Can we reliably measure social work communication skills?

Development of a scale to measure child and family social work direct practice

Charlotte E. Whittaker

Tilda Goldberg Centre, University of Bedfordshire, Luton, United Kingdom

Donald Forrester

CASCADE – Centre for Children's Social Care Research and Development, Cardiff University, Cardiff, United Kingdom

Michael Killian

School of Social Work, University of Texas at Arlington, Arlington, United States

Rebecca K. Jones

Tilda Goldberg Centre, University of Bedfordshire, Luton, United Kingdom

Corresponding author's address: Tilda Goldberg Centre, University of Bedfordshire, c/o Charlotte E. Whittaker, Park Square Luton LU1 3JU, United Kingdom.
E-mail address: charlotte.whittaker@beds.ac.uk.

Abstract

Few attempts have been made to define and measure the effectiveness of social work communication skills. This paper describes a coding scheme for rating seven dimensions of skilled communication in child and family social work practice and presents an empirical evaluation of whether the dimensions can be coded for reliably. Four dimensions of skill were adapted from the Motivational Interviewing Treatment Integrity (MITI) code. A further three dimensions, primarily related to appropriate use of authority, were developed in consultation with key stakeholders. The seven dimensions were used to score 133 audio recordings of direct practice. Of these, 28 (21%) were scored by three independent raters in order to test inter-rater reliability (IRR). IRR was assessed using Krippendorff's α and Intra-class correlation (ICC). Results indicate that it is possible to reliably measure key elements of skilled communication, with Krippendorff's α scores ranging from .461 (good) to .937 (excellent) and ICC ranging from .731 (good) to .967 (excellent). Establishing reliability provides a foundation for exploring the validity of the measure and the relationship between these skills and outcomes, as well as for further research looking at the impact of training, supervision or other methods of professional development on skills in practice. The problems and potential contribution of using such an approach are discussed.

Keywords: child protection, communication skills, rating scale, reliability, Motivational Interviewing, MITI

Introduction

Effective communication is fundamental to social work, yet to date there has been little empirical research on direct practice. This paper describes the development of a coding scheme for rating key dimensions of skilled communication in child and family social work, and an empirical investigation of whether the dimensions developed can be coded for reliably.

While the importance of skilled communication is universally acknowledged, in comparison to fields such as mental health and education, social work has been slow to develop an empirical basis for defining 'good practice'. In 2004 a Social Care Institute for Excellence (SCIE) review of the literature on teaching and learning communication skills in social work education identified "an urgent need to develop a robust methodology, particularly with regard to defining and measuring the effectiveness of communication skills with service users" (Diggins, 2004, p.15). More recently, Forrester, Kershaw, Moss and Hughes (2008) and Ferguson (2011) have commented on the continued lack of research on direct practice. The absence of such knowledge is problematic. Without an empirical basis for thinking about social work communication, we have no way of knowing what skills, if any, are linked to outcomes for service users. Whilst studies in other disciplines such as psychotherapy have identified therapist qualities and techniques that positively influence the therapeutic alliance (Ackerman & Hilsenroth, 2003), we know little about how skills such as these translate into the context of statutory social work where working relationships are often non-voluntary and undoubtedly more complex.

One of the main reasons there has been difficulty developing an empirically based approach to measuring social work communication skills is the absence of research that examines direct social work practice. Here we use the term 'direct social work practice' to describe professional encounters between social workers and their clients, aimed at protecting them from harm and improving outcomes in their lives (British Association of Social Workers, 2016). These encounters take place in a variety of settings including offices, in the community and at the client's home.

Over the past decade, there has been increased interest in this area. Ferguson (2011) for example, took an ethnographic approach to understanding what has previously been the private domain of the home visit. He spent six months shadowing child and family practitioners in order to understand what social workers do, how they relate to families and the context in which this challenging work takes place. Ferguson identified that workers displayed varying levels of skill which he attributes to two key factors; organisational pressures which limit the time they had for quality direct work with families, and the personal qualities of individual workers' which affected how confident they felt interacting with children and families. His work provides rich insight, illustrated with detailed examples, into how social workers interact with families in challenging circumstances.

Hall, Juhila, Matarese and Nijnatten (2014) draw together a collection of discursive studies which explore day-to-day interactions between social workers and clients. Using audio recordings of direct practice, they examine the features of social work talk within the professional context. Like Fergu-

son, Hall et al emphasise the role of institutional practices in shaping social work interaction. Rather than drawing rigid conclusions about the components of good or bad social work communication, they emphasise the use of discursive methods to "make visible the richness and skilfulness of face-to-face interaction in real life social work" and advocate for the routine use of recording to inform professional development.

Qualitative explorations of social work encounters contribute greatly to our understanding of the complex processes involved in direct social work practice. In particular, they shed light on the context in which communication skills are applied and the ways in which this might influence the application of practice skills. It is more difficult to evaluate practice skills using solely qualitative methods. To identify generalizable links between skills and outcomes a quantitative contribution seems most appropriate. If we can reliably code for levels of skill, then we can begin to explore the links between levels of skill and outcomes.

Some attempts have been made to measure practice within the field of social work education. A literature review undertaken by Bogo, Regehr, Hughes, Power and Globerman (2002) identified three scales that have been developed; 1) the Practice Skills Inventory (PSI; O'Hare & Collins, 1997), a self-report measure which is intended to capture how frequently skills are applied, 2) a checklist developed by Wilson (1981) which assesses students' practice by evaluating their process notes and 3) a 25-item rating scale developed by Koroloff and Rhyne (1989), designed for use by students and field instructors. An important limitation of all three scales is that none were evaluated for measuring observed practice. For instance, Koroloff

and Rhyne commented that assessors often did not have time to observe practice and therefore had to rely on data provided by students on their performance. Whilst the review was undertaken over a decade ago, we were unable to identify any other studies that have reviewed existing instruments for skill measurement in social work other than those outlined below.

Bogo, Regehr, Logie, Katz, Mylopoulos and Regehr (2011) have developed a measure of observed social work practice for use in Objective Structured Clinical Examinations (OSCE's), an assessment method originally used in the medical field but adapted for social work education. In this study, OSCE's were used to assess student competence through the use of five simulated practice scenarios involving an actor playing a client. These scenarios were assessed using a rating tool which identifies competencies on a 1 (low) to 5 (high) scale. A key adaptation for the social work context has been the introduction of an additional scale to measure a post-encounter reflective dialogue with the examiner for assessing 'meta-competencies' as well as behavioural skills. The researchers explored the reliability and construct validity of the measure. Their findings indicate that the tool was able to distinguish between experienced and inexperienced practitioners and demonstrated promising internal consistency between the two rating scales.

Further research undertaken by Bogo, Regehr, Katz, Logie, Tufford and Litvack (2012) evaluated the adapted OSCE for assessing student performance at the end of the first semester in a Masters in Social Work (MSW) programme. The researchers were interested in the extent to which the method predicts how well students perform in a field setting with real clients. Students

were assessed using a single simulated practice scenario, rather than the five-scenario approach used previously. An Online Practice-Based Evaluation Tool was also used by field instructors to assess students' field performance at the midterm and end of the field practice placement. The tool assesses practice across six dimensions, with a score between 1 and 5 allocated for each one.

Findings indicate that the OSCE method is able to capture variability in student competence but that the relationship between OCSE scores and scores in field evaluation is complex. For example, some students who struggled in the OSCE performed well in field practice. Crucially, the authors highlight that they could not be certain that the ratings given by field instructors when using the field evaluation tool were actually based on direct observations of student performance in their practice settings.

There therefore seems to be a key gap in the literature: we could identify no studies that attempted to grade, rate or quantitatively categorise *direct* social work practice. Furthermore, we have found no published research on whether social work skills can be measured reliably. This seems an important gap in the literature as validity cannot be investigated until reliability of coding or marking has been achieved.

This paper sets out an approach to measuring some of the key dimensions of skilled communication involved in such work using a sample of audio recordings of practice meetings between a social worker and carers for a child. We describe the development of the coding scheme; however our focus is the evaluation of inter-rater reliability in coding for seven dimensions of skilled communication in child and family social work practice. Our primary research question is thus: *Can key dimensions of skilled social work communi-cation be reliably coded by independent raters using recordings of direct practice encounters?*

This study focuses specifically on identifying and measuring key elements of communication which are important to child and family social work. In the UK, this refers to the work undertaken by qualified professionals who operate within legal frameworks to protect and support vulnerable children and their families (Department for Education, 2014). As such, child and family social workers need communication skills which will enable them to engage parents and form helpful relationships whilst also talking about and managing risk.

This study forms part of a broader programme of work undertaken by Forrester and colleagues over the last 10 years. The reliable coding of practice forms a foundation for studies exploring the relationship between practice skills and outcomes and for research on the individual and organisational factors that influence level of worker skill. Studying such relationships is only possible if skills can reliably be identified.

Background to the development and operationalization of the seven dimensions

Our programme of work has explored Motivational Interviewing (MI) as a communication style within the field of child and family social work (Forrester et al., forthcoming a; Forrester et al., forthcoming b; Westlake, Killian & Forrester, 2014; Whittaker, Forrester & Antonopoulou, 2015). MI has a well-developed body of research focussed on the relationship between practice skills and outcomes for clients and an established behavioural coding system: the Motivational Interviewing Treatment In-

tegrity (MITI) code (Moyers, Martin, Manuel, Miller and Ernst, 2010). This coding system formed the starting point for the measures of skilled communication developed as part of this study. We therefore provide a brief description of MI and the MITI as well as outlining the rationale for using it in describing key elements of social work practice.

MI is "a collaborative conversation style for strengthening a person's own motivation and commitment to change" (Miller & Rollnick, 2013, p. 12). MI involves skilled communication including the use of reflective statements, open questions and affirmations to elicit and reinforce a person's own reasons for change. MI has been of particular interest to us in understanding key elements of skilled social work communication, not least because the values and principles underpinning MI seem highly compatible with social work values (Hohman, 1998; Wahab, 2005; Watson; 2011). Furthermore, the MITI is a valid and reliable integrity measure (Forsberg, Käll-

mén, Hermansson, Berman and Helgason, 2007; Moyers, Martin, Manuel, Hendrickson and Miller, 2005; Pierson et al., 2007) which assesses how well a practitioner is demonstrating these core skills and values in their interactions with clients.

The MITI measures five dimensions of practice which are outlined in Table 1 (Moyers et al., 2010). Each dimension is measured on a scale of 1 (low) to 5 (high), based on a randomly selected 20 minute segment of a practice recording. The scores for Evocation, Collaboration and Autonomy are then averaged to provide an overall Global Spirit Rating which is the primary measure of MI skill. Empathy and Direction are considered important elements of skill but are not unique to MI; therefore the scores for these dimensions do not contribute to the overall Spirit rating.

Extensive research across different settings suggests that MI produces small but significant effects across a range of problem behaviours (Rubak, Sanboek, Lauritzen

Table 1. Dimensions of practice measured by the MITI

Global Dimension	Description
Evocation	'the extent to which the clinician conveys an understanding that motivation for change, and the ability to move toward that change, reside mostly within the client and therefore focuses efforts to elicit and expand it'
Collaboration	'the extent to which the clinician behaves as if the interview is occurring between two equal partners, both of whom have knowledge that might be useful in the problem under consideration'
Autonomy	'the extent to which the clinician supports and actively fosters client perception of choice as opposed to attempting to control the client's behaviour or choices'
Direction	'the degree to which clinicians maintain appropriate focus on a specific target behaviour or concerns directly tied to it'
Empathy	'the extent to which the clinician understands or makes an effort to grasp the client's perspective and feelings'

& Christensen, 2005; Lundahl, Kunz, Brownell, Tollefson & Burke, 2010). However, research exploring MI in child and family social work is still in its infancy. Yet in some senses the focus on MI is incidental; our broader focus has been on understanding key elements of good practice and how they link to outcomes. MI and the MITI have provided a helpful starting point by offering a well-established measure of practice competencies which appear to be relevant to the field of child and family social work.

Findings from early studies in our programme of work suggested that the MITI might provide a useful framework for thinking about effective communication in child and family social work. Studies across different UK Local Authority settings identified differences in social worker skill in simulated interviews after a training course in MI when the MITI was used as a measure of practice skill (Forrester et al, forthcoming a; Forrester, McCambridge, Waissbein, Emlyn-Jones & Rollnick, 2008; Westlake et al., 2014; Whittaker et al., 2015). Furthermore, Forrester et al. (2013) identified significant differences between worker skills in two local authorities with different organisational models of practice. Westlake, Forrester, Killian and Whittaker (forthcoming) also identified a statistical link between MITI skills and self-reported violence or threatening behaviour from clients, with higher MI skill being associated with less reported resistance from parents. Taken together these studies suggest the MITI can be used in simulated interviews in ways that suggest links with "real world" outcomes such as the impact of training, differences between authorities and most importantly self-reported experiences with families. The skills captured by the MITI therefore provide a helpful starting point in deciding what skills should be included in the measure developed as part of this study.

The coding scheme described in this paper was developed over several years of research. Here the key stages in the development of the scheme are described. As discussed, we have used the MITI extensively in studies looking at the impact of training in MI. It rapidly became apparent that "Direction" was not a useful dimension for our purposes. Conceptually, "Direction" was problematic because unlike the other dimensions – all of which were uni-directional with higher scores being considered to be "better" – high levels of Direction could be good or bad, depending on the ways the worker was being directive. Pragmatically, there was almost no variation in our samples: social workers were always directive when compared to counsellors and therefore little was added by coding for Direction. This left us with the three dimensions that comprise MI skill (collaboration, evocation and autonomy) and a key foundational skill (empathy).

A key criticism we faced from workers and managers in using the MITI as a measure of practice skill, was that it missed important aspects of the social work role associated with authority and the appropriate exercise of power. While there was a general sense that skills such as demonstrating empathy or working collaboratively were important, some workers suggested that a focus on empathy and parental autonomy might be dangerous if it led to a failure to focus on the child or be clear about concerns. This echoed findings from Serious Case Reviews which highlight the ease with which the focus on the child can be lost in child protection work (Office for Standards in Education, Children's Services and Skills [Ofsted], 2010). Furthermore, it

has been argued that moving to strengths based approaches which emphasise the client's autonomy and capacity to identify and achieve goals, may result in a failure to be clear about power, authority and a focus on the child (Oliver, 2012). This seemed an important limitation for us to address. At the least we wanted to ensure that a focus on effective engagement skills with parents did not place children at risk.

We therefore developed further dimensions that attempted to provide a description of key elements of child and family social work that encompassed both the more collaborative elements (captured by the MITI) and the appropriate use of authority (Ferguson, 2011). We started by reviewing key social work textbooks and identifying key elements in the good use of authority. We then carried out four seminars which included a wide range of academics, practitioners and practice leaders. These sessions involved an iterative process of refining first the number of dimensions and then the descriptors for different levels of skill. The aim was not to capture every element of social worker skill in direct practice but to identify key elements that would allow us to describe and code for skills associated with both care and control.

The development and refinement of the approach was greatly helped by our involvement in setting-up a new practice-based postgraduate social work programme known as Frontline (see Maxwell et al., 2015). Most of the grades for the Frontline course involved direct observation and assessment of practice. The dimensions we had developed in research fed in to an adapted system for grading students on this course. The process of developing the approach to coding and trying it out on 704 recordings of direct practice, was crucially

helpful and fed-back into the development of the dimensions discussed in the current paper. We were able to refine our understanding of the key skills required in direct social work practice and of these skills, which could be coded for. See Domakin and Forrester (forthcoming) for a description of assessment of direct practice on the Frontline course.

Through this process three further dimensions of practice were identified. These were:

1. Purposefulness: The extent to which the social worker sets out and maintains a focus for the session whilst demonstrating flexibility in response to the client's agenda. To some extent purposefulness is a measure of the degree to which as a listener we feel clear about the point of the session being graded.

2. Clarity about concerns: The extent to which the social worker is clear about the reasons for professional involvement and is able to engage in meaningful dialogue with the client about issues or concerns.

3. Child focus: The extent to which the social worker ensures that the child is meaningfully integrated into the discussion in order to enhance the parents understanding of the child's needs.

For consistency, the new dimensions of skill were rated on a similar five point Likert-type scale to the MITI where 1 indicates low skill and 5 high skill.

During piloting it became apparent that one important variation to the MITI coding approach would be needed. While most of the dimensions could be coded using 20 minute segments of recorded practice, *Clarity about Concerns* and *Child Focus* required listening to a whole interview.

Finally, the dimensions and descriptors as described in the current study were tested on a sample of simulated interviews. We were left with a coding scheme based on four dimensions of the MITI plus three new dimensions focussing on appropriate use of authority. This scheme is not intended to capture all there is to direct practice. However, we developed it in order to explore whether we could reliably identify key elements of good social work practice, in the hope that if we could, this might allow the exploration of the impact of these key skills on outcomes. Here we describe whether the skills could be reliably coded for.

Method

Participants

Participants for this study were qualified social workers involved in the 'Engaging Parents and Protecting Children' study; a randomised controlled trial to test the effectiveness of training and supervision in Motivational Interviewing in child and family social work (Forrester et al., forthcoming a; Forrester et al., forthcoming b). In this study, all parents or carers whose child had an allocated social worker over a period of seven months were asked whether they consented to having a session with their social worker observed, audio recorded and analysed by a researcher. One hundred and thirty three parents or carers who had more than three visits from their worker agreed which resulted in 133 audio recordings of direct practice sessions from a total of 51 qualified social workers. Social workers were 80% female with a mean age of 36.9 and an average of 6.4 years post qualifying experience. All were employed in one London local authority Children's Services department.

Audiotaped sessions of direct practice

Sessions varied considerably, dependent primarily on the nature of the concerns. The types of issues discussed in the sessions included; domestic abuse, substance misuse, parental mental health, physical abuse, behaviour management, conditions in the home, housing and financial support. The majority of sessions took place in the family home, although some took place at the worker's office or in the community. The sessions varied considerably in length (from 11 minutes through to 2.5 hours).

Procedure for training coders

Coders were the first author and two Research Assistants at the University of Bedfordshire. The first author is a qualified social worker and has been trained in MI and coding using the MITI 3.1.1. Neither Research Assistant had a background in either MI or social work. Coders were trained in two stages, first to code for the dimensions of evocation, collaboration, autonomy and empathy as outlined in the MITI. Then, in coding for the new dimensions of purposefulness, clarity about concerns and child focus.

Training to code using the MITI involved participating in workshops facilitated by the first author, using audio files and DVDs to familiarise coders with the style of MI. Coders were then trained using audio recordings of simulated client interviews

from a previous study involving the authors (Forrester et al, forthcoming a). Tapes from this study had been scored by 'expert coders' from two independent specialist MITI coding services which provided a 'gold standard' against which to compare scores. Coders were deemed reliable when they were able to score each of the four global dimensions within a margin of 1.0 from the 'gold standard' scores on 80% of 10 consecutive simulated interview tapes. This method replicated the one used by Moyers et al. (2005) when developing the MITI.

Once competence in using the MITI had been achieved, training then proceeded to coding all seven dimensions using direct practice audio recordings from a pilot study undertaken by some of the authors (Forrester et al., 2013). These recordings were initially coded by the first author to provide a standard against which the trainees could compare scores. As the trainee coders did not have a background in social work, the new dimensions were introduced by listening to and discussing the direct practice recordings. The trainees then followed the same process as before until they were able to code 80% of 10 consecutive direct practice recordings within a margin of 1.0 on each of the 7 dimensions of the coding tool. In total, it took approximately 60 hours of training for coders to reach inter-rater reliability on all seven domains of skill. Throughout the three month duration of the study coders met for weekly sessions in which a tape was coded collectively in order to prevent drift.

Sampling strategy

One hundred and thirty-three audio recordings were randomly allocated for coding to either the first author or one of the two trainee coders. Recordings were scored on each of the seven dimensions of practice using the coding scheme outlined above. The process of coding happened in stages, with 10 recordings at a time being allocated for coding. Of the 10, 20% were randomly selected to be scored by all three coders in order to test for inter-rater reliability. Scores were then checked and a further 10 recordings were distributed, repeating the same process. This allowed for any issues with reliability to be addressed through discussion and further training at an early stage. In total, 28 (21%) of the 133 audio recordings were scored by all three coders.

Data analysis

Inter-rater reliability was calculated using two analytic tests. Inter-rater reliability was first assessed using Krippendorff's α (Hayes & Krippendroff, 2007). This statistic of IRR has the advantage of being able to estimate reliability at any level of measurement (nominal, ordinal, interval/ratio data), any number of raters, and regardless of missing data. Values greater than .40 are considered good, and values greater than .75 are considered excellent (Krippendorff, 1987; Lombard, Snyder-Duch & Bracken, 2002; Reeves, Mullard & Wehner, 2008).

Intra-class correlation (ICC) was calculated using a two-way and agreement based variant (Hallgren, 2012). This commonly reported statistic for IRR is able to provide a statistic given nominal, ordinal or interval/ratio data from two or more raters. ICC, when using a consistency variant, computes a reliability statistic based on variation between raters and not absolute agreement in scores. Larger disagreement between raters,

or greater variation among raters, will result in lower estimates of reliability and lower scores. Generally, ICC scores of less than .40 are considered to have poor IRR, scores between .40 and .59 are considered to have fair IRR, scores between .60 and .74 are considered good IRR, and scores over .75 indicate excellent IRR (Cicchetti, 1994). Prior reliability analyses for MITI have used the same ICC analysis and variant (Moyers et al., 2005; Moyers, Rowell, Manuel, Ernst, & Houck, 2016).

Results

Results from the reliability analyses including means and standard deviations for each coder are provided in Table 2. Inter-rater reliability between coders was all good to excellent. For core dimensions of MITI, each demonstrated good ($\alpha > .40$) and excellent ($\alpha > .75$) Krippendroff's α scores with a range between .731 to .796. Similarly, the ICC scores indicated excellent inter-rater reliability with scores ranging from .897 to .924. The coders demonstrated excellent reliability with the MI Spirit Skill scores with

Krippendroff's α = .937 and ICC = .967 for this measure.

Indicators of inter-rater reliability for the new dimensions of social work skill were lower, although scores indicated good to excellent reliability among raters. Krippendroff's α ranged from .461 to .649, and ICC ranged from .731 to .853.

Evocation can only be coded when a behaviour change issue is discussed. Social work interviews often cover a wide range of issues and it was not always obvious whether a behaviour change issue had been discussed in sufficient detail to allow coding for evocation. The decision by the raters to code Evocation was therefore analysed. In 24 of the 28 cases, all three raters agreed on appropriateness of coding for Evocation, or an agreement rate of 86% among all three raters. Where coders decided to score for evocation, it was possible to identify a behavioural change issue that related to the individual being spoken to. In these instances there was usually evidence of the worker trying to change a parent or carers behaviour either through persuasion and education (low scoring) or through evocative techniques which draw on the parent's

Table 2. Inter-rater reliability (IRR) among MITI+ domains (n=28)

Domain	Coder 1 mean (SD)	Coder 2 mean (SD)	Coder 3 mean (SD)	Krippendroff's α	Intra-Class Correlation
Evocation	2.52 (0.87)	2.71 (1.01)	2.62 (0.80)	.731	.897
Collaboration	2.88 (0.97)	2.84 (0.99)	2.72 (1.06)	.796	.918
Autonomy	2.96 (1.04)	2.83 (0.96)	2.67 (0.82)	.741	.930
Empathy	2.69 (1.16)	2.81 (1.23)	2.85 (1.01)	.797	.924
MITI Skill	2.81 (0.89)	2.79 (0.90)	2.67 (0.85)	.937	.967
Purposefulness	3.00 (0.63)	3.04 (0.60)	3.19 (0.69)	.461	.731
Concerns	3.00 (0.89)	3.00 (0.98)	2.88 (0.99)	.589	.846
Child Focus	3.04 (0.77)	2.92 (0.69)	2.92 (0.80)	.649	.853

intrinsic motivation (high scoring). Examples of the sorts of change issues that were discussed included managing children's behaviour, school attendance, substance misuse, domestic abuse and improving home conditions.

Discussion

This paper has described the development of a coding scheme to measure seven key dimensions of skilled communication in child and family social work practice. The primary aim of this study was to test whether these dimensions could be reliably coded by three independent raters using audio recordings of direct practice. The findings from this study were promising. Krippendorff's α scores ranged from .461 (good) to .937 (excellent) and ICC ranged from .731 (good) to .967 (excellent) across all seven dimensions, suggesting that raters were able to reach a high level of agreement around the quality of practice in relation to different elements of skilled social work communication. Inter-rater reliability was slightly higher for the dimensions adopted from the MITI than the newly developed ones. However, this is somewhat unsurprising given that the reliability of the MITI has been long established (Moyers et al., 2005; Pierson et al., 2007).

Inter-rater reliability for the global spirit rating and dimension of empathy was in a range that is comparable to, if not higher than, in other settings (Moyers et al., 2005; Forsberg et al., 2007; Pierson et al., 2007; McCambridge, Day, Thomas & Strang, 2011; Seng and Lovejoy, 2013; Spohr, Taxman, Rodriguez & Walters, 2015). This was a welcome finding as the reliability of the MITI has not previously been established in child and family social work settings. There are few equivalent studies in the field of social work from which to draw comparisons about inter-rater reliability. However, it is possible to draw from research within the field of Medicine where the assessment of communication skills is commonplace (Comert et al, 2016; Zill et al, 2014). Findings from a systematic review of measures of physician-patient communication suggest that out of ten studies that examined observer instruments, only two received a 'good' score for inter-rater reliability (Zill et al, 2014). Whilst these studies examine practice in the context of medical settings, they do suggest that the level of inter-rater reliability achieved in this study was extremely promising for an instrument measuring direct practice communication skills.

The reliable coding of key dimensions of skilled communication provides an important foundation for further work exploring the validity of the measures. Ultimately, our broader aim is to understand whether any of the dimensions of skilled communication described in this paper predict client outcomes. For instance, are highly empathic or purposeful workers more likely to create client change than workers who are less empathic or purposeful? Research on the effectiveness of social work communication skills is an area that is strikingly absent from the social work literature. Reliability is an important step in this process, but the most important test of worker skill is client outcomes (Bogo et al, 2002). In a series of on-going studies we explore links between skills and outcomes, using the reliability of the measures, as a foundation.

The reliable coding of skills will also allow researchers to explore is what factors influence social worker skills. At present we are undertaking studies comparing dif-

ferent local authorities and teams within local authorities, evaluating training programmes and analysing different approaches to supervision. Reliable measures of worker skills provide a way for social work research to explore key factors that might influence the quality of practice.

This work is founded on the ability to reliably measure practice. However, the process of doing so has also identified challenges for such an approach. A key issue is the complexity of evaluating practice, and in particular the importance of the interplay between *practice* and *context*. Specifically, to what degree should worker skill be conceptualized as dependent on context and to what degree can key elements be seen to be universal? Huntley et al (2012) argue that assessment of communication should be based on whether the approach 'works' in any given context, rather than whether or not a clinician is able to evidence a predefined set of skills. Furthermore, in their study of OSCE's, Bogo et al (2011) found weak correlations between scores from any two randomly selected simulated client interviews (out of a possible five), suggesting that the application of social work skills may be context specific.

In this study, raters raised concerns with regards to making judgements about some dimensions of skill in the absence of context. For example, could it be appropriate to focus exclusively on the parent's needs if a previous discussion focused extensively on the needs of the child? These decision making processes would not necessarily be reflected in the scores. However, raters also suggested that some dimensions of practice such as collaboration, empathy and purposefulness seemed less dependent on context, in that it is hard to imagine interviews where good practice would be characterised by a lack of purpose, empathy or collaboration. Our learning from this study suggests that consideration of such contextual factors is likely to provide the best indication of what best practice is in any given situation. Establishing a reliable coding system for key dimensions of practice provides a foundation for empirically exploring some of these contentions.

The issue of context touches on another element of the coding approach that requires further thought. In this study, the decision was made not to code for evocation on 25% of the sample because no behaviour change issue was discussed. This may be for legitimate reasons (for instance because the focus was solely on assessment or discussion of a social problem). However, it may also be because the worker does not successfully create a conversation about what appears to be a legitimate change issue. The complexity involved in these considerations is an area which requires further consideration.

Yet, analysis of conversation is not solely about understanding specific interactions in context – important as that is. At a wider level we need to be able to define key elements of good practice so that we can develop better descriptions of what good social work is, provide social work education and assessment based on this, and build organisations more able to deliver such good practice. It is in this area that we need to build generalizable models of good practice. A key step in such a programme of work is to ensure that the models can be reliably recognized.

Limitations

There are several potential limitations to this study. The research was carried out in one London based Children's Services department, and while it is currently being replicated in seven other authorities, for the reported findings this may affect generalizability. Secondly, the interviews recorded were predominantly toward the beginning of social work contact, and while representative of all cases allocated, most were from families identified by workers as being low risk. It is possible that the pattern of social work communication is different in important ways in longer-term work and with cases where there are higher level concerns.

At a more fundamental level, our study focuses on skills in direct communication with parents. There is far more to good social work than this, including, amongst other things, talking to children and young people, work with whole families, assessment and decision-making, liaising with other organisations, managing multi-professional meetings, writing up reports, not to mention phone and email contact and numerous other aspects of practice. We are therefore reporting solely on one element of social work – albeit an element that is widely perceived to be of considerable importance.

Finally, establishing reliability does not mean these dimensions should be reified. Indeed, some of the dimensions seem so closely related to one another that it may not be worth separately coding for them. More importantly, the ultimate test is of the validity of the dimensions. It is in their ability to predict key elements of parental experience and child and family outcome that these dimensions must ultimately be judged. The research reported here simply constitutes a first step in a process of developing our understanding of how to code and quantitatively research social worker skills.

Recommendations

As mentioned, an important next step is exploring the predictive validity of the measures: do these skills influence outcomes for children or families? And if they do, which skills and to what degree? Such considerations should feedback into further refinement and development of our understanding of worker skills and the way that the level of skill is best operationalized.

Once skills can be reliably coded for and linked to outcomes the factors that shape the level of skill require study. These could include variation between worker, between organisations and the impact of specific interventions such as training.

Further research is also needed to explore the degree to which skills are consistent across sessions. We have established that different raters listening to the same session can achieve a high degree of agreement. However, we do not know the degree of consistency for individual workers across different sessions: to what degree can level of skill be thought of as something we ascribe to individual workers, and to what degree might it be best thought of as varying by context. This is a crucial question for understanding the nature of social worker skills, and a reliable coding system allows it to be analysed empirically.

Related to this, understanding contextual issues and how they shape practice seems a particularly important next stage in refining our understanding of social worker skills. Whilst quantitative measures enable us to make generalizable links between

practice skills and outcomes, a combination of both qualitative and quantitative analyses is likely to generate the most meaningful picture of worker interaction (Ackerman & Hilsenroth, 2003). We would therefore recommend that further research explore the interplay between practice and context in greater depth. This might include exploring variations in skill across different sessions, analysing sessions with the addition of contextual information, or the development of the measure to include more in-depth combinations of observations of decision-making and practice to provide a multi-level view of practice. Ultimately the aim should be to identify some elements of practice that should be universal, others that are context dependent.

Having established that reliability is possible through an in-depth and time-consuming process, it is important to explore whether acceptable reliability is achievable through a streamlined process. This is particularly important if ultimately it is hoped that the codes might be useful in evaluating social workers, students or services. Currently in England it is proposed that child and family social workers will be accredited through simulated interviews, amongst other measures (Department for Education, 2014). Our study suggests that to do so reliably may be possible, but that achieving such reliability is likely to be a painstaking process – as well as one that does not necessarily have evidence for its validity.

Conclusion

Our aim in developing a reliable tool for coding direct practice is thus on the one hand to allow us to begin to explore the complex interplay between worker skill and outcomes for families, and on the other to help us think about what factors influence worker skill. The tool we have reported on in the current paper is therefore a foundational aspect of an on-going programme of work. Yet as such it is simply a starting point in helping us think about what good practice is, what difference it makes and how we can create organisational structures and cultures that support excellent practice.

References

Ackerman, S.J., & Hilsenroth, M.J. (2003). A review of therapist characteristics and techniques positively impacting the therapeutic alliance. *Clinical Psychology Review*, *23*(1), 1-33. doi:10.1016/S0272-7358(02)00146-0

Bogo, M., Regehr, C., Hughes, J., Power, R., & Globerman, J. (2002). Evaluating a measure of student field performance in direct service: testing reliability and validity of explicit criteria. *Journal of Social Work Education, 38*(3). 385-401.

Bogo, M., Regehr, C., Logie, C., Katz, E., Mylopoulos, M., & Regehr, G. (2011). Adapting Objective Structured Clinical Examinations to Assess Social Work Students' Performance and Reflections. *Journal of Social Work Education*, *47*(1), 5-18. doi**:**10.5175/JSWE.2011.200900036

Bogo, M., Regehr, C., Katz, E., Logie, C., Tufford, L., & Litvack, A. (2012). Evaluating the use of an objective structured clinical examination (OSCE) adapted for social work. *Research on Social Work Practice, 22*(4), 428 - 436. DOI: 10.1177/1049731512437557

British Association of Social Workers (2016). *Social Work Careers.* Retrieved from: https://www.basw.co.uk/social-work-careers/

Cicchetti, D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284-290. DOI: 10.1037/1040-3590.6.4.284

Cömert, M., Zill, J.M., Christalle, E., Dirmaier, J., Härter, M., & Scholl, I. (2016). Assessing Communication Skills of Medical Students in Objective Structured Clinical Examinations (OSCE) - A Systematic Review of Rating Scales. *Plos ONE*, *11*(3), 1-15. doi:10.1371/journal.pone.0152717

Department for Education (2014). *Knowledge and skills statement for approved child and family practitioners.* Retrieved from: https://www.gov.uk/government/publications/knowledge-and-skills-statements-for-child-and-family-social-work

Diggins, M. (2004). *Teaching and learning communication skills in social work education.* London: Social Care Institute for Excellence (SCIE)

Domakin, A. Forrester, D. & Killian, M. (forthcoming). *Can Practice Be Graded Reliably? A Study of 700 Observations of Practice*. Unpublished manuscript.

Ferguson, H. (2011). *Child protection practice.* New York: Palgrave Macmillan.

Forrester, D., Kershaw, S., Moss, H., & Hughes, L. (2008). Communication skills in child protection: how do social workers talk to parents? *Child & Family Social Work, 13*(1), 41-51. DOI: 10.1111/j.1365-2206.2007.00513.x

Forrester, D., McCambridge, J., Waissbein, C., Emlyn-Jones, R., & Rollnick, S. (2008). Child Risk and Parental Resistance: Can Motivational Interviewing Improve the Practice of Child and Family Social Workers in Working with Parental Alcohol Misuse? *British Journal of Social Work, 38*(7), 1302-1319. doi:10.1093/bjsw/bcl394

Forrester, D., Westlake, D., McCann, M., Thurnham, A., Shefer, G., Glynn, G., & Killian, M. (2013). *Reclaiming social work? An evaluation of systemic units as an approach to delivering children's services.* Retrieved from https://www.beds.ac.uk/__data/assets/pdf_file/0008/256742/Short-Systemic-Unit-Report-June-2013.pdf

Forrester, D., Westlake, D., Whittaker, C., Thomas, R., Waits, C., Antonopoulou, V., Killian, M & Hutchison, D. (forthcoming a). *Engaging Parenting and Protecting Children: A Randomized Controlled Trial of the Impact of Training Social Workers in Motivational Interviewing for Children and Families.* Manuscript submitted for publication.

Forrester, D., Westlake, D., Whittaker, C., Thomas, R., Waits, C., Antonopoulou, V., Killian, M & Hutchison, D. (forthcoming b). *How Do Worker Skills Impact on Outcomes for Children and Parents?: A Follow-Up Study of 121 Families.* Unpublished manuscript.

Forsberg, L., Källmén, H., Hermansson, U., Berman, A. H., & Helgason, A. R. (2007). Coding counsellor behaviour in motivational interviewing sessions: inter-rater reliability for the Swedish Motivational Interviewing Treatment Integrity Code (MITI). *Cognitive Behaviour Therapy, 36*(3), 162-169. doi:10.1080/16506070701339887

Hall, C., Juhila, K., Matarese, M., & van Nijnatten, C. (Eds.). (2014). *Analysing Social Work Communication: Discourse in Practice.* Abingdon: Routledge.

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Quantitative Methods for Psychology, 8*(1), 23-34.

Hayes, A. F. & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures, 1,* 77-89. Retrieved from http://afhayes.com/public/cmm2007.pdf

Hohman, M. (1998). Motivational interviewing: an intervention tool for child welfare case workers working with substance-abusing parents. *Child Welfare, 77*(3), 275-289. Retrieved from http://www.cwla.org/child-welfare-journal/

Huntley, C. D., Salmon, P., Fisher, P. L., Fletcher, I., & Young, B. (2012). LUCAS: a theoretically informed instrument to assess clinical communication in objective structured clinical examinations. *Medical Education, 46*(3), 267-276. doi:10.1111/j.1365-2923.2011.04162.x

Koroloff, N. M., & Rhyne, C. (1989). Assessing Student Performance in Field Instruction. *Journal of Teaching in Social Work, 3*(2), 3-16. doi: 10.1300/J067v03n02_02

Krippendorff, K. (1987). Association, agreement and equity. *Quality and Quantity, 21,* 109–123.

Lombard, M., Snyder-Duch, J., & Bracken, C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*(4), 587–604.

Lundahl, B., Kunz, C., Brownell, C., Tollefson, D., & Burke, B. (2010). A meta-analysis of motivational interviewing: twenty-five years of empirical studies. *Research on Social Work Practice, 20(*2), 137-160. doi:10.1177/1049731509347850

Maxwell, N., Scourfield, S., Le Zhang, M., Villiers, T., Hadfield, M., Kinnersley, P., Tayyaba, S. (2015). *Independent Evaluation of the Frontline Pilot.* Retrieved from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/509240/DFE-RR507-Frontline-pilot-independent-evaluation.pdf

McCambridge, J., Day, M., Thomas, B. A., & Strang, J. (2011). Fidelity to Motivational Interviewing and subsequent cannabis cessation among adolescents. *Addictive Behaviors*, *36*(7), 49-754. doi:10.1016/j.addbeh.2011.03.002

Miller, W.R., & Rollnick, S. (2013). *Motivational Interviewing: helping people change.* London: Guilford Press.

Moyers, T.B., Martin, T., Manuel, J.K., Hendrickson, S.M., & Miller, W. R. (2005). Assessing competence in the use of motivational interviewing. *Journal of Substance Abuse Treatment, 28*(1), 19-26. doi:10.1016/j.jsat.2004.11.001

Moyers, T. B., Martin, T., Manuel, J. K., Miller, W.R. & Ernst, D. (2010). *Revised Global Scales: Motivational Interviewing Treatment Integrity 3.1.1 (MITI 3.1.1).* Retrieved from http://casaa.unm.edu/download/MITI3_1.pdf

Moyers, T.B., Rowell, L.N., Manuel, J.K., Ernst, D., & Houck, J.M. (2016). The Motivational Interviewing Treatment Integrity Code (MITI 4): rationale, preliminary reliability and validity. *Journal of Substance Abuse Treatment*, *65*, 36-42. doi:10.1016/j.jsat.2016.01.001

Office for Standards in Education, Children's Services and Skills (Ofsted). (2010). *Learning lessons from serious case reviews 2009–2010*. Retrieved from https://www.gov.uk/government/publications/ofsted-learning-lessons-from-serious-case-reviews-2009-2010

O'Hare, T. & Collins, P. (1997) Development and Validation of a Scale for Measuring Social Work Practice Skills. *Research on Social Work Practice.* 7(2), 228-238. doi: 10.1177/104973159700700206

Oliver, C. (2012). What does Strengths-Based Practice Mean for Relationships between Child Protection Workers and their Mandated Adult Clients. *Relational Child & Youth Care Practice, 25(4),* 5-12.

Pierson, H. M., Hayes, S. C., Gifford, E. V., Roget, N., Padilla, M., Bissett, R., & Fisher, G. (2007). Regular article: An examination of the Motivational Interviewing Treatment Integrity code. *Journal of Substance Abuse Treatment, 32*(1), 11-17. doi:10.1016/j.jsat.2006.07.001

Reeves, M., Mullard, A., & Wehner, S. (2008). Inter-rater reliability of data elements from a prototype of the Paul Coverdell National Acute Stroke Registry. *BMC Neurology, 8*(1), 19. **DOI:** 10.1186/1471-2377-8-19

Rubak, S., Sandboek, A., Lauritzen, T., & Christensen, B. (2005). Motivational interviewing: a systematic review and meta-analysis. *British Journal of General Practice, 55*(513), 305-312. Retrieved from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1463134/

Seng, E. K., & Lovejoy, T. I. (2013). Reliability and Validity of a Treatment Fidelity Assessment for Motivational Interviewing Targeting Sexual Risk Behaviors in People Living with HIV/AIDS. *Journal of Clinical Psychology in Medical Settings*, *20*(4), 440-448. doi:10.1007/s10880-012-9343-y

Spohr, S. A., Taxman, F. S., Rodriguez, M., & Walters, S. T., (2016). Motivational Interviewing Fidelity in a Community Corrections Setting: Treatment Initiation and Subsequent Drug Use. *Journal of Substance Abuse Treatment*, *65*(1), 20-25. doi:10.1016/j.jsat.2015.07.012

Wahab, S. (2005). Motivational Interviewing and Social Work Practice. *Journal of Social Work, 5*(1), 45-60. doi:10.1177/1468017305051365

Watson, J. (2011). Resistance is futile? Exploring the potential of motivational interviewing. *Journal of Social Work Practice, 25(4),* 465-479. doi:10.1080/02650533.2011.626653

Westlake, D. Killian, M. & Forrester, D. (2014). *Helping Families and Protecting Children in Local Authority X: Evaluating, Training and Supporting the Children in Need Service.* Luton: Tilda Goldberg Centre.

Westlake, D., Forrester, D., Killian, M. & Whittaker, C. (forthcoming). *Social Work Communication Skills, Worker Characteristics and Experiences of Practice: A study of correlates of skill in simulated practice for 178 workers*. Unpublished manuscript.

Whittaker, C., Forrester, D., & Antonopoulou, V. (2015). *Motivational Interviewing for a Family Support Team: Evaluation of a Skills Development Package*. Luton: Tilda Goldberg Centre.

Wilson, B. (1981). *Field Instruction*. New York: Free Press.

Zill, J. M., Christalle, E., Müller, E., Härter, M., Dirmaier, J., & Scholl, I. (2014). Measurement of Physician-Patient Communication—A Systematic Review. *Plos ONE, 9*(12), 1-20. doi:10.1371/journal.pone.0112637