Received: 22 January 2022 | Revised: 12 June 2022 | Accepted: 14 June 2022

DOI: 10.1002/cpp.2763

RESEARCH ARTICLE

WILEY

# A behaviour-based coding tool for assessing supervisors' adherence and competence: Findings from a motivational interviewing implementation study

Maria Beckman[1] | Sven Alfonsson[1] | Ingvar Rosendahl[1] | Anne H. Berman[1,2] | Helena Lindqvist[1]

[1]Center for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, and Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden

[2]Department of Psychology, Clinical Psychology, Uppsala University, Uppsala, Sweden

**Correspondence**
Maria Beckman, Karolinska Institutet, Department of Clinical Neuroscience, Stockholm 171 77, Sweden.
Email: maria.beckman@ki.se

## Abstract

Supervision seems to be an essential part of therapist training and thus also of implementing evidence-based practices. However, there is a shortage of valid and reliable instruments for objective assessment of supervision competence that include both global measures and frequency counts of behaviour—two essential aspects of supervisory competence. This study tests the internal consistency and inter-rater reliability of an assessment tool that includes both these measures. Additionally, strategies and techniques used by 10 supervisors in 35 Motivational interviewing supervision sessions are described. Codings were conducted after two separate coding training sessions. The internal consistency across the global measures was acceptable ($\alpha = 0.70$; 0.71). After the second training, the inter-rater reliabilities for all frequency counts were in the moderate to good range, except for two that were in the poor range; inter-rater reliability for one of the four global measures was in the moderate range, and three were in the poor range. A prerequisite for identifying specific supervisor skills central to the development of therapist skills, teaching these skills to supervisors and performing quality assurance of supervision, is to create instruments that can measure these behaviours. This study is a step in that direction.

KEYWORDS

clinical supervision, ESSA, fidelity, integrity, motivational interviewing, reliability

## 1 | INTRODUCTION

Clinical supervision and consultation is often recognized as a central part of therapist training and thus also as a cornerstone of implementation and dissemination of evidence-based practice (EBP) (American Psychological Association, 2015; Tugendrajch et al., 2021). The terms are sometimes used interchangeably. However, supervision is more often referred to when clinicians on a continuous basis are supervised within their organization, while the term consultation more often refers to external support provided within dissemination and implementation efforts (Nadeem et al., 2013). Since the term supervision is more generally applied, in this work, that term is used throughout. Historically, few systematic studies have been conducted on the effects of clinical supervision on therapist competence and/or client outcome. However, there is some research, of varying quality, that provides some support for supervision effects on supervisees (Alfonsson et al., 2018; Barrett et al., 2020; Barwick et al., 2012; Bradley & Becker, 2021; de de Roten et al., 2013; Hoge et al., 2011;

Kuhne et al., 2019; Madson et al., 2019; Milne et al., 2011; Park et al., 2019; Schwalbe et al., 2014; Spence et al., 2012). Yet, a recent review of reviews of research on clinical supervision from 1995 to 2019 concluded that the evidence supporting supervision impact on supervisee and client outcomes is weak at best, and that methodological problems are still hampering the field (Watkins, 2020). These methodological problems are often described as a lack of random assignment and control conditions, reliance on self-report measures, as well as a lack of objective observational approaches and supervision instruments that are reliable and valid (Bearman et al., 2017; Madson et al., 2019; Watkins, 2020).

### 1.1 | Measures of supervisor competence

A plethora of instruments measuring different aspects of supervision practice and process exists (Gonsalvez, 2021; Ottman et al., 2020; Wheeler & Barkham, 2014). Most of these instruments have been used in only one study and focus on only one or a few aspects (e.g., knowledge, attitudes, technical/relational skills, working alliance, meta-competencies, reflective practice and/or professional well-being) (Gonsalvez, 2021). Additionally, most measures consist of questionnaires designed to be used by supervisees, peer supervisors and/or supervisors themselves, and only a few of them have established psychometric properties (Gonsalvez, 2021; Gonsalvez et al., 2020). However, some observational tools for evaluating supervision have been reported: *Teacher's PETS* (Milne et al., 2002), *the Supervisory Competency Scale* (SCS) (Kennerley et al., 2010), *the short Supervision: Adherence and Guidance Evaluation: SAGE* (Short-SAGE) (Reiser et al., 2018) and *The Moeller, Moerch, Rosenberg Supervision Scale* (MMRSS) (Moeller et al., 2020). Most of these instruments are extensive and contain many behavioural categories and items, which can make them difficult to use. Additionally, none of them include any frequency counts of behaviour.

Objective behavioural frequency counts and the more subjective ratings of categories in global rating scales both have their advantages and disadvantages. Frequency counts simply mean tallying every time a behaviour occurs during a certain period of time. They are relatively easy to use and can reflect occurrence of behaviours in an intuitive, straightforward way. Frequency counts can also provide a direct measure of the amount of a behaviour performed over a period of time—usually central for training of supervisors as the goal is often to increase or decrease a certain behaviour. Global rating scales instead involve classifying several different behaviours into discrete categories, in order to judge overall performances in each category in accordance with a predetermined scale. The methods estimates are often highly correlated (Kramer Schmidt et al., 2019; Moyers & Manuel, 2016; van der van der Vleuten et al., 2010), but research also shows that frequency counts tend do better in attaining good reliability (Gill et al., 2019; Kramer Schmidt et al., 2019; van der van der Vleuten et al., 2010) and that global ratings usually better discriminate between levels of expertise (Norman, 2005; van der van der Vleuten & Schuwirth, 2005; van der van der Vleuten et al., 2010).

**Key Practitioner Message**

- There is a lack of valid and reliable instruments for objective assessment of supervision that include both global measures and frequency counts of behaviour.
- *The Evaluation of Supervisory Skills and Adherence manual* (ESSA) is a relatively short supervision assessment tool with acceptable psychometric properties that can be used within all forms and types of supervision.
- The instrument captures key elements in evidence- and competence-based frameworks for supervision and can be used as a self-assessment tool for supervisors or as an objective measure of supervision integrity in teaching, research and quality assurance of supervision.

In order to operationalize, evaluate and standardize specific supervisor competencies, valid, reliable and user friendly assessment tools that include both global measures and frequency counts of behaviour are highly needed. *The Evaluation of Supervisory Skills and Adherence manual* (ESSA) is a new behaviour-based coding system that provides both these measures, developed as a tool for evaluating supervisors' skills (i.e., the skillfulness with which the supervisory intervention is delivered) and adherence (i.e., the degree to which the supervisory intervention is delivered as intended). The primary aim of the current study was to test the internal consistency and inter-rater reliability of ESSA, and to describe the strategies and techniques used by supervisors in a Swedish Motivational interviewing (MI) implementation study.

## 2 | METHOD

### 2.1 | Procedure

Data in this study were retrieved from a MI implementation study conducted in five county councils across Sweden (ClinicalTrials.gov: NCT01197027). In the original study (Beckman, Forsberg, et al., 2017b), 174 practitioners were randomized to regular county council MI training or regular county council MI training followed by six monthly sessions of individual telephone MI supervision. Written informed consent was obtained from all practitioners (the Regional Ethical Review Board in Stockholm, Sweden; 2012/2195–31/5). The MI supervision group ($n = 98$) recorded seven 20-min telephone MI sessions together with actors playing standardized clients, followed by supervision after six of these. The supervision sessions were 30 min long and manual-based. The manual comprised an introduction, a collaboratively sat agenda, reviewing of homework and activities for supervisees' reflection and practice. Details of the manual are described elsewhere (Beckman, Forsberg, et al., 2017b). A total of 12 trained MI supervisors at the Motivational Interviewing Quality Assurance (MIQA) group at Karolinska Institutet in Sweden worked with the monthly supervision sessions. At monthly intervals

throughout the study period, based on self-selected recordings of their supervision sessions, the MI supervisors also received group supervision lead by a senior MI supervisor at MIQA. This study used a random sample of 35 of the self-selected supervision session recordings, provided by 10 of the MI supervisors.

## 2.2 | Instrument

The ESSA manual was developed as a tool for structured feedback to MI supervisors, both for group and individual sessions, and as a measure of supervision integrity in clinical MI-trials. However, the manual can be used within all forms and types of supervision. It includes both global measures and frequency counts of supervisors' behaviour, and is based on the components included in evidence- and competence-based frameworks for supervision (e.g., American Psychological Association, 2015; Borders et al., 2014; Falender & Shafranske, 2014; Martino et al., 2008; Milne & Reiser, 2012; Roth & Pilling, 2007; The Psychology Board of Australia, 2013). The conceptualization and development of the manual were also inspired by the MI coding instruments produced by the University of New Mexico Center on Alcoholism, Substance Abuse and Addictions at The University of New Mexico (e.g., Moyers & Manuel, 2005, 2016). ESSA assesses two dimensions of supervisors' skills and adherence: (1) *Specific Supervisor Behaviours*, which reflect the frequency of the supervisors' utterances coded in eight behavioural categories: *Objective Monitoring* (when the supervisor refers to actual behaviours in some form); *Educating* (when the supervisor provides method-related knowledge); *Prompting* (when the supervisor nudges the supervisee in the right direction); *Modelling Skills* (when the supervisor provides demonstrations of skills); *Eliciting Skills or Potentials* (supervisor's questions in order to promote self-reflection in areas of proficiency); *Positive Feedback* (when the supervisor verbally reinforces supervisee behaviours); *Corrective Feedback*, (when the supervisor highlights behaviours that can be improved); and *Other Utterances* (all supervisor utterances that do not fall under any other category); and (2) *Global Supervisor Behaviours*, which reflect the overall assessment of the supervisor's performance on a five-point scale in four different categories: *Structures and Directs the Session* (how the supervisor directs the session structure and content); *Specifies a Training Focus* (how the supervisor clarifies and organizes supervisee practice); *Performs Active Training* (how the supervisor performs active leaning activities); and *Promotes a Learning Environment* (how the supervisor works with the supervisees' sense of competence, intention to use the method, and aspiration to continuously develop and improve). Both the *Specific Supervisor Behaviours* and the *Global Supervisor Behaviours* are coded during a single review of a recorded supervision session. If the recording is 20–30 min long, the entire session is coded. If the recording is longer, a maximum of 30 min of the session is coded. The length of the coded segment is specified in the protocol. The ESSA manual and protocol can be downloaded and used without cost at: https://www.miqagruppen.org/pdf/manualer-protokoll/manual-essa-19-eng/ (Appendix A: manual) and https://www.miqagruppen.org/pdf/manualer-protokoll/protocol-essa-19-eng/ (Appendix B: protocol).

## 2.3 | Training of raters and codings of supervision sessions

Three coders from MIQA worked with the codings of the 35 manual-based 30 min long supervision recordings. First, in order to promote a common understanding of the instrument, the coders received three half-days of formal training focused on the ESSA manual, plus half a day of independent coding tasks practices. Then, 6 months later, partly due to poor inter-rater agreement following the first training, the same three coders received three additional full days of training. The additional training included a repetition of the ESSA manual, intertwined with group coding tasks practices. Following the first training, the three coders independently coded a random sample of 20 recordings. Following the second training, the three coders independently coded the remaining 15 recordings. During all these codings, the coders were provided with the ESSA manual. If the supervision sessions were a couple of minutes longer than 30 min, the first 30 min of the recordings were coded.

**TABLE 1** The three coders inter-rater reliability after the first and the second training ($n$ = number of recordings)

| ESSA variable | ICC 1($n = 20$) | ICC 2($n = 15$) |
|---|---|---|
| Objective monitoring[a] | .61 | .77 |
| Educating[b] | .77 | .67 |
| Prompting[c] | .65 | .68 |
| Modelling skills[d] | .51 | .18 |
| Eliciting skills or potentials[e] | .87 | .73 |
| Positive feedback[f] | .56 | .68 |
| Corrective feedback[g] | .75 | .87 |
| Other utterances[h] | .40 | .19 |
| Structures and directs[i] | .26 | .30 |
| Specifies a training focus[j] | .09 | .23 |
| Performs active training[k] | .53 | .72 |
| Promotes a learning environment[l] | .02 | .44 |

Abbreviations: ESSA, The Evaluation of Supervisory Skills and Adherence manual; ICC, intraclass correlation coefficient.
[a]When the supervisor refers to actual behaviours in some form.
[b]When the supervisor provides method-related knowledge.
[c]When the supervisor nudges the supervisee in the right direction.
[d]When the supervisor provides demonstrations of skills.
[e]Supervisors questions in order to promote self-reflection in areas of proficiency.
[f]When the supervisor verbally reinforces supervisee behaviours.
[g]When the supervisor highlights behaviours that can be improved.
[h]All supervisor utterances that do not fall under any other category.
[i]How the supervisor directs the session structure and content.
[j]How the supervisor clarifies and organizes supervisee practice.
[k]How the supervisor performs active leaning activities.
[l]How the supervisor works with the supervisees sense of competence, intention to use the method, and aspiration to continuously develop and improve.

### 2.4 | Data analysis

Descriptive statistics were generated and presented as means (SD, range) for the frequencies of strategies and techniques used by the supervisors at the two assessment points. Cronbach's alpha was used to calculate internal consistency of the *Global Supervisor Behaviours* subscale. Since the items on the *Specific Supervisor Behaviours* subscale reflect the frequency of supervisors' utterances (i.e., are not part of an underlying construct), internal consistency was not calculated across these items. The inter-rater agreements of the ESSA codings were estimated by calculating intraclass correlations (ICCs), inserting single measures and employing the two-way mixed model with absolute agreement. The ICCs were interpreted according to the recommendations of Koo and Li (2016): <0.50 as poor; 0.50–0.75 as moderate; 0.75–0.90 as good; and >0.90 as excellent. All analyses were performed using the Statistical Package for the Social Sciences (version 25, SPSS Inc., Chicago, IL, USA).

## 3 | RESULTS

### 3.1 | Internal consistency and inter-rater reliability

The internal consistency across the *Global Supervisor Behaviours* subscale was acceptable at both assessment points: $\alpha = 0.70$ after the first training, and $\alpha = 0.71$ after the second training. All inter-rater reliability estimates are presented in Table 1. After the first training, four of the 12 items were in the poor range, five were in the moderate range, and three were in the good range. After the second training, all frequency counts were in the moderate to good range, except for two that were in the poor range; inter-rater reliability for one of the four global measures was in the moderate range, and three were in the poor range.

### 3.2 | Strategies and techniques practiced by the supervisors at the two assessment points

The strategies and techniques practiced by the supervisors are presented in Table 2. During both assessment points, the skills most frequently practiced by the supervisors were *Eliciting Skills or Potentials*, followed by *Objective Monitoring*, *Prompting*, *Positive Feedback* and *Other Utterances*. The skills least frequently practiced were *Modelling Skills*, followed by *Educating*. These two skills, together with *Corrective Feedback*, were also the only three skills absent in one or more sessions. As for the *Global Behaviours*, the ranges were somewhat limited for three out of four variables at the first assessment, and all means, except for *Structures and Directs* at the first assessment, were in between three and four.

## 4 | DISCUSSION

The primary aim of the current study was to test the internal consistency and inter-rater reliability of the ESSA manual, and to describe the strategies and techniques used by the supervisors in a Swedish MI implementation study. The internal consistency across the four *Global Supervisor Behaviours* was acceptable at both assessment points. After the second training, the inter-rater reliability was in the moderate range for one of the *Global Supervisor Behaviours*, and three were in

**TABLE 2** Strategies and techniques practiced by the 10 supervisors at the two assessment points ($n$ = number of recordings)

| | Assessment 1($n = 20$) | | Assessment 2($n = 15$) | |
|---|---|---|---|---|
| **ESSA variable** | **Means (SD)** | **Min-max** | **Means (SD)** | **Min-max** |
| Specific supervisor behaviours (counts) | | | | |
| Objective monitoring | 10.9 (2.9) | 5.0–18.0 | 11.0 (4.1) | 4.0–20.0 |
| Educating | 7.4 (3.9) | 2.0–19.0 | 7.1 (4.0) | 0.0–16.0 |
| Prompting | 11.6 (4.9) | 1.0–28.0 | 13.0 (4.5) | 5.0–25.0 |
| Modelling skills | 0.6 (0.9) | 0.0–4.0 | 1.1 (1.0) | 0.0–7.0 |
| Eliciting skills or potentials | 15.0 (4.9) | 6.0–26.0 | 15.3 (3.8) | 5.0–22.0 |
| Positive feedback | 13.2 (4.2) | 6.0–28.0 | 11.7 (3.7) | 5.0–19.0 |
| Corrective feedback | 3.1 (2.2) | 0.0–10.0 | 4.2 (3.6) | 0.0–16.0 |
| Other utterances | 11.9 (3.2) | 6.0–24.0 | 11.5 (3.2) | 3.0–26.0 |
| Global supervisor behaviours (1–5) | | | | |
| Structures and directs | 2.6 (0.9) | 1.0–4.0 | 3.4 (0.4) | 1.0–5.0 |
| Specifies a training focus | 3.8 (0.4) | 2.0–5.0 | 3.7 (0.5) | 1.0–5.0 |
| Performs active training | 3.3 (0.9) | 1.0–5.0 | 3.5 (1.2) | 1.0–5.0 |
| Promotes a learning environment | 3.8 (0.5) | 2.0–5.0 | 3.8 (0.7) | 1.0–5.0 |

Abbreviations: ESSA, The Evaluation of Supervisory Skills and Adherence manual; SD, standard deviation.

the poor range. As for the *Specific Supervisor Behaviours*, the inter-rater reliabilities for all frequency counts were in the moderate to good range, except for two that were in the poor range.

### 4.1 | Internal consistency and inter-rater reliability

The internal consistency for the *Global Supervisor Behaviours* was acceptable, suggesting that the four variables are related to each other and at the same time contribute unique information (Sijtsma, 2009). However, in line with previous research on the relative difficulty of achieving reliable codings for global scales (Gill et al., 2019; Kramer Schmidt et al., 2019; van der van der Vleuten et al., 2010), only one of the four *Global Supervisor Behaviours* (i.e., Performs Active Training) reached the moderate range. Yet, the ICC for the *Global Supervisor Behaviours* improved overall after training two. Possibly, revising the global variables and extended or more effective training could have made the coders more reliable. However, the low ICC could also reflect the lack of global variable variability (Koo & Li, 2016): The range was somewhat restricted for three of the *Global Supervisor Behaviours* at the first assessment, and the mean was between three and four for all four variables at both assessment points except for *Structures and Directs* at the first assessment. This lack of variability could be explained by the study's highly trained supervisors, who also followed a strict manual (Beckman, Forsberg, et al., 2017b). A more heterogeneous sample might thus have produced a more reliable result for the *Global Supervisor Behaviours*.

As for the *Specific Supervisor Behaviours*, the coders' inter-rater reliability scores also improved overall after the second training. *Other Utterances* and *Modelling Skills* were the *Specific Behaviours* with lowest ICC after both trainings. This result is somewhat expected for *Other Utterances*, as it is the category for all supervisor utterances during the session that are not sorted into any other category (please see Appendix A). With such a mixture of form and content (e.g., statements, questions, reflections, information, instructions and agreements), single utterances can easily be missed. The form and content mixture also makes that category less significant. Still, a category for each utterance can keep the coders focused and thereby promote coding, as opposed to having some of the utterances not sorted into any category. The poor results for *Modelling Skills* are more unpredicted and noteworthy. This category is used when the supervisors provide demonstrations of skills for the supervisees to learn from (please see Appendix A). One explanation may be that the supervisors' demonstration of skills during role-plays differed from the more subtle demonstrations during other parts of the supervision sessions, which could possibly be better captured with either a revision of the manual or expanded and more effective training of coders. Additionally, *Modelling Skills* was also the least frequent *Specific Behaviour* at both assessment points, which could have made that category difficult to code reliably. Sometimes called a micro-skill, modelling is one of the active learning techniques proposed to promote supervisees' learning (Bailin et al., 2018; Bearman et al., 2017; Caron et al., 2021). Modelling has also, together with role-play, predicted fidelity in an EBP implementation study (Bearman et al., 2013; Caron et al., 2021), and, together with role-play and corrective feedback, increased CBT fidelity and competence in a controlled experiment (Bearman et al., 2017). However, in another recent study of 162 supervision sessions involving 27 therapists, modelling did not increase therapists' own active learning processes (Caron et al., 2021).

### 4.2 | Strategies and techniques practiced by the supervisors

*Eliciting Skills or Potentials* was the most common *Specific Supervisor Behaviour* at both assessment points, followed by *Prompting*, *Positive Feedback* and *Objective Monitoring*. This shows that the supervisors closely followed the manual. More surprising was the low levels of *Modelling Skills* and *Corrective Feedback*, as also these behaviours were clearly defined in the supervision manual. However, the ICC for *Modelling Skills* was in the poor range after the second training, since one of the coders consistently coded more of that category. The lower levels of *Corrective Feedback* are consistent with previous findings on supervisors withholding corrective feedback to avoid negative supervisee reactions (Bailin et al., 2018; Beckman et al., 2019). However, performance-related feedback has shown to be an effective supervisor skill in both clinical practice (Bailin et al., 2018; Bradley & Becker, 2021; Caron et al., 2021; Ivers et al., 2012) and other organizational settings (Sleiman et al., 2020). Additional findings show that supervisees can handle corrective feedback well and that it does not negatively affect the supervisor–supervisee working alliance or supervisee skill acquisition (Beckman et al., 2019, 2021; Ellis, 2010; Ladany et al., 2013). *Objective Monitoring* is as another important feature of clinical supervision as it promotes accurate and specific assessment of supervisee behaviour (Bailin et al., 2018; Caron et al., 2021). Weak correlations between practitioners and observers have been shown in both manualized research treatments and routine practices (Hogue et al., 2015). Preliminary findings also point to objective feedback based on monitoring of sessions as an effective supervision component for facilitating supervisee learning (Bearman et al., 2017; Caron et al., 2021; Caron & Dozier, 2019; Eiraldi et al., 2018; Martino et al., 2016; Webster-Stratton et al., 2014; Weck et al., 2017). The fact that the behaviours *Eliciting Skills or Potentials*, *Prompting* and *Positive Feedback* were all more frequent behaviours than *Education* indicates that the supervisors conducted the supervision sessions in a manner consistent with MI. An MI approach to supervision has the advantage of relying on theoretical principles and empirical knowledge of factors facilitating behaviour change and is also consistent with proposed components of the supervisory relationship (Beckman, Bohman, et al., 2017a). One last interesting thing to note regarding the *Specific Supervisor Behaviours* is the limited number of behaviours coded in the category *Other Utterances*—only 16.1% of the behaviours at the first assessment, and 15.3% at the second, fell outside the classified categories. However, as the ICC for *Other Utterances* was in the poor range, this result too must be interpreted with caution.

For the *Global Supervisor Behaviours*, the range was somewhat restricted and the mean was between three and four for almost all measures, which again shows that the supervisors followed the manual. However, after the second training, only *Performs Active Training* was in the moderate range, and the remaining three categories were in the poor range, so also these results must, again, be interpreted with caution.

## 4.3 | Strengths and limitations

This study tested the internal consistency and inter-rater reliability of a supervisor assessment tool that includes both global measures and frequency counts of behaviour—two essential aspects of supervisory competence. However, ESSA does not include all competencies listed in current frameworks (i.e., normative/managerial, formative/pedagogical, and restorative/supportive). Instead, ESSA focuses on how supervisors conducts supervision during sessions. Since there are no other established validated global and frequency counts measures of supervisor behaviour, a concurrent validity test could not be conducted. Moreover, the small sample size of the current study did not allow for the relationship between items to be examined in a factor analysis. Another limitation of the present study is that all supervision session where provided by highly trained MI supervisors following a strict manual (Beckman, Forsberg, et al., 2017b), which might be difficult to replicate and could limit the generalizability of the findings. Future research should use a larger sample and a wider supervision context (e.g., CBT supervision) to provide additional psychometric validation.

## 4.4 | Conclusions

Supervision is an essential part of therapist training and is thus also a fundamental part of implementing EBPs. However, despite recent years' conceptualization of supervisor competence, there is still a shortage of valid and reliable instruments for objective assessment of supervision competence that include both global measures and frequency counts of behaviour. Competence measures that include frequency counts could also be used as self-assessment tools for supervisors, as a way to learn to more accurately estimate levels of adherence and competence following supervision training and practice. Caron and Dozier (2021) explored therapists' self-assessment of practice and argued that behaviour-based coding likely promotes reliability and reduces self-enhancement biases of one's own performance. This could also be an interesting future direction for supervision research. In any case, a prerequisite for determining specific supervisor skills central to the development of therapist skills, teaching these skills to supervisors in an efficient way and performing quality assurance of supervision sessions is to create instruments that can measure these behaviours. This study is a step in that direction.

## ACKNOWLEDGEMENTS

The Swedish National Board of Health and Welfare provided financial support for parts of this work (2.4-59457/2012). The authors would also like to thank Nina Lindroos for the ICC analyses and Anna-Maria Ahlgren, Carl Lovén and Linda Ljunggren for valuable feedback during the development of the ESSA manual.

## CONFLICT OF INTEREST

The authors report no potential conflict of interest.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ORCID

*Maria Beckman* https://orcid.org/0000-0002-9370-1863
*Sven Alfonsson* https://orcid.org/0000-0003-4570-5891
*Ingvar Rosendahl* https://orcid.org/0000-0002-1325-913X
*Anne H. Berman* https://orcid.org/0000-0002-7709-0230
*Helena Lindqvist* https://orcid.org/0000-0002-3634-4396

## REFERENCES

Alfonsson, S., Parling, T., Spannargard, A., Andersson, G., & Lundgren, T. (2018). The effects of clinical supervision on supervisees and patients in cognitive behavioral therapy: Systematic review. *Cognitive Behaviour Therapy*, *47*(3), 206–228. https://doi.org/10.1080/16506073.2017.1369559

American Psychological Association. (2015). Guidelines for clinical supervision in health service psychology. *American Psychologist*, *70*(1), 33–46. https://doi.org/10.1037/a0038112

Bailin, A., Bearman, S. K., & Sale, R. (2018). Clinical supervision of mental health professionals serving youth: Format and microskills. *Administration and Policy in Mental Health*, *45*(5), 800–812. https://doi.org/10.1007/s10488-018-0865-y

Barrett, J., Gonsalvez, C. J., & Shires, A. (2020). Evidence-based practice within supervision during psychology practitioner training: A systematic review. *Clinical Psychologist*, *24*(1), 3–17. https://doi.org/10.1111/cp.12196

Barwick, M. A., Bennett, L. M., Johnson, S. N., McGowan, J., & Moore, J. E. (2012). Training health and mental health professionals in motivational interviewing: A systematic review. *Children and Youth Services Review*, *34*(9), 1786–1795. https://doi.org/10.1016/j.childyouth.2012.05.012

Bearman, S. K., Schneiderman, R. L., & Zoloth, E. (2017). Building an evidence base for effective supervision practices: An analogue experiment of supervision to increase EBT fidelity. *Administration and Policy in Mental Health*, *44*(2), 293–307. https://doi.org/10.1007/s10488-016-0723-8

Bearman, S. K., Weisz, J. R., Chorpita, B. F., Hoagwood, K., Ward, A., Ugueto, A. M., Bernstein, A., & Research Network on Youth Mental H. (2013). More practice, less preach? The role of supervision processes and therapist characteristics in EBP implementation. *Administration and Policy in Mental Health*, *40*(6), 518–529. https://doi.org/10.1007/s10488-013-0485-5

Beckman, M., Bohman, B., Forsberg, L., Rasmussen, F., & Ghaderi, A. (2017a). Supervision in motivational interviewing: An exploratory study. *Behavioural and Cognitive Psychotherapy*, *45*(4), 351–365. https://doi.org/10.1017/S135246581700011X

Beckman, M., Forsberg, L., Lindqvist, H., Diez, M., Eno Persson, J., & Ghaderi, A. (2017b). The dissemination of motivational interviewing in

Swedish county councils: Results of a randomized controlled trial. *PLoS ONE*, *12*(7), e0181715. https://doi.org/10.1371/journal.pone.0181715

Beckman, M., Forsberg, L., Lindqvist, H., & Ghaderi, A. (2019). Providing objective feedback in supervision in motivational interviewing: Results from a randomized controlled trial. *Behavioural and Cognitive Psychotherapy*, *48*, 1–12. https://doi.org/10.1017/S1352465819000687

Beckman, M., Lindqvist, H., Ohman, L., Forsberg, L., Lundgren, T., & Ghaderi, A. (2021). Ongoing supervision as a method to implement motivational interviewing: Results of a randomized controlled trial. *Patient Education and Counseling*, *104*, 2037–2044. https://doi.org/10.1016/j.pec.2021.01.014

Borders, L. D., Glosoff, H. L., Welfare, L. E., Hays, D. G., DeKruyf, L., Fernando, D. M., & Page, B. (2014). Best practices in clinical supervision: Evolution of a counseling specialty. *The Clinical Supervisor*, *33*(1), 26–44. https://doi.org/10.1080/07325223.2014.905225

Bradley, W. J., & Becker, K. D. (2021). Clinical supervision of Mental health services: A systematic review of supervision characteristics and practices associated with formative and restorative outcomes. *The Clinical Supervisor*, *40*(1), 88–111. https://doi.org/10.1080/07325223.2021.1904312

Caron, E., & Dozier, M. (2019). Effects of fidelity-focused consultation on clinicians implementation: An exploratory multiple baseline design. *Administration and Policy in Mental Health and Mental Health Services Research*, *46*(4), 445–457. https://doi.org/10.1007/s10488-019-00924-3

Caron, E., & Dozier, M. (2021). Self-coding of fidelity as a potential active ingredient of consultation to improve clinicians fidelity. *Administration and Policy in Mental Health and Mental Health Services Research*, *49*(2), 237–254. https://doi.org/10.1007/s10488-021-01160-4

Caron, E. B., Lind, T. A., & Dozier, M. (2021). Strategies that promote therapist engagement in active and experiential learning: Micro-level sequential analysis. *The Clinical Supervisor*, *40*(1), 112–133. https://doi.org/10.1080/07325223.2020.1870023

de Roten, Y., Zimmermann, G., Ortega, D., & Despland, J. N. (2013). Meta-analysis of the effects of MI training on clinicians' behavior. *Journal of Substance Abuse Treatment*, *45*(2), 155–162. https://doi.org/10.1016/j.jsat.2013.02.006

Eiraldi, R., Mautone, J. A., Khanna, M. S., Power, T. J., Orapallo, A., Cacia, J., … Jawad, A. F. (2018). Group CBT for externalizing disorders in urban schools: Effect of training strategy on treatment fidelity and child outcomes. *Behavior Therapy*, *49*(4), 538–550. https://doi.org/10.1016/j.beth.2018.01.001

Ellis, M. V. (2010). Bridging the science and practice of clinical supervision: Some discoveries, some misconceptions. *The Clinical Supervisor*, *29*(1), 95–116. https://doi.org/10.1080/07325221003741910

Falender, C. A., & Shafranske, E. P. (2014). Clinical supervision: The state of the art. *Journal of Clinical Psychology*, *70*(11), 1030–1041. https://doi.org/10.1002/jclp.22124

Gill, I., Oster, C., & Lawn, S. (2019). Assessing competence in health professionals' use of motivational interviewing: A systematic review of training and supervision tools. *Patient Education and Counseling*, *103*, 473–483. https://doi.org/10.1016/j.pec.2019.09.021

Gonsalvez, C. J. (2021). A short scale to evaluate supervision and supervisor competence-the SE-SC8. *Clinical Psychology & Psychotherapy*, *28*(2), 452–461. https://doi.org/10.1002/cpp.2510

Gonsalvez, C. J., Hamid, G., Savage, N. M., & Livni, D. (2020). The supervision evaluation and supervisory competence scale: Psychometric validation. *Australian Psychologist*, *52*(2), 94–103. https://doi.org/10.1111/ap.12269

Hoge, M. A., Migdole, S., Farkas, M. S., Ponce, A. N., & Hunnicutt, C. (2011). Supervision in public sector behavioral health: A review. *The Clinical Supervisor*, *30*(2), 183–203. https://doi.org/10.1080/07325223.2011.604276

Hogue, A., Dauber, S., Lichvar, E., Bobek, M., & Henderson, C. E. (2015). Validity of therapist self-report ratings of fidelity to evidence-based practices for adolescent behavior problems: Correspondence between therapists and observers. *Administration and Policy in Mental Health*, *42*, 229–243.

Ivers, N., Jamtvedt, G., Flottorp, S., Young, J. M., Odgaard-Jensen, J., French, S. D., O'Brien, M. A., Johansen, M., Grimshaw, J., & Oxman, A. D. (2012). Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev*(6), CD000259. https://doi.org/10.1002/14651858.CD000259.pub3

Kennerley, H., Butler, G., & Clohessy, S. (2010). Supervisor competency scale. In D. Westbrook (Ed.), *The Oxford Guide to Surviving as a CBT Therapist*. Oxford University Press.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting Intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kramer Schmidt, L., Andersen, K., Nielsen, A. S., & Moyers, T. B. (2019). Lessons learned from measuring fidelity with the motivational interviewing treatment integrity code (MITI 4). *Journal of Substance Abuse Treatment*, *97*, 59–67. https://doi.org/10.1016/j.jsat.2018.11.004

Kuhne, F., Maas, J., Wiesenthal, S., & Weck, F. (2019). Empirical research in clinical supervision: A systematic review and suggestions for future studies. *BMC Psychol*, *7*(1), 54. https://doi.org/10.1186/s40359-019-0327-7

Ladany, N., Mori, Y., & Mehr, K. E. (2013). Effective and ineffective supervision. *The Counseling Psychologist*, *41*(1), 28–47. https://doi.org/10.1177/0011000012442648

Madson, M. B., Villarosa-Hurlocker, M. C., Schumacher, J. A., Williams, D. C., & Gauthier, J. M. (2019). Motivational interviewing training of substance use treatment professionals: A systematic review. *Substance Abuse*, *40*(1), 43–51. https://doi.org/10.1080/08897077.2018.1475319

Martino, S., Gallon, S., Ball, S., & Carroll, K. M. (2008). A step forward in teaching addiction counselors how to supervise motivational interviewing using a clinical trials training approach. *Journal of Teaching in the Addictions*, *6*(2), 39–67. https://doi.org/10.1080/15332700802127946

Martino, S., Paris, M. Jr., Anez, L., Nich, C., Canning-Ball, M., Hunkele, K., Olmstead, T. A., & Carroll, K. M. (2016). The effectiveness and cost of clinical supervision for motivational interviewing: A randomized controlled trial. *Journal of Substance Abuse Treatment*, *68*, 11–23. https://doi.org/10.1016/j.jsat.2016.04.005

Milne, D., James, I., Keegan, D., & Dudley, M. (2002). Teacher's PETS: A new observational measure of experiential training interactions. *Clinical Psychology & Psychotherapy*, *9*(3), 187–199. https://doi.org/10.1002/cpp.308

Milne, D., & Reiser, R. P. (2012). A rationale for evidence-based clinical supervision. *Journal of Contemporary Psychotherapy*, *42*(3), 139–149. https://doi.org/10.1007/s10879-011-9199-8

Milne, D. L., Sheikh, A. I., Pattison, S., & Wilkinson, A. (2011). Evidence-based training for clinical supervisors: A systematic review of 11 controlled studies. *The Clinical Supervisor*, *30*(1), 53–71. https://doi.org/10.1080/07325223.2011.564955

Moeller, S. B., Rosenberg, N. K., Hvenegaard, M., Straarup, K., & Austin, S. F. (2020). A new tool for rating cognitive behavioural supervision – Preliminary findings in a clinical setting. *The Cognitive Behaviour Therapist*, *13*. https://doi.org/10.1017/s1754470x2000001x

Moyers, M., Manuel, H., & Miller, W. (2005). Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment*, *28*(1), 19–26. https://doi.org/10.1016/j.jsat.2004.11.001

Moyers, R., Manuel, E., & Houck, J. (2016). The motivational interviewing treatment integrity code (MITI 4): Rationale, preliminary reliability and validity. *Journal of Substance Abuse Treatment*, *65*, 36–42. https://doi.org/10.1016/j.jsat.2016.01.001

Nadeem, E., Gleacher, A., & Beidas, R. S. (2013). Consultation as an implementation strategy for evidence-based practices across multiple contexts: Unpacking the black box. *Administration and Policy in Mental Health*, *40*(6), 439–450. https://doi.org/10.1007/s10488-013-0502-8

Norman, G. (2005). Editorial--checklists vs. ratings, the illusion of objectivity, the demise of skills and the debasement of evidence. *Advances in Health Sciences Education: Theory and Practice*, *10*(1), 1–3. https://doi.org/10.1007/s10459-005-4723-9

Ottman, K. E., Kohrt, B. A., Pedersen, G. A., & Schafer, A. (2020). Use of role plays to assess therapist competency and its association with client outcomes in psychological interventions: A scoping review and competency research agenda. *Behaviour Research and Therapy*, *130*, 103531. https://doi.org/10.1016/j.brat.2019.103531

Park, E. H., Ha, G., Lee, S., Lee, Y. Y., & Lee, S. M. (2019). Relationship between the supervisory working Alliance and outcomes: A Meta-analysis. *Journal of Counseling & Development*, *97*(4), 437–446. https://doi.org/10.1002/jcad.12292

Reiser, R. P., Cliffe, T., & Milne, D. L. (2018). An improved competence rating scale for CBT supervision: Short-SAGE. *The Cognitive Behaviour Therapist*, *11*. https://doi.org/10.1017/s1754470x18000065

Roth, A., & Pilling, S. (2007). A competence framework for the supervision of psychological therapies. Retreved from: https://www.researchgate.net/publication/265872800_A_competence_framework_for_the_supervision_of_psychological_therapies

Schwalbe, C. S., Oh, H. Y., & Zweben, A. (2014). Sustaining motivational interviewing: A meta-analysis of training studies. *Addiction*, *109*(8), 1287–1294. https://doi.org/10.1111/add.12558

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Sleiman, A. A., Sigurjonsdottir, S., Elnes, A., Gage, N. A., & Gravina, N. E. (2020). A quantitative review of performance feedback in organizational settings (1998-2018). *Journal of Organizational Behavior Management*, *40*(3–4), 303–332. https://doi.org/10.1080/01608061.2020.1823300

Spence, S. H., Wilson, J., Kavanagh, D., Strong, J., & Worrall, L. (2012). Clinical supervision in four Mental health professions: A review of the evidence. *Behaviour Change*, *18*(3), 135–155. https://doi.org/10.1375/bech.18.3.135

The Psychology Board of Australia. (2013). *Guidelines for supervisors and supervisor training providers.*

Tugendrajch, S. K., Sheerin, K. M., Andrews, J. H., Reimers, R., Marriott, B. R., Cho, E., & Hawley, K. M. (2021). What is the evidence for supervision best practices? *The Clinical Supervisor*, *40*(1), 68–87. https://doi.org/10.1080/07325223.2021.1887785

van der Vleuten, C. P., & Schuwirth, L. W. (2005). Assessing professional competence: From methods to programmes. *Medical Education*, *39*(3), 309–317. https://doi.org/10.1111/j.1365-2929.2005.02094.x

van der Vleuten, C. P., Schuwirth, L. W., Scheele, F., Driessen, E. W., & Hodges, B. (2010). The assessment of professional competence: Building blocks for theory development. *Best Practice & Research: Clinical Obstetrics & Gynaecology*, *24*(6), 703–719. https://doi.org/10.1016/j.bpobgyn.2010.04.001

Watkins, C. E. (2020). What do clinical supervision research reviews tell us? Surveying the last 25 years. *Counselling and Psychotherapy Research*, *20*(2), 190–208. https://doi.org/10.1002/capr.12287

Webster-Stratton, C. H., Reid, M. J., & Marsenich, L. (2014). Improving therapist fidelity during implementation of evidence-based practices: Incredible years program. *Psychiatric Services*, *65*(6), 789–795. https://doi.org/10.1176/appi.ps.201200177

Weck, F., Kaufmann, Y. M., & Höfling, V. (2017). Competence feedback improves CBT competence in trainee therapists: A randomized controlled pilot study. *Psychotherapy Research*, *27*(4), 501–509. https://doi.org/10.1080/10503307.2015.1132857

Wheeler, S., & Barkham, M. (2014). A core evaluation battery for supervision. In C. E. Watkins, Jr. & D. L. Milne (Eds.), *The Wiley International Handbook of Clinical Supervision* (pp. 365–385). John Wiley & Sons Ltd. https://doi.org/10.1002/9781118846360

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Beckman, M., Alfonsson, S., Rosendahl, I., Berman, A. H., & Lindqvist, H. (2022). A behaviour-based coding tool for assessing supervisors' adherence and competence: Findings from a motivational interviewing implementation study. *Clinical Psychology & Psychotherapy*, *29*(6), 1942–1949. https://doi.org/10.1002/cpp.2763