The Utility and Development of the Competencies of Professional Psychology Rating Scales (COPPR)

Kylie Rice[1]*, Nicola S. Schutte[1], Suzanne M. Cosh[1], Adam J. Rock[1], Stephanie E. Banner[1] and Jade Sheen[2]

[1] School of Psychology, University of New England, Armidale, NSW, Australia, [2] School of Psychology, Deakin University, Burwood, VIC, Australia

Competency-based professional psychology training is now common practice in many countries. An implication of competency-based training is the need to assess practitioner competence across multiple domains of practice; however, standardized measures of competence are limited. In Australia, currently there is no standardized, quantitative measure of professional competencies at registered psychologist level. The absence of a measure has implications for education, training, practice, and research in professional psychology. To address this gap, this article provides a conceptual overview of the utility and development of the *Competencies of Professional Psychology Rating* scales (COPPR), including the process of initial pre-test, pilot, and review. This developmental process resulted in the thematic identification of competencies within 11 domains of practice, and the creation of both COPPR-Self report and COPPR-Observer report versions. The pre-test provided content validity for the COPPR, and the initial results of the pilot test suggest strong convergent and divergent validity. The measure differentiated between novice and experienced practitioners, suggesting the scale is appropriate for use across career stages. The COPPR scales address the need for a standardized and quantitative measure across multiple domains of practice at registered psychologist level in Australia. The COPPR scales are intended to have utility across professional psychology student and supervisee performance evaluation, self-reflection for psychologists in practice, educational evaluation at professional psychology level, and various research contexts.

Keywords: Competencies of Professional Psychology Rating scales (COPPR), competency, measure, psychologist, OSCE

INTRODUCTION

The Utility and Development of the Competencies of Professional Psychology Rating Scales

Professional psychology training has moved to a competency-based framework in many countries (Rubin et al., 2007; Rodolfa et al., 2014). Competency-based training "is designed to ensure that the learner attains a predetermined and clearly articulated level of competence in a given domain or professional activity" (Kaslow, 2004, p. 777). The development of the Competency Cube

(Rodolfa et al., 2005) and the *Competency Benchmarks* (Fouad et al., 2009) provided a platform for understanding and assessing competence within professional psychology in the United States. This focus is replicated internationally, although there is no consensus on the structure of professional psychology competencies, with varying models, clusters, and frameworks available (Rodolfa et al., 2014). Thus, despite the 'culture of competence' being widely adopted, there are many issues that are unresolved in professional training, in particular the assessment of competence (beyond previously used measures such as hours accumulated) across domains of practice (Rodolfa and Schaffer, 2019).

In Australia, the *Accreditation Standards for Psychology Programs* (henceforth called the "Standards") (Australian Psychology and Accreditation Council [APAC], 2019) introduced considerable change across all levels of psychology training, and in particular, across the professional training streams. In these Standards, *Professional competencies* (Level 3) provide the expected standards at the professional level for general registration (Australian Psychology and Accreditation Council [APAC], 2019). These core competencies are, subsequently, built upon in *Professional competencies for specialized areas of practice* (Level 4), which stipulates the specializations or endorsement areas, such as clinical psychology, which require further post-graduate training and practice.

Collectively, the Standards now apply a competency-based model of professional training. The Standards "are designed to ensure students acquire the knowledge, skills and attributes required to practice psychology competently and safely" (Australian Psychology and Accreditation Council [APAC], 2019, p. 5). This focus is consistent with a significant shift in the profession, away from a tasks-achieved approach (based on the completion of set tasks and placement hours) toward the development and achievement of competence across multiple domains of practice (e.g., assessment, intervention) (Gonsalvez et al., 2016). In order to train students, identify and address performance difficulties, strengths and weaknesses, and to be accountable as a profession, clear competencies need to be established (Hatcher et al., 2013). An implication, and challenge, of the competency-based model is the measurement and assessment of competence (Rodolfa and Schaffer, 2019). This challenge is exacerbated by the lack of standardized measures of professional psychology competencies. Currently, there is no known multi-dimensional measure of professional psychology competence designed for an Australian context at registered psychologist level, which has implications for training, practice, and research. This paper addresses this gap, by providing an overview of the development of an emerging, multi-dimensional measure of psychologist competencies, in accordance with the APAC Standards (Level 3).

Development of Competence

The seminal article by Dreyfus and Dreyfus (1980) yielded a five stage model of competence development; novice, advanced beginner, competent, proficient and expert. While these stages were derived philosophically and established within a computer science area, they have been adapted to the acquisition of expertise across a multitude of vocational areas (Sharpless and Barber, 2009). While there has been debate about whether the stages can account for the complex nature of implicit and explicit clinical skills (Peña, 2010), the model has been applied usefully within many healthcare disciplines (Sharpless and Barber, 2009), and provides a framework for learning needs at each stage (Benner, 2004). The stages have been specifically applied in competency development within professional psychology practicums, and offer a trajectory for the acquisition of expertise (see Hatcher and Lassiter, 2007). This trajectory provides a progression of skill acquisition over time, and as such a practitioner is not necessarily at the same stage for each skill domain at each timepoint (Benner, 2004). Some skill areas are likely to be foundational and develop earlier in the trajectory (e.g., ethics and basic interviewing), while a higher level of mastery and professional experience is required for other domains (e.g., complex interventions) (Hatcher and Lassiter, 2007; Nash and Larkin, 2012; Gonsalvez et al., 2015). Furthermore, despite the model providing categorical descriptors, the development of competence is recognized to be a dimensional construct within the benchmarks of these stages (Sharpless and Barber, 2009). In addition, competency development is not necessarily stepwise (Deane et al., 2017); it is dynamic and fluid (Nash and Larkin, 2012), allowing for progression or regression regarding any specific skill over time.

Measurement of Competence

In general terms, competence refers to "*professional skills across numerous domains. Competency, on the other hand, is used to refer to the particular skills that sit within these domains*" (Stevens et al., 2017, p. 175). Previously there has been no consensus in Australia over the required competencies of psychologists (Lichtenberg et al., 2007); however, the Standards (Australian Psychology and Accreditation Council [APAC], 2019) have now usefully identified the core skills at each level. The identification of these skills leads to the methods of assessing competence.

There are many different assessment tools in psychology that can be applied to evaluate skills and knowledge, such as essays, exams and supervisor reports (Lichtenberg et al., 2007). However, many of the traditional tools lack ecological validity, fidelity to practice, generalizability and inter-rater reliability (Lichtenberg et al., 2007). Furthermore, these tools are typically site and subject specific. More recently, the field has adopted Objective Structured Clinical Examinations (OSCEs) as a form of competency-based assessment (Sheen et al., 2015; Yap et al., 2021) focusing on the demonstration of clinical skills. OSCEs have been used in medicine for many years, preparing students for practice, as formative learning, to rate performance on clinical skills, and to provide feedback to students (Kelly et al., 2016), and are now used in the assessment of competence in other disciplines. As Harden (2016) states, "There are 'good' OSCEs and 'not so good' OSCEs. Reliability and validity are related to how the OSCE is implemented" (p. 379). Assessors must consider content validity, face validity and reliability for each OSCE task (Ward and Barratt, 2005), and these usually differ between institutions and tasks. Brannick et al. (2011) found that overall OSCEs were not particularly reliable in medical settings, and reported

a large variability between raters, especially for communication skills. The selection of standardized scoring rubrics is essential to the objectivity and structure of OSCEs (Khan et al., 2013). As such, competency-based assessments, including OSCEs, require standardized approaches to marking skills demonstrations, in order to increase reliability and consistency. Thus, it is vital that appropriate rating scales are developed to enable clinical skills demonstrations to be objectively scored, to increase consistency between raters, tasks, and institutions. The creation of quantitative competency measures at the registered psychologist level could assist with the standardization of ratings for competency-based tasks, including OSCEs.

Competence in Professional Training

A quantitative measure of competence could have widespread utility in professional training. For example, a measure of placement performance at Level 4 APAC Standards, called the Clinical Psychology Practicum Competencies (CΨPRS), is widely used in evaluating performance on clinical psychology placements across Australia (Gonsalvez et al., 2015, 2016; Deane et al., 2017; Stevens et al., 2017). This measure consists of 60 items within 10 domains, which are rated on a four-point scale of Beginner to Competent (Hitzeman et al., 2020). However, this measure is restricted to the evaluation of Clinical Psychology practicums, specified as Level 4 in the APAC Accreditation Standards (2019), rather than the Professional Psychology competencies identified in Level 3. A similar measure that can standardize the evaluation of performance on professional psychology placements within Level 3 APAC Accredited programs is needed. This measure could also be used within the professional training at registered Psychologist level, such as evaluating OSCE performance, and for educational institutions to undertake unit outcomes, performance comparisons, and program evaluations.

Concerns about leniency in supervisor ratings has been noted in the clinical psychology specialization (Gonsalvez and Freestone, 2007). Leniency has been observed with early career supervisees in clinical psychology, with supervisors often rating students in placements as 'competent' (Gonsalvez et al., 2015), which is the endpoint of the CΨPRS scale. Ratings of 'competent' early in training may produce a ceiling effect that leaves minimal room for growth throughout the career (Gonsalvez et al., 2015), and may produce early career clinicians who have an inflated belief in their skills (Gonsalvez and Freestone, 2007). These issues highlight the need to assess practitioners within their developmental stage for each skill area, as per Dreyfus and Dreyfus (1980) model, which goes beyond 'competent.' Nonetheless, despite these limitations, quantitative measures help to standardize the assessment of competence for supervisees on placements in clinical psychology and other disciplines. Supervisors rate the 'lack of objective measures for competence and incompetence' as a key factor contributing to bias (Gonsalvez et al., 2015, p. 26). Furthermore, such leniency effects may be mitigated by increasing the rating scale response options, providing greater discrimination between response ratings, and clearly indicating the performance required to achieve that rating. Increasing the scale response options is consistent with the

Dreyfus and Dreyfus (1980) stages of competence development model, which recognizes expertise beyond 'competent,' in options including 'proficient' and 'expert.' Including such scale points above the rating of 'competent' could facilitate an ongoing assessment of performance post-training.

Competence in Professional Practice

Assessment of competence is not limited to student performance and educational institutions. A central responsibility and professional accountability for all healthcare practitioners is the self-regulation of professional development (Sheridan, 2021). The Code of Ethics mandates that psychologists only practice within their areas of competence (Australian Psychological Society, 2007), an ethical obligation that requires practitioners to be acutely aware of their strengths and competencies, as well as areas outside their competence. "Without reviewing and assessing their own practice, therapists risk becoming increasingly incompetent without being aware of it" (Loades and Myles, 2016, p. 3). Competence is considered to be dynamic rather than static, and an ongoing obligation even after professional registration is achieved (Nelson, 2007; Rodolfa et al., 2013). Furthermore, investigating competence enables motivated practitioners to attain and maintain high standards, and be effective clinicians who are able to facilitate positive change for clients (Sharpless and Barber, 2009). In order to continue skill development, practitioners need to implement a regime of deliberate practice, which includes feedback and repetition (Ericsson et al., 1993). This deliberate practice needs to be extended and lifelong in order to become expert performers (Ericsson and Charness, 1994), and to maintain a high level of domain-specific performance (Krampe and Ericsson, 1996). Without feedback, optimal learning will not occur, and repetition will not necessarily result in skill improvement (Ericsson et al., 1993). Deliberate practice is a focused approach of intentional repetition with performance monitoring and review, toward a defined goal (Duvivier et al., 2011). A crucial strategy for the feedback in self-directed learning and performance development is self-assessment (Sheridan, 2021).

Practitioners' ability to self-assess is essential in order to improve client outcomes, maintain competence, and identify the need for further training (Loades and Myles, 2016), and measures of competence can assist this self-assessment (Hatcher et al., 2013). These assessments need to enable practitioners to engage in continual learning and development at all stages of their career (Roberts et al., 2005), not just while training. As Rubin et al. (2007) stated in their history of the competence movement in psychology, "the profession must increase the focus on self-assessment, self-monitoring, reflection, and self-awareness, which not only reflect ethical behavior but are critical to the assessment of competence" (p. 459). These authors also identified that the "challenge of the next decade is for the profession to devise, implement, and evaluate the effectiveness of more comprehensive, developmentally informed competency assessments throughout the professional life span" (Rubin et al., 2007, p. 460). More than a decade on, this challenge remains. Currently, in Australia practitioners rely on self-reflection to determine the level of their own skills across practice domains.

Self-reflection is referred to in the Standards (Australian Psychology and Accreditation Council [APAC], 2019) and is an essential practice skill throughout the career (Roberts et al., 2005). However, the sole reliance on unstructured self-reflection is problematic as the effectiveness of self-reflection is variable and subjective, and may not cover the full scope of psychological competencies. Nonetheless, self-reflection can inform accurate self-assessment, which is a crucial skill in developing and maintaining competence (Loades and Myles, 2016). One key strategy to improve the accuracy of self-assessment is utilizing an assessment tool, which can be utilized to guide learning and professional development (Sheridan, 2021). As such, a multi-dimensional measure could enable psychologists to undertake systematic evaluations of their competence across domains, and track the development of these competencies over time and with development opportunities (such as training courses).

The Present Research

In response to the above identified need, the first author KR, in consultation with NS and JS, developed a rating scale to measure performance on professional psychology competencies, called the *Competencies of Professional Psychology Rating* scales (COPPR). This measure provides the first known quantitative measure of professional competencies in Australia based on the Accreditation Standards (Level 3, Australian Psychology and Accreditation Council [APAC], 2019 Standards). The COPPR includes a self-report version for students and clinicians to rate their perceived competence across the domains (COPPR-Self Report), and an observer-rated version for supervisors to rate the performance demonstrated by their supervisees (COPPR-Observer). These scales are intended to have utility across the following contexts: (1) professional psychology student placement evaluation, (2) supervisee development on internships, (3) psychologists in practice for self-reflection and professional development, (4) educational evaluation of professional psychology training courses, placements, and units, (5) OSCE evaluation, and (6) various research contexts and outputs. This paper outlines the development of the COPPR scales, and the initial process of pre-test, refinement, and initial pilot testing.

MATERIALS AND METHODS

Scale Development

The measure was developed through a systematic series of scale development procedures (informed by DeVellis, 2003; Morgado et al., 2017). As outlined below, this procedure included definition of the conceptual domain (Stage 1), followed by generating an item pool (Stage 2), assessing content validity (Stage 3), determining the format and structure of the measure (Stage 4), and then pre-testing the measure (Stage 5). Stages 1–4 were informed by a comprehensive literature review of competence assessment, as well as stakeholder review. Stakeholders included students, academics, clinicians, and clinical supervisors, providing representation across application contexts, given that the COPPR scales are intended to be used across education, placement, clinical practice, and professional practice settings. Stakeholder consultation was an iterative process. Initial item wording and subsequent refinements of the measure were made by the first author (KR), based on the APAC Standards, stakeholder feedback and consensus between the authors KR, NS, and JS. All refinements were reviewed by stakeholders at each stage in the scale development process to ensure the technical precision, content validity, and suitability of the items and measure. Given the intended breadth of the COPPR scales and its intended users, this co-design process was favored during development. The following sections detail this process.

Stage One: Concept Definition

The construct of competence was operationalized based on the Australian Psychology and Accreditation Council [APAC] (2019) Level 3 standard's definitions and range of competencies. To create and refine the overall conceptual framework for the measure, the first author KR, in consultation with NS and JS, evaluated the Level 3 APAC Standards for core components. In this process, each item in the Standards was reduced into separate components, which were then grouped into the 11 domains of Scientist-Practitioner, Cultural Responsiveness, Working across the Lifespan, Professional Communication and Liaison Skills, Clinical Interviewing, Counseling Micro-skills, Formulation and Diagnosis, Assessment, Intervention, Ethics, and Self-Reflective Practice. In this process, two Level 3 Standards were removed from the scales as they were not deemed to be relevant for the measure. The first, "Demonstrate successful (prior or concurrent) achievement of pre-professional competencies" (3.1) is a pre-requisite for entry into a professional training program, and deemed unnecessary for the practice-based measure. The final Standard "Investigate a substantive individual research question relevant to the discipline of psychology" (3.17) was not included as it is not related to professional practice. Removal of these standards was supported by the stakeholders as neither of these Standards are relevant to competencies in practice.

Stage Two: Item Pool Generation

The item pool was generated by a combined inductive and deductive process (Morgado et al., 2017). That is, items were initially generated based on the components of the APAC Standards. The Standards were converted to create competency-based statements for the item pool with the wording of items informed deductively, by the literature review, as well as inductively, from information regarding the construct and its measurement provided by stakeholders. The individual items were developed and modified multiple times by the authors (KR in consultation with NS and JS). Meetings were held to reach agreement on items, until the items were representative of the core competencies covered in the Level 3 Standards. For example, Standard 3.12 "*Operate within the boundaries of their professional competence, consult with peers or other relevant sources where appropriate, and refer on to relevant other practitioners where appropriate*" (Australian Psychology and Accreditation Council [APAC], 2019, p. 14) was separated into three components of acting within areas of competence, consulting others, and referring. These components were,

subsequently, expanded into three separate competency-based items within the 'Ethics' domain: 'Only practices within areas of professional competence,' 'Appropriately seek consultation with supervisor or peers as needed,' and 'Appropriately refers to other clinicians, professionals or services as needed.' Based on this Standard (3.12), an item 'Consult peers, supervisor or others as needed' was added to the *Professional communication and liaison skills* domain.

Each domain was, subsequently, analyzed for competency completeness (i.e., to assess if there was a skill gap missing from the domain) and to ensure relevant Standards were thoroughly captured. Based on this review, several items were added to the domains to address competency gaps. For example, 'Comply with legal and ethical requirements in practice' was added to the *Ethics* domain. The wording of each item was reviewed and modified by the authors and stakeholders to ensure clarity and technical quality, and the total item pool consisted of 81 items.

Stage Three: Content Validity

After the 81 individual items in the item pool were generated, stakeholders were asked to review the item pool to assess each item in terms of its relevance to the construct of psychologist competence, as well as to ensure that all aspects of competency were being assessed by the items in the pool. While various criteria for determining acceptability of items have been described in the literature (e.g., 80% agreement or universal agreement; see Polit and Beck, 2006; Zamanzadeh et al., 2015), universal agreement is deemed preferable where the sample is five or less (Lynn, 1986). Given that the stakeholders represented various user groups (i.e., students, academics, clinicians, and clinical supervisors), and each group consisted of fewer than five, the more conservative approach of universal agreement was favored. After this consultation with the stakeholders, all items were deemed to have content validity, and thus retained, with the COPPR scales also then deemed to have content validity as all items were rated as relevant.

Stage Four: Measure Structure

The 11 domains of practice were established during the construct definition process. It was intended that the domains could represent distinct subscales, which could be used together as a global measure of overall competence, or separately to evaluate specific tasks and skill areas with higher scores indicating greater competence. The structure was determined by the authors KR, NS, and JS in line with the APAC standards, with consensus reached on the overall structure from the stakeholders.

Response format was initially considered based on the literature review, and then in consultation with the stakeholders. The literature review indicated that prior competence measures have used a four-point scale to indicate stage of competence development (e.g., CΨPRS; Hitzeman et al., 2020) along with open ended comment sections and overall ratings of placement progress. The four-stage response format was designed for assessing competence during placement training with stage four indicative of competence at the level of a generally registered psychologist. However, this response format was deemed limited by the stakeholders for the assessment of competence across the spectrum of training and experience. That is, the COPPR is aimed at assessing competence during training, but also throughout professional practice, thus, a more robust response format with additional categories was deemed necessary to reflect development of competence over the duration of careers.

While five-point Likert scales have often been recommended for self-report measures (Revilla et al., 2014), this format was also deemed by the stakeholders to be too limited for reporting competency and a larger range was considered necessary. A larger range of response options is also suggested in other research (e.g., Gonsalvez and Freestone, 2007), and may help to mitigate supervisor leniency in ratings. Further, more points between dimension poles may result in more accurate metrics and a better psychometric assessment of a construct (Wu and Leung, 2017). Thus, a seven-point format was selected, as previous research suggested that a five-point and six-point scale was insufficient to reduce supervisor leniency (Gonsalvez and Freestone, 2007; Gonsalvez et al., 2013). Seven-point Likert scales have been found to be accurate and easy to use while providing a more sensitive and nuanced assessment (Finstad, 2010; Debets et al., 2020). A score of five was selected to indicate 'competence' as per a registered psychologist, and to enable multiple levels below a rating of 'competent' for students and those in the early stages of skill development.

Some differences in perspective across the stakeholder group were noted regarding the proposed response scale format. Two versions of the response scales were, therefore, developed for review during the pre-test. The first version measured items across a seven-point Likert scale, with an anchor at each end from "not yet competent" to "expert."

The second version of the rating scale included descriptors at each of the seven points, and behavioral indicators for each level of competence. The anchor points were considered to provide guidance without being too prescriptive or limiting of responses, while other stakeholders indicated a preference for clarity and behavioral descriptors across the rating scales, with a view that this might produce greater consistency across users of the measure and allow for better comparison. This scale was consistent with the five stages of skill acquisition framework (Dreyfus and Dreyfus, 1986), that have been used in competence assessments in many disciplines including nursing (Benner, 2004) and psychology (Hatcher and Lassiter, 2007).

Stage Five: Pre-test

Following the initial stages of scale development, a pre-test was undertaken. Pre-testing or piloting scales with stakeholders is deemed critical as a final stage of scale development (Carpenter, 2018). The pre-test was conducted to assess the technical quality of the scale and obtain additional feedback on items and the scale format from experiential experts or end users, including psychologists and students (i.e., those who would be users of the measure; see Schilling et al., 2007). In outlining how to ascertain content validity, Lynn (1986) proposed a minimum of three people for expert panel review of content validity, but suggested that no more than 10 were required. There continues to be a lack of universal consensus on sample size for stakeholder panels; however, between five and 10 continue to be recommended

(Yaghmale, 2003; Zamanzadeh et al., 2015). A review of content validity assessment panels showed that, of the psychometric tools examined, expert panels had consisted of between 2 and 15 members (Polit and Beck, 2006), with expert panels of between five and 10 commonly used in assessing content validity of psychological measures (e.g., Halek et al., 2017; Ghazali et al., 2018; Usry et al., 2018).

Following ethics approval from the University's Human Research Ethics Committee, purposive sampling was used to recruit psychologists and students through professional networks, with an aim to recruit a mixed sample representing diverse professional backgrounds and varied career stages. Thirteen comprehensive responses were obtained (with three participants providing feedback by email), with representation across several employment sectors (e.g., Health, Academia, and Private Practice). Participants' years of registration ranged from 1 year, through to 22 years of psychology registration, with one respondent who declined to answer (M = 13.3 years, SD = 7.4 years). The varied ranges of experience and professional backgrounds allowed for experiential experts across the target audience of the measure (i.e., students, supervisors, psychologists of varying experience levels), which has been argued to provide a more thorough assessment (Schilling et al., 2007).

Respondents provided informed consent to participate, and were provided with a link to the *Level 3 Australian Psychology Accreditation Council* (*APAC*) 2019 *Standards for Psychology Programs* and then asked to review the COPPR measure in an online survey. The online survey presented each domain of the measure, with open ended questions to obtain feedback from respondents. Respondents were guided to comment on technical precision, such as item wording and appropriateness of item content. Given that each group of participants in the pre-test (e.g., academics, Health employees, students etc.) consisted of five or fewer participants, the more conservative approach of universal agreement was, again, favored (see Lynn, 1986), meaning that the authors reviewed all items which were commented upon in the pre-test, and these items were then refined accordingly with some additional stakeholder review.

In the pre-test, additional questions were asked with regard to measure format (e.g., Likert response format, inclusion of open-ended questions), including the clarity of the instructions provided in the scale. Respondents were also presented with both Likert-scale response formats for the items and asked to indicate their preferred response format. An open-ended question also allowed respondents to elaborate reasons for their preferences. Respondents were then asked to indicate a preference for including open ended questions (a) at the end of each domain section, (b) only at the end of the rating scale, or (c) not including open ended questions; and were also provided with an open-ended question to outline reasons for their response.

Results of the Pre-test
Initial Instructions
The overall instructions for completing the scale were considered to be clear during the pre-test. There was one suggestion to enhance the clarity of the instructions for completing the

items in the *Scientist-practitioner* domain; with these instructions modified for the final version of the scale (changed from "This section refers to your ability to work effectively as a scientist-practitioner. To what extent did you demonstrate each of the following" to "This section refers to your ability to work effectively as a scientist-practitioner. Please rate the extent to which you currently:").

Items and Domains
Respondents provided feedback on each domain, as well as the individual items within. No comments were made suggesting any changes or modifications to the domains themselves, thus these were retained unmodified. Respondents then provided feedback regarding each item within each of the domains. All suggestions made during the pre-test were reviewed by authors KR, NS, and JS, with refinements made accordingly. Across the domains, small grammatical or phrasing edits were suggested for several items, with these items modified as suggested to enhance readability and precision.

Response Format
During the pre-test, feedback was sought with regard to which of the two response formats was considered preferable. Feedback indicated that the majority of the stakeholders preferred the Likert scale with behavioral descriptors throughout, thus, they were retained as the final response format.

Some modifications were, however, made to these descriptors and response format based on respondent feedback. Feedback suggested that refining the behavioral indicators would increase consistency, inter-rater reliability and usability, and it was suggested that the behavioral anchors at the 'competent' end of the scale be further developed. Based on this feedback, these descriptors for each anchor were reviewed and modified to provide greater clarity on the requirement for achievement at this level, in an effort to increase inter-rater reliability. The level indicative of 'competent' performance was also moved from a rating of '5' on the scales provided in the pre-test, to a rating of '4' on the final scale. It is anticipated that the increased options will provide greater clarity and room for growth at the 'competent' to 'expert' end of the scale, which will be useful for tracking development of practitioners. These labels are consistent with Dreyfus and Dreyfus' (1986) stages of skill acquisition framework descriptors (novice, advanced beginner, competent, proficient, and expert), that have been applied to professional psychology competency development on practicums (Hatcher and Lassiter, 2007). The scale point of 'advanced' was also included to ensure there is sufficient discrimination possible in the ongoing assessment of competence for practitioners post-training. The scale point of 'not yet competent' was also included for use if competence is not yet achieved (Sharpless and Barber, 2009). This response format may help to reduce the ceiling effects observed in early training (e.g., Gonsalvez et al., 2015) that render it difficult to assess change later in training. See **Figure 1**.

During the pre-test, respondents were also asked to comment on whether the measure should contain sections for open-ended comments: (a) at the end of each domain section, (b) only at the end of the rating scale, or (c) should not contain comment.

Not yet Competent 1	Beginner 2	Advanced Beginner 3	Competent 4	Proficient 5	Advanced 6	Expert 7
Not yet demonstrated appropriately.	Demonstrated but beginning stage of development; Numerous inconsistencies and inaccuracies in performing this skill.	Demonstrated but early stage of development. Some inconsistencies and inaccuracies in performing this skill.	Demonstrated meeting standard for competence. No inconsistencies or inaccuracies in performing this skill.	Demonstrated above standard competence for this skill. More accomplished and practiced, demonstrated with ease.	Demonstrated competence as exceeding standard with fluency, accuracy and finesse.	Demonstrated proficiency as expected of a very highly experienced registered psychologist. Stand out in the field in performing this skill. Demonstrates outstanding knowledge, skill and delivery of this competence. Very few students achieve this rating.

FIGURE 1 | Modified response format. Please use the following scale to rate level of competence for each item. Competence is defined as the level expected of a registered psychologist (Rating = 4).

Six respondents indicated that comment sections should be included at the end of each domain, and three said that comment sections should only be used at the end of the rating scale (other respondents did not report a preference). Given that the domains are designed to be able to be used separately as individual subscales, upon review by authors KR, NS and JS it was determined that a comments sections at the end of each domain would be the most practical. It was determined to allow these sections to be optional, as they are not included in the quantitative scoring or the self-report version.

Measure Structure

Open-ended responses during the pre-test indicated that all respondents endorsed the categorization of items according to the domains. No respondent suggested changes to the structure of the measure.

Development of the Self-Report and Observer Report Versions

Following the pre-test and confirmation of the wording of the items and domains, authors KR, NS, and JS were able to create a second version of the COPPR for observer report (COPPR Observer; COPPR-O). This process involved rephrasing the instructions for each section in order to reflect observation rather than self-report. The actual items remained the same, only the instructions were altered, to reflect observations of the 'trainee' rather than the self. The self-report version of the scale was henceforth called the COPPR Self-report (COPPR-S) to clearly specify the version.

PILOT STUDY

The aim of the pilot study was to assess initial psychometric properties of the COPPR.

Method

Participants

Following University Ethics Committee approval, the study was made available online, for both postgraduate professional psychology students and practitioners holding any type of psychologist registration in Australia. The first 85 participants who provided complete responses on the COPPR-S were included in this pilot test. The majority of participants identified as female ($N = 69$), which is representative of the Australian workforce, where the majority of the Psychologists are female (79.7% in 2019; Australian Institute of Health and Welfare, 2022).

There was distribution of professional experience, with representation from provisional psychologists and post-graduate psychology students ($N = 27$), psychologists with general registration ($N = 29$), and psychologists with one or more practice endorsement ($N = 29$). The number of years practicing as a psychologist ranged from less than one year through to 40 years, with an average of 12.8 years of experience ($SD = 11.26$). Fifty-two percent of participants had 10 years or less experience, and the remaining 48% had more than 10 years experience.

Materials

Competencies of Professional Psychology Self-Report Scale

As outlined above, the COPPR-S is a newly created self-report scale of psychologist competencies based on the APAC Standards (Level 3; 2019). This scale consists of 81 items across 11 domains of practice, and practitioners rate their perceived competence for each item on a seven-point Likert scale.

Psychologist and Counselor Self-Efficacy Scale (Watt et al., 2019)

The PCES is a 31-item self-report scale measuring counselor and psychologist self-efficacy across a range of competencies. The scale was based on the commonalities of several different Australian competency frameworks for counselors and psychologists, and items assess self-efficacy regarding research, ethics, legal matters, assessment and measurement, and

intervention components. Instructions of "How confident are you that you have the ability to . . .", are provided, with a five-point scale from one (not at all) to five (extremely) (Watt et al., 2019). The PCES reliability in this sample was estimated using Cronbach's alpha (0.94).

Career Optimism

The Career Futures Inventory (CFI; Rottinghaus et al., 2005) contains three subscales that focus on dimensions of employability and career self-management. The original scale has been effectively shortened to a nine-item version (McIlveen et al., 2013). For the purposes of this pilot study, the Career Optimism and Career Adaptability subscales were utilized, as both are strengths-based constructs that focus on self-regulation of career management (Rottinghaus et al., 2005). Both Adaptability and Optimism have shown strong reliability in the short form version, with Cronbach's alphas of 0.82 and 0.84, respectively (McIlveen et al., 2013). Strong Cronbach's alphas were also computed in this sample (Adaptability = 0.81 and Optimism 0.94).

Procedure

Each participant provided informed consent, and completed the survey online via Qualtrics (Provo, UT, United States).

Analysis and Results

As presented in **Table 1**, the average item score for the COPPR-S was 4.90 (SD = 0.87, range 1 to 7), which is between the rating of 'Competent' and 'Proficient.' Given the registration type of participants (with 68% of respondents being fully registered or registered with a practice endorsement), this average item response is in the expected range, and clustering above the midpoint. These results demonstrate that participants were not just agreeing or disagreeing by responding at the endpoints of the scale (Clark and Watson, 1995). The mean of the PCES was similar, clustering above the scale midpoint of the five-point scale (M = 3.71, SD = 0.52, range 1–5), similar to the results reported in Watt et al. (2019).

Validity

To assess convergent validity, a Pearson's product-moment correlation was performed to evaluate the relationship between the COPPR-S total score and the PCES total score. The result was statistically positively significant, $r(83)$ = 0.73, p < 0.001. According to Cohen's (1988) conventions, a correlation coefficient of 0.5 or greater is considered 'large.' Therefore, this statistically significant and large correlation of

0.73 demonstrates strong convergent validity for the COPPR-S. To assess divergent validity, a Pearson's correlation was performed to evaluate the relationship between the COPPR-S total score and Optimism. The result was not statistically significant, $r(83)$ = −0.02, p = 0.89. The relationship between the COPPR-S total score and Adaptability was also not statistically significant, $r(83)$ = 0.02, p = 0.84. Taken, together these results provide initial evidence of divergent validity for the COPPR-S.

To assess the relationship between practitioners' years of experience and the COPPR-S total score, a Pearson's correlation was computed. The relationship was statistically positively significant, $r(83)$ = 0.46, p < 0.001. Based on Cohen's (1988) conventions, the strength of this relationship was moderate to strong.

A one-way between groups analysis of variance (ANOVA) was computed to determine if there was a statistically significant difference between registration type (i.e., provisional, general, and endorsement) regarding COPPR-S total score. The result was statistically significant, $F(2,82)$ = 14.59, p < 0.001, η^2 = 0.26. Tukey *post hoc* tests revealed that provisional psychologists or those who were undertaking post-graduate study had significantly lower COPPR-S scores (M = 347.70) than psychologists with general registration (M = 404.21, p = 0.003, 95% CI = [−95.68, −17.33]), or those with practice endorsements (M = 435.38, p < 0.001, 95% CI = [−126.85, −48.50]). There was no statistically significant difference on COPPR-S scores between those with general registration and those with practice endorsements (p = 0.14), which could be expected given the COPPR-S items are based on the *Professional Competencies* (Level 3; Australian Psychology and Accreditation Council [APAC], 2019), which are common competencies for psychologist registration with or without practice endorsements. The COPPR-S does not include *Professional competencies for specialized areas of practice* specified at Level 4 in the Standards (Australian Psychology and Accreditation Council [APAC], 2019). Thus, the COPPR-S scale demonstrated a very good ability to discriminate between provisional psychologists and those who were fully registered. An *a posteriori* statistical power analysis using G*Power (Faul et al., 2007) revealed that the ANOVA effect size (f = 3.11) was associated with sufficient power (0.99) (see Cohen, 1988).

Reliability

Reliability analysis for this sample was performed using Cronbach's alpha. The total COPPR-S score in this sample had a Cronbach's alpha of 0.99 (81 items), which indicates a very high level of reliability according to Nunnally's (1978) frequently cited cut-off scores, whereby an alpha above 0.95 is considered to be optimal in research. However, this level of reliability needs to be interpreted with caution as it can also be indicative of redundancy (Tavakol and Dennick, 2011), and reducing the number of items on the measure is generally suggested. The items on the COPPR-S are narrow and specific in focus, which is also related to a higher Cronbach's alpha (Panayides, 2013). However, the items on the COPPR-S are based on competencies specified in the APAC Standards (2019), and as such, items cannot be removed without potentially omitting important aspects of practice, and

TABLE 1 | Average item score for each psychologist registration type.

	Mean item score on the COPPR-S (*SD*) Response range 1–7	Mean item score on the PCES (*SD*) Response range: 1–5
Provisional psychologists/post graduate students (*N* = 27)	4.29 (1.01)	3.48 (0.62)
Psychologist with general registration (*N* = 29)	4.99 (0.56)	3.67 (0.39)
Psychologist with practice endorsement (*N* = 29)	5.38 (0.65)	3.96 (0.41)
Total sample	4.90 (0.87)	3.71 (0.52)

the narrow focus of the questions is required to specify the competencies of each practice domain. Furthermore, there is overlap and redundancy in the competencies specified in the APAC Standards, and given the COPPR-S must conform to these Standards, a level of redundancy is required as the COPPR-S's content validity is relative to the APAC Standards. The Cronbach's alpha for the PCES total score (0.94) was similar to that of the COPPR-S total score in this sample, further supporting that measures assessing these constructs have very high reliability due to the narrow focus of the items related to perceived competence. For example, responses between items (particularly within each domain) will likely be similar to each other due to the inter-relationship of competencies and underlying domains. However, this is not to assume that the items are unidimensional, rather, that that each item is measuring something similar to others (Taber, 2018). Thus, it is likely that high reliability, as measured by Cronbach's alpha, is a natural phenomenon for these types of scales, with necessary redundancy from competency clusters and the Standards.

Given these constraints and the sensitivity of Cronbach's alpha to be inflated with increased items on a scale, split-half reliability was calculated as an alternative estimate (Yang and Green, 2011). The items were split via IBM SPSS Statistics for Windows, Version 28 (IBM Corp, 2021) and yielded a Spearman-Brown coefficient of 0.97. When entered manually as odd and even questions, the Spearman-Brown coefficient increased to 0.99. According to conventions, a Spearman-Brown coefficient of between 0.70 and 0.80 is considered very good with higher coefficients being more desirable when the aim is to use test scores to make decisions regarding individuals (see Salkind, 2010).

DISCUSSION

This paper detailed the development of the Competencies of Professional Psychology Rating Scale (COPPR), to address the need for a measure of Professional Psychology competencies (Level 3, Australian Psychology and Accreditation Council [APAC], 2019). The COPPR includes a self-report version for students and clinicians (COPPR-Self Report), and an observer-rated version for supervisors (COPPR-Observer). The scale has utility in that it is based on the Australian Psychology and Accreditation Council [APAC] (2019) Professional Competencies, and represents the first known quantitative measure of these in Australia. The domains and items were created thematically from the APAC Standards (Level 3; 2019), and this developmental process resulted in the identification of competencies across 11 Domains: *Scientist-Practitioner*, *Cultural Responsiveness*, *Working across the Lifespan*, *Communication and Liaison*, *Clinical Interviewing*, *Counselling Micro-Skills*, *Formulation and Diagnosis*, *Assessment*, *Intervention*, *Ethics*, and *Self-Reflective Practice*. Each of the domains can be applied separately, for rating of discrete tasks, such as OSCE evaluation. The items were also mapped to the Standards, providing consistency between the content of the scale and the required graduate competencies at the registered psychologist level.

The pre-test provided content validity, and the pilot test demonstrated strong convergent and divergent validity for the

COPPR-S. The COPPR-S demonstrated the ability to differentiate between provisionally registered practitioners and both generally registered practitioners and endorsed practitioners. Furthermore, the COPPR-S was significantly positively related to years of experience. Taken together, these results provide initial evidence that the scale is appropriate for use across the lifespan of careers, and may be useful to assess perceived competence for student and provisional psychologists, through to those who are more experienced practitioners. These results are consistent with the stages of skill development (Dreyfus and Dreyfus, 1986) and also the development of professional expertise through deliberate practice throughout the career (Ericsson and Charness, 1994; Duvivier et al., 2011).

These scales extend the current methods of competency-based assessment at registered psychologist level by providing a multidimensional, quantitative measure to standardize assessment. The self-report and observer rated versions provide wide application for the scales in professional psychology in Australia, within both practice and training. In education and training, the scales may be useful for student placement evaluation (at 5th year), supervisee performance on internships and in practice, evaluation of professional psychology training courses or units, and standardized rubrics (e.g., OSCE evaluation). In practice, the scales may be useful for self-reflective practice, as a tool for clinicians to assess their competence across all domains of practice at registered psychologist level. As a self-assessment tool, the COPPR-S can facilitate a systematic, multidimensional self-evaluation of competence to establish professional development goals and training plans to address competence gaps. Practitioners and supervisors can use the COPPR scales to track competencies over time and with development opportunities (such as training courses). Both the self-report and observer-rated versions have various research applications, such as facilitating the understanding of competence development and the relationship between domains.

The response format, with the 7-point rating scale including behavioral anchors, may provide sufficient delineation between ratings of competence to allow supervisors and practitioners to discriminate the level of performance on each item. Despite this utility, Likert scales do have inherent limitations. Items measured using Likert scales seldom provide contexts, and respondents may experience difficulties interpreting words that comprise items (Ambrose et al., 2003). Further, Likert scales exhibit multidimensionality, where respondents must assess both the content of an item and their level of intensity regarding an item, increasing cognitive complexity (Hodge and Gillespie, 2003). Despite these limitations, it is worthy to note that Likert methodology is one of the most common in research across many disciplines (Carifio and Perla, 2008; Pescaroli et al., 2020), that the use of Likert scales has been consistently justified through sufficient strength and quality of data (Wigley, 2013), and that Likert ratings have demonstrated higher test reliability compared to other types of scales (Waples et al., 2010). In competency assessment specifically, Likert scales are user-friendly, economical, easy to administer, easy to score, and able

to assess a range of competencies (Gonsalvez et al., 2013, 2015). As such, while recognizing the limitations, Likert rating scales offer value in the assessment of competence. Likert scales that have a greater number of rating points are likely to be more advantageous in assessing the range of performance achievement, with evidence suggesting more accuracy and nuanced detail from 7-point Likert scales than those with five or fewer response options (Finstad, 2010; Debets et al., 2020). Furthermore, use of a greater number of points between the poles of the rating dimension, such as seven rather than three, better approximates continuous data and the underlying metric and concept (Wu and Leung, 2017).

Future Research

A thematic structure has been established in the development of these scales, based on the Standards (Level 3; Australian Psychology and Accreditation Council [APAC], 2019), although variations to this structure may emerge in future evaluations. It is possible that the core domains of practice are not independent (Gonsalvez and Crowe, 2014), and there may be clusters of competencies (either at skill or domain level) that are interrelated. Subsequent investigation of the scales will provide a clear understanding of the utility of the factor structure and overall number of items, and further psychometric properties. Comparing responses on the self-report version of the scale (COPPR-S) and the observer report version (COPPR-O) will provide insight into the self-assessment of competence and bi-directional learning (Gonsalvez and Crowe, 2014).

CONCLUSION

This paper has outlined the development and initial pilot test of the COPPR. These scales provide the first standardized, quantitative measures of professional psychology competencies in Australia, and show promise as an emerging measure of competence for use in practice, training, and research settings. Future research will further evaluate the psychometric properties of these scales. It is hoped that this measure will provide an initial platform to assess competence across multiple domains of practice at registered psychologist level in Australia.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of New England's Human Research Ethics Committee. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KR conceived the measure, led the research from creation to completion, drafted the manuscript, and created the COPPR scales with consultation from NS and JS. KR and NS undertook ethics approval and collected the pre-test data. KR, NS, and JS adapted the measure from stakeholder review in the pre-test. SC and JS wrote the pre-test analysis section. KR and SC undertook a second ethics approval for an initial data collection. KR and SB undertook a further ethics approval and collected the pilot test data. KR and AR analyzed the pilot test data and completed the revisions for the final version of the manuscript. AR, JS, SC, and NS contributed to editing and writing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

REFERENCES

Ambrose, R., Philipp, R., Chauvot, J., and Clement, L. (2003). A web-based survey to assess prospective elementary school teachers' beliefs about mathematics and mathematics learning: an alternative to Likert scales. *Int. Group Psychol. Math. Educ.* 2, 33–40.

Australian Institute of Health and Welfare (2022). *Mental Health Services in Australia Report*. Available online at: https://www.aihw.gov.au/reports/mental-health-services/mental-health-services-in-australia/report-contents/mental-health-workforce (accessed on March 28, 2022).

Australian Psychology and Accreditation Council [APAC] (2019). *Accreditation Standards for Psychology Programs*. Available online at: Version 1.2 https://www.psychologycouncil.org.au/sites/default/files/public/Standards_20180912_Published_Final_v1.2.pdf (accessed on January 1, 2019).

Australian Psychological Society (2007). *Code of Ethics. Author*. Melbourne: Australian Psychological Society.

Benner, P. (2004). Using the Dreyfus model of skill acquisition to describe and interpret skill acquisition and clinical judgment in nursing practice and education. *Bull. Sci. Technol. Soc.* 24, 188–199. doi: 10.1177/0270467604265061

Brannick, M. T., Erol-Korkmaz, H. T., and Prewett, M. (2011). A systematic review of the reliability of objective structured clinical examination scores. *Med. Educ.* 45, 1181–1189. doi: 10.1111/j.1365-2923.2011.04075.x

Carifio, J., and Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. *Med. Educ.* 42, 1150–1152. doi: 10.1111/j.1365-2923.2008.03172.x

Carpenter, S. (2018). Ten steps in scale development and reporting: a guide for researchers. *Commun. Methods Meas.* 12, 25–44. doi: 10.1080/19312458.2017.1396583

Clark, L. A., and Watson, D. (1995). Constructing validity: basic issues in objective scale development. *Psychol. Assess.* 7, 309–319. doi: 10.1037/1040-3590.7.3.309

Cohen, J. (1988). *Statistical Power Analysis*, 2nd Edn. New Jersey: US: Erlbaum.

Deane, F. P., Gonsalvez, C. J., Joyce, C., and Britt, E. (2017). Developmental trajectories of competency attainment amongst clinical psychology trainees across field placements. *J. Clin. Psychol.* 74, 1641–1652. doi: 10.1002/jclp.22619

Debets, M. P., Scheepers, R. A., Boerebach, B. C., Arah, O. A., and Lombarts, K. (2020). Variability of residents' ratings of faculty's teaching performance measured by five- and seven-point response scales. *BMC Med. Educ.* 20:325–325. doi: 10.1186/s12909-020-02244-9

DeVellis, R. F. (2003). *Scale Development: Theory and Applications*, 2nd Edn. Thousand Oaks: Sage Publications.

Dreyfus, H. L., and Dreyfus, S. E. (1986). *Mind Over Machine*. Mumbai: Free Press.

Dreyfus, S. E., and Dreyfus, H. L. (1980). *A Five-Stage Model of the Mental Activities Involved in Directed Skill Acquisition*. Norwalk: Storming Media.

Duvivier, R. J., van Dalen, J., Muijtjens, A. M., Moulaert, V., Van der Vleuten, C., and Scherpbier. (2011). The role of deliberate practice in the acquisition of clinical skills. *BMC Med. Educ.* 11:101. doi: 10.1186/1472-6920-11-101

Ericsson, K. A., Krampe, R. T., and Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychol. Rev.* 100, 363–406. doi: 10.1037/0033-295x.100.3.363

Ericsson, K. A., and Charness, N. (1994). Expert performance: its structure and acquisition. *Am. Psychol.* 49, 725–747. doi: 10.1037/0003-066X.49.8.725

Faul, F., Erdfelder, E., Lang, A.-G., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/bf03193146

Finstad, K. (2010). Response interpolation and scale sensitivity: evidence against 5-point scales. *J. Usability Stud.* 5, 104–110.

Fouad, N. A., Grus, C. L., Hatcher, R. L., Kaslow, N. J., Hutchings, P. S., Madson, M. B., et al. (2009). Competency benchmarks: a model for understanding and measuring competence in professional psychology across training levels. *Train. Educ. Prof. Psychol.* 3, S5–S26. doi: 10.1037/a0015832

Ghazali, N., Nordin, M. S., Hashim, S., and Hussein, S. (2018). Measuring content validity: students' self-efficacy and meaningful learning in massive open online course (MOOC) scale. *Adv. Soc. Sci. Educ. Hum. Res.* 115, 128–133. doi: 10.2991/icems-17.2018.2

Gonsalvez, C. J., Bushnell, J., Blackman, R., Deane, F., Bliokas, V., Nicholson-Perry, K., et al. (2013). Assessment of psychology competencies in field placements: standardised vignettes reduce rater bias. *Train. Educ. Prof. Psychol.* 7, 99–111. doi: 10.1037/a0031617

Gonsalvez, C. J., and Crowe, T. P. (2014). Evaluation of psychology practitioner competence in clinical supervision. *Am. J. Psychother.* 68, 177–193. doi: 10.1176/appi.psychotherapy.2014.68.2.177

Gonsalvez, C. J., Deane, F. P., Knight, R., Nasstasia, Y., Shires, A., Nicholson-Perry, K., et al. (2015). The hierarchical clustering of clinical psychology practicum competencies: a multisite study of supervisor ratings. *Clin. Psychol. Sci. Pract.* 22, 390–403. doi: 10.1111/cpsp.12123

Gonsalvez, C. J., Deane, F. P., and Caputi, P. (2016). Consistency of supervisor and peer ratings of assessment interviews conducted by psychology trainees. *Br. J. Guid. Couns.* 44, 516–529. doi: 10.1080/03069885.2015.1068927

Gonsalvez, C. J., and Freestone, J. (2007). Field supervisors' assessments of trainee performance: are they reliable and valid? *Aust. Psychol.* 42, 23–32. doi: 10.1080/00050060600827615

Halek, M., Holle, D., and Bartholomeyczik, S. (2017). Development and evaluation of the content validity, practicability and feasibility of the innovative dementia-oriented assessment system for challenging behaviour in residents with dementia. *BMC Health Ser. Res.* 17:554. doi: 10.1186/s12913-017-2469-8

Harden, R. M. (2016). Revisiting 'assessment of clinical competence using an objective structured clinical examination (OSCE)'. *Med. Educ.* 50, 376–379. doi: 10.1111/medu.12801

Hatcher, R. L., and Lassiter, K. D. (2007). Initial training in professional psychology: the practicum competencies outline. *Train. Educ. Prof. Psychol.* 1, 49–63. doi: 10.1037/1931-3918.1.1.49

Hatcher, R. L., Fouad, N. A., Grus, C. L., Campbell, L. F., McCutcheon, S. R., and Leahy, K. L. (2013). Competency benchmarks: practical steps toward a culture of competence. *Train. Educ. Prof. Psychol.* 7, 84–91. doi: 10.1037/a0029401

Hodge, D. R., and Gillespie, D. (2003). Phrase completions: an alternative to Likert scales. *Soc. Work Res.* 27, 45–55. doi: 10.1093/swr/27.1.45

Hitzeman, C., Gonsalvez, C. J., Britt, E., and Moses, K. (2020). Clinical psychology trainees' self versus supervisor assessments of practitioner competencies. *Clin. Psychol.* 24, 18–29. doi: 10.1111/cp.12183

IBM Corp (2021). *IBM SPSS Statistics for Windows, Version 28.0*. Armonk, NY: IBM Corp.

Kaslow, N. J. (2004). Competencies in professional psychology. *Am. Psychol.* 59, 774–781. doi: 10.1037/0003-066X.59.8.774

Kelly, M. A., Mitchell, M. L., Henderson, A., Jeffrey, C. A., Groves, M., Nulty, D. D., et al. (2016). OSCE best practice guidelines –applicability for nursing simulations. *Adv. Simul.* 1:10. doi: 10.1186/s41077-016-0014-1

Khan, K. Z., Ramachandran, S., Gaunt, K., and Pushkar, P. (2013). The Objective Structured Clinical Examination (OSCE): AMEE guide no. 81. part I: an historical and theoretical perspective. *Med. Teach.* 35, 1437–1446. doi: 10.3109/0142159X.2013.818634

Krampe, R. T., and Ericsson, K. A. (1996). Maintaining excellence: deliberate practice and elite performance in young and older pianists. *J. Exp. Psychol. Gen.* 125, 331–359. doi: 10.1037/0096-3445.125.4.331

Lichtenberg, J. W., Portnoy, S. M., Bebeau, M. J., Leigh, I. W., Nelson, P. D., Rubin, N. J., et al. (2007). Challenges to the assessment of competence and competencies. *Prof. Psychol. Res. Pract.* 38, 474–478. doi: 10.1037/0735-7028.38.5.474

Loades, M. E., and Myles, P. J. (2016). Does a therapist's reflective ability predict the accuracy of their self-evaluation of competence in cognitive behavioural therapy? *Cogn. Behav. Therap.* 9, 1–14. doi: 10.1017/S1754470X16000027

Lynn, M. R. (1986). Determination and quantification of content validity. *Nurs. Res.* 35, 382–385.

McIlveen, P., Burton, L. J., and Beccaria, G. (2013). A short form of the Career Futures Inventory. *J. Career Assess.* 21, 127–138. doi: 10.1177/1069072712450493

Morgado, F. F., Meireles, J. F., Neves, C. M., Amaral, A. C., and Ferreira, M. E. (2017). Scale development: ten main limitations and recommendations to improve future research practices. *Psicol. Reflex. Crít.* 30, 1–20. doi: 10.1186/s41155-016-0057-1

Nash, J. M., and Larkin, K. T. (2012). Geometric models of competency development in specialty areas of professional psychology. *Train. Educ. Prof. Psychol.* 6, 37–46. doi: 10.1037/a0026964

Nelson, P. D. (2007). Striving for competence in the assessment of competence: psychology's professional education and credentialing journey of public accountability. *Train. Educ. Prof. Psychol.* 1, 3–12. doi: 10.1037/1931-3918.1.1.3

Nunnally, J. C. (1978). *Psychometric Theory*. New York, NY: McGraw-Hill.

Panayides, P. (2013). Coefficient alpha: interpret with caution. *Eur. J. Psychol.* 9, 687–696. doi: 10.5964/ejop.v9i4.653

Peña, A. (2010). The Dreyfus model of clinical problem-solving skills acquisition: a critical perspective. *Med. Educ. Online* 15, 10.3402/meo.v15i0.4846. doi: 10.3402/meo.v15i0.4846

Pescaroli, G., Velazquez, O., Alcántara-Ayala, I., Galasso, C., Kostkova, P., and Alexander, D. (2020). A likert scale-based model for benchmarking operational capacity, organizational resilience, and disaster riskreduction. *Int. J. Disaster Risk Sci.* 11, 404–409. doi: 10.1007/s13753-020-00276-9

Polit, D. F., and Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? critique and recommendations. *Res. Nurs. Health* 29, 489–497. doi: 10.1002/nur.20147

Revilla, M. A., Saris, W. E., and Krosnick, J. A. (2014). Choosing the number of categories in agree-disagree scales. *Soc. Methods Res.* 43, 73–97. doi: 10.1177/0049124113509605

Roberts, M. C., Borden, K. A., Christiansen, M. D., and Lopez, S. J. (2005). Fostering a culture shift: assessment of competence in the education and careers of professional psychologists. *Prof. Psychol. Res. Pract.* 36, 355–361. doi: 10.1037/0735-7028.36.4.355

Rodolfa, E., Baker, J., DeMers, S., Hilson, A., Meck, D., Schaffer, J., et al. (2014). Professional psychology competency initiatives: implications for training, regulation, and practice. *S. Afr. J. Psychol.* 44, 121–135. doi: 10.1177/0081246314522371

Rodolfa, E., Bent, R., Eisman, E., Nelson, P., Rehm, L., and Ritchie, P. (2005). A cube model for competency development: implications for psychology educators and regulators. *Prof. Psychol. Res. Pract.* 3, 347–354. doi: 10.1037/0735-7028.36.4.347

Rodolfa, E., Greenberg, S., Hunsley, J., Smith-Zoeller, M., Cox, D., Sammons, M., et al. (2013). A competency model for the practice of psychology. *Train. Educ. Prof. Psychol.* 7, 71–83. doi: 10.1037/a0032415

Rodolfa, E., and Schaffer, J. (2019). Challenges to psychology education and training in the culture of competence. *Am. Psychol.* 74, 1118–1128. doi: 10.1037/amp0000513

Rottinghaus, P. J., Day, S. X., and Borgen, F. H. (2005). The Career Futures Inventory: a measure of career-related adaptability and optimism. *J. Career Assess.* 13, 3–24. doi: 10.1177/1069072704270271

Rubin, N. J., Bebeau, M., Leigh, I. W., Lichtenberg, J. W., Nelson, P. D., Portnoy, S., et al. (2007). The competency movement within psychology: an historical perspective. *Prof. Psychol. Res. Pract.* 38, 452–462. doi: 10.1037/0735-7028.38.5.452

Salkind, N. J. (2010). *Encyclopedia of Research Design*, Vol. 1. Thousand Oaks: Sage.

Schilling, L. S., Dixon, J. K., Knafl, K. A., Grey, M., Ives, B., and Lynn, M. R. (2007). Determining content validity of a self-report instrument for adolescents using a heterogeneous expert panel. *Nurs. Res.* 56, 361–366. doi: 10.1097/01.NNR.0000289505.30037.91

Sharpless, B. A., and Barber, J. P. (2009). A conceptual and empirical review of the meaning, measurement, development, and teaching of intervention competence in clinical psychology. *Clin. Psychol. Rev.* 29, 47–56. doi: 10.1016/j.cpr.2008.09.008

Sheen, J., Mcgillivray, J., Gurtman, C., and Boyd, L. (2015). Assessing the clinical competence of psychology students through objective structured clinical examinations (OSCEs): student and staff views. *Aust. Psychol.* 50, 51–59. doi: 10.1111/ap.12086

Sheridan, B. (2021). How might clinical pharmacists use a profession developed competency based self-assessment tool to direct their professional development? *Aust. J. Clin. Educ.* 10, 1–32. doi: 10.53300/001c.24868

Stevens, B., Hyde, J., Knight, R., Shires, A., and Alexander, R. (2017). Competency-based training and assessment in Australian postgraduate clinical psychology education. *Clin. Psychol.* 21, 174–185. doi: 10.1111/cp.12061

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res. Sci. Educ.* 48, 1273–1296. doi: 10.1007/s11165-016-9602-2

Tavakol, M., and Dennick, R. (2011). Making sense of Cronbach's alpha. *Int. J. Med. Educ.* 2, 53–55. doi: 10.5116/ijme.4dfb.8dfd

Usry, J., Partington, S. W., and Partington, J. W. (2018). Using expert panels to examine the content validity and inter-rater reliability of the ABLLS-R. *J. Dev. Phys. Disabil.* 30, 27–38. doi: 10.1007/s10882-017-9574

Waples, C. J., Weyhrauch, W. S., Connell, A. R., and Culbertson, S. S. (2010). Questionable defeats and discounted victories for Likert rating scales. *Ind. Organ. Psychol.* 3, 477–480.

Ward, H., and Barratt, J. (2005). Assessment of nurse practitioner advanced clinical practice skills: using the objective structured clinical examination (OSCE): Helen Ward and Julian Barratt examine how OSCEs can be developed to ensure a robust assessment of clinical competence. *Prim. Health Care* 15, 37–41. doi: 10.7748/phc2005.12.15.10.37.c563

Watt, H. M. G., Ehrich, J., Stewart, S. E., Snell, T., Bucich, M., Jacobs, N., et al. (2019). Development of the psychologist and counsellor self-efficacy scale. *High. Educ. Skills Work Based Learn.* 9, 485–509. doi: 10.1108/HESWBL-07-2018-0069

Wigley, C. J. (2013). Dispelling three myths about Likert scales in communication trait research. *Commun. Res. Rep.* 30, 366–372. doi: 10.1080/08824096.2013.836937

Wu, H., and Leung, S. O. (2017). Can Likert scales be treated as interval scales?—A simulation study. *J. Soc. Serv. Res.* 43, 527–532. doi: 10.1080/01488376.2017.1329775

Yaghmale, F. (2003). Content validity and its estimation. *J. Med. Educ.* 3, 25–27.

Yang, Y., and Green, S. B. (2011). Coefficient alpha: a reliability coefficient for the 21st century? *J. Psychoeduc. Assess.* 29, 377–392. doi: 10.1177/0734282911406668

Yap, K., Sheen, J., Nedeljkovic, M., Milne, L., Lawrence, K., and Hay, M. (2021). Assessing clinical competencies using the objective structured clinical examination (OSCE) in psychology training. *Clin. Psychol.* 25, 260–270. doi: 10.1080/13284207.2021.1932452

Zamanzadeh, V., Ghahramanian, A., Rassouli, M., Abbaszadeh, A., Alavi-Majd, H., and Nikanfar, A. R. (2015). Design and implementation content validity study: development of an instrument for measuring patient-centered communication. *J. Caring Sci.* 4, 165–178. doi: 10.15171/jcs.2015.017