-education-programs.pdf?sfvrsn=7f6e77b5_13

Carkhuff, R. R. (1969). Critical variables in effective counselor training. *Journal of Counseling Psychology*, *16*(3), 238–245. https://doi.org/10.1037/h0027223

Cole, B. S. (1991). Legal issues related to social work program admissions. *Journal of Social Work Education*, *27*(1), 18–24. https://doi.org/10.1080/10437797.1991.10672165

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO PI-R professional manual*. Psychological Assessment Resources, Inc.

Council for Accreditation of Counseling and Related Educational Programs. (2009). *CACREP 2009 standards*. http://www.cacrep.org/wp-content/uploads/2017/07/2009-Standards.pdf

Council for Accreditation of Counseling and Related Educational Programs. (2015). *CACREP 2016 standards*. http://www.cacrep.org/forprograms/2016-cacrep-standards

Crawford, M., & Gilroy, P. (2013). Professional impairment and gatekeeping: A survey of master's level training programs. *The Journal of Counselor Preparation and Supervision*, *5*(1). https://doi.org/10.7729/51.0030

Demos, G. D., & Zuwaylif, F. H. (1966). Characteristics of effective counselors. *Counselor Education and Supervision*, *5*(3), 163–165. https://doi.org/10.1002/j.1556-6978.1966.tb02062.x

Dugger, S. M., & Francis, P. C. (2014). Surviving a lawsuit against a counseling program: Lessons learned from *Ward v. Wilbanks*. *Journal of Counseling & Development*, *92*(2), 135–141. https://doi.org/10.1002/j.1556-6676.2014.00139.x

Engram, B. E., & Vandergoot, D. (1978). Correlation between the Truax and Carkhuff scales for measurement of empathy. *Journal of Counseling Psychology*, *25*(4), 349–351. https://doi.org/10.1037/0022-0167.25.4.349

Epstein, M. H., & Synhorst, L. (2008). Preschool behavioral and emotional rating scale (PreBERS): Test–retest reliability and inter-rater reliability. *Journal of Child and Family Studies*, *17*(6), 853–862. https://doi.org/10.1007/s10826-008-9194-1

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378–382. https://doi.org/10.1037/h0031619

Flynn, S. V., & Hays, D. G. (2015). The development and validation of the Comprehensive Counseling Skills Rubric. *Counseling Outcome Research and Evaluation*, *6*(2), 87–99. https://doi.org/10.1177/2150137815592216

Frame, M. W., & Stevens-Smith, P. (1995). Out of harm's way: Enhancing monitoring and dismissal processes in counselor education programs. *Counselor Education and Supervision*, *35*(2), 118–129. https://doi.org/10.1002/j.1556-6978.1995.tb00216.x

Freeman, B. J., & Garner, C. M. (2017). Professional Dispositions Competency Assessment, Revised. Unpublished instrument, ScholarWorks.

Freeman, B. J., Garner, C. M., Fairgrieve, L. A., & Pitts, M. E. (2016). Gatekeeping in the field: Strategies and practices. *Journal of Professional Counseling: Practice, Theory & Research*, *43*(2), 28–41. https://doi.org/10.1080/15566382.2016.12033954

Freeman, B. J., Garner, C. M., Scherer, R., & Trachok, K. (2019). Discovering expert perspectives on dispositions and remediation: A qualitative study. *Counselor Education and Supervision*, *58*(3), 209–224. https://doi.org/10.1002/ceas.12151

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).* In *Ideas and research you can use: VISTAS 2016*. American Counseling Association. https://www.counseling.org/knowledge-center/vistas/by-year2/vistas-2016/docs/default-source/vistas/article_5235f227f16116603abcacff0000bee5e7

Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters* (4th ed.). Advanced Analytics.

Halinski, K. H. (2010). *Predicting beginning master's level counselor effectiveness from personal characteristics and admissions data: An exploratory study* [Doctoral dissertation, University of North Texas]. https://digital.library.unt.edu/ark:/67531/metadc11038

Herman, J. L., Aschbacher, P. R., & Winters, L. (1992). *A practical guide to alternative assessment.* Association for Supervision and Curriculum Development.

Hernández, T. J., Seem, S. R., & Shakoor, M. A. (2010). Counselor education admissions: A selection process that highlights candidate self-awareness and personal characteristics. *Journal of Counselor Preparation and Supervision*, *2*(1), 74–87. https://doi.org/10.7729/21.2010

Hutchens, N., Block, J., & Young, M. (2013). Counselor educators' gatekeeping responsibilities and students' first amendment rights. *Counselor Education and Supervision*, *52*(2), 82–95. https://doi.org/10.1002/j.1556-6978.2013.00030.x

Johnson, W. B., & Campbell, C. D. (2002). Character and fitness requirements for professional psychologists: Are there any? *Professional Psychology: Research and Practice*, *33*(1), 46–53. https://doi.org/10.1037/0735-7028.33.1.46

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, *2*(2), 130–144. https://doi.org/10.1016/j.edurev.2007.05.002

Kerl, S., & Eichler, M. (2005). The loss of innocence: Emotional costs to serving as gatekeepers to the counseling profession. *Journal of Creativity in Mental Health, 1*(3–4), 71–88. https://doi.org/10.1300/J456v01n03_05

Kerl, S. B., Garcia, J. L., McCullough, C. S., & Maxwell, M. E. (2002). Systematic evaluation of professional performance: Legally supported procedure and process. *Counselor Education and Supervision*, 41(4), 321–334. https://doi.org/10.1002/j.1556-6978.2002.tb01294.x

Koerin, B., & Miller, J. (1995). Gatekeeping policies: Terminating students for nonacademic reasons. *Journal of Social Work Education*, *31*(2), 247–260. https://doi.org/10.1080/10437797.1995.10672261

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/10.1016/j.jcm.2016.02.012

Kopenhaver Haidet, K., Tate, J., Divirgilio Thomas, D., Kolanowski, A., & Happ, M. B. (2009). Methods to improve reliability of video-recorded behavioral data. *Research in Nursing & Health*, *32*(4), 465–474. http://doi.org/10.1002/nur.20334

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lumadue, C. A., & Duffey, T. H. (1999). The role of graduate programs as gatekeepers: A model for evaluating student counselor competence. *Counselor Education and Supervision*, *39*(2), 101–109. https://doi.org/10.1002/j.1556-6978.1999.tb01221.x

McAdams, C. R., III, Foster, V. A., & Ward, T. J. (2007). Remediation and dismissal policies in counselor education: Lessons learned from a challenge in federal court. *Counselor Education and Supervision*, *46*(3), 212–229. https://doi.org/10.1002/j.1556-6978.2007.tb00026.x

McCaughan, A. M., & Hill, N. R. (2015). The gatekeeping imperative in counselor education admission protocols: The criticality of personal qualities. *International Journal for the Advancement of Counseling, 37*, 28–40. https://doi.org/10.1007/s10447-014-9223-2

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

Miller, S. (2006). *Professionalisation, ethics and integrity systems: The promotion of professional ethical standards, and the protection of clients and consumers*. A report for the Professional Standards Councils, Centre for Applied Philosophy and Public Ethics, Australia.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, *9*, 129–144. https://doi.org/10.1016/j.edurev.2013.01.002

Pope, V. T. (1996). *Stable personality characteristics of effective counselors: The Counselor Characteristic Inventory* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global (Order No. 9625345).

Randolph, J. J. (2005). Free-marginal multirater kappa (multirater κ_{free}): An alternative to Fleiss' fixed-marginal multirater kappa. *Department of Computer Science*, 1, 1–20. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.59.8776&rep=rep1&type=pdf

Redekop, F., & Wlazelek, B. (2012). *Counselor dispositions: An added dimension for admission decisions*. In *Ideas and research you can use: VISTAS 2012.* American Counseling Association. https://www.counseling.org/knowledge-center/vistas/by-year2/vistas-2012/docs/default-source/vistas/vistas_2012_article_17

Rosen, J., Mulsant, B. H., Marino, P., Groening, C., Young, R. C., & Fox, D. (2008). Web-based training and interrater reliability testing for scoring the Hamilton Depression Rating Scale. *Psychiatry Research*, *161*(1), 126–130. https://doi.org/10.1016/j.psychres.2008.03.001

Schuermann, H., Avent Harris, J. R., & Lloyd-Hazlett, J. (2018). Academic role and perceptions of gatekeeping in counselor education. *Counselor Education and Supervision*, *57*(1), 51–65. https://doi.org/10.1002/ceas.12093

Smaby, M. H., Maddux, C. D., Richmond, A. S., Lepkowski, W. J., & Packman, J. (2005). Academic admission requirements as predictors of counseling knowledge, personal development, and counseling skills. *Counselor Education and Supervision*, *45*(1), 43–57. https://doi.org/10.1002/j.1556-6978.2005.tb00129.x

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/10.1002/j.2161-007X.2012.00011.x

Swank, J. M., Lambie, G. W., & Witta, E. L. (2012). An exploratory investigation of the counseling competencies scale: A measure of counseling skills, dispositions, and behaviors. *Counselor Education and Supervision*, *51*(3), 189–206. https://doi.org/10.1002/j.1556-6978.2012.00014.x

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/10.1002/j.1556-6978.2014.00048.x

Taub, D. J., Servaty-Seib, H. L., Wachter Morris, C. A., Prieto-Welch, S. L., & Werden, D. (2011). Developing skills in providing outreach programs: Construction and use of the POSE (Performance of Outreach Skills Evaluation) rubric. *Counseling Outcome Research and Evaluation*, *2*(1), 59–72. https://doi.org/10.1177/2150137811401019

Truax, C. B., & Carkhuff, R. (1967). *Toward effective counseling and psychotherapy: Training and practice*. Aldine.

U.S. Department of Justice, Civil Rights Division. (2010). Americans with Disabilities Act Title III Regulations: Part 36 Nondiscrimination on the Basis of Disability in Public Accommodations and Commercial Facilities (CRT Docket No. 106). https://www.ada.gov/regs2010/titleIII_2010/titleIII_2010_regulations.htm

Utley Buensuceso, J. M. (2008). *The Sixteen Personality Factor Questionnaire and ratings of counselor effectiveness* (Order No. 3341140) [Doctoral dissertation, Azusa Pacific University]. ProQuest Dissertations and Theses Global.

Ward v. Wilbanks. (2010). No. 09-CV-112 37, 2010 U.S. Dist. WL 3026428 (E.D. Michigan, July 26, 2010).

Williams, J. L., Williams, D. D., Kautzman-East, M., Stanley, A. L., Evans, W. J., & Miller, K. L. (2014). *Assessing student dispositions in counselor training programs: Implications for supervision, program policy, and legal risk management* [PowerPoint slides]. DocPlayer. https://docplayer.net/2862339-Assessing-student-dispositions-in-counselor-training-programs-implications-for-supervision-program-policy-and-legal-risk-management.html

nbcc