A Systematic Review of Reflective Practice Questionnaires and Scales for Healthcare Professionals: A narrative synthesis

Su Min Ooi*, Paul Fisher & Siân Coker

Department of Clinical Psychology and Psychological Therapies, University of East Anglia, Norwich, UK

*Corresponding author: ivyooi@ymail.com

Biographical notes:

Dr Su Min Ooi is a clinical psychologist who has more than 10 years of clinical experience in healthcare settings. She has developed interests around clinical supervision and reflective practice when she underwent her doctorate training at University of East Anglia (UEA).

Dr Paul Fisher is a Clinical Psychologist, Senior Clinical Lecturer and the Programme Director of the Psychological Wellbeing Practitioner (PWP) training at UEA. His teaching and research interests include professional practice issues for Clinical psychologists such as formulation and reflective practice. He has expertise in the use of qualitative research methods, and this often informs his research.

Professor Siân Coker is the Deputy Head of Department Clinical Psychology and Psychological Therapies at UEA. Her research interests include perfectionism and and its' role in the management of a number of health conditions. She is also a research enthusiast who has significantly contributed in the teaching of research methods, ethical practice, and ethics in research on the doctoral programme in Clinical Psychology.

A Systematic Review of Reflective Practice Questionnaires and Scales for Healthcare Professionals: A narrative synthesis

Abstract

Reflective Practice (RP) is considered a crucial component of personal and professional learning. RP is regarded as a way that professionals learn from experience to understand and enhance their practice by responding appropriately to self-reflection. Despite playing a crucial role in healthcare settings, there is little agreement on how to assess RP. This study aims to systematically review self-rating instruments that assess RP in healthcare professionals. Peer review articles assessing RP in healthcare professionals using a self rating instructment, published in English between 1998 and 2018 from PubMed, CINAHL, and PsychInfo databases, were considered for inclusion. 18 papers were appraised, the strengths and weaknesses of the measures were discussed in accordance with an adapted critical appraisal checklist. In general, all self-report instruments included in this review were potentially generalisable to healthcare professionals or health science programmes with some adaptation. Given the limited evidence for other measurement scales, the Reflective Questionnaire and Self-Reflection and Insight Scale are recommended for measuring RP within healthcare settings. Future research developing a standardised tool for the review of mixed-method, heterogeneous, questionnaire studies is strongly recommended.

Keywords: reflective practice; reflection; measure; questionnaire; scale; healthcare professional

Introduction

Reflection or reflective practice (RP) is a crucial component of personal and professional learning (Nguyen, Fernandez, Karsenti, & Charlin, 2014). Dewey (as cited in Finlay, 2008) stated that reflective thinking encourages reflective action which involves careful and critical consideration of knowledge by moving away from conventional thinking or action. For Dewey, individuals learn through both thinking and doing (as cited in Finlay, 2008); they do not merely think about what they are doing but also the rationale for their actions. The concept of reflection has gained attention from researchers and professionals in various disciplines, including education, medicine, nursing, midwifery, social work, and other health science professions (Nguyen et al., 2014; Sweet, Bass, Sidebotham, Fenwick & Graham, 2019; Tummons, 2011).

Kolb (1984) proposed the experiential learning cycle which emphasises the role of reflection in learning (Figure 1). In this model learning begins with an individual having a concrete experience or encounter that leads to an observation. A reflection on the observation leads to the formation of new ideas (or modification of an existing abstract concept) which can be applied to future situations resulting in new experiences. Given this, reflection has a vital role in changing a person's concrete experiences to abstract meanings.

[Figure 1 near here]

Reflection became the focus of further attention when Schön, (1983) coined the term 'Reflective Practice'. For Schön (1983), the person on the high ground can see a range of possible routes and plan a suitable path to get to the destination. Someone else, in the 'swampy lowlands' (p.42) is unable to do the same and this lowland person learns from their mistakes through trial and error, enabling them to navigate through the swamp, which is regarded as a reflective approach by Schön. For Schön critical reflection in practice with

other forms of scientific evidence is crucial in decision making because professionals are often required to make quick and complex decisions without access to all available resources (cited in Fisher, Chew, & Leow, 2015).

In recent years there has been greater research interest around this topic (Harford & MacRuairc, 2008; Mann, Gordon, & MacLeod, 2009; Ruch, 2005). The demonstration of RP is also necessitated by various health professional accrediting bodies such as Health and Care Professions Council (HCPC, 2015), British Psychological Society (BPS, 2017), American Psychological Association (APA, 2018), General Medical Council (GMC, 2018), and Nursing and Midwifery Council (NMC, 2010).

Despite the importance placed on RP, there has been little consensus on the definition of reflection (Bassot, 2015; Fisher et al., 2015; Nguyen et al., 2014). Nguyen and colleagues (2014) believe that the lack of shared understanding of reflection has hampered the development of practical methods to analyse, teach and assess RP. In order to enhance mutual understanding of reflection, Nguyen and colleagues (2014) operationally defined reflection as "The process of engaging self in attentive, critical, exploratory and iterative interaction with one's thoughts and actions, and their underlying conceptual frame, with a view to changing them and a view on the change itself" (p.1176). Given that RP is fluid and contingent in nature (Tummons, 2011), it remains a complex concept to be pragmatically operationalised.

Health professionals are expected to have the capacity to reflect upon their clinical work to sustain professional growth (O'Reilly & Milner, 2015). However, only a limited number of instruments have been developed to assess RP. Due to the lack of a unified definition of RP, the inherent implication is a similar lack of unified assessment. Boenink, Oderwald, De Jonge, Van Tilburg and Smal (2004) outlined scales to measure different aspects of reflection. These include moral reasoning, teaching and learning, as well as

professional competency. Although the existing instruments claim to measure self-reflection or RP, they were developed for a variety of different purposes. For instance, some scales target the assessment of reflective learning process (Phan, 2009; Sobral, 2001), whereas others emphasised the level of involvement in RP (Aukes, Geertsma, Cohen-Schotanus, Zwierstra, & Slaets, 2007; Grant, Franklin, & Langford, 2002; Priddis & Rogers, 2018). Based on current research, it would be difficult for a professional or academic to know which tools designed to measure reflection are appropriate.

Overall, despite the importance placed on the concept of RP there is comparatively little evidence-based research focusing on its measurement and existing published research has not been the subject of a systematic review. Such a review would enable an appraisal of the quality of existing tools and promote the integration of a potentially disparate body of literature. The current paper aims to identify and systematically review self-rating instruments that assess RP within healthcare professionals.

Research Questions

What self-report questionnaires and scales are available to assess reflective practice in qualified healthcare professionals?

Methodology

Data Sources and Searches

In order to review the published literature and the assessment of RP in healthcare professions, the PubMed, CINAHL, and PsychInfo databases were searched using the '*' symbol (wildcard) to replace some letters in keywords. Search terms were "reflective practice", reflective, reflection*, self-reflection, self-reflective, self-awareness, self-perception* AND measure*, assess*, scale, questionnaire AND "healthcare professional*", "health care professional*", "healthcare worker*", "health care worker*", nurse*, medical doctor*, doctor*, "occupational therap*", "physical therap*", physiotherap*, "social worker*",

dietitian*, dietician*, "speech and language therap*", "speech therap*", psychology, and psychologist*.

Eligibility Criteria

The titles and abstracts of identified studies were screened by the lead reviewer (S.M.) according to inclusion and exclusion criteria. Both electronic and hand searches were limited to the following inclusion criteria: 1) journal articles published in English language, 2) peer reviewed articles, 3) published between 1998 to 2018, 4) articles that discussed or used a self-rating instrument to measure RP, and 5) the instrument is used to measure RP in healthcare professionals. Exclusion criteria included 1) commentaries, 2) personal reviews or reflections, 3) book reviews, 4) papers that did not describe RP or the use of self-rating instrument for RP, and 5) translation of an existing scale into another language.

Study Selection

The initial search conducted on 22nd January 2019 identified 778 journal articles. After duplicates were removed, 509 studies remained from the electronic search. These articles were screened for suitability with reference to the inclusion and exclusion criteria. Articles that did not meet the inclusion criteria were removed (n= 497). The lead reviewer hand-searched the reference lists of the identified papers (n=12) to identify additional records meeting the criteria which generated six further articles. A second reviewer (P.F.) independently reviewed the full-text of the remaining studies against the inclusion and exclusion criteria, and no discrepancies arose. This resulted in the inclusion of 18 papers for full-text review and critical appraisal. A PRISMA flowchart (see Figure 2) illustrates the screening process.

[Figure 2 near here]

Data Extraction

Information extracted from the 18 papers including authors, study location, study population, sample size, study design, measures used, and brief summary of results are shown in Table 1. Study participants were mainly healthcare professionals or healthcare students, and the sample size ranged from 11 to 1664 participants. Cross-sectional surveys were predominantly used with a mixture of cohort studies and multimethod studies. Five papers described the use of the Self-Reflection and Insight Scale (SRIS) and the Reflection Questionnaire (RQ) respectively, and two studies used the Reflection-in-Learning Scale (RLS). Other instruments used included: Reflective Practice Questionnaire (RPQ, N=1), Reflective Learning and Interaction Model Questionnaire (RLIMQ, N=1), Groningen Reflection Ability Scale (GRAS, N=1), Critically Reflective Work Behaviour (CRWB, N=1), 10-item scale (N=1), and 37-item scale (N=1).

All relevant data was extracted from the selected papers by the lead reviewer and critically appraised. A critical appraisal checklist was used to evaluate various aspects of the questionnaire based research. Given the absence of a robust single checklist that could be used to critically appraise heterogeneous studies with diverse study designs and different questionnaire types, an adapted version of three measures: Critical Appraisal of a Questionnaire Study (Roever, 2016), the Critical Appraisal Checklist for a Questionnaire Study (National Institute for Health and Care Excellence, NICE clinical guideline 143, 2012; p.143-144), and the Critical Appraisal of Qualitative Studies (Centre for Evidence Based Medicine, n.d.) criteria was initially trialled and used to assess the papers. To critically appraise different aspects of the identified multimethod studies, 24 items were selected from these measures (see Appendix A).

Quality Assessment and Data Analysis

Appraisal focused on the issue of adequate coverage in the following domains: Research aim and study design; sampling; format; piloting; psychometric properties; distribution,

administration and response; analysis; discussion and conclusion; and ethics. A consensus

was reached among the authors on items to be included in the critical appraisal checklist as

well as a calculation of the quality of the paper based on the criteria developed by Roever

(2016, see Appendix A for details). The quality of the 18 papers was appraised by the lead

reviewer according to the 24-item checklist (Appendix A). The third author (S.C.) rated four

out of the total of 18 papers and an agreement was reached for all ratings with one exception;

which was resolved via discussion to achieve 100% agreement for the final ratings. A

summary table of quality appraisal measure and ratings is shown in Appendix A.

Given the heterogeneity of the studies included, particularly with regard to the

measures used, a narrative synthesis is provided.

Results

Fifteen studies are of high quality and the remaining are rated to have acceptable quality (see

Overall Quality column, Appendix A). The characteristics of all 18 papers and findings

extracted from the individual studies are briefly outlined (see Table 1). This is followed by a

synthesis of different instruments used to measure reflection or RP and the quality of the

measures used.

Narrative Synthesis

Six of the included papers (studies 1, 4, 6, 8, 13 and 18) were validation studies, four papers

(studies 2, 7, 10 and 14) aimed to explore an integrated reflective model in a specific study

context, five papers (study 3, 5, 11, 12 and 17) attempted to investigate the relationship

between reflection and other variables, and three papers studied the impact of a programme

or a tool on reflective ability. Ten studies used convenience sampling and targeted healthcare

students (i.e., psychology, occupational and physical therapy, medical, nursing, and dietetic).

Nonetheless, every instrument used by the respective studies can be adapted and used for a

variety of healthcare professionals, according to the authors of the papers including in this review. In terms of study design, all papers included in this review conducted a survey study in either cross-sectional or longitudinal study design. Appropriate methodology was applied for all of the studies, for example validation studies were carried out to check the psychometric properties of the instrument used.

Measures

Nine different measures of RP or self-reflection were identified. The strengths and weaknesses of the different questionnaires such as user-friendliness, psychometric properties, generalisability, sampling method and the nature of sample were identified and considered (see Appendix B). The RQ and SRIS were the more frequently used measures investigated and these will be appraised in more detail together with a brief description of the other measures identified in the review.

Reflective Questionnaire (RQ)

Kember and colleagues (2000) developed a user-friendly and readily interpretable, four-scale, 16-item questionnaire to measure the extent to which health sciences students engage in reflective thinking during their educational programme. The RQ was developed based primarily on Mezirow's reflective thinking framework (i.e., Habitual Action, Understanding, Reflection, and Critical Reflection; as cited in Kember et al., 2000). Mezirow (as cited in Kember et al., 2000), described *Habitual Action* as a frequently used, learnt action which has become an automatic activity that requires little conscious thought such as riding a bicycle or typing on a keyboard. *Understanding* is regarded as a type of thinking that makes use of existing knowledge without trying to appraise that information (e.g., 'learning from books'). *Reflection* is interpreted as 'validity testing' that involves the review of assumptions on the process of content to further make sense of individuals' experience. Finally *Critical*

Reflection, is described as a higher level of reflective thinking which involves awareness of the way individuals perceive, think, feel, or behave in certain ways.

In the development of the RQ, the respondents were asked to rate on a 5-point Likert scale, ranging from definitely agree to definitely disagree, higher scores indicated greater agreement with engaging in the specific reflective thinking each scale assessed. The RQ is primarily used as a tool to examine the effect of the teaching and learning environment on reflective thinking. The authors proposed that the instrument could be used to explore the study patterns of individual students, to investigate the inter-relationship between reflective thinking and other constructs, and to compare groups of students that were subjected to different treatments or conditions.

The psychometric properties of the RQ (Kember et al., 2000) had satisfactory reliability for each scale (α ranging from 0.62 to 0.76). Good validity was also established through confirmatory factor analysis and a good fit to the intended factor structure (χ^2 = 179.3, df = 100) with a comparative fit index (CFI = 0.903) was shown. Given that it was designed for an academic settings, some modification would be required for the RQ to be used to measure reflective thinking in healthcare settings (Kember et al., 2000).

Self-Reflection and Insight Scale (SRIS)

The SRIS (Grant et al., 2002) was developed to examine levels of self-reflection and insight. The authors assert that self-reflection is a metacognitive factor that contributes to a purposeful and directed change, hence they developed the SRIS to inform individuals' performance by monitoring their reflective thinking and insight. It is a self-administered, 20-item questionnaire which is categorised into three subscales, Engagement in Self-Reflection, Need for Self-Reflection, and Insight. Respondents rate items on a 6-point Likert scale (1 = strongly disagree, 6 = strongly agree), with higher scores indicating higher self-reflection and insight. The SRIS was originally designed and constructed to be an advance on the Private

Self-Consciousness Scale (PrSCS; as cited in Grant et al., 2002), and was used to investigate the relationship between self-reflection and insight with other variables (in study 8, 9 and 10).

Good internal consistency (α = 0.91 for self-reflection; α = 0.87 for insight), test-retest reliability, and convergent and discriminant validity was reported (Grant et al., 2002; Roberts & Starks, 2008). Although the SRIS has been validated with a small and homogeneous sample, it has been adapted for use by various healthcare disciplines (study 7, 8, 9 and 10). It is therefore expected that the SRIS can be adapted and generalised to suit other healthcare professionals.

Other measures

Seven other self-rating instruments were discussed in the remaining eight studies. Two studies used the Reflective-in-Learning Scale (RLS), and the remaining studies each discussed a reflective measure such as Reflecive Practice Questionnaire (RPQ), Reflective Learning and Interaction Model Questionnaire (RLIMQ), Groningen Reflection Ability Scale (GRAS), Critically Reflective Work Behaviour (CRWB), 10-item scale (created during the research project MIRROR, as cited in Renner et al., 2014), and a 37-item scale (i.e., 21, 5-point Likert scale; 13 open- and three closed-ended free test questions; as cited in O'Reilly & Milner, 2015). These scales were validated or used within an education or healthcare population. Some measures focused on self-reflection and learning (RLS and GRAS) whereas others aimed to understand the relationship between RP and other constructs (CRWB and RPQ), and to investigate the effectiveness of technology-based reflective tools (RLIQ, 10-item scale and 37-item scales).

The reliability and various validity analyses reported within the studies for the measures are reported in Table 1. The reliability of each instrument is reported based on the Hair, Black, Babin, Anderson and Tatham (2006)'s cut-off (i.e. Cronbach's alpha between .60 to .70 is acceptable, and .70 and above is considered good reliability, see Table

1). Although the validity of some instruments was not reported or had yet to be established, a more robust validation study with larger sample size was recommended by some studies included in this review (Lowe et al., 2007; O'Reilly & Milner, 2015; Priddis & Rogers, 2018).

[Table 1 near here]

Based on the review the RQ and SRIS, which were more extensively used and had been evaluated in previous research with different populations, are likely to be the most useful in measuring RP within healthcare settings.

Discussion

This section briefly summarises the findings and discusses the relationship between RP and other relevant constructs. Current trends in the body of research and the implications for professional and continuous learning are discussed. Finally, the limitations and recommendations are highlighted.

Summary of Findings

18 papers with acceptable to high quality ratings were included in this review. Nine instruments were reviewed, and a majority demonstrated satisfactory to good internal consistencies and validity. In general, all self-measure instruments included in this review were potentially generalisable to various healthcare professionals or health science education programmes with further adaptation.

The RQ and SRIS were the most used and reported measures. These measures were shown to have adequate to good psychometric properties from various studies but did rely on homogeneous sample and purposeful or convenience sampling methods. Although the quality of the remaining studies fell within the acceptable to high range, some limitations were noted such as: unclear psychometric properties; not being readily validated, and small

sample sizes. Given limited evidence for other measurement scales, the RQ and SRIS would be recommended for use in measuring RP within healthcare professionals.

RP and Reflective Measures

In order to effectively manage the fast changing and complex healthcare environment, RP has gained increasing attention in education and professional practice settings in the last two decades (Levine, 2014; Mamede & Schmidt, 2004; O'Reilly & Milner, 2015). Given the fluid and contingent nature of the concept of RP (Tummons, 2011), the review found that different models of RP were used to further develop the reflective measures. Despite being a registration requirement of various professional accreditation bodies, this review found limited works focusing on how healthcare professionals or agencies develop, cultivate and more crucially measure the use of RP. This may relate to the lack of consensus on how reflection is understood and assessed (Nguyen et al 2014).

To unify and clarify the concept of RP, it is recommended that intra-disciplinary collaboration is considered to generate consensus on an appropriate reflective measure for various healthcare settings. It also needs to be noted that the different conceptions of RP across cultures makes the development of a measure that can be used cross culturally challenging. A combination of qualitative and quantitative methods would be useful to capture the richness of RP and to explore the in-depth contextual nature of reflective thinking (Phan, 2009). This might enable researchers to not always seek a unified concept of RP, with the inherent difficulties this raises, but instead to be able to contextualised what RP means within the specific healthcare/ academic/ cultural context under exploration.

Reflective measures were not only used to measure the level of engagement in RP, but also to understand the interaction between RP and other variables to establish a wider perspective on reflective thinking. Given that different studies have different background assumptions about what RP is, the scale they use is often paired with a range of variables

such as learning approaches, achievement goal orientations, academic performance, clinical competence, learning effectiveness, and self-directed change to understand their interaction. For example, reflective measures were used to identify learning strategies that help enhance reflective learning (Levine, 2014), understand factors that influence the development of clinical competency in healthcare professionals (Pai, 2016) as well as examine the experiences, advantages, and potential pitfalls of RP (Priddis & Rogers, 2018).

It is recommended that the RQ and SRIS be adapted for future RP studies within healthcare professionals given adequate to good psychometric properties reported in various validation studies. For future validation studies, expansion from homogeneity to heterogeneity sample increase the generalisability of the RQ and SRIS across professionals. As the RQ was initially developed for use in academic settings, some modifications would be necessary for use in wider settings.

Limitations and Future Research

One limitations was that the electronic search included for the literature search only generated two thirds of the final papers. One-third of the studies were identified through hand-search. More comprehensive search terms and a wider use of databases might have captured these papers. Since this review was carried out in early 2019, new papers published since the search are not included in the analysis.

Given the lack of a standardised critical appraisal checklist for the use of multimethod, heterogeneous studies the appraisal checklist used in this review was adapted from different studies. This allowed for a comprehensive and relevant appraisal of the identified studies that was relevant to the research aims. Whilst deemed appropriate for the purposes of the current review it would have been helpful to have formally piloted the current adapted checklist. Future research focusing on developing a standardised tool for the review of mixed-method, heterogeneous, questionnaire studies is strongly recommended. The

establishment of a pool of critical appraisal items that covers various research designs, methods, and sampling strategies would be a valuable research initiative.

Conflict of Interest

The authors declare that they have no conflicts of interest in the study.

References

American Psychological Association. (2018). *APA guidelines for clinical supervision in health service psychology*. Retrieved January 2019 from https://www.apa.org/about/policy/guidelines-supervision.pdf

Aukes, L. C., Geertsma, J., Cohen-Schotanus, J., Zwierstra, R. P., & Slaets, J. P. J. (2007). The development of a scale to measure personal reflection in medical practice and education. *Medical Teacher*, *29*(2–3), 177–182.

Bassot, B. (2015). *The reflective practice guide: An interdisciplinary approach to critical reflection*. Oxfordshire, UK: Routledge.

Boenink, A. D., Oderwald, A. K., De Jonge, P., Van Tilburg, W., & Smal, J. A. (2004). Assessing student reflection in medical practice. The development of an observer-rated instrument: Reliability, validity and initial experiences. *Medical Education*, *38*(4), 368–377.

British Psychological Society. (2017). *Standards for the accreditation of doctoral programmes in clinical psychology*. Retrieved January 2019 from https://www.ucl.ac.uk/clinical-psychology-doctorate/sites/clinical-psychology-doctorate/files/Appendix4r2010.pdf

Centre for Evidence Based Medicine. (n.d.). *Critical Appraisal of Qualitative Studies*.

Retrieved January 2019 from https://www.cebm.net/wp-content/uploads/2019/01/Qualitative-Studies.pdf

Dunn, L., & Musolino, G. M. (2011). Assessing reflective thinking and approaches to learning. *J Allied Health*, *40*(3), 128–136.

Finlay, L. (2008). Reflecting on "Reflective practice." *PBPL Paper*, *52*, 1–27.

Fisher, P., Chew, K., & Leow, Y. J. (2015). Clinical psychologists' use of reflection and reflective practice within clinical work. *Reflective Practice*, *16*(6), 731–743.

General Medical Council. (2018). Outcomes for graduates. Retrieved 2020 from https://www.gmc-uk.org/-/media/documents/outcomes-for-graduates-a4-6_pdf-78952372.pdf

Grant, A. M., Franklin, J., & Langford, P. (2002). The Self-Reflection and Insight Scale: A new measure of private self-consciousness. *Social Behavior and Personality: An International Journal*, *30*(8), 821–836.

Groot, E. de, Jaarsma, D., Endedijk, M., Mainhard, T., Lam, I., Simons, R.-J., & Beukelen, P. Van. (2012). Critically reflective work behavior of health care professionals. *Journal of Continuing Education in the Health Professions*, *32*(1), 48–57.

Hair, J.E., Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R.L. (2006). Multivariate data analysis. Pearson-Prentice Hall, New Jersey.

Harford, J., & MacRuairc, G. (2008). Engaging student teachers in meaningful reflective practice. *Teaching and Teacher Education*, *24*(7), 1884–1892.

Health & Care Professions Council. (2015). Standard of proficiency: Practitioner psychologists. Retrieved June 2017 from https://www.hcpc-uk.org/globalassets/resources/standards/standards-of-proficiency---practitioner-psychologists.pdf

Kember, D., Leung, D. Y., Jones, A., Loke, A. Y., Mckay, J., Sinclair, K., … Yeung, E.

(2000). Development of a questionnaire to measure the level of reflective thinking. *Assessment & Evaluation in Higher Education*, *25*(4), 381–395.

Kolb, D. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice-Hall.

Lethbridge, K., Andrusyszyn, M. A., Iwasiw, C., Laschinger, H. K. S., & Fernando, R. (2013). Assessing the psychometric properties of Kember and Leung's Reflection Questionnaire. *Assessment and Evaluation in Higher Education*, *38*(3), 303–325.

Levine, T. C. (2014). The use of blogging in tertiary healthcare educational settings to enhance reflective learning in nursing leadership. *Journal for Nurses in Professional Development*, *30*(6), 287–295.

Lowe, M., Rappolt, S., Jaglal, S., & Macdonald, G. (2007). The role of reflection in implementing learning from continuing education into practice. *Journal of Continuing Education in the Health Professions*, *27*(3), 143–148.

Mamede, S., & Schmidt, H. G. (2004). The structure of reflective in medicine. *Medical Education*, *38*(12), 1302–1308.

Mann, K., Gordon, J., & MacLeod, A. (2009). Reflection and reflective practice in health professions education: A systematic review. *Advances in Health Sciences Education*, *14*(4), 595–621.

National Institute for Health and Care Excellence. (2012). Critical appraisal checklist for a questionnaire study. Retrieved December 2018 from https://www.nice.org.uk/guidance/cg143/evidence/appendix-e-evidence-tables-pdf-186634335

Nguyen, Q. D., Fernandez, N., Karsenti, T., & Charlin, B. (2014). What is reflection? A conceptual analysis of major definitions and a proposal of a five-component model. *Medical Education*, *48*(12), 1176–1189.

Nursing and Midwifery Council. (2010). Standards for competence for registered nurses. Retrieved June 2020 from https://www.nmc.org.uk/globalassets/sitedocuments/standards/nmc-standards-for-competence-for-registered-nurses.pdf

O'Reilly, S. L., & Milner, J. (2015). Transitions in reflective practice: Exploring student development and preferred methods of engagement. *Nutrition & Dietetics*, *72*(2), 150–155.

Pai, H. C. (2015). The effect of a self-reflection and insight program on the nursing competence of nursing students: A longitudinal study. *Journal of Professional Nursing*, *31*(5), 424–431.

Pai, H. C. (2016). An integrated model for the effects of self-reflection and clinical experiential learning on clinical nursing performance in nursing students: A longitudinal study. *Nurse Education Today*, *45*, 156–162.

Peltier, J. W., Hay, A., & Drago, W. (2005). The reflective learning continuum: Reflecting on reflection. *Journal of Marketing Education,* 27(3), 250-263.

Phan, H. P. (2009). Exploring students' reflective thinking practice, deep processing strategies, effort, and achievement goal orientations. *Educational Psychology*, *29*(3), 297–313.

Priddis, L., & Rogers, S. L. (2018). Development of the reflective practice questionnaire: Preliminary findings. *Reflective Practice*, *19*(1), 89–104.

Renner, B., Kimmerle, J., Cavael, D., Ziegler, V., Reinmann, L., & Cress, U. (2014). Web-based apps for reflection: A longitudinal study with hospital staff. *Journal of Medical Internet Research*, *16*(3), 1–9.

Roberts, C., & Stark, P. (2008). Readiness for self-directed change in professional behaviours: Factorial validation of the Self-reflection and Insight Scale. *Medical*

Education, *42*(11), 1054–1063.

Roever, L. (2016). Critical appraisal of a questionnaire study. *Evidence Based Medicine and Practice*, *1*(2), e110.

Ruch, G. (2005). Reflective practice in contemporary child-care social work: The role of containment. *British Journal of Social Work*, *37*(4), 659–680.

Scaife, J. (2010). *Supervising the reflective practitioner*. East Sussex, UK: Routledge.

Schön, D. (1983). *The reflective practitioner*. New York: Basic Books.

Sobral, D. T. (2000). An appraisal of medical students' reflection-in-learning. *Medical Education*, *34*(3), 182–187.

Sobral, D. T. (2001). Medical students' reflection in learning in relation to approaches to study and academic achievement. *Medical Teacher*, *23*(5), 508–513.

Sweet, L., Bass, J., Sidebotham, M., Fenwick, J., & Graham, K. (2019). Developing reflective capacities in midwifery students: Enhancing learning through reflective writing. *Women and Birth,* 32(2), 119-126.

Tricio, J., Woolford, M., & Escudier, M. (2015). Dental students' reflective habits: Is there a relation with their academic achievements? *European Journal of Dental Education*, *19*(2), 113–121.

Tummons, J. (2011). "It sort of feels uncomfortable": Problematising the assessment of reflective practice. *Studies in Higher Education*, *36*(4), 471–483.

List of Tables and Figures

Figures

Tables

Figure 1. Kolb's (1984) Experiential Learning Cycle. Adapted from "Frameworks Supporting RP" in J. Scaife, 2010, *Supervising the Reflective Practitioner*, p.26.

Figure 2. PRISMA flowchart

Identification

Records identified through database searching

(Pubmed=283; PsychInfo=243; CINAHL=252)
(n = 778)

Duplicates removed
(n = 269)

Records after duplicates removed
(n = 509)

Screening

Records screened
(n = 509)

Excluded **(n = 497)** for the following reasons:

- Commentary (n = 14)
- Personal reflection (n = 28)
- Book review (n = 4)
- Not reflective practice (n = 327)
- No reflective practice measure (n = 119)
- Not self-rating scales/questionnaire (n = 3)
- Translation of existing scale (n=2)

Eligibility

Full-text articles assessed for eligibility
(n = 12)

Additional records identified through other sources
(n = 6)

Included

Studies included for review
(n = 18)
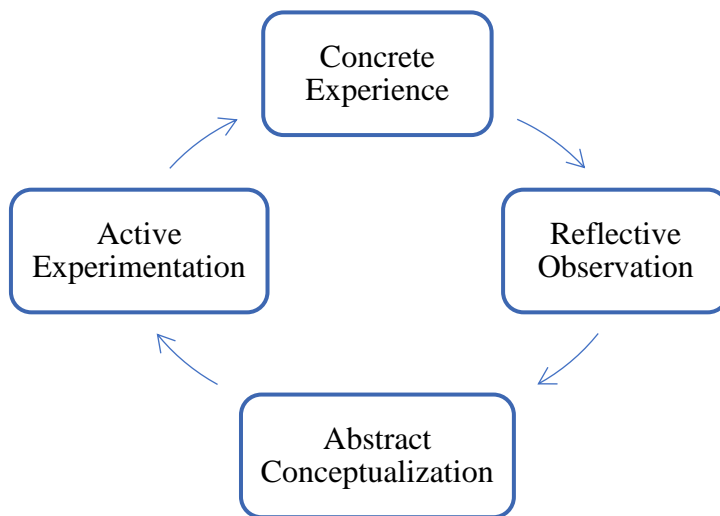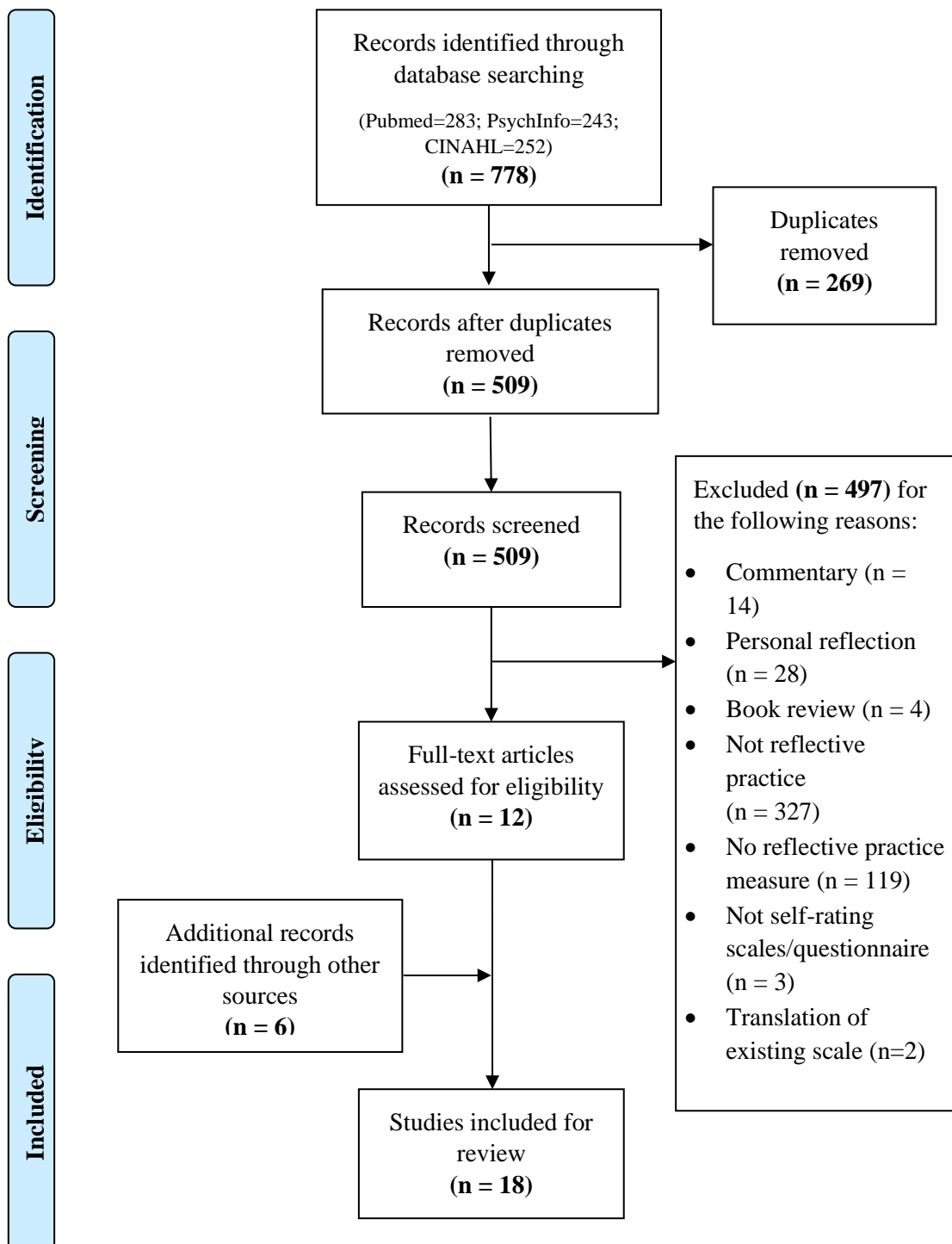
Table 1. Characteristics and Results of Individual Study

	Author(s), Date & Country	Population & Sample Size (N)	Type of Study and Design	Measures used (psychometric properties)	Results of Individual Study
1	Kember et al. (2000), Hong Kong	Under- and postgraduate students from Health Science Faculty (N = 303)	Cross-Sectional Survey	RQ (Acceptable to good internal reliability and acceptable construct validity)	The RQ was developed and validated to measure the level of reflective thinking. A principal use of the RQ is to examine the effects of the teaching and learning environment on reflective thinking. Modification is required if it is intended to be used in various professional practices.
2	Phan (2009), Fiji	Undergraduate educational psychology students (N = 347)	Cross-Sectional Survey	RQ (Good internal reliability)	The RQ was used to test a conceptual model comprising deep processing strategies, effort, mastery and performance-approach goals, reflection, and critical thinking. The evidence suggested that mastery and performance-approach goals, reflection, and critical thinking are the determinants of students' learning and academic achievement.
3	Dunn & Musolino (2011), United States	Occupational and physical therapy graduate students (N = 125)	Online Survey, Cohort Study	RQ (Low to acceptable internal consistency and moderate to good test-retest reliability)	The reliability and responsiveness of RQ and Revised Study Process Questionnaire (RSPQ-2F) were assessed. The stability and responsiveness of both instruments for assessing changes in reflective thinking and learning approaches was supported.
4	Lethbridge, Andrusyszyn, Iwasiw, Laschinger, & Fernando (2013), Canada	Baccalaureate nursing students (N = 538)	Survey, Cohort Study	RQ (Low to good internal consistency and acceptable construct validity)	This study examined the psychometric properties of the RQ. The 'Understanding' and 'Reflection' dimensions were the most commonly used approach among the four-level reflective skills. Reliability and validity of RQ were established.

5	Tricio, Woolford, & Escudier (2015), United Kingdom	Dentistry students (N = 324)	Cross-Sectional online Survey	RQ (Acceptable to good internal consistency)	The study explored the levels of reflection and the relationship between reflection and academic performance. Students with more experience demonstrated higher reflective habits. Most engaged in 'Understanding' and 'Reflection' approaches, and those with high 'Understanding' score tend to have good reflective scores.
6	Grant et al. (2002), Australia	Undergraduate psychology students (N = 260 + 28 + 121)	Cross-Sectional and Cohort Survey	SRIS (Good internal consistency, test-retest reliability, and construct validity)	The SRIS was developed and validated to measure self-reflection and insight. The study found an ambiguous relationship between self-reflection and insight scale, and that journal keeping is not correlated with increased self-reflection and insight. Two types of self-reflection: solution-focused and self-focused were discussed.
7	Lowe, Rappolt, Jaglal, & Macdonald (2007), Canada	Occupational therapists (N = 41 + 33 + 10)	Multimethod Cohort Study	SRIS (Not applicable, cited Grant's, 2002 study)	The study examined the putting into practice reflection learnt from a short course. Two models were generated with the use of the SRIS and Commitment to Change (CTC) statements. Participants were found using reflection pre-, during, and post-course, and this was associated with the course, practice context and the individual factor.
8	Roberts & Stark (2008), United Kingdom	Medical students (N = 1214)	Cross-Sectional Survey Study	SRIS (Good internal consistency and construct validity)	The SRIS was utilised to explore self-reflection and insight in the context of purposeful, self-regulated changes in professional behaviours. Self-reflection was related to the need for positive role models whereas insight was related to the need for reflection or motivation. Attending to feelings was found to be an important, integral aspect of self-reflection and insight.
9	Pai (2015), Taiwan	Nursing students (N = 245)	Correlational Cohort Study	SRIS (Good internal consistency and content validity)	The SRIS, was used to design and evaluate a self-reflection practice programme that incorporated clinical competence, self-reflection, and stress. The self-reflection learning exercise helped improve self-reflection and perceived practice stress that affect clinical competence.

10	Pai (2016), Taiwan	Nursing students (N = 80)	Correlational Cohort Study	SRIS (Good internal reliability and acceptable construct validity)	The SRIS, was used to develop an integrated model exploring the interrelationship among anxiety, self-reflection, and learning effectiveness. The study found that self-reflection with insight and clinical experience are helpful in deflecting anxiety.
11	Sobral (2000), Brazil	Medical students (N = 103)	Survey, Cohort Study	RLS (Good internal consistency, acceptable test-retest reliability and construct validity)	The 10-item version RLS was used to investigate the reflection-in-learning profile of medical students' clinical apprenticeship. The level of reflection-in-learning was significantly correlated with self-perceived competence. The study also reported that greater effort of reflection was associated with more positive learning experience.
12	Sobral (2001), Brazil	Medical students (N = 196)	Survey, Cohort Study	RLS (Good internal consistency, acceptable concurrent validity)	The 14-item version RLS was used to explore the relationship between reflection and study approaches, perceived learning outcome, and academic achievement. Findings suggested that high achievers tend to show stability or positive change in the RLS with stronger personal efficacy in self-reflection. The RLS is a useful tool in appraising the dimensions of learning processing and self-monitoring in students' reflective profile.
13	Aukes et al. (2007), Netherlands	Medical students (N = 1664)	Multimethod Cross-Sectional Study	GRAS (Good internal consistency, acceptable content validity)	The GRAS was developed to measure the personal reflection ability. The scale consists of three aspects of personal reflection: Self-reflection, Empathetic Reflection and Reflective Communication. GRAS can be used in combination with other scales to cover the richness of reflection.
14	Groot et al. (2012), Netherlands	Veterinarians (N = 1290)	Cross-Sectional Survey Study	CRWB (Acceptable internal consistency and construct validity)	The study suggested that Perceived for Lifelong Learning, but not workplace quality, predicts CRWB. Four factors that reflect on the CRWB model are 1) Individual CRWB, 2) CRWB social interaction, 3) cross-checking of information, and 4) openness to new findings.

15	Levine (2014), United States	Nurse managers (N = 11)	Cross-Sectional online Survey	RLIMQ (Cited Peltier, Hay, & Drago's paper, 2005).	The RLIMQ was used to evaluate the effectiveness of blogging in nursing leadership. The blog group and the traditional learning group did not differ significantly on reflective learning dimensions, the mean scores from both groups showed a reflective experience.
16	Renner et al. (2014), Germany	Neurological hospital staff (N = 334)	Survey, Cohort Study	10-item scale (Good internal consistency and acceptable construct validity)	This study examined the effect of software applications (apps) in supporting reflection in hospital staff. The findings showed an increase in collaborative reflection after introduction of the apps. Positive correlation between collaborative reflection and job satisfaction was found.
17	O'Reilly & Milner (2015), Australia	Undergraduate dietetic students (N = 45)	Multimethod Cross-Sectional online Survey	37-item scale (Content validity established)	The study investigated students' experience of different RP activities. Students with more clinical experience preferred more autonomous methods such as e-journaling and engaged in reflection for non-assessment reasons. They also reported fewer barriers and more comfortable engagement in RP.
18	Priddis & Rogers (2018), Australia	General Australian population & mental health professionals (N = 188 & 45)	Cross-Sectional Survey	RPQ (Acceptable to good internal consistency)	The RPQ was developed to measure the experiences, benefits, and potential pitfalls of RP and reflective supervision. RP was not only found to enhance confidence and self-improvement but also increase uncertainty and stress in some individuals. Positive reflective supervision is associated with greater reflection, desire for improvement, and confidence.

List of Appendices

Appendices

Appendix A. Summary Table of Quality Appraisal Measure and Rating

Study	Reference (year)	Scale Used	Research Aim & Study Design		Sampling	
			Was the study aim clearly stated in this study?	Is questionnaire an appropriate study design in this study?	Was the sampling sufficiently large and representative in this study?	Was the sampling approach appropriate in this study?
1	Kember & Leung (2000)	RQ	Y	Y	Y	Y
2	Phan (2009)	RQ	Y	Y	Y	Y
3	Dunn & Musolino (2011)	RQ	Y	Y	Y	Y
4	Lethbridge et al. (2013)	RQ	Y	Y	Y	Y
5	Tricio et al. (2015)	RQ	Y	Y	Y	Y
6	Grant et al. (2002)	SRIS	Y	Y	N	Y
7	Lowe et al. (2007)	SRIS	Y	Y	N	Y
8	Roberts & Stark (2008)	SRIS	Y	Y	Y	Y
9	Pai (2015)	SRIS	Y	Y	Y	Y
10	Pai (2016)	SRIS	Y	Y	N	Y
11	Sobral (2000)	RLS	Y	Y	Y	Y
12	Sobral (2001)	RLS	Y	Y	Y	Y
13	Aukes et al. (2007)	GRAS	Y	Y	Y	Y
14	Groot et al. (2012)	CRWB	Y	Y	Y	Y
15	Levine (2014)	RLIMQ	Y	Y	N	Y
16	Renner et al. (2014)	10-item	Y	Y	Y	Y
17	O'Reilly & Milner (2015)	37-item	Y	Y	Y	Y
18	Priddis & Rogers (2018)	RPQ	Y	Y	N	Y

Appendix A. Summary Table of Quality Appraisal Measure and Ratings (cont'd)

| Study | Format | | | | Piloting |
	Was the title of questionnaire appropriate?	Were instructions for completion adequate in this study?	Were example questions provided in this study?	Were questions clear and easy to understand?	Was the questionnaire adequately piloted/reported? (method, administration, representativeness)
1	Y	Y	Y	Y	Y
2	Y	Y	Y	Y	N/A
3	Y	Y	N	Unclear	N/A
4	Y	Y	Y	Y	Y
5	Y	Y	Y	Y	N/A
6	Y	Y	Y	Y	N
7	Y	Y	N	Y	N/A
8	Y	Y	Y	Y	N/A
9	Y	Y	N	Y	N/A
10	Y	Y	N	Y	N/A
11	Y	Y	Y	Y	N/A
12	Y	Y	Y	Y	N/A
13	Y	Y	Y	Y	Y
14	Y	Y	Y	Y	N
15	Y	Y	Y	Y	N/A
16	Unclear	Y	Y	Y	N
17	Unclear	Y	Y	Y	Y
18	Y	Y	Y	Y	Y

Appendix A. Summary Table of Quality Appraisal Measure and Ratings (cont'd)

Study	Psychometric Properties			Distribution, Administration and Response		
	Was the origin of construct clearly stated in this study?	Have claims for validity been made and justified in this study?	Have claims for reliability been made and justified in this study?	Was the method of distribution and administration reported in this study?	Were response rates reported in this study?	Have any potential response biases been discussed in this study?
1	Y	Y	Y	Y	Y	N
2	Y	N	N	Y	N	N
3	Y	Y	Y	Y	Y	N
4	Y	Y	Y	Y	N	Y
5	Y	Y	Y	Y	Y	Y
6	Y	Y	Y	Y	N	Y
7	N	Y	Y	Y	Y	N
8	Y	Y	Y	Y	Y	Y
9	Y	Y	Y	Y	Y	Y
10	Y	Y	Y	Y	Y	N
11	Y	Y	Y	Y	Y	N
12	Y	Y	Y	Y	Y	N
13	Y	Y	Y	Y	N	N
14	Y	Y	Y	Y	Y	Y
15	Y	N	Y	Y	N	Y
16	Y	Y	Y	Y	Y	N
17	Y	N	N	Y	Y	Y
18	Y	N	Y	Y	N	Y

Appendix A. Summary Table of Quality Appraisal Measure and Ratings (cont'd)

Study	Analysis			Discussion & Conclusion			
	Was the type of analysis appropriate in this study?	Were both significant and non-significant results reported in this study?	Were qualitative results been adequately interpreted and justified in this study?	Was appropriate link between the data and conclusion drawn in this study?	Are recommendations justified in this study?	Can the questionnaire be used for healthcare professionals?	Were conflicts of interests declared in this study?
1	Y	Unclear	N	Y	Y	Y	N
2	Y	Y	N/A	Y	Y	Y	N
3	Y	Y	N/A	Y	Y	Y	N
4	Y	Y	N/A	Y	Y	Y	N
5	Y	Y	Unclear	Y	Y	Y	N
6	Y	Y	N/A	Y	Y	Y	N
7	Unclear	N	Y	Y	Y	Y	N
8	Y	Y	N/A	Y	Y	Y	Y
9	Y	Y	N/A	Y	Y	Y	Y
10	Y	Y	N/A	Y	Y	Y	N
11	Y	Y	N/A	Y	Y	Y	N
12	Y	Y	N/A	Y	Y	Y	N
13	Y	Y	Unclear	Y	Y	Y	N
14	Y	Y	N/A	Y	Y	Y	N
15	Y	Y	N/A	Unclear	Y	Y	N
16	Y	Y	N/A	Y	Y	Y	Y
17	Y	Y	Y	Y	Y	Y	Y
18	Y	Y	N/A	Y	Y	Y	Y

Appendix A. Summary Table of Quality Appraisal Measure and Ratings (cont'd)

| Study | Ethics | | Quality Rating [a] | Overall Quality [b] | First Reviewer | Checker |
	Was the ethical approval stated in this study?	Is the role of researcher clearly described in this study?				
1	N	N	75%	++ [c]	S.M.	
2	N	N/A	71%	+ [d]	S.M.	S.C.
3	Y	N	77%	++	S.M.	
4	Y	N	87%	++	S.M.	
5	Y	N	87%	++	S.M.	
6	N	N	74%	+	S.M.	
7	N	N	61%	+	S.M.	S.C.
8	Y	Y	100%	++	S.M.	
9	Y	N/A	95%	++	S.M.	
10	Y	N/A	81%	++	S.M.	
11	N	N/A	86%	++	S.M.	
12	N	N/A	86%	++	S.M.	S.C.
13	N	N	75%	++	S.M.	
14	N	N	83%	++	S.M.	
15	Y	N/A	76%	++	S.M.	S.C.
16	Y	Y	87%	++	S.M.	
17	Y	Y	88%	++	S.M.	
18	N	N	78%	++	S.M.	

[a] Quality rating is calculated by dividing the total number of Y by the total number of checklist items (e.g., 24; excluding the number of N/As)

[b] Overall quality is derived from the percentage rating: ≥75% = High Quality (++); ≥50% and <75% = Acceptable (+); ≥25% and <50% Low Quality (-); <25% = Reject (o), with written permission from L. Roever.

[c] ++ = Majority of criteria met, little or no risk of bias. [d] + = Most criteria met, some flaws in the study with an associated risk of bias.

Appendix B. Strengths and Limitations of the Questionnaire Studies

Instruments	Strengths	Limitations
RQ	• Simple and user-friendly • Acceptable to good psychometric properties • Can be adapted to suit different study contexts or professional practices	• Homogeneous sample. • Convenience or purposive sampling that may introduce self-selection bias
SRIS	• Simple and user-friendly • Good psychometric properties • Can be adapted to suit different study contexts or professional practices • Measure engagement in reflection and insight	• Homogeneous sample • Convenience or purposive sampling
RLS	• Simple and user-friendly • Acceptable to good psychometric properties • Can be adapted to measure reflection in learning for different disciplines	• Homogeneous sample • Purposive sampling
GRAS	• Simple and user-friendly • Acceptable to good psychometric properties • Can be adapted to measure personal reflection for different study contexts or professional practices	• Homogeneous sample • Convenience or purposive sampling
CRWB	• Simple and user-friendly • Acceptable reliability • Generalisability to other healthcare professionals	• Homogeneous sample • Convenience or purposive sampling

RLIMQ	• Simple and user-friendly • Generalisability to other healthcare professionals • Other: Measure reflection and interaction between learner and instructor	• Convenience sampling • Homogeneous sample • Unclear psychometric properties
10-item scale	• Simple and user-friendly • Acceptable to good psychometric properties • Generalisability to other healthcare professionals • Heterogeneous sample	• Convenience or purposive sampling • Other: Questionnaire title not available
37-item scale	• Generalisability to other healthcare professionals • Other: Mixed method scale (i.e. Likert rating scale and free text input)	• Unclear user-friendliness • Not fully validated • Homogeneous sample • Convenience or purposive sampling
RPQ	• Simple and user-friendly • Acceptable to good internal consistency • Generalisability to other healthcare professionals • Heterogeneous sample • Other: Tapped on various aspects of reflective practice including appraisal of supervision measure	• Not fully validated • Convenience sampling

Note. RQ = Reflection Questionnaire; SRIS = Self-Reflection and Insight Scale; RLS = Reflective-in-Learning Scale; GRAS = Groningen Reflection Ability Scale; CRWB = Critically Reflective Work Behaviour; RLIMQ = Reflective Learning and Interaction Model Questionnaire; 10-item scale (as cited in Renner et al., 2014); 37-item scale (as cited in O'Reilly & Milner, 2015); and RPQ = Reflective Practice Questionnaire.