



CREaTE

Canterbury Research and Theses Environment

Canterbury Christ Church University's repository of research outputs

<http://create.canterbury.ac.uk>

Please cite this publication as follows:

Butler, Catherine. A., Lask, J., Sheils, L., Crossley, J., Joscelyne, T. and Pote, H. (2018) Measuring competence in systemic practice: development of the 'systemic family practice – systemic competency scale' (SPS). *Journal Of Family Therapy*. ISSN 0163-4445.

Link to official URL (if available):

<http://dx.doi.org/10.1111/1467-6427.12251>

This version is made available in accordance with publishers' policies. All material made available by CReaTE is protected by intellectual property law, including copyright law. Any use made of the contents should comply with the relevant law.

Contact: create.library@canterbury.ac.uk



Measuring Competence in Systemic Practice:

Development of the 'Systemic Family Practice – Systemic Competency Scale' (SPS)

Abstract

Ensuring that practitioners are competent in the therapies they deliver is important for training, therapeutic outcomes and ethical practice. The development of the Systemic Practice Scale (SPS) is reported - a measure to assess the competence of novice systemic practitioners trialed by Children and Young Person's Improving Access to Psychological Therapies (CYP-IAPT) training courses. Initial reliability assessment of the SPS with twenty-eight supervisors of systemic practice evaluating students' competence using an online recording of a family therapy session is detailed. The SPS was found to be a reliable measure of systemic competence across training settings. Rating variability was noted, with training and benchmarking to improve rating consistency recommended. Further research using the SPS to further establish the reliability and validity of the scale is required.

Practitioner Points

- SPS represents an important tool, particularly for the supervision and development of more junior staff or students
- Initial reliability for use of the SPS as a formative tool has been established. Further benchmarking is required if using the tool in a summative manner.

Introduction

A cornerstone of systemic therapy practice is the delivery of effective, ethical and collaborative interventions to diverse clients by competent therapists. Thorough systemic training to develop competences together with skills in effective outcome monitoring have been essential in

delivering the highest standards of systemic therapy to our clients. Whilst the evidence base for systemic interventions continues to grow (Lebow, 2016) a variety of opinions remain about the validity and usefulness of competence and outcome measures within the systemic community (Moran, 2017). Some are concerned that accurately measuring competences and outcomes may be impossible given the multiple layers of context, multiple client perspectives, therapist variables and a myriad of treatment outcomes that are involved in systemic therapy. Any 'snapshot' of practice may lead to a limiting or reductionist view of practice. Further, a post-modern, social constructionist philosophy raises important questions about what should be measured and how this can be achieved effectively whilst valuing the voice and individual experiences of clients (Boston, 2000).

The public and educational sectors in the U.K. have seen a marked shift towards measuring both competence and outcome. For example, there has been an increasing emphasis on the delivery of evidence-based therapies and the use of routine outcome measures. Family therapists, many of whom work in the public sector in the UK, have recognised that post-modern assumptions are not incompatible with research into outcomes and have responded by developing philosophically coherent frameworks for measuring change (France & Uhlin, 2006). SCORE, for example, is a measure that was developed to evaluate outcome and process changes within families (Stratton et al. 2010). This consideration of 'process' is of particular importance, as the changes targeted in systemic therapy focus on the whole system and are often concerned with larger changes in family functioning such as coping and relationship changes rather than simple symptom improvement in the referred person. Perhaps due to SCORE's face validity, it has become widely used by family therapists in the UK in a broad range of settings over the past decade (Carr & Stratton, 2017).

Assessing Therapist Competences

Related to the emphasis on outcome measurement, a growing priority, particularly for therapy training courses, has been the need to measure, articulate and improve therapist competence, especially in the delivery of research-based approaches. Competence can be defined as the knowledge and skills required to deliver a treatment to the standard needed to achieve its expected effects (Fairburn & Cooper, 2011). As a result, competency-based teaching has become a fundamental requirement for many professional training courses including those for psychology and psychotherapy. For example, accreditation in Clinical Psychology training in the UK is dependent on trainees demonstrating a range of core skills in working therapeutically (The British Psychological Society, 2015). This requirement presents new challenges for course staff and supervisors on how to evaluate clinical competencies in students (Tweed et al., 2010). Psychology training has typically utilised numerous methods to assess learners including case reports and supervisor feedback. However, these methods have been criticised as not being specific enough to distinguish more subtle differences between levels of practice (Tweed et al., 2010). Gallichan and Mitchell (2008) cite in vivo assessment as the most rigorous method to assess student competence and a number of rating scales have been developed for this purpose. Scales such as the Cognitive Therapy Scale-Revised (CTS-R) (Blackburn et al., 2001), Clinical Skills Assessment Rating Form (Tweed et al., 2010) and the Cognitive Behaviour Therapy Scale for Children and Young People (CBTS-CYP) (Stallard, et al., 2014) are increasingly being used by training courses and have been established as reliable and valid.

One limitation of these rating scales is they typically assess students' competence in providing individual therapy, with the majority focusing on Cognitive Behavioural Therapy (CBT). Measures to assess a broader range of evidence-based models such as systemic therapy were required to support training and continuing professional development. It was hoped that such a scale would provide a framework and common language for formative assessment of therapist systemic skills

in practice, which could be used in dialogue with the supervisee to highlight areas of strength and skill development.

Another important motivation for a scale measuring competence in systemic practice came from recent changes in mental health policy within the UK. For the past 15 years, there has been a policy level drive from the UK central government to enhance the availability, accessibility and effectiveness of psychological therapies within mental health care (DoH, 2011). This led to the introduction of the Children and Young People's Improving Access to Psychological Therapies (CYP IAPT) service framework in 2011, with the inclusion of systemic family practice as one of the evidence-based therapies. The systemic family practice CYP-IAPT course trains experienced practitioners (from a variety of different professions) to an intermediate level of competence in work with families (Systemic Family Practice) which is normally the level reached after two of the four years of Association for family Therapy accredited family therapy training.

The new CYP-IAPT training also meant that systemic trainers were positioned in a culture where micro measurement of therapy, in both formative and summative assessment, was the norm. CBT colleagues already had a range of methods available for assessing CBT competence (Muse & McManus, 2013), and were training practitioners to deliver tightly manualised therapies with a strong emphasis on adherence to the manual. It therefore became necessary for systemic therapists¹ to develop a tool that would fit the requirements for the CYP IAPT training and assess an intermediate level of practice, while also remaining in keeping with the strong systemic training tradition and philosophy. There was also a political imperative to show that both training

¹ The terms 'therapist' and 'therapy' are used throughout the paper to refer to those conducting systemic therapy at intermediate, qualifying or qualified level, unless we are particularly referring to CYP-IAPT courses, in which case the term 'practitioner' is used in line with CYP-IAPT accepted practice.

and practice were effective in order to secure further government funding.

Systemic therapy manuals have sought fidelity and consistency with the systemic principles of working in ways that fit with and are useful to families, and have therefore embraced a flexible approach that left space for therapist creativity and decision making (Escudero, 2012; Fruggeri, 2012). Value is placed on a range of practice skills but there is also an acknowledgement that the way choices are made and interventions are linked together is also an important dimension for the developing therapist. An example of this is the Leeds Systemic Family Therapy Manual (Pote et al., 2003). Individual systemic training courses have also developed their own assessment frameworks, typically reflecting broad areas of practice rather than the micro assessment of a particular session. Furthermore, the increasing emphasis on postmodern approaches including social constructionist, narrative and dialogic approaches over the last 25 years has reduced the attention given, on some systemic courses to some of the more 'first order' interventions, however, these interventions were central to the CYP-IAPT curriculum because of their strong evidence-base. Examples of such interventions are the attention given to helping parents and other caregivers to work as a team and in-session interventions to help family members achieve more effective communication (Henggeler & Sheidow, 2012; Diamond et al., 2010; Simic & Eisler, 2012). There was therefore a complex array of factors to take into consideration when developing a measure for use in systemic training within CYP IAPT.

Creating the Systemic Family Practice-Systemic Competency Scale (SFP-SCS)

The creation of the Systemic Family Practice-Systemic Competency Scale (SFP-SCS – now known as the Systemic Practice Scale) was led by one of the authors of this paper (JL) in collaboration with a panel of experts consisting of senior trainers, clinicians and researchers who were part of the CYP IAPT Systemic Curriculum group, and other directors of Systemic courses in the UK.

Guided by research into common factors in psychological therapies (Wampold, 2015), the group based the measure on the structure of the CTS-R (Blackburn et al., 2001), which was already widely used in CBT training and was a key assessment for the CBT modules of the CYP IAPT course. It was informed by intermediate level systemic competencies developed by the Association for Family Therapy (AFT) (AFT, 2015), and by the competency map for Systemic Family Therapy (Roth and Pilling, 2008). The scale development also took into account the summative and formative measures in current use by systemic training courses. The range of competencies that comprise the Systemic Practice Scale (SPS) were agreed by the above expert group with the first draft modified in response to feedback by CYP IAPT trainers. The scale was specifically designed to assess the competencies outlined in the CYP IAPT curriculum for systemic family practice. Following feedback, twelve different areas of competence such as 'collaboration', and 'working with power and difference' were developed. A 7-point scale ranging from 0 for inappropriate use to 6 for highly competent use was added to be comparable with the CTS-R. For example a '2' would be given to a supervisee where there was some evidence of competency, but also examples of unhelpful practice and lack of consistency. And so, for Item 8 'Exploring and managing emotions', the benchmark for a score of 2 is:

'Some questioning about emotions and appropriate reaction and some notice of emotional response in session but inconsistent or limited to particular emotions or family members.'

This would not be considered competent practice at an intermediate level. Whereas to score 6, which would be considered very competence, the student must demonstrate:

'Works positively with a range of emotions in a number of different ways even when the emotional atmosphere in the session is challenging and some family members may want to stifle the discussion. Maintaining a good therapeutic relationship.'

The SPS was introduced to CYP-IAPT training in February 2014, with the aim of reviewing its applicability and identifying areas for development. It was well received within the UK systemic field, which was perhaps not surprising as it was based on extensive work by experienced therapists, researchers and trainers.

Correspondence with CYP-IAPT providers showed that the SPS is currently being used for both formative and summative feedback (including formal pass/fail marking). For example, course tutors and supervisors were regularly using the scale to rate therapy tapes, and trainees on the courses were also using the scale for self-rating. Correspondence from systemic course tutors on Clinical Psychology training courses suggests that courses are using the SPS as a formative tool for use by supervisors and trainees, with only one programme using it for summative purposes –the University of Bath (UoB), where the scale is used as part of their intermediate systemic practitioner accreditation. The UoB requires all students completing intermediate training (which is embedded within their Clinical Psychology training) to submit the SPS passed at 50% as rated by their placement supervisor. The course provides training on the use of the scale to supervisors, both as a live CPD event, and as an audio-recording available to supervisors unable to attend the CPD event. Students might work with their supervisor on aspects of the scale, using it in a formative way to highlight areas of improvement, prior to obtaining a summative assessment that is then submitted to the course. AFT noted the rigor used in assessment methods such as this when they accredited the course. Two courses, (Glasgow and Edinburgh Universities) have also adapted the scale with stakeholder involvement in order to reflect systemic practice that is not family related – for example working with a care home staff team and this new version has had good feedback from supervisors in their region.

In January 2016, an initial usability study of the scale was conducted by JL funded by AFT. In this study, CYP-IAPT staff who had been using the scale for more than 18 months were invited to comment on each of the 12 domains within the SPS and rate their satisfaction with each domain, as well as their satisfaction with the measure overall. Twelve staff responded and confirmed the face validity of the 12 domains. Their feedback resulted in some minor changes to rater instructions and descriptors for different levels of competence

By the time of the current study the scale had been used in CYP-IAPT training for over two years to rate trainees' systemic clinical competence, and it could therefore be said to hold face validity across a variety of training settings. However, whether or not the SFP-SCS was reliable, that is whether it consistently replicated a measurement of competency (DeVon et al., 2007), remained untested at this point.

Aims

This project, therefore, aimed to assess the reliability of the Systemic Practice Scale (SPS) and also assessed the possibility of the scale being useful for other courses, particularly in the training of Clinical Psychologists in the UK. The project was funded by the Division of Clinical Psychology, British Psychological Society.

Design

There were two phases to the study: a pilot and a main project. The pilot phase took place with 10 participants on an SPS training day at the host university. The pilot aimed to test and receive feedback on the assessment task that the main project participants would be asked to complete. During this phase, verbal feedback on the scale itself was also gathered. The main project was conducted online from a national pool of participants. The main project used mixed-methods:

scores on the scale provided quantitative data and comment boxes collected qualitative data.

Because the pilot and main project used different therapy videos the quantitative data from these two phases could not be pooled but is reported separately.

Method

Participants.

Pilot: Participants were recruited from an SPS training event ($n = 10$). The sample consisted of regional supervisors of Clinical Psychology trainees and CYP-IAPT training staff.

Main Project: An invitation email with the participant information sheet was distributed amongst three different groups: supervisors of Clinical Psychology trainees, CYP-IAPT training staff, and trainers of systemic therapy students. Fifteen Intermediate Courses were contacted reaching 34 training staff, and an email was also sent to the clinical psychology supervisors mailing list of the host university (with approximately 350 recipients). Fifty information sheets were also handed out at the annual AFT conference. Several contacts personal to the research team also further distributed the Participant Information Sheet, therefore it is not possible to identify the exact number of individuals who read this information.

A total of 58 individuals expressed interest in the research, 45 accessed the online survey and of these, 28 individuals completed the study. Seven participants had used the SPS before and participants had a range of systemic qualifications ($n = 2$ had no formal qualification, $n = 1$ Foundation level qualification (Year 1), $n = 3$ Intermediate level qualification (Year 2), $n = 19$ Qualifying level qualification (Year 3 & 4) and $n = 3$ in training at Qualifying level). All participants had previous experience of supervising students' (clinical psychology or CYP-IAPT) practice in systemic therapy.

Measures

The SPS is a questionnaire investigating the 12 following areas: interpersonal effectiveness; convening and managing the session; collaboration; conveying a systemic view of family life and the wider context; conceptual integration; use of questioning; use of feedback; intervening in process during the session; working with power and difference; managing emotions in sessions; use of change techniques and incorporating the outside world. Each item has a seven-point descriptive Likert rating scale using whole numbers from 0 (where the practitioner does not demonstrate that skill) to 6 (where a high level of the skill is demonstrated). Each Likert point includes a general description of the competency and some guidance/illustration about what might be seen at different levels.. For example for a '4' in the area of use of questioning there might be good circular questioning and questions that are differentiated for different family members, whereas a '2' might be given to a therapist using only closed and interrogatory questions Raters are required to score the therapist on each of the 12 items. For those using the scale for summative information the 'pass' score for intermediate level of training would be to receive a 3 or above in all areas with no area scoring less than 2.

Procedure

Pilot: Individuals attending a training event in use of the SPS were given an information sheet, consent form and time to ask questions. All individuals watched an hour of systemic family therapy during the training session and then individually rated the therapist (a clinical psychology trainee) using the SPS. Those who wished to participate in the pilot study completed the consent form and provided the research team with their completed SPS; all workshop participants consented to be part of the pilot.

Following discussion and feedback during the pilot, changes were made to the SPS:

- The benchmark description for scoring was improved, highlighting that most practitioners will score 3 or 4, and an average score of 3 should be considered the minimum for students reaching the level of clinical competence required to complete a CYP-IAPT/Intermediate systemic course. A '3' would describe a competent practitioner with some problems or inconsistencies in their practice.
- The scale instructions outlined that a score of 3 should be used if an item is appropriately not covered at all. This was limited to twice in one session.
- The items on the 'collaboration' scale were altered.
- 'Use of Questioning' was moved to appear earlier.
- An item was renamed, from 'Exploring and managing emotions in sessions' to 'Eliciting and managing Emotions'.
- An item was renamed, from 'Conceptual Integration' to 'Conceptualisation'.

An online version of the questionnaire was created with the changes above for the main stage of the research project.

Main Project: Data was collected throughout June to September 2016. Material was presented online and included an audio recording of the SPS training event held at the host university, a 60 minute video recording of a systemic therapy session with a therapist in his third year of training in Systemic Family Therapy and a link to an online version of the SPS (the consent form, survey and debrief information).

Potential participants were sent an email with the participant information sheet and encouraged to contact the researcher if they had questions or wanted to discuss the research. Participants

read and signed a Responsible Use Agreement (RUA). The RUA was designed to ensure the therapy video on the password protected platform was only viewed by the participant, not downloaded or shared with others and not discussed with anyone other than the researchers. Once the RUA had been signed, the participant was allocated a time limited username and password to access the research material.

Participants completed an online consent form, listened to the training audio, watched the 60-minute therapy recording of the trainee family therapist, answered demographic questions and completed the online SPS. Participants also had an optional question asking about their thoughts and experiences of completing the SPS. Once participants had completed the tasks they accessed the Debrief Information, which included the researcher's contact details again. Participants were offered a £25 voucher to be sent to them after data collection. Participants wanting to receive the voucher completed a separate online survey to enter their name and address. The voucher details were stored separately to ensure anonymity of the research data.

Ethics

Ethical permission was granted by the University of Bath's Department of Psychology Ethics Committee (reference number, 15-247). The main ethical issue was protecting the video family's confidentiality which was ensured by the methods described above.

Results

Pilot

Pilot data involved 10 ratings of the single video shown at the SPS training day. The mean total score for the SPS was 29.15 (SD 9.72, minimum score = 10 and maximum score = 39). Figure 1:

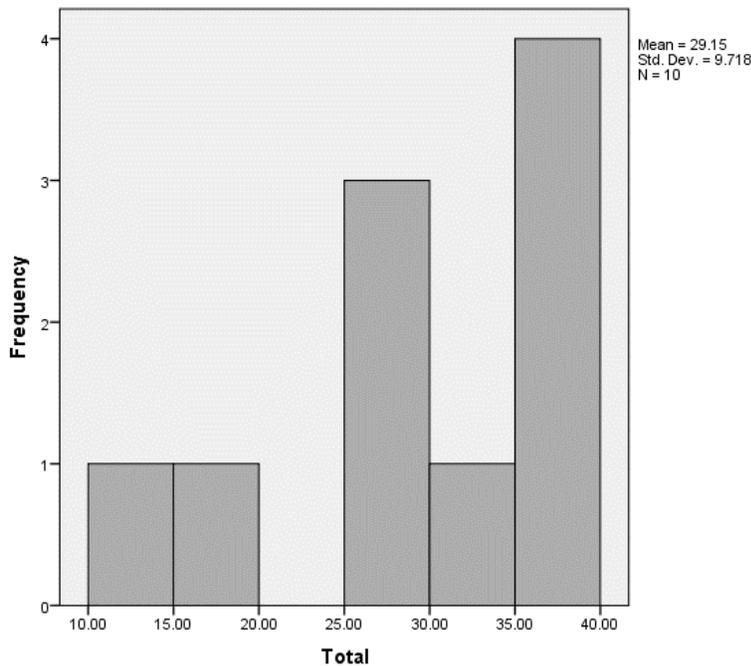


Figure 1. Histogram of scores across 10 raters for the pilot.

Lowest score obtainable = 0, highest score obtainable = 72.

40% would have given the therapist a pass score with a further 10% a borderline score.

Cronbach's Alpha was .96 indicating high internal reliability (this value is the same as consistency Intraclass Correlation (ICC) with Average Measures). Cronbach's Alpha did not improve if any of the 12 items were removed. When several raters (10 participants) assess the same target (the therapy video), an ICC with two-way mixed effects and absolute agreement is appropriate (Shrout and Fleiss, 1979). A high degree of inter-rater reliability was found. The Average Measures ICC was .92 with a 95% confidence interval from .82 to .98 ($F(9, 99) = 25.16, p < .001$)

Although the scale was found to have good inter-rater reliability, there was some variation in scores between raters for each item, as illustrated in Figure 2:

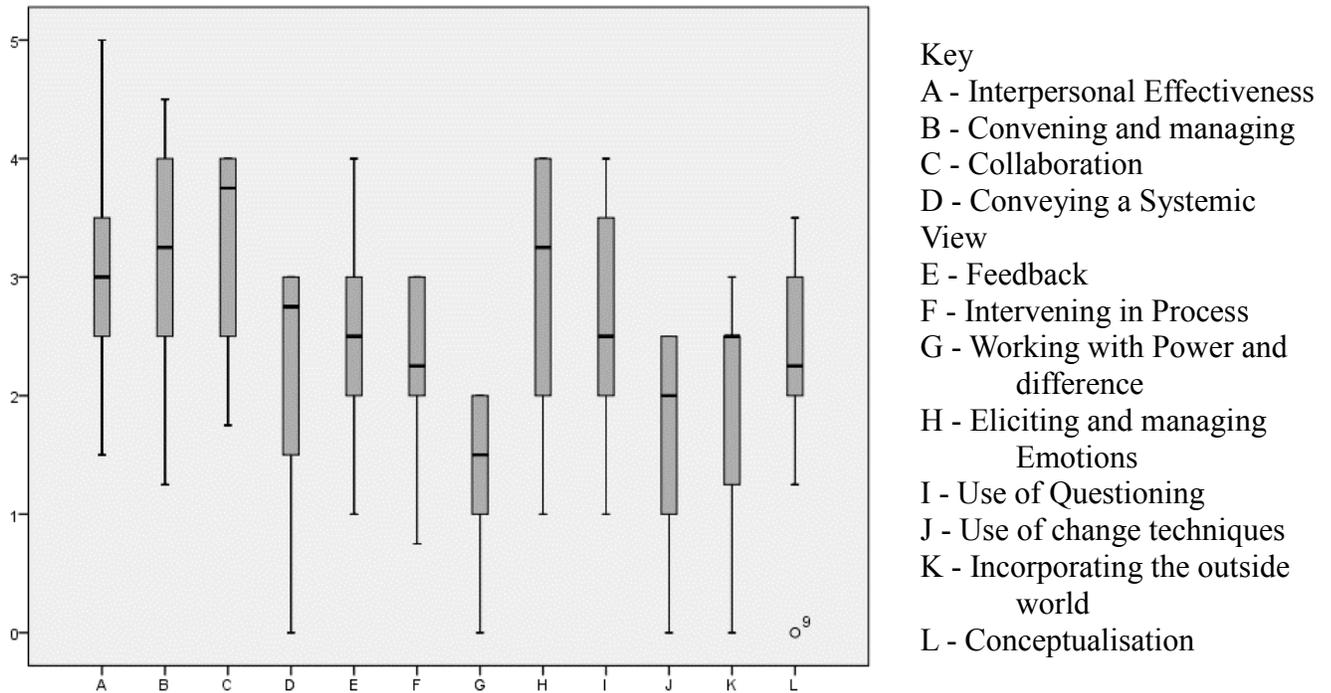


Figure 2. Scores across items across raters for the pilot.

Main Project – Quantitative Results.

Data from 28 participants was obtained. The mean total score for the SPS was 41.60 (SD = 9.24, minimum score = 20.0 and maximum score = 61.0), Figure 3 shows the spread of scores across raters:

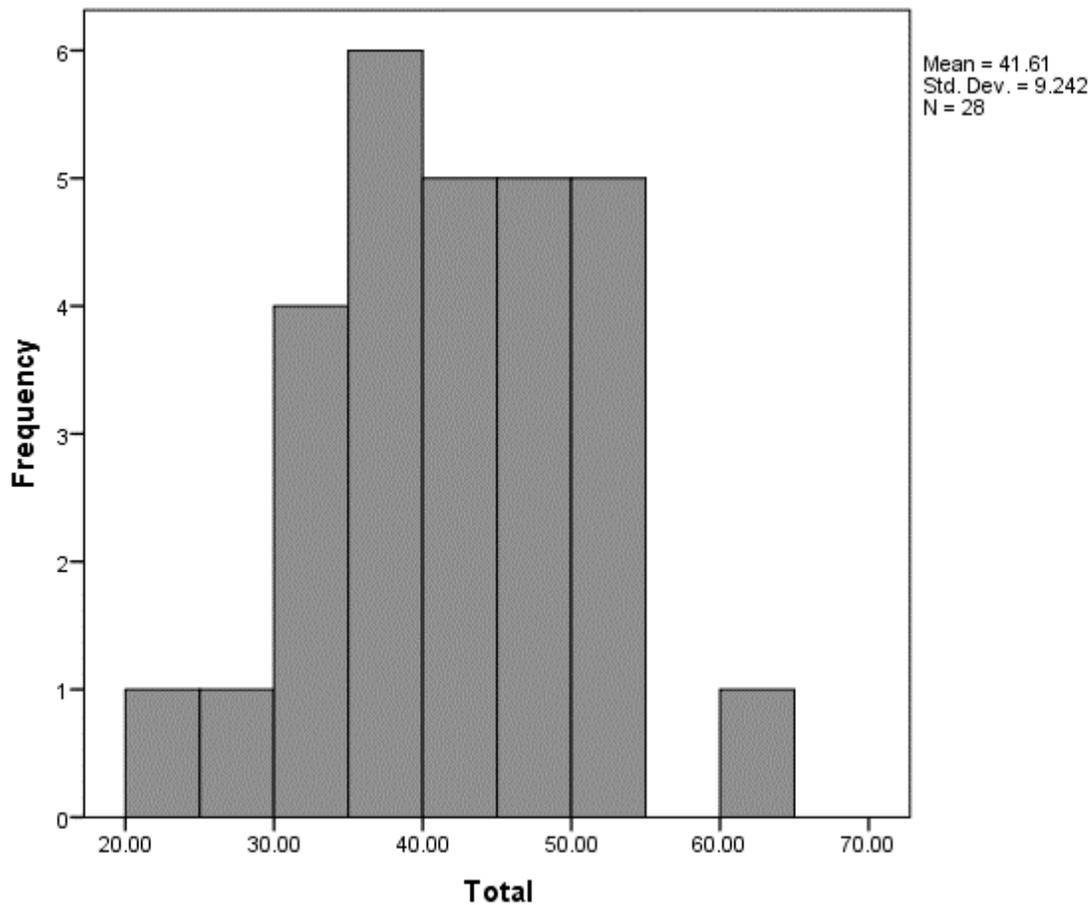


Figure 3: Histogram of scores across the 28 raters for the main study.

Lowest score obtainable = 0, highest score obtainable = 72. 78% of raters would have given the therapist a pass with 14% borderline.

Cronbach's Alpha was .95 indicating a high degree of internal reliability. Cronbach's Alpha did not improve if any of the 12 items were removed. An Intraclass Correlation (ICC) was conducted (two-way mixed effects with absolute agreement). A high degree of inter-rater reliability was found. The Average Measures ICC was .94 with a 95% confidence interval from .89 to .97 ($F(27, 297) = 20.36, p < .001$).

Similar to the findings during the pilot there was variation in scores between raters for each item, as illustrated in Figure 4:

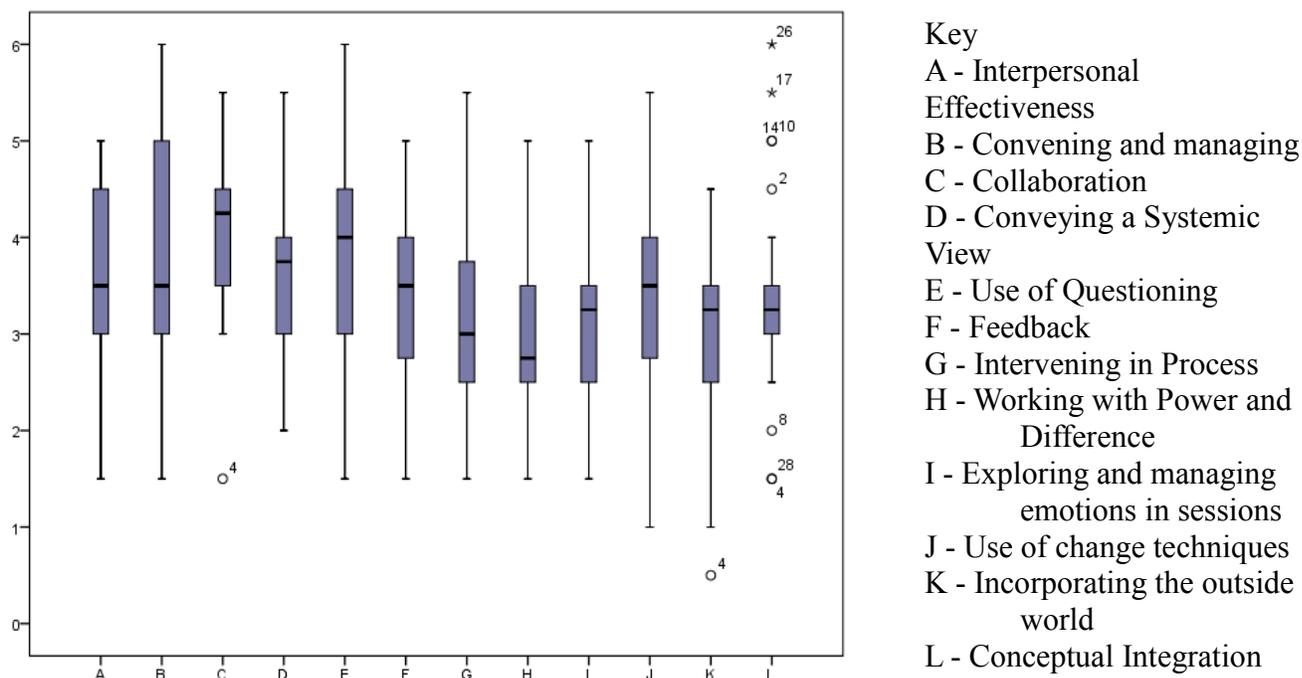


Figure 4. Scores across items across raters for the main study.

Due to the small number of participants, statistical tests were not conducted to see if there was a relationship between scores on the SPS and the level of participants' training.

Main Project - Qualitative Results

There were responses from 23 participants on the optional free text box. Inductive Thematic Analysis (Braun and Clarke, 2006) was conducted independently by two members of the research team. The two researchers discussed their findings and agreed upon 3 main themes: useful, improvements, fit for purpose.

Theme 1: Useful

Within this theme several sub-themes emerged. Participants indicated that the SPS was clear and an easy measure to use: 'It is compact and easy to complete'. Participants also suggested that the scale provides structure, 'helpful in structuring my feedback and enable me to be analytical in a systematic way'. Furthermore, the SPS appeared to support evaluation and reflection, 'it has been a useful exercise to reflect on my own clinical and supervision practice' and one participant stated the scale would be 'helpful for trainees to complete their own rating scale as comparison with supervisor analysis'.

Amongst these useful benefits participants also expressed their desire to apply the scale in their usual practice, 'in future for supervision of clinical practice for my trainees or clinical colleagues'.

Theme 2: Improvements

Within this theme, participants made a range of suggestions including rewording some areas of the SPS for increased clarity, 'conceptual integration could have come earlier since it is the road map that should guide the session'. Additionally, some participants indicated the need for feedback beyond the scale, 'it felt strange to just give a numerical mark without written feedback'. Participants also mentioned the need for clarity regarding the inclusion of the reflecting team in the therapy video, 'I scored the team rather than the individual therapist as they provided the idea for the change technique and intervened quite a lot in the session'. Participants did, however, explain that 'if it [the SPS] were both valid and reliable I would be happy to use it'.

Theme 3: Fit for Purpose?

Within this theme, participants were contemplating whether the scale would be suitable for all systemic models, 'If this scale were to be used with a wide variety of family therapy models I would think that attention to the descriptors of each mark would be necessary'. Furthermore, they highlighted the need for training and benchmarking in using the SPS effectively, 'if I was to continue using it, I would probably try to team up with a colleague and co-rate initially to have some idea of a suitable baseline'. One participant explained 'if I hadn't watched the teaching video I would have struggled to know what was 'good enough' in this context'.

Changes to the scale

Following the qualitative feedback, a comment box was added after each item. Additionally, the conceptual map item was moved further up (after items on relationship and general systemic reframing). Wording of some descriptors was made clearer and the role of the supervising team was made explicit in the scoring section of the scale: the team's practice was not scored but an optional qualitative box was added to the end of the scale to comment on the practitioner's ability to effectively make use of supervisory comments and interventions from the reflecting team and/or co-therapist.

Discussion

The SPS was found to be a reliable measure of systemic competence in both the pilot and main stage of the current study. However, there was significant variation in scores between raters for each item with outliers at both ends of the scale. This was not surprising given the wide variation in training and experience amongst the raters. However, this is also a similar finding to Stallard et

al.'s (2014) study, where the total scores of the 12 raters assessing the same video ranged from 19%-68% on the CBTS-CYP and 24%-70% on the CTS-R. The authors suggest that reading the scoring instructions alone does not appear sufficient to ensure agreement of level between raters. Stallard et al. (2014) showed that through consensual group conversations, raters raised or lowered their scores by considerable margins. Stallard therefore suggest that future studies on the CBTS-CYP should explore the effect of consensual group discussions on reducing variations between raters. This finding was similar amongst our sample, with participants expressing the need for benchmarking. We would therefore recommend that when first using the scale, raters discuss scores with a colleague based on the criteria laid out in the descriptors in order to calibrate their ratings. A further way of helping with benchmarking would be for training sessions on the SPS to include material of therapists with different levels of skill so that supervisors become familiar with agreed levels of competency before using the scale. This advice would be particularly pertinent when the scale is used for summative evaluation as with the CYP-IAPT courses particularly if different markers are marking different students. In fact, until markers become used to the scale, it may be advisable to have double marking, or use a moderator who has access to a sample of the original student video material.

The qualitative feedback indicates that when using the SPS in training supervisors must retain a flexible approach to competence assessment, using it both formatively and summatively.

Approaches to training which focus on competence assessments are not without limitations.

Questions remain about the validity of judging trainees on therapy competences due to the lack of evidence demonstrating that therapist competence is actually associated with positive therapy outcomes. A recent meta-analysis found that neither adherence to nor competence in a model of therapy was related to patient outcome (Webb, DeRubeis & Barber, 2010). This review was across 36 studies and included a range of therapy modalities including CBT, Interpersonal Therapy, and

Dynamic Therapy. The authors do however emphasise that the methodological limitations of the individual studies, as well as the process of pooling effect sizes across studies, may be at least partly responsible for the lack of a positive relationship.

The authors also recognise the potential impact of therapist responsiveness on patient outcome. Potentially, therapists may reduce their levels of adherence to a model in order to increase their responsiveness to patients, rather than delivering predetermined interventions. It is argued that this may lessen the level of adherence and competence that is measured, but may actually improve outcome. This argument highlights the importance of measures of competence defining characteristics of good practice in a manner that also permits the flexibility to respond in ways that are helpful and fit with the family, as is consistent with systemic principles (Cecchin, 1992). The SPS has been developed to encourage such flexibility, because therapists are scored higher for being collaborative and listening and working with feedback (for example, under Feedback competencies a 6 includes 'good flexibility in adapting to family style'). It will therefore be important for future research to investigate how scores on the SPS relate to clinical outcome. Arguably, the SPS may also become a measure which could become useful to those researching the link between competence and therapy outcomes as previously there were few observational measures to assess competency in systemic skills.

Supervisors should also use the SPS critically. Within the field of systemic therapy specifically, there has also been discontent with attempts to quantify elements of systemic therapy because of the requirement to be adaptable, work from feedback and fit the therapy with needs of the client(s), taking into account culture and family dynamics. There is thus an emphasis on creativity and encouraging trainee therapists to find their own style and voice. This has contributed to high levels of suspicion about the emphasis on applying evidence-based methods for competence

assessment, despite key leaders in this field emphasising its importance and necessity for the profession (i.e. Stratton, 2016). In further assessments of the SPS it will be important for supervisors to note if any competences and creative practices are being lost by the focus on twelve specific systemic competence domains.

The twelve items on the SPS have been designed to be broad and flexible enough to fit with a vast array of systemic models, including models such as narrative and dialogical therapy which embrace a social constructionist epistemology and eschew more a directive approach. Items such as 'working with power and difference' and 'collaboration' fit readily here, and even less immediately obvious items such as 'use of change techniques' can be understood to be consistent with the decentred and influential stance of narrative therapy (White, 1997). It may merely be necessary to adapt the examples of competency at each Likert point, for example to shift the focus from circular questions to relative influence questions within the 'use of questioning' item in order to be consistent with a narrative approach (White & Epston, 1990). The item 'conceptual integration' in many ways epitomises the aim of the scale, which is not to allow one specific model to be followed but rather a coherent approach to be established, whether that be 'first order' or postmodern.

Strengths and Limitations

A significant challenge during this research study was recruitment. The original design was to recruit via the DClinPsy supervisors' mailing list of the research team, 15 systemic intermediate training courses and all CYP-IAPT courses. Due to minimal responses the Association of Family Therapy (AFT) sent information about the research to their mailing list and an announcement was made at the Annual AFT conference. Despite interest from 58 individuals only 28 completed the research. Thus, funding and pragmatic constraints of this research project meant that the data

collection phase was shorter than ideal and included a wider variation of training and experience with the SPS than originally planned. Future research would benefit from engaging multiple training centres over a longer period of time to increase the sample size and representativeness of the sample. With a larger sample, further analysis could explore if the scale was used differently by therapists with different qualifications and experience.

Future Research & Recommendations

Despite the small sample size, the SPS was used on two different therapy videos and similar Cronbach's Alpha and ICC were found. This indicates that the scale could be reliable across different therapists, as well as with different raters. However, given that there were some outlier ratings in both studies (either lower or higher scoring than the main body of participants) it may be that future research could ascertain if there were some supervisors and trainers for whom the scale would not be suitable. This would be particularly pertinent within Clinical Psychology training, where supervisors without full training in systemic therapy may be doing the rating.

Given the current variability in scores between raters, the authors suggest that if the scale is used for formal examination purposes with parallel markers, raters should spend time agreeing benchmarking before independently rating students' competencies. Training in the scale ideally should include live examples of therapists at different skill levels. The raters should have an opportunity to discuss their marking and reach a consensus before returning marks to trainees. For example for parallel raters, we recommend that discussions happen prior to marking to distinguish points close together on the scale. For example to discuss what might constitute 'minimal' (score 5) vs 'minor' (score 4) problems in 'applying a range of creative change techniques'. As with most marking, there is a degree of subjectivity in judging competencies and

markers who work closely together frequently are more likely to become aligned in their rating scores.

On the other hand, the scale in its present form is very helpful in providing a way for trainers and supervisors to give feedback on skills in a summative way, to notice gaps in skill areas, and for trainees to be able to rate themselves as a way of self-assessing their practice.

The overall difference in average scores between the pilot and main studies in which the two video therapists differed in experience and training, does suggest that the scale is able to distinguish between therapist on levels of competence. In this study, the therapist with three years training was rated by the majority of raters as achieving a pass score, whereas the therapist with less formal training was only passed by fewer than half the raters. However, future research on this would be useful, for example by having the same group of participants rate three therapist videos (Foundation, Intermediate and Qualified systemic therapists). This research would be helpful in order to know if, despite some variability in the exact scores between raters, there was a reliable difference in scores for therapists at different levels or stages in training.

In addition, although reliability is a necessity, validity also needs to be measured (Cook and Beckman, 2006) as it may be that raters can agree on scores, but that the scale does not in fact rate therapist competence, but some other related variable such as therapist intelligence. For example predictive validity could measure trainees' scores on future achievement in their courses, and construct validity could explore in more detail if the rated qualities would be predicted from systemic theory in the area of training – for example there may be some therapy competences that are more universal across therapy models, whereas others would only be present in systemic therapists.

The scale was also developed with a particular model of training, and it would be helpful to know if it was relevant to different theoretical models of systemic practice and therefore applicable to a wide range of systemic training courses.

One possible use of the scale in a wider future research project might be to add to the knowledge about how therapist's skills and adherence are related to therapy outcome. The lack of suitable scales to measure these factors in systemic therapy has meant that the literature has been skewed towards individual therapy. With CYP-IAPT services using more outcome measures as routine, there may be an opportunity to match therapists' skills to outcome in ways that would not have been possible previously.

In addition, the SPS includes assessment of the therapist's adaptability to the context and incoming information (something that is considered essential in many systemic theoretical ideas). This 'adaptability' score might be used to tease apart the variables of therapist responsiveness and therapist adherence to help researchers understand more about the therapist variables associated with therapy outcome.

Conclusion

In conclusion, we have established the SPS to have face validity and be a reliable measure of systemic competence. We have improved the face validity of the scale based on qualitative analysis of participants' feedback. Considering the findings of the current study, we recommend the use of the SPS in the assessment of systemic students' competences during training particularly as a reflective/feedback tool. If used for summative assessment, we suggest that

courses consider some further benchmarking by parallel markers discussing and highlighting examples of practice at different levels before marking.

References

Association of Family Therapy (2015). *Blue Book 4th edition*. www.aft.org.uk

Blackburn, I., James, I. A., Milne, D. L., Baker, C., Standart, S., Garland, A., et al. (2001). The revised cognitive therapy scale (CTS-R): Psychometric properties. *Behavioural and Cognitive Psychotherapy*, 29, 431–446

Boston, P. (2000). Systemic family therapy and the influence of post-modernism. *Advances in Psychiatric Treatment*, 6, 450-457

Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-10

Carr, A. & Stratton, P. (2017). The score family assessment questionnaire: a decade of progress. *Family Process*, 56(2), 285-301

Cecchin, C. (1992). Constructing therapeutic possibilities. In S. McNamee & K. Gergen (Eds.), *Therapy as social construction*. Newbury Park, CA: Sage

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *The American journal of medicine*, 119(2), 166-e7

Department of Health (2011). *No health without mental health*. London: Department of Health

DeVon, H. A., Block, M. E., MoyleWright, P., Ernst, D. M., Hayden, S. J., Lazzara, D. J., & Kostas-

Polston, E. (2007). A psychometric toolbox for testing validity and reliability. *Journal of Nursing scholarship*, 39(2), 155-164

- Diamond, G. S., Wintersteen, M. B., Brown, G. K., Diamond, G. M., Gallop, R., Shelef, K., & Levy, S. (2010). Attachment-based family therapy for adolescents with suicidal ideation: A randomized controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 49, 122-131
- Escudero, V. (2012). Reconsidering the 'heresy' of using treatment manuals. *Journal of Family Therapy*, 34 (1), 106-113
- Fairburn, C.G. & Cooper, Z. (2011). Eating disorders, DSM-5 and clinical reality. *The British Journal of Psychiatry*, 198(1), 8-10
- France, C.M. & Uhlin, B.D. (2006). Narrative as an outcome domain in psychosis. *Psychology and Psychotherapy: Theory, Research & Practice*, 79, 53-67
- Fruggeri, L. (2012). Different levels of psychotherapeutic competence. *Journal of Family Therapy*, 34(1), 91-105
- Gallichan, D., & Mitchell, A. (2008). Making the case for *in vivo* assessments of clinical competence. *Clinical Psychology Forum*, 184, 18-22
- Lebow, J. (2016) Editorial. Special Issue: Empirically Supported Treatments in Couple and Family Therapy. *Family Process*, 55(3), 385-389
- Henggeler, S., & Sheidow, A. (2012). Empirically supported family-based treatments for conduct disorder and delinquency in adolescents. *Journal of Marital and Family Therapy*, 38, 30-58
- Moran P. (2017) Selecting Self Report Outcome Measures for use in Family and Systemic Therapy. *Journal of Family Therapy*, 39, 41-46
- Muse, K. & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review*, 33, 484-499
- Pote, H., Stratton, P., Cottrell D., Shapiro, D., & Boston, P., (2003). Systemic Family Therapy can be manualized: Research Process and findings. *Journal of Family Therapy*, 25, 3, 236-252

- Roth, A. D., & Pilling, S. (2008). *A competence framework for the supervision of psychological therapies*. Retrieved from http://www.ucl.ac.uk/Clinical-Psychology/CORE/Competence_frameworks.htm, December, 18, 2016
- Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428
- Simic, M. & Eisler, I. (2012). Family and multifamily therapy. In J. Fox & K. Goss (Eds.), *Eating and its disorders*. Oxford, Willey-Blackwell
- Stallard, P., Myles, P., & Branson, A. (2014). The cognitive behaviour therapy scale for children and young people (CBTS-CYP): Development and psychometric properties. *Behavioural and cognitive psychotherapy*, 42(03), 269-282
- Stratton, P, Bland, J., Janes, E & Lask, J. (2010) Developing a practicable outcome measure for systemic family therapy: The SCORE. *Journal of Family Therapy*, 32, 232-258
- Stratton, P. (2016). *The evidence bases of family therapy and systemic practice*. United Kingdom: Association of Family Therapy
- The British Psychological Society. (2015). *Standards for the accreditation of Doctoral programmes in clinical psychology*. Leicester: British Psychological Society
- Tweed, A., Graber, R., & Wang, M. (2010). Assessing trainee clinical psychologists' clinical competence. *Psychology Learning & Teaching*, 9(2), 50-60
- Wampold, E. (2015) How important are common factors in psychiatry – an update. *World Psychiatry*, 14, 270-277
- Webb, C.A., DeRubeis, R.J. & Barber, J.P. (2010). Therapist adherence / competence and treatment outcome: a meta-analytic review. *Journal of Consulting and Clinical Psychology*, 78(2), 200-211
- White, M (1997) *Narratives of Therapists' Lives*. Adelaide: Dulwich Centre Publications
- White, M. & Epsom, D. (1990). *Narrative means to therapeutic ends*. New York: Norton

1. In its initial validation the scale was titled the Systemic Family Practice – Systemic Competency Scale but is now referred to as the Systemic Practice Scale (SPS). [↑](#)