

ORIGINAL RESEARCH

Reliability of the Reflective Essay Marking Scale (REMS): a scale for marking of students' reflective essays

Craig Chigwedere^{1*} , Brian Fitzmaurice¹ and Richard Thwaites² 

¹Department of Psychiatry, School of Medicine, Trinity College Dublin, University of Dublin, Dublin, Ireland and ²CNTW, National Health Service, UK

*Corresponding author. Email: cchigwedere@stpatsmail.com

(Received 1 June 2022; revised 8 December 2022; accepted 11 December 2022)

Abstract

Personal practice (PP) is widely practised and a requirement across major psychology and psychotherapy organizations and modalities. However, one of the challenges for training institutions is how to assess the quality of such PP. The Reflective Essay Marking Scale (REMS) was developed to improve standardization of marking reflective essays in cognitive behavioural psychotherapy (CBT) training. A small sample of 16 expert CBT participants recruited by email used the REMS to rate two mock reflective essays in a within-subjects design. The internal consistency of REMS was acceptable (Cronbach's $\alpha = .73$) with excellent inter-rater reliability. Across the raters, it sufficiently differentiated quality ($t_{12} = 4.91$; $p < .0001$). Although these are the results of a preliminary and very small study with a small sample using mock essays, the REMS may be a useful scale, allowing CBT courses to account for students' reflective work in a standardized way. A larger validation study is required in the future.

Key learning aims

- (1) To improve the thinking about what raters should focus on when rating the reflective essays of trainee therapists.
- (2) To describe the development of the scale and how its reliability was tested.
- (3) To improve the transparency and objectivity in assessing and rating reflective practice.

Keywords: CBT; essays; marking; rating scale; reflection

Introduction

Historically, therapist self-development through personal therapy has played an important role in psychotherapy training and therapist development. Bennett-Levy and Finlay-Jones, (2018) proposed the term personal practice (PP), defined as, 'formal psychological interventions and techniques that therapists engage with self-experientially over an extended period of time (weeks, months or years) as individuals or groups, with a reflective focus on their personal and/or professional development'. PP encompasses a range of practices that therapists engage with in their professional development, including reflective, meditative and personal therapy practices. However, therapists experiencing their own personal therapy is the most dominant form of PP (Geller *et al.*, 2005).

Internationally, major psychology and psychotherapy organizations acknowledge the need for therapist PP (American Psychological Association, 2012; British Psychological Society, 2016; European Certificate of Psychotherapy, 2017). For reasons such as theoretical standpoint, equivocal empirical support and acceptability, cognitive behavioural psychotherapy (CBT) has historically not prioritized formal PP (Geller *et al.*, 2005). However, Bennett-Levy and colleagues (Bennett-Levy, 2006; Bennett-Levy *et al.*, 2009) proposed self-practice/self-reflection (SP/SR) as an aid to therapist development through self-practising therapy techniques in CBT. Although both are forms of PP, SP/SR differs from personal therapy, which usually involves therapy with the aid of a therapist. In SP/SR, therapists engage in the self-application (self-practice) of techniques they apply with patients and reflect (self-reflection) on them alone, in pairs or in groups, thereby creating a deeper sense of knowing of CBT (Bennett-Levy *et al.*, 2001). SP/SR manuals have now been developed for use in CBT (Bennett-Levy *et al.*, 2015), compassion-focused therapy (Kolts *et al.*, 2018) and schema therapy (Farrell and Shaw, 2018). Furthermore, research evidence supports SP/SR's utility as a method of therapist development with a range of professional and personal effects (Chigwedere, 2019; Gale & Schröder, 2014; Pakenham, 2015; Scott *et al.*, 2020). Bennett-Levy (2006) has proposed the explanatory declarative-procedural-reflective (DPR) model, which was later elaborated upon in the personal practice model (PPM) (Bennett-Levy and Finlay-Jones, 2018). The PPM proposes that PPs may differentially impact a therapist's professional or therapist-self (TS) and private or personal-self (PS), with reflection as the bridge which transfers learning between the two (Chigwedere *et al.*, 2021). The PPM may provide a guiding theoretical framework for choosing a PP to meet desired therapist training and development goals, including in CBT.

SP/SR is now practised widely as an important component of CBT training across different countries including the United Kingdom, Ireland, Sweden, Australia and New Zealand (Haarhoff *et al.*, 2015). A challenge for CBT training institutions promoting PP may be how to assess its quality. To date, two SP/SR measures have been validated, the Self-focused Practice Questionnaire (SfPQ; Chigwedere *et al.*, 2017) and the Self-Reflective Writing Scale (SRWS; So *et al.*, 2018). The SfPQ is a self-report measure of the overall self-perceived impact of SP/SR on the personal- and therapist-self domains. It is a measure of change in perception with good inter-rater and internal consistency reliability (Chigwedere *et al.*, 2017). However, it is not a measure of the quality of personal practice *per se*. The SRWS is an assessor-rated measure of quality of reflective summaries after a self-practice exercise, with acceptable inter-rater reliability and validity (So *et al.*, 2018). Both the SfPQ and SRWS are useful and important, but some CBT training institutions (e.g. those at Trinity College Dublin and University College Cork in Ireland), assess trainees through the writing of reflective essays at the end of an SP/SR module. The SRWS, which is an excellent measure of summaries of specific exercises, may not rate essays describing overall learning from SP/SR or other forms of PP. The Reflective Essay Marking Scale (REMS) is intended as a measure of learning from reflections on any form of PP. It is worth noting that guidelines exist for the completion of case reports (British Association of Behavioural and Cognitive Psychotherapy, 2021), but not for the completion of the type of reflections discussed in the current paper. The subjective and personal nature of such reflective writing and essays makes assessment of their quality a challenge, risking unacceptable rater bias. For this reason, we developed the REMS.

This paper reports the preliminary results of a small validation study to assess the REMS's internal consistency, item correlations and discrimination of a good from a poorer quality essay. Our main goal was to evaluate whether the REMS could discriminate essay quality across a range of raters. Our hypothesis, based on a previous trial usage of the REMS, was that a range of REMS-naïve raters would be able to discriminate two reflective CBT essays on quality, even with minor qualitative differences.

Method

Sixteen cognitive behavioural therapy trained participants, all with previous experience of using SP/SR (9 males and 7 females; $\chi^2 = .250$; d.f. = 1; $p = .62$) were recruited by email. Initial recruitment was from attendees to an annual skills workshop conducted by the Trinity College Dublin (TCD) CBT department but most participants declined to participate. The main reason for declining to participate was being unqualified to rate an academic essay. For this reason, purposive and snowball sampling approaches were adopted, first targeting those on the original list who consented, then asking them if they would inform known colleagues who had practised SP/SR or personal therapy, and were likely to be interested. Potential participants could contact the lead researcher (C.C.), who would then email them the formal invitation letter, study information sheet and consent form as necessary.

Of the 16 participants, 11 (68.75%) had practised SP/SR, two (12.5%) general self-reflection and three (18.75%) personal therapy with a therapist. They ranged in years since qualification and in active practice from 1 to 26. Participant primary practice modalities were CBT ($n = 13$), cognitive analytic therapy ($n = 1$) integrative ($n = 1$) and one did not disclose a practice modality. Six participants identified at least one secondary modality including acceptance and commitment therapy, eye movement desensitization reprocessing, emotion focused therapy, human givens and mindfulness, with two identifying multiple modalities. Nine (56.25%) of the participants described their professional background as mental health nursing, five (31.25%) as psychology, one (6.25%) as psychiatry and one as social care. All participants returned scores but three were unusable because there were no scores visible on Survey Monkey.

Measure

Measure development

REMS was developed for the purpose of rating reflective essays, based on known theories of self-practice/self-reflection. The items were developed by experts in SP/SR, initially constructed by the lead researcher (C.C.), and then discussed with R.T. and B.F. Initially the scale was a 5-item scale with the fifth item titled 'self-as-learner'. The self-as-learner item assessed how well a participant had provided evidence of application of reflection in their own life and how this would influence learning in the future. The scale was sent to two further experts: James Bennett-Levy, the developer of SP/SR and Suzanne Ho-Wai So, for further review. It was considered that the focus of the fifth item was sufficiently covered by the remaining four items, and also, that it risked predicting the future. As such, it was dropped. The four remaining items were retained, because they encourage participants to write about experiences already reflected upon. The REMS was trialled by two raters [R.T. and Yvonne Tone (Y.T.)] on 20 real trainee assignments. They independently rated the same essays, then met with C.C. and B.F. to review the scale's applicability, resulting in a change to some of the scale's wording.

The final REMS is a brief, 4-item scale (see Table 1 for item description), rating reflective writing essays. Each item has a descriptor and is rated on a 6-point Likert-style scale, ranging from 0 (absence of feature, or highly in appropriate reflection) to 6 (excellent reflection on the feature). The items are as follows:

- (1) *Personal-self*. Does the reflective writing demonstrate evidence of learning about the self, including development of self-awareness, links to early developmental experiences, and experience of personal change due to the SP/SR practice?
- (2) *Therapist-self*. Does the writing demonstrate evidence of impact of SP/SR on factors such as knowledge of concepts and procedures, use of specific skills, empathy and other interpersonal skills?

- (3) *Evidence of bridging.* Does the writing demonstrate evidence of the integration of learning from reflection on the personal and therapist-self, as well as how SP/SR experience links to clinical and personal practice?
- (4) *Understanding of reflective process.* Does the writing show awareness, understanding and application of SP/SR theories and suggested best practices, as well as how they inform the writer's own reflective practice?

Procedure

Participants were recruited by email from the list of those consenting to be invited to participate in research projects as per GDPR guidelines. On receipt of the consent form, a Survey Monkey link was forwarded to the consenting participant. The link opened the study page, containing study instructions, demographic questionnaire, a description of the REMS, and the two essays, each with a rating scale. With adaptations, the essays were brief, 600-word mock composites based on ones rated highly (i.e. good) or low (i.e. poor) by R.T and Y.T. in the trial described in the 'Measure development' section above. Ethics approval for the study was granted by the TCD School of Medicine Ethics Committee.

Data analysis

Data analyses were conducted using SPSS 24 (SPSS, Inc., Chicago, IL, USA). All available data were used for the analysis with a listwise approach being applied. The level of missing data was acceptable (18.75%).

First, we wanted to assess the factor structure of REMS. We performed a principal components analysis (PCA) with a Promax Rotation and maximum 25 iterations for convergence. Rotation converged in four iterations. This was based on an acceptable but low Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (MSA = .50) [$>.50$ (miserable), $>.60$ (mediocre), $>.70$ (middling), $>.80$ (meritorious) and $>.9$ (marvelous)] (Kaiser and Rice, 1974) and a significant Bartlett's test of sphericity ($\chi^2 = 41.222$; d.f. = 6; $p < .0001$) for the good quality essay. Item coefficients were significant if $r > .4$ (Field, 2005; Stevens, 1992). Component choice was based on the Kaiser principle (i.e. an eigen value $>.1$; Kaiser, 1960; Kaiser, 1970).

Secondly, we wanted to examine the REMS's internal consistency reliability. This was based on an overall Cronbach's alpha (α ; Cronbach, 1951), as well as alpha if item deleted with α ranges of $<.5$ (unacceptable), $.5-.6$ (poor), $.6-.7$ (questionable), $.7-.8$ (acceptable), $.8-.9$ (good) and $>.9$ (excellent) (George and Mallery, 2003).

A third goal was to assess the scale's inter-rater reliability. We calculated the inter-item correlation coefficient (ICC) using a two-way mixed model and testing for absolute agreement 'type'. ICC ranges were based on $<.5$ (poor), $.5-.75$ (moderate), $.75-.90$ (good) and $>.9$ (excellent) (Koo and Li, 2016).

A fourth goal was to assess item correlations. Pearson's correlation coefficient (r) was calculated to test the correlations amongst the items, with ranges $\pm.70$ (strong), $\pm.50$ (moderate), $\pm.30$ (weak) and 0 (no relationship) (Ramsey, 2011).

Lastly, although the study used mock essays, these were based on composites of essays by real students. We therefore considered it a useful task to assess the REMS's utility for discriminating essay quality. To do this, paired-samples t -tests were performed on the total scores of the two essays.

Results

Component loadings

The items loaded onto three components (Table 1), with the Therapist-self and Evidence of Bridging forming a single component (Professional Development) with an eigenvalue of 2.337, explaining

Table 1. Component loadings of the items of the REMS

Item	Component		
	1	2	3
1. Personal-self		.998	
2. Therapist-self	.984		
3. Evidence of bridging	.919		
4. Understanding of reflective process			.990

Table 2. Internal consistency of the items of the REMS

	Scale mean if item deleted	Scale variance if item deleted	Cronbach's alpha if item deleted
Item 1	12.38	7.256	.81
Item 2	13.31	5.564	.62
Item 3	13.31	4.40	.50
Item 4	13.15	4.64	.70

$n = 13$; item 1, Personal-self; item 2, Therapist-self; item 3, evidence of bridging; item 4, understanding of reflective process.

Table 3. Inter-item correlations

Essay/item	1	2	3	4	5	6	7	8
1. 1/Personal-self	1							
2. 1/Therapist-self	.000	1						
3. 1/Evidence of bridging	.393	.831**	1					
4. 1/URP	.165	.530	.505	1				
5. 2/Personal-self	.191	.104	.162	.155	1			
6. 2/Therapist-self	.108	-.388	-.196	-.030	.627*	1		
7. 2/Evidence of bridging	.000	.000	.105	-.177	.766**	.777**	1	
8. 2/URP	.000	-.355	-.199	.022	.713**	.939**	.781**	1

** $p < .01$; * $p < .05$; URP, understanding of reflective process.

58.42% of the variance, Personal-self, with an eigenvalue of 1.01, explaining 25.241% of the variance and the Understating of Reflective Process component with a low eigen value of .575, explaining 14.36% of the variance. Component correlations were .206 (components 1 and 2), .145 (components 2 and 3) and .525 (components 1 and 3). For theoretical reasons (see Discussion) we maintained the four separate items, without creating a 3-factor scale.

Internal consistency and inter-rater reliability

Table 2 shows that the four items of the scale had acceptable internal consistency reliability (Cronbach's $\alpha = .73$). However, removing the first item (Personal-self) would improve the scale's internal consistency from acceptable to good ($\alpha = .81$). The scale's inter-rater reliability was excellent (ICC = .93; 95% CI = .84–.98).

Item correlations

Correlation differences were observed between the two essays (Table 3). Essay 1 (the good-quality essay) significant correlations were only observed between the Therapist-self and Evidence of bridging items ($r = .83$, $p < .0001$).

However, with Essay 2 (the poor-quality essay) all items were significantly correlated. No significant correlations were observed between the two essays. This suggests that the REMS

could discriminate correlated items from those that were not, when the essay was of good quality but not when it was not.

Differences in quality between the two essays

To determine the REMS's discrimination of a good- and poor-quality essay, paired samples *t*-tests were conducted using the items and total REMS scores. Differences were observed between the total scores of the essays, such that the score for the good (mean = 17.62; *SD* = 2.90) was significantly higher than that of the poor (mean = 8.54; *SD* = 5.62) essay (mean difference = 9.08, *SD* = 6.66; 95% *CI* = 5.05–13.10; $t_{12} = 4.91$, $p < .0001$). Significant item level differences were observed for all items, suggesting that the REMS may sufficiently discriminate essays by quality.

Discussion

We aimed to evaluate the likelihood that REMS-naïve expert therapists would be able to use the scale to rate the quality of two reflective CBT essays. We hypothesized that using the REMS, it would be possible to discriminate between a good quality and a poorer quality essay. As such, we wrote two short mock reflective CBT essays, which were independently reviewed for quality by 13 expert CBT therapists with previous experience of PP. Results suggested that the four items of the REMS may fall into three factors, rating Professional development (Therapist-self and Evidence of bridging), Personal-self and Understanding of reflective process. Acceptable internal consistency and excellent inter-rater reliability were observed. Item correlations were different between the good and poor quality essays and the total scores were significantly different between the two essays.

The REMS may be a reliable scale for rating CBT reflective essays. Although based on a small sample, the REMS exhibited acceptable internal consistency reliability, within three possible factors. Furthermore, it may support the theory proposed in the DPR (Bennett-Levy, 2006) and PPM (Bennett-Levy and Finlay-Jones, 2018) models. The models propose that the Therapist-self develops from professional learning, while the Personal-self exists prior to therapy training. In SP/SR, existing declarative and procedural knowledge is personally experienced and reflected upon, thereby informing the Therapist-self. This may explain the findings of the current study. SP/SR was developed as a therapist development tool, although it also impacts the personal-self. As such, item correlations suggest a strong association between Therapist-self and Evidence of bridging, which may represent a single factor of Professional development through personal practice. However, that relationship may represent the influence of reflection on the therapist-self in CBT SP/SR, through bridging personal experience into professional knowledge. Understanding of Reflective Process and Therapist-self showed a trend towards significance, perhaps suggesting the importance of knowledge of reflective theory in CBT SP/SR. As such, training institutions may need to improve teaching of reflective skills and theory. As such, the REMS may help to address academic conventions beyond quality of reflective essays. For example, by having a scale that is theory congruent, trainees and trainers may be clearer about learning outcomes and expectations, making PP more systematic and marking more efficient. By being consistent with the theory, learning reflection as a skill might be enhanced, thereby improving gains for therapists. The REMS may improve the use and rating of the quality of reflective learning beyond learning as exemplified by Chigwedere (2019).

The Personal-self item, which is based on knowledge and learning developed prior to the existence of the professional-self, was not significantly correlated with the other items. However, Bennett-Levy and Finlay-Jones (2018) propose that there is a key distinction

between the Personal-self and Therapist-self and that reflection bridges learning from personal to the professional and vice versa. Without systematic reflection, SP/SR as PP may not sufficiently influence the private-self, while personal therapy may not sufficiently influence the professional-self (Bennett-Levy and Finlay Jones, 2018). The current findings may support the proposed SP/SR theories and models, pointing to SP/SR in CBT as primarily a therapist development approach. However, these results raise the interesting question of whether reflective essays from personal therapy participants may result in a reverse of current findings (i.e. significant correlations between Personal-self and understanding of Reflective process).

Interestingly, all items for the poor-quality essay were significantly correlated, possibly because all the items were scored low and were therefore poorly discriminated. This hypothesis is supported by significant findings of the paired samples *t*-test, which showed that with the REMS, it was possible to discriminate the good from the poor essay.

Limitations

Although practice of SP/SR is growing, it remains confined to a relatively small number of countries, training and clinical institutions and therapists. As such, although we tried to recruit a larger sample of raters, participant recruitment was challenging, resulting in a small sample size. The KMO measure of sampling adequacy was only just acceptable. As such, small effects may have been missed. Therefore, the results should be interpreted with caution. Due to the small *N*, we considered not publishing the results. However, they seem theoretically sound and important as a preliminary step in the validation of the REMS. Larger sample studies, preferably with real reflective essays from trainees, will be needed in future.

It is notable that the study used vignettes, simulating reflective essays. Although the vignettes were based on real essays, they may not be entirely representative of the reflections of real students post-SP/SR. Furthermore, we chose to rate only two essays, using a number of raters rather than more essays with fewer raters. Although unlikely, the current results may be a function of the way the essays were written, rather than the effectiveness of the REMS as a marking tool. With a larger sample of essays, the REMS may not effectively distinguish different levels of quality. As such, in future, it will be important to assess a range of essays with fewer raters.

Although most participants were senior therapists with known experience in CBT training in academic institutions, recruited using a snowball sampling approach from UK, Ireland, New Zealand and Australia, there were some who were not experienced trainers. Such participants may not have had experience of rating academic writing. However, we aimed to test the ease of use of the REMS and had hoped to recruit a larger sample than we managed to get in the end. In future, it may be important to repeat this study with a greater number of essays, rated by fewer, experienced raters.

Conclusion

The REMS may be a valid and reliable tool for marking reflective CBT essays. One of the challenges for trainers may be the motivation of trainees engaging in PP. Having a rating scale for assessing reflective essays at the end of training may motivate trainees to engage with, and prioritize specified aspects of PP. Without such a tool, trainees may not only fail to understand how PP may be important to their coursework, but what to attend to, and prioritize in their learning. As such, the REMS may offer a framework for both

trainers and trainees, which may help to standardize practice in support of the SP/SR and CBT theory.

Key practice points

- (1) Due to the subjectivity of reflective practice, it is important to rate reflective essays for assessment, using a validated tool that is reliable.
- (2) The REMS may be a reliable tool for rating reflective essays.
- (3) It may be challenging to recruit a sufficiently large sample size to reliability test a marking scale for reflection.

Further reading

Bennett-Levy, J., & Finlay-Jones, A. (2018). The role of personal practice in therapist skill development: a model to guide therapists, educators, supervisors and researchers. *Cognitive Behaviour Therapy*, *47*, 185–205.

So, H.-W.S., Bennett-Levy, J., Perry, H., Wood, D. H., & Wong, C. (2018) The Self-Reflective Writing Scale: a new measure to assess self-reflection following self-experiential cognitive behaviour therapy training. *Reflective Practice*, *19*, 505 – 521. <https://doi.org/10.1080/14623943.2018.1536652>

Data availability statement. Data are available on request from the authors, and the REMS is available from the lead author.

Acknowledgements. We thank Professor James Bennett-Levy, Professor Suzanne Ho-Wai So and Mrs Yvonne Tone.

Author contributions. **Craig Chigwedere:** Conceptualization (equal), Data curation (lead), Formal analysis (lead), Methodology (lead), Project administration (lead), Writing – original draft (lead), Writing – review & editing (equal); **Brian Fitzmaurice:** Conceptualization (equal), Formal analysis (supporting), Methodology (supporting), Visualization (supporting), Writing – original draft (supporting), Writing – review & editing (equal); **Richard Thwaites:** Conceptualization (equal), Visualization (supporting), Writing – original draft (supporting), Writing – review & editing (equal).

Financial support. None.

Conflicts of interest. Craig Chigwedere, Brian Fitzmaurice and Richard Thwaites were involved in the development of the REMS. Richard Thwaites is Editor of *the Cognitive Behaviour Therapist*. He was not involved in the review or editorial process for this paper, on which he is listed as an author.

Ethical standards. The authors have abided by the Ethical Principles of Psychologists and Code of Conduct as set out by the BACP and BPS. Ethical approval was granted by the Trinity College Dublin School of Medicine Research Ethics Committee (Application Number: 20190403).

References

American Psychological Association (2012). Revised competency benchmarks in professional psychology. Available at: <http://www.apa.org/ed/graduate/benchmarks-evaluation-system.aspx> (accessed 25 May 2017).

Bennett-Levy, J. (2006). Therapist skills: a cognitive model of their acquisition and refinement. *Behavioural and Cognitive Psychotherapy*, *34*, 57–78. <https://doi.org/10.1017/S1352465805002420>

Bennett-Levy, J., Thwaites, R., Haarhoff, B., & Perry, H. (2015). *Experiencing CBT from the Inside Out. A Self-Practice/Self-Reflection Workbook for Therapists*. New York: Guilford.

Bennett-Levy, J., & Finlay-Jones, A. (2018). The role of personal practice in therapist skill development: a model to guide therapists, educators, supervisors and researchers. *Cognitive Behaviour Therapy*, *47*, 185–205. <https://doi.org/10.1080/16506073.2018.1434678>

Bennett-Levy, J., McManus, F., Westling, B., & Fennell, M. (2009). Acquiring and refining CBT skills and competencies: which training methods are perceived to be most effective? *Behavioural and Cognitive Psychotherapy*, *37*, 571–583. <https://doi.org/10.1017/S1352465809990270>

Bennett-Levy, J., Turner, F., Beaty, T., Smith, M., Paterson, B., & Farmer, S. (2001). The value of self-practice of cognitive therapy techniques and self-reflection in the training of cognitive therapists. *Behavioural and Cognitive Psychotherapy*, *29*, 203–220.

- British Association of Behavioural and Cognitive Psychotherapy** (2021). Case study marking criteria. Available at: <https://babcp.com/case-study-marking-criteria> (accessed 8 December 2022).
- British Psychological Society** (2016). Standards for the accreditation of Doctoral programmes in clinical psychology. Available at: <http://www.bps.org.uk/careers-education-training/accredited-courses-training-programmes/useful-accreditation-documents/clinical-psychology/clinical-psychology>
- Chigwedere, C.** (2019). Writing the 'self' into self-practice/self-reflection (SP/SR) in CBT: learning from autoethnography. *the Cognitive Behaviour Therapist*, 12, e35. <https://doi.org/10.1017/s1754470x19000217>
- Chigwedere, C., Bennett, Levy, J., Fitzmaurice, B., & Donohoe, G.** (2021). Personal practice in counselling and CBT trainees: the self-perceived impact of personal therapy and self-practice/self-reflection on personal and professional development. *Cognitive Behaviour Therapy*, 50, 422–438. <https://doi.org/10.1080/16506073.2020.1846608>
- Chigwedere, C., Fitzmaurice, B., & Donohoe, G.** (2017). The Self-focused Practice Questionnaire (SfPQ): preliminary psychometric properties of a measure of therapist self-focused practice. *Behavioural and Cognitive Psychotherapy*, 45, 497–509. <https://doi.org/10.1017/>
- Cronbach, L. J.** (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334.
- European Certificate of Psychotherapy** (2017). Available at: https://www.europsyche.org/app/uploads/2019/05/ECP-document-version-7-0-voted-AGM-Vienna-Feb-2017_offic.pdf
- Farrell, J. M., & Shaw, I. A.** (2018). *Experiencing Schema Therapy from the Inside Out: A Self-Practice/Self-Reflection Workbook for Therapists*. Guilford.
- Field, A.** (2005). *Discovering Statistics Using SPSS* (3rd edn). London: Sage.
- Gale, C., & Schröder, T.** (2014). Experiences of self-practice/self-reflection in cognitive behavioural therapy: a meta-synthesis of qualitative studies. *Psychology and Psychotherapy: Theory, Research & Practice*, 87, 373–392. doi: 10.1111/papt.12026
- Geller, J. D., Norcross, J. C., & Orlinsky, D. E.** (eds) (2005). *The Psychotherapist's Own Psychotherapy: Patient and Clinician Perspectives*. Oxford University Press.
- George, D., & Mallery, P.** (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference. 11.0 Update* (4th edn). Boston: Allyn & Bacon.
- Haarhoff, B., Thwaites, R., & Bennett-Levy, J.** (2015) Engagement with self-practice/self-reflection as a professional activity: the role of therapist beliefs. *Australian Psychologist*, 50, 322–328.
- Kaiser, H. F.** (1960). The application of electronic computers to factor analysis. *Educational Psychological measurement*, 20, 141–151.
- Kaiser, H. F.** (1970). A second generation little jiffy. *Psychometrika*, 35, 401–415.
- Kaiser, H.F., & Rice, J.** (1974) Little jiffy mark IV. *Educational and Psychological Measurement*, 34, 111–117.
- Kolts, R., Bell, T., Bennett-Levy, J., & Irons, C.** (2018). *Experiencing Compassion-Focused Therapy from the Inside Out: A Self-Practice/Self-Reflection Workbook for Therapists*. Guilford Press.
- Koo, T. K., & Li, M. Y.** (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15, 155–163.
- Pakenham, K. I.** (2015). Investigation of the utility of the acceptance and commitment therapy (ACT) framework for fostering self-care in clinical psychology trainees. *Training and Education in Professional Psychology*, 9, 144–152. <https://doi.org/10.1037/tep0000074>
- Ramsey, D. J.** (2011). *Statistics for Dummies* (2nd edn). Hoboken, NJ: Wiley Publishing.
- Scott, J., Yap, K., Bunch, K., Haarhoff, B., Perry, H., & Bennett-Levy, J.** (2020). Should personal practice be part of cognitive behaviour therapy training? Results from two self-practice/self-reflection cohort control pilot studies. *Clinical Psychology & Psychotherapy*. <https://doi.org/10.1002/cpp.2497>
- So, H.-W.S., Bennett-Levy, J., Perry, H., Wood, D. H., & Wong, C.** (2018) The Self-Reflective writing scale: a new measure to assess self-reflection following self-experiential cognitive behaviour therapy training. *Reflective Practice*, 19, 505–521. <https://doi.org/10.1080/14623943.2018.1536652>
- Stevens, J. P.** (1992). *Applied Multivariate Statistics for the Social Sciences* (2nd edn). Hillsdale, NJ: Erlbaum.

Cite this article: Chigwedere C, Fitzmaurice B, and Thwaites R. Reliability of the Reflective Essay Marking Scale (REMS): a scale for marking of students' reflective essays. *The Cognitive Behaviour Therapist*. <https://doi.org/10.1017/S1754470X22000617>