An improved competence rating scale for CBT Supervision: Short-SAGE

Robert P. Reiser[1]*, Tom Cliffe[2] and Derek L. Milne[3]

[1] *UCSF Department of Psychiatry*
[2] *University of Leeds*
[3] *Morpeth*, *Northumberland*, *UK*

Abstract. Recent developments have led the UK government to deem clinical supervision 'essential' to a safe and effective national health service. Cognitive behavioural therapy (CBT) supervision has been increasingly operationalized and manualized, but there are few psychometrically sound observational instruments with which to measure CBT supervision. This paper reports the factor analysis of a promising 23-item instrument for observing competence in CBT supervision (Supervision: Adherence and Guidance Evaluation: SAGE). $N =115$ qualified mental health practitioners (supervisors and their supervisees) rated the same supervision session by completing SAGE. A principal components analysis indicated that a two-factor solution, identified as the 'Supervision Cycle' and the 'Supervisee Cycle' components, accounted for 52.8% of the scale variance and also demonstrated high internal reliability ($\alpha = .91$ and $\alpha = .81$, respectively). These findings provide the basis for a shorter, 14-item version of SAGE, clarify the factor structure of SAGE, ease implementation, and afford more succinct feedback. Short-SAGE also improves implementation yield, taking half the time to complete as the original 23-item scale. These conceptual and practical improvements strengthen the role of SAGE as a promising observational instrument for evaluating CBT supervision, complementing self-report assessments of competent CBT supervision with an instrument that can fulfil the distinctive functions that are provided through direct observation.

Key words: clinical supervision, evidence-based practice, cognitive therapy training, therapist competence

Introduction

SAGE (Supervision: Adherence and Guidance Evaluation) is a 23-item, direct observation instrument, developed to meet the growing need for competence monitoring and evaluation in

*Author for correspondence: Robert Reiser, 945 Butterfield Road, San Anselmo, CA 94960, USA (email: robert.reiser@gmail.com)

cognitive behavioural therapy (CBT) supervision (Milne *et al.*, 2011). Since the publication of SAGE, clinical supervision has gained greater recognition as an essential mechanism for developing competent clinicians, and as a procedure for promoting safe and effective therapy. A tragic illustration of what can go wrong in the absence of sound supervision was demonstrated by the Morecambe Bay investigation (Department of Health, 2016). This report indicated that fatal clinical errors and failures in providing compassionate care had taken place in midwifery services in this region of England, partly attributed to ineffectual supervision arrangements. These arrangements failed to identify poor practice, and it was only through complaints by patients that unsafe practices were highlighted. Under guidance from the Department of Health, part of the solution has been to develop a new model of midwifery supervision, one which includes personal action for continuous quality improvement (Nursing and Midwifery Council, 2017). Although not linked explicitly to CBT supervision, the Department's strong message about the need for effective supervisory and monitoring arrangements carries clear implications for CBT and other theoretical orientations to supervision. This includes the need to define and monitor supervisory competence, linked to assessments of the effectiveness of supervision. SAGE can contribute to these objectives.

Similar concerns have been raised latterly about unethical supervision, including poor supervisory boundaries, a lack of consistent formative feedback, and inadequate documentation of problems in supervision. Such concerns have prompted experts to consider related improvements to supervision arrangements in the USA (Ellis *et al.*, 2017), including the publication of guidelines (American Psychological Association, 2015), and a manual for evidence-based CBT supervision (Milne and Reiser, 2017), supplementing other international developments (Watkins and Milne, 2014). In summary, clinical supervision is emerging belatedly from 'the swampy lowlands' of professional practice (Schon, 1983), becoming increasingly recognized as an essential competency that must be monitored. The present paper makes a contribution to these improvements through developing a better instrument for measuring and monitoring CBT supervision. Improving measurement is a critical pathway towards dragging ourselves up to the higher and harder professional ground of evidence-based practice (e.g. by providing valid feedback on their supervision to supervisors).

Despite these developments in clinical supervision, there remains a significant problem regarding its measurement. In a landmark review, Ellis and Ladany (1997) concluded that there were no instruments designed to measure competence in clinical supervision that they could recommend. A subsequent review of 233 studies by Ellis and colleagues (Inman *et al.*, 2014), although not as methodological in focus, indicated that little progress had been made in measuring supervision (e.g. '... a clear lack of longitudinal data ...', p. 87). More recently, Gonsalvez *et al.* (2017) reached a similar conclusion, arguing for greater psychometric rigour in the development of instruments, but some progress has been noted (Watkins and Milne, 2014), including a 'core outcome battery' to address the lack of cumulative progress (Wheeler and Barkham, 2014). However, the resulting 'toolkit' of six instruments, selected to support routine data collection in practitioner-led research, are all self-report questionnaires. As argued in the original SAGE paper (Milne *et al.*, 2011), it is desirable within research and practice to triangulate self-report questionnaires with complementary measures, the 'multi-method' approach (Muse and McManus, 2013). Examples include standardized role-plays, vignettes, permanent products (archival data, such as clinical outcomes), audits and direct observation. Direct observation is a distinctive emphasis within CBT, an established method within supervision for evaluating the competence of trainees, and is considered 'especially

effective' in training therapists (Watkins, 1997, p. 337), as it provides for a relatively rigorous and objective evaluation.

Another reason to supplement the toolkit of Wheeler and Barkham (2014) is that it does not include any instruments that explicitly measure competence in supervision. The concept of competence has become part of current professional training and licensing (including supervision: Watkins and Milne, 2014), linked to the commissioning of training and of clinical services, and providing the basis for developing measurable, accountable, evidence-based clinical services (Epstein and Hundert, 2002). For such reasons, we believe that supervision-specific instruments for observing competence are needed. We do, however, agree with the need to make instruments practitioner-friendly (Wheeler and Barkham, 2014), and acknowledge that the original 23-item SAGE may have been too time-consuming for routine use, outside of academic and research settings. To our knowledge, to date SAGE has only been used regularly within university-based, postgraduate supervisor training programmes. Therefore, a shorter version may improve uptake within health services and aid its routine implementation, although there may well be other barriers to implementation.

A review of observational tools located and scrutinized 10 observational instruments, concluding that there was a need for something like SAGE (Milne and Reiser, 2011). Since then we are aware of four new instruments that assess competence directly, though none represents progress in rating CBT supervision. Specifically, the 'supervisory competence scale' (Gonsalvez *et al*., 2017) relies on supervisees' ratings, which the authors acknowledged to be a biased and simplistic measure. A similar problem exists with the self-ratings made by supervisors receiving training in supervision (Newman-Taylor *et al*., 2012). These trainees completed a self-rating of their own supervisory competence, in terms of the 18 supervision competencies within the Roth and Pilling (2007) competence framework. The competence ratings were made by the participants themselves, on a user-friendly three-point 'traffic light' scale (red = 'not/barely achieved', amber = 'partially achieved', green = 'well/fully achieved'). However, this short and user-friendly instrument was developed especially for the study and no psychometric data were reported. The authors acknowledged that such data were necessary, and that self-ratings may have been inflated. A lack of peer review affects The Supervisory Competence Scale (Rakovshik, 2015), which also lacks a detailed manual to guide its application. Similar concerns limit the suitability of The Supervisor Evaluation Scale (Corrie and Worrell, 2012). In summary, although there has been growing interest in measuring the competence of supervisors, there remains a need for an instrument like SAGE, preferably a brief version.

The paper on SAGE (Milne *et al*., 2011) extended the psychometric criteria used by Ellis and Ladany (1997) by adding an emphasis on the practicalities and benefits of alternative instruments (i.e. pragmatic criteria). This was achieved by supplementing consideration of the 'design' of an instrument (i.e. psychometric considerations, such as reliability and validity) with 'implementation' and 'yield' dimensions (the DIY criteria), as advocated by Barkham *et al*. (1998). We agree with these authors that developments in supervisory practice require due consideration of important practical determinants, as these influence the utilization of instruments (e.g. cost and availability; the need for user training; the utility of the obtained data). The importance of considering an instrument's implementation and yield, as advocated by Barkham *et al*. (1998), has grown in recognition through the field of implementation science (e.g. Lewis *et al*., 2015), broadly termed 'pragmatic properties'. This field shares many of the same measurement problems as found in CBT generally, and especially in

CBT supervision, such as a reliance on single-use or adapted instruments; dependence on instruments with uncertain reliability and validity; and the scarcity of instruments that assess theoretically important constructs. For instance, the extensive review reported by Muse and McManus (2013) concluded that there were 'significant limitations' (p. 496) to existing assessment methods within CBT, and called for more psychometric work. They also noted important implementation considerations (e.g. the need for assessor training, benchmarks, and corrective feedback). These measurement problems need to be overcome if we are to identify critical variables, such as the moderators, mediators and mechanisms that help us to understand and enhance CBT supervision.

The extent of the measurement problem in implementation science was specified in a systematic review of 104 instruments (Lewis *et al*., 2015). These instruments were concerned with a range of evidence-based implementation variables (e.g. acceptability, adoption and feasibility), as used within mental or behavioural health (e.g. parenting strategies). Of particular relevance to the present study, 'feasibility' is the extent to which a new intervention or instrument can be disseminated, successfully used, and sustained within a clinical setting. Critical aspects of dissemination include an instruments availability, length and convenience (e.g. user-friendliness). The successful use of an instrument includes what Barkham *et al*. (1998) termed 'yield': the extent to which it furnishes information that has some practical value or utility (e.g. corrective feedback to supervisees; outcome benchmarking within clinical services). Lastly, measurement feasibility includes the sustained and routine use of instruments, which will be influenced by considerations such as the associated cost, the need for new assessors to be trained, and continued organizational support. Lewis *et al*. (2015) found that details of the psychometric and pragmatic properties of the surveyed instruments was limited and variable, leading them to conclude that these instruments were under-developed, hampering progress. Together, these psychometric and pragmatic considerations need to be balanced if instruments are to become cost-effective (Muse and McManus, 2013). In summary, our rationale for shortening SAGE is to enhance implementation (by removing barriers and by improving boosters) and aid feedback (by simplifying the factors and reducing the items).

Aims

Recent concerns about ineffectual or harmful supervision have heightened the need for reliable, accessible and user-friendly instruments with which to evaluate supervision. Such instruments need to balance psychometric and pragmatic considerations if they are to prove feasible. To date, there appear to be no observational instruments for rating the competence of CBT supervision that combine fully the necessary psychometric and pragmatic properties (Milne and Reiser, 2011). In this present paper we describe the results of a principal components factor analysis, designed to simplify SAGE and provide a more pragmatic, accessible and briefer version of the original instrument.

Method

Participants

Participants were 115 supervisors and supervisees in the USA ranging in education levels, degrees, years of supervision practice, and theoretical orientation (see Table 1), who agreed

Table 1. *Sample characteristics*

		Supervisor	Supervisee
Gender:			
	Male	35%	90.2%
	Female	65%	9.8%
Average age (years):		47	31
Current degree:			
	PhD	69.1%	9.8%
	MA	7.3%	36.1%
	MSW	7.3%	1.6%
	MD	3.6%	4.9%
	BA	0%	29.5%
	BS	0%	3.3%
	Other	12.7%	14.8%
Current license (supervisors only)			
	Psychologist	80.0%	
	Licensed clinical social worker	7.3%	
	Licensed physician	3.6%	
	Licensed marriage and family counsellor	3.6%	
	Other	5.5%	

An extremely low response rate made ethnicity data uninterpretable.

to complete an online survey. Supervisors and supervisees constituted a convenience sample and were recruited based on listserv notifications and invitations to participate in supervision research. While settings included doctoral level training clinics, community mental health centres, hospitals and college counselling centres, a large portion of the sample was recruited from mainly post-doctoral level training programmes that were contacted by electronic mail for inclusion in the project. Selection of supervisors was not restricted by theoretical orientation and there were no exclusion criteria apart from requiring that supervisors were providing clinical supervision of psychotherapy. In the survey, supervisors were asked to select a 'typical' or 'representative' supervisee.

A total of 125 non-duplicative responses were received and 115 surveys (92%) provided complete data that could be interpreted and were part of the final factor analysis. 52.2% of the respondents were supervisors and 47.8% of the respondents were supervisees. Supervisors were experienced, reporting an average of 16 years of clinical experience, and 11 years of supervisory experience. Supervisors were mainly licensed psychologists (80.0%). Supervisees had spent an average of 8 months in supervision with their current supervisor.

Measure

The original 23-item version of SAGE was an empirically derived instrument designed to assess the competence of supervisors, based upon direct observation. These 23 items were drawn from prior research and relevant theory that had been found to contribute to

the effective supervision of therapy (Milne *et al.*, 2011). These items were assumed to be relevant in varying degrees across the full range of therapists' (supervisees') proficiency. For instance, within SAGE, supervision is perceived as a leadership activity, but the style of leadership would be expected to reflect the developmental needs of the participants (ranging appropriately from a 'master-apprentice' relationship, early in a professional career, to co-construction in the supervision of experienced therapists). Items 1–4 were termed 'The common factors', intended to assess relationship variables, such as the supervisory alliance; items 5–18 were termed 'The supervision cycle', intended to assess the supervisors 'technical skills' or competencies; and items 19–23 were termed 'The supervisee's learning', and were designed to assess the supervisee's engagement in experiential learning. The rationale was that within a collaborative supervisory relationship, an effective supervisor utilizes a range of supervisory behaviours responsively (e.g. goal-setting and corrective feedback), enabling the supervisee to engage in experiential (i.e. a combination of action, reflection, conceptualization and experiencing, as set out in Kolb, 1984). Consistent with Kolb (1984), we understand these cycles to be akin to the vicious and virtuous cycles in therapy, namely a number of interacting variables that operate in a complex manner to dampen or facilitate the experiential learning process (e.g. a supervisor's questioning aiding a supervisee's reflection). This process is expected to promote desirable outcomes (e.g. acquiring competence). In Table 3 we have simplified these cyclical processes to ease comprehension, to make the link to the SAGE scale straightforward, and to be consistent with standard practice. The circumplex and other devices for conveying the idea of a cyclical process can be found in Milne (2017).

SAGE was intended to be used to evaluate supervision sessions through direct observation (audio- or video-tape recordings). In the hands of a competent SAGE rater, approximately 75–90 min was required to complete the instrument ratings and comments, assuming that a full 60 min session of supervision was being reviewed. Each SAGE item was defined within an observation manual, together with a 7-point, bipolar competence rating scale, ranging from 'incompetent' to 'expert', based on the Dreyfus model of expertise (Dreyfus and Dreyfus, 1986). As a direct observation instrument, SAGE was an 'event recording' or 'global rating' tool (i.e. ratings were based on observing a sample of behaviours, then providing an overall judgement about the quality of the sample). SAGE also invited the observer to record the details of the observed supervision (date, participants, etc.), and to make qualitative suggestions on ways to improve the observed sample of supervision. SAGE data were used primarily as corrective feedback to supervisors, especially those participating in postgraduate training programmes in clinical supervision, but could also be used as an outcome measure (dependent variable) in research, or as an audit tool in clinical services.

Procedure

Supervisors were contacted via email and asked if they were interested in participating in a supervision research programme. The email contained a link to the internet-based consent form and survey. Supervisors who agreed to participate in the research were asked to nominate one current supervisee of theirs, who was then contacted via email, informed about the research project, and asked to participate and review the online consent form. Supervisees were informed that participation was voluntary, and that non-participation or withdrawing from the project would have no negative consequences in terms of their supervision. By clicking on an 'I agree' button on the internet survey consent form, participants actively agreed

to participate in the study (or chose to decline by clicking the 'I do not agree' button, at which point they exited the survey, with no data collected). Within 2 weeks of entering the study, consenting participants were instructed to rate a single supervision session with the SAGE instrument, immediately after the conclusion of the session. Supervisors and supervisees were instructed to complete the survey privately. The original 23-item SAGE was used in the present study, accompanied by an abbreviated version of the SAGE manual. Participants were asked to rate each item, which was accompanied by a corresponding description for each of these 23 items. Participants were not trained in the use of SAGE for the present study. The SAGE measure and a demographic questionnaire were made available to participants in a secure online format via SurveyMonkey. Participants were also asked to provide demographic information. Scores were entered into an SPSS database on a password protected computer with two levels of password protection. This study was approved by the Palo Alto University IRB.

Results

Sample size and characteristics

There is considerable debate around the requisite sample size for conducting a factor analysis and whether the criteria of the total sample size or the subjects-to-variables ratio should be used (MacCallum *et al*., 1999; Velicer and Fava, 1998). Sample size recommendations can range from 100 (or less) to 1000 (Comrey and Lee, 1992; MacCallum *et al*., 1999), and the subjects-to-variables ratio from 20:1 to 2:1 (Hogarty *et al*., 2005; Kline, 1979). The sample size of $N = 115$ with a subjects-to-variable ratio of 5:1 was deemed adequate to conduct a factor analysis based on both statistical and practical research which suggests that sample sizes of less than 100 and smaller subjects-to-variable ratios can lead to good factor structure recovery (Arrindell and van der Ende, 1985; Costello and Osborne, 2005; MacCallum *et al*., 1999).

The sample contained a range of supervisors and supervisees (see Table 1) who completed SAGE to provide ratings of supervisory competence in relation to a recently completed supervision session. This is consistent with the recommendation that samples contain participants who will give a range of responses regarding the construct being measured (Gaskin *et al*., 2017). The use of self-rated supervision competencies (in the case of the participating supervisors) has been reported previously (Newman-Taylor *et al*., 2012).

Factor analysis

A principal component analysis (PCA) was conducted using an oblique rotation (Direct Oblimin). PCA has been recommended as a suitable way of shortening measures (Stanton *et al*., 2002). Data demonstrated multicollinearity, and due to an assumption of underlying components not being independent, an oblique rotation method was utilized. The Kaiser–Meyer–Olkin measure of sampling adequacy was determined (KMO = .902), which could be deemed as superb (Field, 2009). Bartlett's test of sphericity was significant at $p < 0.001$ [$\chi^2 (253) = 1624.978$]. The Kaiser criterion of eigenvalues greater than one suggested five components with a total variance of 69.57% explained. The fourth and fifth factors were just over 1, at 1.16 and 1.00, respectively. However, there was considerable overlap between items

and components. Additionally, the scree plot was more ambiguous, demonstrating inflexions that could justify retaining between two and five components.

Therefore, 2-, 3- and 4-factor solutions were also tested. The 2-factor solution contained the least amount of overlap between items and components. This solution also made the most conceptual sense, as it was entirely consistent with the underlying 'tandem' model of leadership (a supervision or leader cycle) linked to experiential learning (a supervisee or follower cycle: Milne and James, 2005), although the original 'common factors' component (including the supervision alliance) did not emerge from this analysis. Therefore, these two components were retained, accounting for 52.80% of the variance in SAGE scores. Table 2 shows the pattern matrix and structure matrix for this 2-factor solution of the SAGE measure.

Revisions to SAGE

Due to the fairly small sample size, items were retained if they scored above .60 on a single component and below .40 on another component (MacCallum *et al.*, 1999). The pattern matrix was utilized for the final interpretation (Field, 2009). The retained items were allocated to one of two components, which were given the titles 'Supervision Cycle' and 'Supervisee Cycle'. The Supervision Cycle component accounted for 43.22% of the variance in SAGE scores and contains items relating to the supervisor's behaviours in supervision, including how they manage supervision, evaluate their supervisee, and provide feedback (see Table 3). Extensive definitions of all items, together with examples, are provided in the SAGE manual (available on request from the first author). The Supervisee Cycle component accounted for 9.58% of the variance and contains items concerning the supervisee's level of reflection, conceptualization, planning and experiencing, as per Kolb's (1984) account of experiential learning. Based on this procedure, a short version of SAGE was developed, consisting of 14 items (see Table 3) with high internal consistency ($\alpha = .91$). The Supervision Cycle (10 items) and Supervisee Cycle subscales (4 items) also demonstrated high internal reliability ($\alpha = .91$ and $\alpha = .81$, respectively).

Discussion

In this paper we described a principal components factor analysis of the original 23-item SAGE measure, intended to develop a briefer version of the instrument to simplify how we might construe, evaluate and provide feedback on CBT supervision. The factor structure of the shortened version of the SAGE measure is consistent with two of the major components of the original 'tandem' model, a front wheel or cycle steered by the supervisor (the leader), facilitating learning; and a back wheel or cycle related to the supervisee's consequent engagement in experiential learning (Milne and James, 2005; Milne and Reiser, 2017). Kolb's (1984) theory of experiential learning provided the operationalization for the Supervisee Cycle (i.e. reflecting, conceptualizing, etc.). This conclusion therefore provides strong continuity with the tandem model, which was the underlying conceptual framework that guided the original version of SAGE (Milne *et al.*, 2011). By contrast, the present factor analysis has eliminated the original 'Common Factors' component (items: relating, collaborating, managing, and facilitating). This had been included originally due to the dominant conception of supervision as a collaborative relationship or 'alliance' (Watkins and Scaturo, 2013), and was operationalized in the original SAGE as akin to the role of

Table 2. *Pattern and structure matrices for the 2-factor solution of the SAGE measure*

		Rotated factor loadings	
Matrix	Item	Supervision cycle	Supervisee cycle
Pattern matrix	Formulating	.832	−.047
	Evaluating	.804	−.107
	Teaching	.799	−.097
	Prompting	.799	−.079
	Agenda-setting	.78	−.094
	Feedback receiving	.775	−.164
	Training/experimenting	.742	.015
	Questioning	.73	.103
	Managing	.671	.127
	Demonstrating	.614	.166
	Discussing	.59	.288
	Feedback giving	.583	.274
	Experiencing	.465	.35
	Collaborating	.45	.442
	Listening	.416	.349
	Observing	.401	−.053
	Experimenting	.337	.239
	Conceptualizing	−.066	.806
	Supervisee experiencing	−.103	.803
	Reflecting	.009	.757
	Planning	−.014	.686
	Facilitating	.399	.51
	Relating	.459	.496
Structure matrix	Formulating	.812	.309
	Questioning	.774	.415
	Prompting	.766	.263
	Evaluating	.758	.236
	Teaching	.758	.244
	Training/experimenting	.749	.332
	Agenda-setting	.74	.239
	Managing	.725	.413
	Discussing	.713	.54
	Feedback (receiving)	.705	.167
	Feedback (giving)	.7	.523
	Demonstrating	.685	.428
	Collaborating	.638	.634
	Experiencing	.614	.549
	Listening	.565	.527
	Experimenting	.439	.383
	Observing	.378	.118
	Conceptualizing	.278	.778
	Reflecting	.332	.761
	Supervisee experiencing	.24	.759
	Relating	.67	.692
	Facilitating	.616	.68
	Planning	.279	.68

Table 3. *The 14 items retained in Short-SAGE*

SAGE item 'Supervision cycle'	Definition Supervisor:
1. Agenda-setting	Defining session objectives (e.g. prioritizing topics)
2. Managing	Leading supervision ('Scaffolding' learning)
3. Formulating	Analysing, synthesizing, explaining (e.g. case reformulation)
4. Questioning	Gathering information; raising awareness (e.g. open questions)
5. Prompting	Reminding and cueing (e.g. reminding; rephrasing)
6. Demonstrating	Modelling competence (e.g. live or video-modelling)
7. Teaching	Informing/discussing/educating (e.g. 'information transmission')
8. Training/experimenting	Facilitating experiential learning (e.g. role-play; simulations)
9. Evaluating	Monitoring and giving supervisee feedback (e.g. on closing gaps)
10. Feedback	Seeking feedback on supervision (e.g. clarifying improvements)
'Supervisee cycle'	Supervisee:
11. Reflecting	Summarizing and understanding subjective material (e.g. incident)
12. Conceptualizing	Integrating objective material (e.g. grasping relevant theory)
13. Planning	Problem-solving, decision-making (e.g. action planning)
14. Experiencing	Processing feelings (e.g. emotionally self-aware)

the common factors in psychotherapy (Horvath *et al.*, 2011). In the shortened instrument, these common factors did not appear to be supported as a distinct third factor, although the item 'managing', together with some features of that original alliance factor, are retained within the Supervisor Cycle (e.g. item 1: collaborating over the supervision agenda; item 2: managing the session; item 8: facilitating learning). The resulting 14-item SAGE also eliminated the following Supervision Cycle items: relating, facilitating, collaborating, giving feedback, discussing, experiencing, listening, and observing. Additionally, the Supervisee Cycle item 'experimenting' was eliminated. This elimination procedure was conducted in order to reduce item overlap, enhance the conceptual clarity, speed up the rating (fewer SAGE items), and enable the provision of more succinct feedback (requiring less time and effort to record or discuss). While these results should be seen as preliminary and subject to confirmation through an additional factor analysis, the supervision model suggested here is consistent with CBT supervision (e.g. Liese and Beck, 1997; Reiser, 2014). For instance, the central role of the educational aspects of supervision in the Supervisor Cycle component are consistent with Goodyear's account of supervision (2014) and with the latest 'expertise' perspective on supervision and training (Rousmaniere *et al.*, 2017).

The elimination of the original 'Common Factors' component within SAGE is certainly not consistent with the continued emphasis of the importance of the supervisory alliance in the field, which some researchers have considered 'the quintessential integrative variable in psychotherapy supervision' (Watkins and Scaturo, 2013, p. 151). However, the same authors also recognized that the supervision alliance was treated quite differently across the six supervision approaches they compared (Watkins and Scaturo, 2013), and that in CBT supervision the alliance (together with other common factors) is subsidiary to the teaching of CBT skills and practical forms of collaboration, a relationship sometimes termed a 'task alliance', with correspondingly less emphasis on the relational and emotional bonds between supervisee and supervisor (e.g. creating a 'safe base'). This is also consistent with

the empirically derived Roth and Pilling (2008) supervision framework where the generic competence 'Forming a good supervisory alliance' contains several structured elements including 'ability to structure supervision', helping supervisees present their work, helping the supervisee reflect on their work and the usefulness of supervision, and 'giving accurate and constructive feedback'. These elements appear to go beyond the traditional supervisory alliance conception of a bond and a sense of common goals and tasks.

This relatively narrow definition of the supervision alliance in CBT supervision is also consistent with the clinical application of CBT, and with recent research on supervision and clinical outcomes. For example, a poor alliance has actually been associated with good clinical outcomes, while 'supervisor agreeableness' has been significantly negatively associated with client change (Rieck *et al.*, 2015). These authors conjectured that their findings could be explained by supervisors being challenging and directive, which appeared to contribute to improved outcomes at the expense of the supervisory alliance. Such research, combined with relevant theory and expert consensus, led Milne and Reiser (2017) to define the CBT supervision alliance as a primarily educational relationship that was highly structured, offered a professional role model and was consistently collaborative. Short-SAGE embodies this more educationally focused definition (e.g. items 6–10 in Table 3), while reducing emphasis on the more traditional supervisory alliance conception of a bond, as represented in our original SAGE item 'relating'. Because of the some of the limitations noted below, we should be cautious about drawing conclusions about which aspects of the supervisory alliance are fundamental to improving supervision outcomes, awaiting further studies replicating this finding.

Limitations of the study

Our sample of supervisors and supervisees was a relatively small and narrow convenience sample limited largely to doctoral-level supervisees and PhD level psychologists in the USA, with limited cultural diversity. It follows that their responses might not be generalizable to more representative, interdisciplinary samples. For example, there is a risk that this sample may have biased the findings towards an expert-novice power imbalance, a bias that might disappear in samples including more experienced supervisees receiving post-qualification supervision (CPD supervision). Secondly, our sample received no training in completing SAGE, and only received brief descriptions of the scale items, derived from a shortened manual. This may have introduced errors and biases that affected how supervision sessions were rated, and may have confounded the factor analysis. Thirdly, the surveys completed by the participating supervisors were based on self-report, and systematic reviews have indicated that clinicians tend to provide inflated ratings of their adherence to guidelines (and proficiency), resulting in large overall differences between objective measures and self-report assessments (Adams *et al.*, 1999). However, there was no significant difference between the total SAGE ratings of supervisors ($M = 106.98$, $SD = 14.01$) and supervisees ($M = 107.72$, $SD = 19.75$), $t(92.42) = -.23$, $p = .822$), suggesting to us that in this instance self-rated SAGE data were similar to the observational data. However, this does not rule out the possibility that both sets of scores were inflated, and indeed the item means listed in Table 4 clearly indicate this. A final related limitation to note was that the obtained ratings were based on supervisors with a range of theoretical orientations likely in most cases to be markedly less experiential than evidence-based CBT supervision (Milne, 2008; Milne and Reiser, 2017). It follows that

Table 4. *Short-SAGE descriptive statistics*

SAGE item 'Supervision cycle'	Mean	SD
Agenda-setting	4.2	1.3
Managing	4.7	0.9
Formulating	4.6	1.3
Questioning	4.8	1.1
Prompting	4.6	1.0
Demonstrating	4.7	1.3
Teaching	4.6	1.4
Training/experimenting	4.1	1.8
Evaluating	4.3	1.3
Feedback	4.4	1.4
'Supervisee cycle'		
Reflecting	5.0	0.8
Conceptualizing	4.8	0.9
Planning	4.9	0.9
Experiencing	4.9	0.9

a survey of the latter might yield a factorial structure with a greater experiential emphasis, with supervisor activities like 'feedback' and 'observation' (and supervisee 'experimenting') re-emerging as SAGE items.

Recommendations for future research

Future studies should validate Short-SAGE with larger, more representative samples, and also against widely accepted supervision guidelines, like the Roth and Pilling competency framework (Roth and Pilling, 2007). This would help to integrate the different tools and recommendations (although all the competencies in that framework are at least assessed by the 23-item SAGE: Milne and Reiser, 2017). Also consistent with recent supervision rating scales (Newman-Taylor *et al*., 2012), the present 7-point competence rating scale within SAGE should ideally be reduced to fewer rating points. This would further enhance the feasibility of Short-SAGE, and ease rater training (we have appended a 3-point rating scale to the Short-SAGE manual, for training purposes). The 7-point scale seems excessively detailed for Short-SAGE, and may reduce feasibility (e.g. too time-consuming to be used routinely; unnecessarily discriminating). On the other hand, there is reason to believe that, while two- or three-point scales may be quick, easy to use, yet still effective in discriminating between competence levels, research suggests that a 7-point scale is more precise and probably optimal (allowing judgements to be expressed fully: Preston and Colman, 2000). We hope to evaluate the relative value of the 7-point SAGE rating scale alongside an attempted validation of the 14-item SAGE scale. In the meantime, we have prepared a new, Short-SAGE manual which retains the original 7-point scale for research and other purposes, with a 3-point scale appended for training purposes (adopting the 'traffic light' approach for feasibility reasons: Newman-Taylor *et al*., 2012). A copy of the new manual may be obtained free from any of the authors.

Conclusions

In summary, we achieved the objective of shortening SAGE, thereby providing a more user-friendly and conceptually neat observational instrument for rating the competence of CBT supervision. Clinical supervision in general is emerging belatedly from 'the swampy lowlands' of professional practice (Schon, 1983), becoming increasingly recognized, monitored and evidence-based. The present paper represents a contribution to these improvements through developing a better instrument for measuring CBT supervision. Better measurement is one way of dragging ourselves up to the higher and harder professional ground of evidence-based practice, although we recognize that there may be other barriers to the use of SAGE.

Main points

(1) In theory, supervision provides basic assurances as to the quality, safety and effectiveness of clinical services.

(2) Establishing competence has become a fundamental part of current professional training and licensing (including supervision: Watkins and Milne, 2014), linked to the commissioning of training and of clinical services, and providing the basis for developing measurable, accountable, evidence-based clinical services (Epstein and Hundert, 2002).

(3) There are few available psychometrically sound observational instruments with which to measure CBT supervision.

(4) This article discusses the development of a shorter user-friendly version of SAGE (Supervision: Adherence and Guidance Evaluation) designed to assist in the direct observation, monitoring and evaluation of clinical supervision.

Acknowledgements

Conflicts of interest

Ethical statement

The authors have abided by the Ethical Principles of Psychologists and Code of Conduct as set out by the APA. Authors also confirm that ethical approval was granted by the Palo Alto University IRB for this research.

Financial support

Recommended follow-up reading:

Milne D (2009 and 2017). *Evidence-Based Clinical Supervision: Principles and Practice*. Oxford: BPS Blackwell.

Milne DL, Reiser RP (2011). Observing competence in CBT supervision: a systematic review of the available instruments. *the Cognitive Behaviour Therapist* **4**, 89–100.

Milne DL, Reiser RP (2017). *A Manual for Evidence-Based CBT Supervision.* Chichester: Wiley-Blackwell.

Milne DL, Reiser RP, Cliffe T, Raine R (2011). SAGE: an instrument for observing competence in CBT supervision. *the Cognitive Behaviour Therapist* **4**, 123–138.

Roth AD, Pilling S (2007). A competence framework for the supervision of psychological therapies. Available at: http://www.ucl.ac.uk/CORE/

References

Adams AS, Soumerai SB, Lomas J, Ross-Degnan D (1999). Evidence of self-report bias in assessing adherence to guidelines. *International Journal for Quality in Health Care* **11**, 187–192.

American Psychological Association (2015). Guidelines for clinical supervision in health service psychology. *American Psychologist* **70**, 33–46.

Arrindell WA, van der Ende J (1985). An empirical test of the utility of the observations-to-variables ratio in factor and components analysis. *Applied Psychological Measurement* **9**, 165–178.

Barkham M, Evans C, Margison F, McGrath G, Mellor-Clark J, Milne DL, Connell J (1998). The rationale for developing and implementing core outcome batteries for routine use in service settings and psychotherapy outcome research. *Journal of Mental Health* **7**, 35–47.

Comrey AL, Lee HB (1992). *A First Course in Factor Analysis.* Hillsdale, NJ: Lawrence Erlbaum.

Corrie S, Lane DA (2015). *CBT Supervision.* London: Sage.

Corrie S, Worrall M (2012). The Supervisor Evaluation Scale. Unpublished instrument, described in Corrie and Lane (2015) and available from: sarah.corrie@nhs.net

Costello AB, Osborne JW (2005). Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation* **10**, 1–9.

Department of Health (2016). *Proposals for changing the system of midwifery supervision in the UK.* London: Department of Health.

Dreyfus HL, Dreyfus SE (1986). *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer.* Oxford: Blackwell.

Ellis MV, Ladany N (1997). Inferences concerning supervisees and clients in clinical supervision. An integrative review. In CE Watkins (ed), *Handbook of Psychotherapy Supervision* (pp. 447–507). New York: Wiley.

Ellis MV, Taylor EJ, Corp DA, Hutman H, Kangos KA (2017). Narratives of harmful clinical supervision: Introduction to the Special Issue. *The Clinical Supervisor* **36**, 4–19. https://doi.org/10.1080/07325223.2017.1297753

Epstein RM, Hundert EM (2002). Defining and assessing professional competence. *Journal of the American Medical Association* **287**, 226–235.

Field A (2009). *Discovering Statistics Using SPSS.* London: SAGE.

Gaskin CJ, Lambert SD, Bowe SJ, Orellana L (2017). Why sample selection matters in exploratory factor analysis: implications for the 12-item World Health Organization Disability Assessment Schedule **2**.0. *BMC Medical Research Methodology* **17**, 1–9.

Gonsalvez CJ, Hamid G, Savage NM, Livni D (2017). The supervision evaluation and supervisory competence scale: psychometric validation. *Australian Psychologist* **52**, 94–103.

Goodyear RK (2014). Supervision as pedagogy: attending to its essential instructional and learning processes. *The Clinical Supervisor* **33**, 82–99.

Hogarty KY, Hines CV, Kromrey JD, Ferron JM, Mumford KR (2005). The quality of factor solutions in exploratory factor analysis: the influence of sample size, communality, and overdetermination. *Educational and Psychological Measurement* **65**, 202–226.

Horvath AO, Del Re AC, Flückiger C, Symonds D (2011). Alliance in individual psychotherapy. *Psychotherapy* **48**, 9–16. https://doi.org/10.1037/a0022186

Inman AG, Hutman H, Pendse A, Devdas L, Luu L, Ellis MV (2014). Current trends concerning supervisors, supervisees, and clients in clinical supervision. In CE Watkins, DL Milne (eds), *Wiley International Handbook of Clinical Supervision* (pp. 61–102). Chichester: Wiley Blackwell.

Kline P (1979). *Psychometrics and Psychology*. London: Academic Press.

Kolb D. A. (1984). *Experiential Learning: Experience as the Source of Learning and Development* (vol. 1). Englewood Cliffs, NJ: Prentice-Hall.

Lewis CC, Fischer S, Weiner BJ, Stanick C, Mimi Kim M, Martinez RG (2015). Outcomes for implementation science: an enhanced systematic review of instruments using evidence-based rating criteria. *Implementation Science* **10**, 155. doi: 10.1186/s13012-015-0342-x

Liese B. S., Beck J. S. (1997). Cognitive therapy supervision. In C. E. Watkins (ed), *Handbook of Psychotherapy Supervision* (pp.114–133). New York: Wiley.

MacCallum RC, Widaman KF, Zhang S, Hong S (1999). Sample size in factor analysis. *Psychological Methods* **4**, 84–99.

Milne DL (2008). CBT supervision: from reflexivity to specialisation. *Behavioural and Cognitive Psychotherapy* **36**, 779–786.

Milne D (2009 and 2017). *Evidence-Based Clinical Supervision: Principles and Practice*. Oxford: BPS Blackwell.

Milne DL, James IA (2005). Clinical supervision: 10 tests of the tandem model. *Clinical Psychology Forum* **151**, 6–9.

Milne DL, Reiser RP (2011). Observing competence in CBT supervision: a systematic review of the available instruments. *the Cognitive Behaviour Therapist* **4**, 89–100.

Milne DL, Reiser RP (2017). *A Manual for Evidence-Based CBT Supervision*. Chichester: Wiley-Blackwell.

Milne DL, Reiser RP, Cliffe T, Raine R (2011). SAGE: an instrument for observing competence in CBT supervision. *the Cognitive Behaviour Therapist* **4**, 123–138.

Muse K, McManus F (2013). A systematic review of methods for assessing competence in CBT. *Clinical Psychology Review* **33**, 484–489.

Newman-Taylor K, Gordon K, Grist S, Olding C (2012). Developing supervisory competence: preliminary data on the impact of CBT supervision training. *the Cognitive Behaviour Therapist* **5**, 83–92.

Nursing and Midwifery Council (2017). *A-Equip: a model of clinical midwifery supervision*. London: NHS England.

Preston CC, Colman AM (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica* **104**, 1–15.

Rakovshik S (2015). *The Supervisor Competence Scale: Development and Psychometric Properties*. Paper presented at the annual conference of the BABCP, Warwick, UK.

Reiser RP (2014). Supervising cognitive and behavioural therapies. In Jr CE Watkins and DL Milne (eds), *The Wiley International Handbook of Clinical Supervision* (pp. 493–517). Chichester: Wiley-Blackwell. doi: 10.1002/9781118846360.

Rieck T, Callahan JL, Watkins Jr CE (2015). Clinical supervision: an exploration of possible mechanisms of action. *Training and Education in Professional Psychology* **9**, 187.

Roth AD, Pilling S (2007). A competence framework for the supervision of psychological therapies. Available at: http://www.ucl.ac.uk/CORE/

Roth AD, Pilling S (2008) Using an evidence-based methodology to identify the competencies required to deliver effective cognitive and behavioural therapy for depression and anxiety disorders. *Behavioural and Cognitive Psychotherapy* **36**, 129–147.

Rousmaniere T, Goodyear RK, Miller SD, Wampold BE (2017). *The Cycle of Excellence: Using Deliberate Practice to Improve Supervision and Training*. Chichester: Wiley-Blackwell.

Schon D (1983). *The Reflective Practitioner*. London: Temple-Smith.

Stanton JM, Sinar EF, Balzer WK, Smith PC (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology* **55**, 167–194.

Velicer WF, Faver JL (1998). Effects of variable and subject sampling on factor pattern recovery. *Psychological Methods* **3**, 231–251.

Watkins CE (ed) (1997). *Handbook of Psychotherapy Supervision* (pp. 447–507). New York: Wiley.

Watkins CE, Milne DL (eds) (2014). *The Wiley International Handbook of Clinical Supervision*. Chichester: Wiley-Blackwell.

Watkins Jr CE, Scaturo DJ (2013). Toward an integrative, learning-based model of psychotherapy supervision: supervisory alliance, educational interventions, and supervisee learning/relearning. *Journal of Psychotherapy Integration* **23**, 75.

Wheeler S, Barkham M (2014). A Core Evaluation Battery for Supervision. In CE Watkins Jr and DL Milne (eds), *The Wiley International Handbook of Clinical Supervision* (pp. 365–385). Chichester: Wiley-Blackwell. doi: 10.1002/9781118846360.

Learning objectives

(1) Readers will be able to identify common problems in the development of instruments to observe supervision.

(2) Readers will be able to list three key items in the shortened version of SAGE that reflect important supervision competencies.

(3) Readers will be able to identify three key limitations of the study and recommendations for future research.