The Inter-Rater Consistency of Clinician Ratings of Posttraumatic Stress Disorder (PTSD) Therapy Content

Manoj Thogesan[1], David Berle[1,2], Dominic Hilbrink[3], Rachael Kiely[3], Clare Russell-Williams[3], Natasha Garwood[3] & Zachary Steel[1,3,4]

1. School of Psychiatry, UNSW Sydney, NSW, Australia.
2. Discipline of Clinical Psychology, University of Technology Sydney, NSW, Australia.
3. St John of God Health Care, Richmond Hospital, NSW, Australia.
4. Black Dog Institute, Randwick, NSW, Australia.

Address for correspondence:

David Berle
Discipline of Clinical Psychology
University of Technology Sydney
PO Box 123, Broadway NSW 2007
Australia
E: David.Berle@uts.edu.au
Ph: +61 2 9514 4278
(ORCID: 0000-0002-4861-2220)

Acknowledgements: Nil.
Conflicts of Interest: The authors declare that they have no conflicts of interest to declare.
Financial support: Preparation of this paper was supported by a National Health and Medical Research Council (NHMRC) Early Career Fellowship (GNT1122203) awarded to the second author. The NHMRC had no role in the design of the study, data collection, data analysis, interpretation of data, or in the writing or revision of the manuscript.
Research Involving Human Participants: This study was performed in line with the principles of the Declaration of Helsinki. The study was approved by the St John of God Human Research Ethics Committee (Ref. 839) and adheres to the American Psychological Association Ethical Principles of Psychologists and Code of Conduct.

Abstract

Effective communication between clinicians is essential for the success of mental health interventions in multidisciplinary contexts. This relies on a shared understanding of concepts, diagnoses and treatments. A major assumption of clinicians when discussing psychological treatments with each other is that both parties have a shared understanding of the theory, rationale and application of the respective technique. We aimed to determine to what extent there is inter-rater agreement between clinicians in describing the content of group therapy sessions. Pairs of clinicians, drawn from a large multidisciplinary team (13), were asked to provide ratings of the therapeutic content and emphasis of $N = 154$ group therapy sessions conducted during an intensive residential treatment program for post-traumatic stress disorder (PTSD). In most therapeutic content domains there was a moderate level of agreement between clinicians regarding session content (Cohen's Kappa 0.4 to 0.6), suggesting that clinicians have a broad shared understanding of therapeutic content, but that there are also frequent discordant understandings. The implications of these findings on multidisciplinary team communication, patient care and clinical handovers are discussed and directions for further research are outlined.

Keywords: CBT; PTSD; Therapy; Group psychotherapy

When therapists communicate with each other through written case notes or at clinical handover meetings, a key assumption is that there is a shared understanding of what each therapeutic concept refers to and how it is applied in therapy. In the case of cognitive-behaviour therapy (CBT) based treatment approaches for PTSD, there are multiple separate treatment components which can each be described broadly as cognitive or behavioural in focus – or a mixture of the two. For instance, "psychoeducation" – where the clinician provides a detailed overview of the nature of PTSD and the way in which a person's thinking and behaviours may maintain their symptoms, is sometimes considered to be a "cognitive" intervention, in that one of the aims is to change the way the person understands and appraises their PTSD symptoms.

In addition to this, therapists might describe an intervention differently according to the intended purpose of the therapy. Therapeutic exercises might also have more than one intended goal, compounding the confusion about the purpose of a given intervention. For example, consider a scenario where a patient returns to traffic lights where a car crash occurred. One therapist might describe this as exposure based behavioural therapy designed to increase their sense of safety at the traffic lights by reducing association between the lights and danger, whereas another therapist might describe it as a cognitive approach intended to change the person's beliefs and help them to realise that they can cope better with the anxiety at traffic lights. In reality, there may be both cognitive and behavioural processes at work and both goals may be achieved but each therapist may describe the "active ingredient" of the exercise differently. This has implications for how therapists describe treatment interventions and their emphasis on what caused the change.

A good understanding of the components of therapy is important for effective communication as any inconsistencies in understanding of the key therapeutic components can compromise inter-professional communication [1]. Most clinical care settings involve

multiple interactions and patient handovers between numerous practitioners who have varying educational and occupational training. If team members are not communicating effectively, it can complicate the delivery of effective interventions.

Effective communication is highly reliant on the use of a common “language” [2] as this can establish role patterns in addition to integrating and bridging the experiences of team members [3]. In a multidisciplinary team, the different training and education background of individuals is an important variable since health professionals during their training learn to use a specific “register” of language [4]. This register refers to the style of language used in different situations, for instance the register of a physician consists of words he or she has learnt to exhibit among other physicians. Hence, when interacting with health professionals from other disciplines, different words would be more or less available for conceptual use [5]. The presence of different health professional “languages” can lead to different ways in which the content of sessions is understood and communicated, in turn having implications for the clinical handover of clients. The concept of a language register assumes that the words used have the same meaning across teams of clinicians, which may not always hold true, especially considering the potential for variability in therapeutic processes within psychiatry.

To date, there are few studies which have directly investigated the question of consistency in conceptual understandings among professionals. Borst and Nelson [6] reported low levels of agreement between the understanding of terms used by occupational therapists when compared to non-professionals. While this study provides evidence of language ambiguity within allied health, there is an absence of studies looking specifically at terminology used in psychological therapy especially when describing CBT interventions. This is an important research gap to explore since variability within CBT can create an element of ambiguity where one therapeutic term can have different understandings and interpretations of what it entails.

A further question that arises is whether there are differences in understandings of therapeutic content between clinicians from the same profession as opposed to clinicians from two different professions/disciplines. One might imagine that there would be a greater level of understanding within a given professional discipline where training is relatively consistent compared to that between professions, where some of the misunderstandings may arise from the different ways in which professionals of different disciplines have been taught.

The present project then, sought to investigate the consistency of clinician understandings of therapy content. We asked clinicians to provide post-session ratings of therapy content across consecutive groups attending a residential treatment program for PTSD. It was hypothesised that, despite professional differences in training, there would be broad consistency between clinicians in rating the content of recently completed therapy sessions. The second hypothesis was that there would be higher rates of agreement in describing the content of therapy sessions between professionals within the one professional discipline when compared with pairs of clinicians across different disciplines, owing to the discipline specific ways in which health professionals are trained.

Method

Participants

Thirteen clinicians (mean post-registration experience = 11.2 years; median = 9; SD = 8.3) provided independent ratings of the therapeutic content of PTSD group therapy session that they jointly facilitated. The thirteen clinicians included nine who were either psychologists or clinical psychologists (69.2%), one social worker (7.7%), one occupational therapist (7.7%) and two mental health nurses (15.4%).

Treatment program

The therapist rating occurred as part of a four-week residential group treatment program for PTSD with group sessions occurring five days per week between September 2015 and March 2017.

The treatment program was run at the at [Blinded for peer review] accredited psychological intervention program for PTSD. The program provided a CBT-based intervention for individuals meeting DSM-5 diagnostic criteria for PTSD as determined by the Clinician-Administered PTSD Scale for DSM-5 (CAPS-5). The group-based treatment program involved the following components:

(i) psychoeducation about PTSD, including sessions focused on neurobiology of PTSD, moral injury, recovery and posttraumatic growth.

(ii) arousal reduction strategies, including diversionary activities to assist with down-regulation of affect.

(iii) cognitive restructuring, including the identification of the impact of traumatic events on beliefs about self and world and challenging of broadband deployment of 'rules' from the 'trauma world' in the 'now world.'

(iv) exploration of trauma themes such as safety, trust, and power/control consistent with cognitive processing therapy interventions [7] and

(v) discharge planning.

Whilst the manual for the program has not been subject to systematic empirical validation in its own right, we note that each of the components of the program are derived from well-established and evidence-based approaches. The group program comprised morning sessions (from 9am or 9:30am to 12pm) as well as afternoon sessions (from 1pm until 4pm) with breaks for morning and afternoon tea and shorter afternoon sessions on the days when individual therapy sessions were scheduled. Thus, a number of different therapy strategies were at times covered within the one session. One clinician typically led each

session with the other clinician assisting. Clinicians were requested to provide ratings for all group sessions they attended. Clinicians were able to tailor the delivery of the manual such that they could vary the length and emphasis of particular therapy components according to client needs. In addition to the group intervention, participants also attended individual therapy sessions two-to-three times a week where individualised prolonged imaginal and in vivo exposure therapy was undertaken. The exposure exercises conducted in the individual therapy sessions were intended to complement the overall CBT-focus and principles of the group program. The program also included a 3- and 9-month review of client progress.

Procedure

Each of the group sessions was run by two clinicians and typically included 6-8 adults with occupation-related PTSD (predominantly military veterans, ex-police and other emergency services workers and current defence force personnel). In each group session, clinicians worked together rather than in parallel when discussing strategies and techniques, thus ensuring shared awareness of the content of each session. The group sessions were skills-based, but also provided opportunities for group discussions and bonding. While the group program adhered to a treatment manual, clinicians were able to modify and tailor approaches to the presenting concerns of group members, thus allowing for a degree of variation in the ways in which approaches were applied. For instance, the amount of program time devoted to specific techniques varied and the choice of examples and practice exercises were also at times varied in accordance with the particular needs of clients. Clinicians' were asked to provide online ratings of session therapeutic content of each session. Each clinician independently provided ratings of the proportion of overall session time (%) that they spent using each of 16 different therapy approaches (see Level 1 ratings, Table 1) in the respective therapy session. Data was collected from clinicians across 12 residential PTSD group treatment programs delivered across a two-year period.

Data analysis

Data were analysed in the Statistical Package for Social Sciences version 23.0 (SPSS). The final sample that was analysed included 154 group sessions that each had one pair of clinician ratings.

The 16 Level 1 categories (Table 1) were collapsed into five Level 2 categories (behavioral interventions, cognitive interventions, other active intervention, education, other) and three Level 3 categories (behavioral interventions, cognitive interventions, other) respectively to simplify the interpretation of results.

The collapsing of the categories was determined partly by the definition used in systematic reviews such as Mendes, Mello, Ventura, Passarela & Mari [8]. However, for instances where specific categories were not reported in the categorization of Mendes et al [8], we referred to leading treatment manuals for PTSD, such as Resick, Monson & Chard's Cognitive Processing Therapy manual [7], which confirmed that "trauma themes" was in fact a cognitive strategy. We allocated psychoeducation its own category as it formed a significant proportion of the overall program and is derived from both cognitive and behavioural approaches. In order to group other interventions, such as interpersonal skills and lifestyle interventions, which were derived from either a mix of cognitive and behavioural approaches, or which have been inconsistently categorized in the literature, but which were nonetheless "active" interventions where group participants needed to practice or apply some skill, we categorised them as "other active interventions". All other remaining interventions were group under the "other" category.

Third, for interpretative ease, we collapsed percentage ratings into "0" (for 0%) and "1" (for anything >0%) to indicate the presence or absence of each respective approach being used (for any amount of time) in each respective session. Finally, rates of percentage agreement were calculated for each approach, Cohen's Kappa values were calculated for each

of the pairs of clinician ratings (n = 154) and an Area Under the Curve (AUC) value was calculated. In contrast to percentage agreement, Kappa values adjust for chance agreement [9]. Thus, while we report rates of percentage agreement, we focus our discussion on the obtained Kappa and AUC values. While rates of percentage agreement may in some cases be high, corresponding Kappa values may be low, or vice versa [10]. The Kappa calculation is performed based on the amount of actual agreement between clinicians (observed agreement) compared to the amount of agreement that occurred due to chance alone (expected agreement). We interpreted results in accordance with established thresholds for the magnitude of Kappa, that is: Kappa < 0 = "less than chance agreement"; 0.01-0.20 = "slight agreement"; 0.21 to 0.40 = "fair agreement"; 0.41 to 0.60 = "moderate agreement"; 0.61 to 0.80 = "substantial agreement"; and 0.81 to 1.00 = "almost perfect agreement" [11].

The Area Under the Curve (AUC) value provided an additional indication of the level of agreement between clinicians on therapeutic content. Estimates of AUC range from 0.5 to 1. A score of 0.9 or greater indicates near perfect agreement; 0.8–0.9, substantial agreement; 0.7–0.8, moderate agreement; 0.6–0.7 fair agreement; and less than 0.6, slight agreement [12].

Consistent with the hypothesis, analyses were undertaken for comparisons between any two psychologist/clinical psychologists and for comparisons between any psychologist/clinical psychologist and any other professional discipline (e.g., social worker or occupational therapist or nurse). Given that there were proportionately more psychologists/clinical psychologist staff members than other professional disciplines, these analyses were only conducted for psychologist/clinical psychologist vs psychologist/clinical psychologist and psychologist/clinical psychologist vs any other discipline to preserve statistical power.

Results

Due to the low likelihood of agreement expected when clinicians provided any given Level 1 rating in the context of many similar, though slightly different clinical interventions, Level 1 ratings (e.g., in vivo vs other exposure therapy etc), we focus the results of the Level of clinician agreement in relation to Level 2 and Level 3 intervention groupings.

Ratings at Level 2 categorisation

Kappa values for Level 2 of categorisation (Table 1) suggest that clinicians had moderate levels of agreement on ratings for cognitive, education and other active interventions with Kappa values of 0.47, 0.57 and 0.57 respectively. Conversely there was only a slight level of agreement for behavioural (0.13) and the other (0.15) categories.

We then filtered the results to either include only psychologist-psychologist pairings or psychologist-non-psychologist pairings (Table 2). For psychologist-psychologist pairings, results show a moderate level of agreement for cognitive (0.48), education (0.47) and other active interventions (0.52). A fair level of agreement was also observed in our behavioural (0.37) and other (0.22) categories.

Similarly, moderate levels of agreement were observed in the ratings of cognitive (0.44), education (0.56) and other active interventions (0.59) for psychologist-non-psychologist pairings but the agreements for behavioural and other interventions were lower compared to psychologist-psychologist pairings, with only a slight, near chance level of agreement between clinicians.

Ratings at Level 3 categorisation

Kappa values for Level 3 categorisation demonstrate similar results to level 2 (Table 1) with cognitive and other interventions showing moderate levels of agreement at 0.47 and 0.43 respectively. Similarly, clinician agreement on behavioural interventions remained low consistent with the Level 2 analysis, with a kappa value of 0.13 (slight levels of agreement).

When ratings were filtered based on psychologist-psychologist pairings (Table 2), there was an increase in clinician agreement for behavioural interventions observed at 0.37 (fair agreement) while cognitive interventions remained at moderate levels of clinician agreement (0.48).

With psychologist-non-psychologist pairings (Table 2), other interventions show the highest degree of agreement (0.79) with a substantial level of interrater agreement between clinicians. While agreement on cognitive interventions (0.44) was also at a moderate level of agreement, the agreement of behavioural intervention between psychologist and non-psychologist was much lower than psychologist-psychologist pairings, showing only a slight level of agreement (0.01).

Discussion

The aim of our study was to determine whether clinicians have a shared understanding of therapy interventions. Ensuring consistency in this regard is important for effective multi-disciplinary communication and to allow optimisation of treatment for PTSD. To investigate this, we assessed clinician ratings of the content of therapy sessions during the course of a residential PTSD program. To the best of our knowledge, we have conducted the first study that has systematically measured clinician agreement on the content of therapy sessions in the area of mental health. Together, our results indicate that, irrespective of the level of categorization, there is at best only a moderate level of agreement between any two clinicians on the therapeutic content of PTSD group sessions that they facilitated.

When we examined the agreement between clinicians for cognitive, other, other active interventions and education on different categorisation levels, clinician ratings agreed to a moderate level on most of these therapeutic interventions. This suggests that clinicians have a broad level of shared understanding of most components of CBT; however, their understanding of the specifics of particular approaches vary. In this respect, if more specific

interventions were rated (e.g., cognitive restructuring within all cognitive approaches), agreement levels would invariably have been lower.

Our findings also show that behavioural interventions had low levels of agreement. This may have been a consequence of low rates of agreement for prolonged imaginal exposure therapy in particular, as agreement was stronger for in vivo and other forms of behavioural interventions. However, we also note that exposure-based exercises were rarely conducted during the group therapy sessions, thus resulting in considerably fewer ratings for exposure based therapies than for the other intervention categories. This limitation notwithstanding, the low levels of agreement for behavioural intervention may be explained by the fact that exposure therapy is often given in conjunction with other types of CBT-based therapies (e.g. cognitive) which, despite being both grouped under the umbrella of CBT, are derived from different theoretical frameworks. As such, while one clinician might define an exposure based exercise as an intervention intended to precipitate behavioural change (e.g., reduced avoidance), another might describe the aim of the intervention as cognitive (e.g., to reduce fear expectancies which lead to avoidance behaviour). While these distinctions may in some regards be artificial, they nonetheless lead clinicians to describe the same intervention in different ways. Unfortunately, we do not have sufficient data to ascertain whether, and to what extent, this was the case, but this may be a focus for further research.

It is likely that some of the differences in understanding of therapeutic content may affect communication. Client notes, team discussions, and in turn, client care, are likely to reflect differences in the ways in which clinicians understand and, in turn, report the content of therapy sessions. Such differences in therapeutic descriptions can result in misunderstandings between clinicians of what may have been covered in therapy and could lead to unnecessary repetition of therapeutic approaches or otherwise an omission of approaches which the clinician believes has already been sufficiently covered. There may

also be more serious implications, Arora and colleagues [13] found that poor communication resulted in content omission and failure to outline the rationale behind making a decision. Raduma-Tomàs, Flin, Yule and Williams [14] also reported that poor communication is the main cause for inaccurate transfer of information in clinical handovers and is responsible for 60-70% of hospital incidents.

The agreement between any two pairs of clinicians for most therapeutic interventions occurred only at a moderate level assessed by analysis of two measures of agreement, regardless of whether the pairs of clinicians providing ratings were from the same professional background (i.e., psychology/clinical psychology) or from different backgrounds. For behavioural interventions, the psychologist and non-psychologist pairings had close to chance agreement compared to psychologist-psychologist whose agreement were considerably higher. The finding that the profession-concordant pairs had higher agreement levels compared to the profession-discordant pairs on certain therapeutic techniques suggests that clinical terminology may be relatively specific to the clinician's training, particularly so when behavioural interventions are considered. This broadly supports Dickerson, Davis and Staplin [15] who suggested that professional groups such as physicians and occupational therapists have a common language within their groups, but the same words do not have the same meaning between different professional groups. Similar observations were also made by Stallinga et al [16] who commented that language ambiguity allows different interpretations of the same word, with the term "trainability" having a different meaning to different professional groups. Therefore, a possible explanation for low levels of agreement for certain therapeutic techniques between different professional groups may be language ambiguity. This may in turn lead to poor decision making regarding treatment goals and less favourable clinical outcomes [17,18]. We were unable to determine whether this was the case in the present study, however, cognitive behaviour therapists working in multidisciplinary

teams may benefit from adopting clearer and less jargon-laden terminology when communicating with colleagues, such as terms from Five Areas model of CBT [19].

Agreement rates between psychologist pairs were generally greater than between-discipline rates and moderate in magnitude (percent agreement range: 74.4% to 87.2%; Kappa between 0.37 and 0.52). These values are only somewhat lower that is obtained when diagnostic interviews are validated (e.g., see [20]) and may be common when different therapy approaches, with all their conceptual overlaps and variations, are applied in routine clinical practice.

Our findings need to be considered in light of a number of limitations. While many of the ratings were provided promptly following the respective session, the real-world data collection setting of our project meant that in some instances, clinicians were only able to provide their ratings after a delay. In this respect, recall bias may have confounded our findings regarding clinician agreement. Nonetheless, such retrospective recall biases may also affect clinical handover meetings which similarly may not always occur on the same day as the intervention itself. Moreover, not every group session was rated by both clinicians and the routine nature of the clinical setting meant that clinical work took priority such that there were numerous sessions which only received a single rating. We therefore cannot be sure whether or not there were systematic patterns in the missingness of data. We suggest future studies implement a daily automated remainder for clinicians which may facilitate greater clinician compliance.

A second limitation is that we had no independent measure of treatment fidelity during the therapeutic session. Measuring treatment fidelity by independent observation would have allowed us to extend the aims of our study to also investigate the *accuracy* of clinician ratings. As it stands, we had no way of knowing whether therapists were actually using the approaches which they reported they were. Unfortunately, for feasibility reasons,

we were not able to record sessions or collect sufficient information to determine treatment fidelity. This also precludes certainty regarding whether discrepancies in ratings reflect different levels of understanding of treatment approaches, or whether discrepancies only reflect differences in the ways in which session content are reported, irrespective of understanding and application.

Third, the "other" category reflected any number of approaches, such that higher rates of agreement would have been relatively easy to achieve for this category, as clinicians may have been considering distinctly different, but nonetheless "other" approaches when providing this rating. Future studies should ideally aim to more thoroughly assess the range of approaches which fall within this category.

Fourth, we did not collect detailed information regarding the training and experience levels of the clinicians, beyond their disciplines and years of post-registration practice. Further studies might aim to collect detailed information regarding the orientation and focus of training of each practitioner, given that therapeutic orientation may account for between-rater differences to almost the same extent as professional affiliation.

Finally, the clustered nature of our data is a limitation because of greater shared variance between sets of ratings provided within any one 4-week therapy group when compared with pairs of ratings provided between two groups. Future research should aim to address this, perhaps by using multilevel modelling approaches.

In conclusion, our findings demonstrate a moderate level of inter clinician agreement for most therapeutic techniques but for certain approaches such as behaviourally based interventions, the inter-clinician agreement was low. Moreover, these differences are most pronounced when pairings are between clinicians from different professional backgrounds when compared to clinicians who are from the same professional background. Poor inter-clinician agreement potentially affects multidisciplinary communication thereby influencing

patient care. While our research provides a crucial first step in demonstrating that clinicians are broadly consistent in their reflections of therapeutic content, it also calls for further research in our identified areas of high discrepancy, such as behavioural interventions, as well as a close examination of the extent to which inconsistencies in ratings correspond to sub-optimal clinical outcomes.

References

1. Hughes R. Professional Communication and Team Collaboration. In: Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: Agency for Healthcare Research and Quality; 2008.
2. Patterson M, Hayes S. Verbal communication between students in multidisciplinary health teams. Med Educ. 1977;11:205–9.
3. Berger PL, Luckmann T. The Social Construction of Reality. New York, NY: Doubleday; 1967.
4. Halliday M, McIntosh A, Strevens P. The linguistic sciences and language teaching. London: Longman; 1964.
5. Mackey WF. Inference, integration, and the synchronic fallacy. Washington, DC: International Center for Research on Bilingualism; 1970.
6. Borst MJ, Nelson DL. Use of Uniform Terminology by Occupational Therapists. Am J Occup Ther. 1993;57:611–8.
7. Resick PA, Monson CM, Chard KM. Cognitive Processing Therapy for PTSD: A comprehensive manual. New York, NY: Guilford Press; 2016.
8. Mendes DD, Mello MF, Ventura P, de Medeiros Passarela C, de Jesus Mari J. A systematic review on the effectiveness of cognitive behavioral therapy for posttraumatic stress disorder. Int J Psychiatry Med. 2008;38(3):241–59.
9. Hadzi-Pavlovic D. Agreement II: Categorical measures. Acta Neuropsychiatr. 2010;22:155–7.
10. McHugh ML. Interrater reliability: The kappa statistic. Biochem Medica. 2012;22:276–82.
11. Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic.

Fam Med. 2005;37:360–3.

12. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data Data for Categorical of Observer Agreement The Measurement. Biometrics. 1977;33:159–74.
13. Arora V, Johnson J, Lovinger D, Humphrey HJ, Meltzer DO. Communication failures in patient sign-out and suggestions for improvement: A critical incident analysis. Qual Saf Heal Care. 2005;14:401–7.
14. Raduma-Tomàs MA, Flin R, Yule S, Williams D. Doctors' handovers in hospitals: A literature review. BMJ Qual Saf. 2011;20:128–33.
15. Dickerson AE, Davis ES, Staplin L. Can clinicians, researchers, and driver licensing officials build a shared vocabulary. Occup Ther Heal Care. 2014;28:188–93.
16. Stallinga HA, ten Napel H, Jansen GJ, Geertzen JH, de Vries Robbé PF, Roodbol PF. Does language ambiguity in clinical practice justify the introduction of standard terminology? An integrative review. J Clin Nurs. 2015;24:344–52.
17. Tempest S, McIntyre A. Using the ICF to clarify team roles and demonstrate clinical reasoning in stroke rehabilitation. Disabil Rehabil. 2006;28:663–7.
18. Fink P, Rosendal M. Recent developments in the understanding and management of functional somatic symptoms in primary care. Curr Opin Psychiatry. 2008;21:182–8.
19. Williams C, Garland A. A cognitive-behavioural therapy assessment model for use in everyday clinical practice. Adv Psychiatr Treat. 2002;8:172–9.
20. Sheehan D V., Lecrubier Y, Sheehan KH, Janavs J, Weiller E, Keskiner A, et al. The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. Eur Psychiatry. 1997;12:232–41.

Table 1. Kappa values (AUC; percentage agreement) at levels 1, 2 and 3 of categorisation.

Level 3		**Level 2**		**Level 1**	
Behavioral interventions	0.13 (0.55; 85.0%)	Behavioral interventions	0.13 (0.55; 85.0%)	Behavioral interventions – prolonged imaginal exposure therapy	-0.03 (0.49; 94.8%)
				Behavioral interventions – in vivo exposure therapy	0.96 (1; 99.3%)
				Behavioral interventions – other than exposure therapy	0.64 (0.78; 90.9%)
Cognitive interventions	0.47 (0.74; 73.4%)	Cognitive interventions	0.47 (0.74; 73.4%)	Cognitive therapy strategies	0.12 (0.55; 74.0%)
				Trauma themes exercise or discussion	0.73 (0.87; 87.0%)
Other	0.43 (0.98; 96.8%)	Other active intervention	0.57 (0.79; 82.4%)	Lifestyle intervention (re: exercise, sleep)	0.25 (0.65; 93.5%)
				Mindfulness-based approaches	0.72 (0.87; 89.0%)
				Interpersonal skills/assertiveness/communication skills	0.21 (0.59; 91.3%)
		Education	0.57 (0.79; 79.3%)	Relapse prevention	0.25 (0.63; 93.5%)
				Education – re: PTSD or mental health generally	0.65 (0.83; 83.7%)
				Diversionary activities (e.g., darts)	0.64 (0.82; 87.7%)
		Other	0.15 (0.57; 77.2%)	Review of skills or review of overall progress	0.13 (0.56; 77.9%)
				General supportive	0.25 (0.65; 65.6%)
				Other	0.51 (0.83; 91.6%)

AUC = Area Under the Curve.

Percentage agreement values (ranging 0 to 100) do not necessarily closely correspond to Kappa values (-1 to 1) given that percentage agreement values do not account for chance agreements.

* EMDR specific and Referral related discussion are not reported in this table due to insufficient numbers of co-ratings or because at least one variable in each two-way table was a constant.

Table 2. Kappa values (AUC; percentage agreement) at levels 1, 2 and 3 of categorisation for ratings between psychologists and between psychologists and other professional discipline groups.

Level 3			**Level 2**			**Level 1**		
	Psychol-Psychol	**Psychol-Non-psychol**		**Psychol-Psychol**	**Psychol-Non-psychol**		**Psychol-Psychol**	**Psychol-Non-psychol**
Behavioral interventions	0.37 (0.71; 87.2%)	0.01 (0.50; 80.0%)	Behavioral interventions	0.37 (0.71; 87.2%)	0.01 (0.50; 80.0%)	Behavioral interventions – prolonged imaginal exposure therapy	*	-0.03 (0.48; 92.9%)
						Behavioral interventions – in vivo exposure therapy	0.84 (0.99; 97.4%)	1 (1; 100%)
						Behavioral interventions – other than exposure therapy	0.72 (0.89; 92.3%)	0.59 (0.75; 87.0%)
Cognitive interventions	0.48 (0.74; 74.4%)	0.44 (0.73; 71.8%)	Cognitive interventions	0.48 (0.74; 74.4%)	0.44 (0.73; 71.8%)	Cognitive therapy strategies	0.34 (0.70; 74.3%)	0.06 (0.52; 73.0%)
						Trauma themes exercise or discussion	0.79 (0.89; 89.7%)	0.66 (0.83; 82.9%)
Other	*	0.79 (0.99; 98.9%)	Other active intervention	0.52 (0.77; 79.5%)	0.59 (0.79; 82.4%)	Lifestyle intervention (re: exercise, sleep)	0.47 (0.74; 94.9%)	0.21 (0.61; 93.0%)
						Mindfulness-based approaches	0.81 (0.92; 92.3%)	0.69 (0.86; 87.0%)
						Interpersonal skills/assertiveness/communication skills	-0.02 (0.47; 87.2%)	0.18 (0.58; 91.8%)
						Relapse prevention	0.66 (0.99; 97.5%)	-0.05 (0.48; 90.6%)
			Education	0.47 (0.74; 74.4%)	0.56 (0.78; 78.8%)	Education – re: PTSD or mental health generally	0.51 (0.75; 76.9%)	0.67 (0.84; 84.7%)

	Other	0.22 (0.63; 87.2%)	0.18 (0.59; 78.8%)	Diversionary activities (e.g., darts)	0.42 (0.73; 79.5%)	0.67 (0.84; 87.6%)
				Review of skills or review of overall progress	-0.15 (0.43; 74.4%)	0.32 (0.66; 81.2%)
				General supportive	0.16 (0.59; 66.7%)	0.28 (0.65; 67.0%)

AUC = Area Under the Curve; Psychol = Psychologist.
Percentage agreement values (ranging 0 to 100) do not necessarily closely correspond to Kappa values (-1 to 1) given that percentage agreement values do not account for chance agreements.
* Other and EMDR specific and Referral related discussion are not reported in this table due to insufficient numbers of co-ratings or because at least one variable in each two-way table was a constant.

Manoj Thogesan is a medical officer working at the Tamworth Hospital, Hunter New England Local Health District, NSW Australia. His research interests include PTSD; Psychotherapy, mood and anxiety disorders.

David Berle is an associate professor and National Health and Medical Research Council (NHMRC) Early Career Fellow in the Graduate School of Health at the University of Technology Sydney (UTS), Sydney, Australia.

Dominic Hilbrink is a social worker and Senior Clinician at St John of God Richmond Hospital in Sydney Australia.

Rachael Kiely is a psychologist who works in private practice in Sydney, Australia.

Clare Russell-Williams is a clinical psychologist based in Sydney, Australia.

Natasha Garwood is a psychologist based in Sydney, Australia.

Zachary Steel is the St John of God Chair in Trauma and Mental Health in the School of Psychiatry at UNSW in Sydney Australia.