Motivational interviewing training for child and family social workers in Finland: An exploratory evaluation study

Elina Aaltio [1,2,*], Kaisa Saurio [3], Matti Heino[3,4],
Kaisa Pasanen[2], Nanne Isokuortti[2], Laura Alasimonen[2],
Johanna Moilanen [2,5], Nelli Hankonen [3],
Donald Forrester [6], and Maija Jäppinen [2]

---

[1]Faculty of Social Sciences, Social Work, Tampere University, Kalevantie 5, Tampere, 33014, Finland
[2]Faculty of Social Sciences, Social Work, University of Helsinki, Unioninkatu 37, University of Helsinki, 00014, Finland
[3]Faculty of Social Sciences, Social Psychology, Tampere University, Kalevantie 5, Tampere, 33100, Finland
[4]Medical Faculty Mannheim, Heidelberg University, Röntgenstraße 7, Mannheim, 68167, Germany
[5]Faculty of Humanities and Social Sciences, Social Work, University of Jyväskylä, Keskussairaalantie 2, Jyväskylä, 40014, Finland
[6]School of Social Sciences, Children's Social Care Research and Development Centre CASCADE, Cardiff University, Maindy Road, Cardiff, CF24 4HQ, UK

*Corresponding author. Faculty of Social Sciences, Social Work, Tampere University, Kalevantie 5, Tampere, 33014, Finland. E-mail: elina.aaltio@tuni.fi

Abstract

Good communication skills are essential for effective social work practice, yet evidence-based communication approaches are rarely applied in child and family social work. Motivational interviewing (MI) has shown promise in enhancing practitioners' skills and outcomes in child and family social work, but its use outside the UK remains

understudied. This study evaluates improvements in social workers' communication skills following MI training tailored to child and family social work in Finland. The training included coaching sessions and a self-practice programme informed by behaviour change science. Thirty-three Finnish social workers participated. Recordings of simulated and real service user meetings were analysed using the Social Work and Interviewing Motivationally (SWIM) coding tool. Of the seven measured communication skills, only empathy showed statistically significant improvement at the group level. Due to the small sample size, statistical power was limited. Notably, person-based analyses identified three subgroups: nine participants improved, four declined, and the rest showed mixed changes. Exploratory survey analyses suggested that differences in practice, self-efficacy, and intentions influenced outcomes. These findings underscore the need to move beyond group averages to examine individual responses to training. Such detailed analyses can help optimize interventions and improve the effectiveness of communication skill development in social work.

Keywords: child and family social work; communication skills; motivational interviewing; social work practitioners; training.

Introduction

Good communication is widely regarded as essential for high-quality and effective social work (Trevithick 2012; Healy 2018; Koprowska 2020), yet relatively few studies have focused on how to train communication skills for social work practitioners or students (see e.g. Reith-Hall and Montgomery 2023). One widely used approach for effective communication is motivational interviewing (MI), which is among the few evidence-based communication approaches applied in social work contexts. However, there is little research on how to effectively train social work practitioners in MI and none that appears to explore variations in how individual practitioners learn MI.

   MI involves skills and tools for strengthening engagement and trust in social work encounters and promoting behaviour change (Forrester et al. 2019). Key skills include empathic listening, evocation (i.e. eliciting and enhancing service users' intrinsic motivation), fostering collaboration, and recognizing and increasing service users' autonomy. Concrete tools involve open questions, affirmations, reflections, summary statements, and identifying and eliciting 'change talk' (motivations or plans for change). There are over 1,000 randomized controlled trials with none we are aware of showing a negative impact and evidence of positive effects in many types of behaviour change discussions. For instance, research has found MI to have a positive effect in services related to substance use, health-related behaviours, and gambling (Lundahl et al. 2010;

Smedslund et al. 2011). When applied by social workers, MI has been mostly applied in healthcare settings to deal with similar problems, and to a lesser extent in social care (Boyle et al. 2019). The one study conducted in social care settings focused on training MI to child and family social workers (Forrester et al. 2018). In this randomized controlled trial (RCT), a training package including supervision improved participants' MI skills, although not enough to change service user outcomes. However, in secondary analysis from the same study, Forrester et al. (2019) found an association between social workers' MI-related communication skills and better service user outcomes, such as parent-reported engagement and perceived well-being. This is important because it is one of the few studies identifying a link between social work communication skills and outcomes for families. In addition, a systematic narrative review on applying MI in child welfare (Shah et al. 2019) found MI useful, for example, in promoting parenting skills, parent/child mental health, and retention in services.

To our knowledge, MI training studies with objectively coded pre-post meetings between service users and social workers in child and family social work have been conducted only in a single research project (Forrester et al. 2018, 2019) worldwide, despite the promise of this approach for service quality. Evaluating the outcomes and processes of such trainings will provide insight into how to further improve training effectiveness.

Adopting MI skills requires not only participation in training sessions but also rehearsal in professional practice and reflective engagement (Forrester et al. 2021; Miller and Rollnick 2023). However, the amount of rehearsal during the training and its association with learned skills are rarely empirically assessed and reported. Similarly, for the trainees to change their behaviour to be MI-consistent, it is important for the training to impart not only the skills, but also a sense of self-efficacy to use those skills (as measured by confidence in one's ability to use the skills in practice) as well as behavioural intention to continue using those skills (Fishbein and Ajzen 2010). Those with higher self-efficacy and stronger intentions to use those skills are likely to maintain MI-consistent behaviours sustainably, as has been shown in other behavioural domains (Fishbein and Ajzen 2010; McEachan et al. 2016).

Our research reports findings from a project designed to improve the effectiveness of direct child and family social work practice by strengthening social workers' communication skills in Finland. The project provided a four-day training course including coaching sessions and a self-practice programme for practitioners. The training was grounded in MI and incorporated three additional skills central to high-quality practice in child and family social work, that is, purposefulness, clarity, and child-focus (Whittaker et al. 2016; Forrester et al. 2019). The present study examines the training outcomes, specifically improvements in

participants' skills, which were assessed by rating real and simulated service user meetings that were recorded before and after the training, and their association with practicing, self-efficacy, and intentions to use those skills.

While intervention studies typically assess group-level changes, such nomothetic analyses can obscure the varied and meaningful ways individuals respond to training (Laursen and Hoff 2006; Saqr et al. 2024). Group averages tell us little about who benefits from an intervention, as factors like baseline skills can lead to diverse learning trajectories, and little to nothing about the processes of change. In addition, in the case of small samples, a few individuals with large changes can disproportionately affect the observed mean effects. This heterogeneity is central to calls for more person-centred, idiographic analysis in behavioural sciences (Molenaar 2007; Heino et al. 2021), for which exploratory data analysis (EDA) is a crucial tool for revealing patterns beyond what standard linear models can detect (Tukey 1977; Heino et al. 2019; Franconeri et al. 2021).

This study evaluates how training participants' MI skills changed from baseline to post-training assessment. The specific research questions are as follows:

1. Do social workers' communication skills change after the training, that is, is there a difference in the *average* communication skills?
2. At the *individual* level, what patterns of changes in skills were observed and what factors were associated with individual-level variations?

Given that this training was being evaluated in a novel context, our primary goal was to first discover and describe the nature of any changes, rather than to test pre-formulated and well-operationalized hypotheses. This approach ensures transparency by clearly separating data-driven discovery from formal statistical testing.

Materials and methods

Design

The study applied an explorative non-controlled repeated-measures design. It was part of a larger mixed-methods study focusing on needs and means to improve communication skills among social work practitioners working in statutory child and family social work. Ethical review was obtained from the Research Ethics Committee in the Humanities and Social and Behavioural Sciences of the University of Helsinki in September 2023. Data were collected before the training (T1), after the

first two training days (T2), and after the entire intervention (T3). The participation was based on informed consent.

Supplementary material can be found in Open Science Framework (OSF): https://osf.io/6c5ju/? view_only=aa5cff8acf8c40189abace978f3835a5

Context

While communication skills are regarded as a key competency in social work, their share in the social work curriculum in Finland is not especially high. Out of 300 credits of bachelor's and master's degrees that together qualify for social work, students typically receive fifteen to thirty credits of teaching and training in social work practice involving communication and other practice skills.

According to the Finnish Social Welfare Act (15 §), social work is described as practice that seeks to promote change. Hence, it was assumed that MI would be a good fit with the Finnish system. Previously, elements of MI have been embedded in social work curricula in some Finnish universities, and additional training has been provided to practitioners in some local organizations. However, this training has typically been short and general in nature, without additional support focusing on how to apply MI in practice, and without a specific focus on child and family social work.

Intervention

To ensure a high-quality and strong theory base for the training, we chose to adapt a training package developed at Cardiff University (https://cascadewales.org/motivational-interviewing-for-childrens-services-training/) to the Finnish context. The researchers translated the training materials and culturally adapted the case examples and exercises with the help of two social work researchers. In addition, some modifications were made, for example, the coaching was provided for small groups rather than for individuals. The training involved four full training days led by three facilitators with substantial experience in either MI-based trainings or social work education. The first two days focused on MI and key MI skills, and days 3 and 4, one month later, focus on applying MI and specific social work skills in child and family social work. The training days included lectures, small group discussions, and real play and role play exercises. To enhance the uptake and maintenance of the new skills, the training incorporated three coaching sessions, two optional online booster sessions, and a training handbook including a self-practice programme informed by behaviour change science. In addition, an adapted version of the SWIM tool (see Methods section) was included in the training handbook to support self-monitoring of skills acquisition.

Table 1. The content and schedule of the training and the data collection.

	Weeks									
	up to –1	1	2	3	4	5	6	7	8	9–14
Intervention										
Training days		1–2					3–4			
Coaching sessions						1				2–3
Rehearsal in Practice										
Peer group meetings			3 meetings (without a facilitator)							
Data Collection										
Recordings (real)	T1									T3
Recordings (simulated)	T1									T3
Survey		T1					T2			T3

Overall, the training took place over thirteen weeks, and the time lag between baseline and post-intervention measures was on average fourteen weeks (Table 1).

Participants

The training was delivered to two training groups in Spring 2024. Originally, thirty-seven social workers from sixteen teams applied for the training, and following drop-outs, thirty-three workers participated in two groups. In total, twenty-two social workers who had participated in recording simulated or real service user meetings both pre- and post-intervention were included in this study. Participants came from the full range of teams, from assessment to long-term, they were relatively experienced with half having six or more years of experience and all but two were licensed social workers. Half had had previous training in MI (for more detailed description of participants, see Supplementary Table S1).

Data

Real service user meetings

Training participants were asked to record their meetings with service users during two recording periods, each lasting two weeks at T1 and T3. Additionally, participants were asked to fill in information sheets covering basic information (length, participants, cancellations, recorded/not recorded, reason for not recording) on all scheduled meetings during these periods.

During the recording period, the social workers were instructed to recruit all service users and professionals participating in meetings that met the eligibility criteria (see Table 2). The form and location of the meeting did not matter. The meetings might include a child, a guardian or parent, and/or other professionals in any composition.

Table 2. The eligibility criteria for the real audio recordings.

Topic area	Inclusion criteria	Exclusion criteria
Meeting participants	In addition to the social worker, the meeting can involve four other individuals, who actively participate in the discussion	Meetings that involve five or more individuals who actively participate in the discussion
Children present	If the meeting is held primarily with children, they need to be plus twelve years old	A meeting held primarily with a child under twelve years old
Language	Finnish, English, Swedish, or Russian	Other languages
Risk assessment	Not applicable	Social worker estimates that the recording might cause harm to the service user's situation

A total of 286 scheduled real service user meetings were reported in information sheets at T1 and 184 at T3, with an average of 1.38 meetings per day per practitioner at T1 and 1.36 at T3. Of these meetings, seventy-eight (27 percent) were recorded at T1 and thirty-four (18 percent) at T3. An additional thirteen meetings were recorded, although not marked in information sheets, resulting in 125 recordings in total.

Meetings varied between 17 and 105 minutes ($M = 56.3$). Most were held in the office (59 percent), 9 percent were home visits, 18 percent phone calls and 9 percent online. The most common issues (not mutually exclusive) discussed in the meetings were school attendance (35 percent), child custody and meeting arrangements (26 percent), child's behaviour (21 percent), and parent's coping (18 percent).

Simulated interviews

All training participants were invited to attend simulated interviews at T1 and T3. The interviews were conducted with a professional actress. All participants were given the same scenario in the first interview, while at T3, there were two scenarios randomly allocated. The scenarios were adapted from the UK to the Finnish context with the help of researcher social workers. The interviews were conducted and recorded via Microsoft Teams. In total, twelve practitioners participated in simulated service user meetings both at T1 and T3, resulting in a total of twenty-four recordings.

Survey

All participants were invited to respond to an electronic survey at T1 focusing on background information, an intermediate survey at T2 focusing

on experiences of the first training days and self-practice, and an end survey at T3 focusing on experiences of the entire training and applying MI in practice. We analyse three categories of responses:

1. Self-reported frequency of practicing the skill during the intervention ('behavioural experiments') was assessed in a survey at T2, with queries about six skills: summaries, reflections, open questions, change talk recognition, change talk evocation, affirmations, spectrum of directiveness.
2. Self-efficacy to use each skill was measured at T3 with an item stem 'How certain are you, that you can use MI skills in your own work?', queries being about thirteen skills: summaries, spectrum of directiveness, rolling with resistance, reflections, open questions, guiding the meeting, elicit-provide-elicit, child focus, change talk recognition, change talk evocation, affirmations, negotiating the agenda, clarity of concerns.
3. Intention to use each of the aforementioned thirteen skills was measured at T3 with the question 'Do you intend to use the following MI skills and interaction techniques in your work?'.

Response options can be found in the Supplementary Material (Statistical procedures section).

Data analysis

SWIM coding

To assess training outcomes, practitioners' communication skills were rated by listening to and coding simulated and real service user meetings using the SWIM coding tool (Moyers et al. 2014; Forrester et al. 2024). SWIM includes four key MI skills, that is, evocation (enhancement of service user's intrinsic motivation), collaboration (working cooperatively with the service user), autonomy (recognizing and increasing service user's autonomy), and empathy (taking a curious and active interest in service user's worldview), and three skills particularly important for child and family social work, that is, purposefulness (maintaining a clear focus for the discussion), clarity about risk and need (raising and exploring issues and concerns), and focus on child (integrating child's needs and worldview into the discussion with the parent). Each of the skills is coded on a five-point scale based on descriptions and verbal anchors specified for each level. A previous study reported a high degree of inter-rater reliability (Whittaker et al. 2016).

The coding process proceeded as follows. Initially, the coding team (E.A., K.P., N.I., J.M., L.A.) trained in coding for thirty hours by listening and coding to real and simulated meetings (not included in the data)

and having reflective discussions. The team also attended a two-day training organized by the developers of the SWIM coding tool.

The coding team proceeded to the actual coding phase by first coding the simulated data. The coding team jointly decided the ratings for 17 percent ($n=4$) of the simulated service user meetings. The remaining (83 percent, $n=20$) of the simulated meetings were coded in pairs with coders first rating the audio recording individually and then resolving the final ratings mutually.

Nineteen social workers recorded real service user meetings both at T1 and T3. The coders started the rating process by listening to the first and last recordings of these practitioners. In twenty-four cases, the recording was not suitable for SWIM coding, and another recording was selected. The main reason for not coding the meeting was that the meeting was held with another practitioner who dominated the discussion. In addition, discussions with children under the age of twelve years were excluded. Finally, fifty-eight real service user meeting recordings were listened to, and thirty-four of them were coded involving seventeen practitioners. Of these, 26 percent ($n=9$) were coded in pairs and 74 percent ($n=25$) by one coder. Throughout the process, the coding team organized meetings to discuss and calibrate coding.

In the whole sample, seven practitioners could be rated based on both real and simulated meeting recordings at T1 and T3, five practitioners only on simulated and ten on only real meeting recordings, giving a total sample of forty-four recordings for twenty-two workers.

The inter-rater reliability was assessed by using intra-class correlation (ICC) based on all real or simulated meeting recordings that had been rated in pairs ($n=29$). ICC ranged from moderate ($\alpha > 0.5$) to good ($\alpha > 0.75$) (Koo and Li 2016).

Two skills were not applicable to all real service user meetings. First, in three cases, the active participant in a meeting was an older child and therefore focus on child was not coded. Second, in some cases, the need for behaviour change was not explicitly brought into discussion at any point during the meeting and therefore evocation could not be coded for.

Change in communication skills

To assess the change in communication skills in real meetings (RQ 1), a paired samples t-test was applied. Corrections were made to the significance level to account for multiple comparisons. To assess the robustness of the results, we also conducted non-parametric Wilcoxon signed-rank test using SPSS 29.

Exploratory data analysis

We used EDA for two purposes: examining the heterogeneity within mean-level changes in SWIM scores and exploring the correspondence

between survey measures and SWIM scores. Our approach to the latter followed this procedure:

First, we began by inspecting intra-individual variation in SWIM-coded skills and differences between T1 and T3 recordings therein. To do this, we examined plots of individual-level skill changes, and the heterogeneity therein. Based on this, we qualitatively grouped each individual into one of the following categories: 'Improvers', 'Decliners', and 'Others' (mixed or stagnant skill development).

Second, we examined the correspondence of this categorization with visualizations accounting for baseline performance. To do this, we plotted mean T1–T3 change scores against initial skill levels of average person-level skills (Supplementary Figure S4). Using a threshold of 0.5 points as indicating meaningful change, we grouped individuals into three categories—as in the person-level analysis in step one—and examined the correspondence between these two grouping methods.

Third, we explored how these SWIM performance-based categories ('Improvers', 'Decliners', 'Others') related to the process measures collected with the survey: practicing between sessions, self-efficacy, and intentions to use MI skills. To visualize these relationships, we created composite variables by averaging the item scores for each of the three survey categories: self-efficacy (13 items), practicing (6 items), and intention (13 items). A successful training experience, as captured by the 'Improvers' category, would conceivably correspond with higher self-reported practicing, self-efficacy, and intentions. Conversely, the 'Decliners' group would be expected to show the lowest scores, with the 'Others' group falling in between.

Lastly, to assess the data-driven hypothesis that emerged from the exploratory analysis, a statistical procedure was conducted. To do this, we first evaluated the psychometric quality of these composite scales. To address the known limitations of Cronbach's Alpha (Crutzen and Peters 2017), we used McDonald's Omega (ω), a more robust estimate of reliability derived from a factor-analytic model (McDonald 1999). Next, the ordered hypothesis that performance groups would align as Decliners < Others < Improvers was formally tested on the three composite variables using the Jonckheere–Terpstra test with 1M permutations (Terpstra 1952; Jonckheere 1954). This non-parametric test is specifically designed to detect a monotonic trend across ordered groups. All analyses were conducted in R version 4.5.0 (R Core Team 2025) using the package *psych* (Revelle 2025) for reliability analyses.

Results

Change in practice skills

The results of the main analysis, a paired samples t-test for the real service user meetings, are presented in Table 3. Out of the seven skills, empathy showed a statistically significant change at the group level.

To ensure the robustness of the findings, we also analysed the simulated meetings separately, as well as combining the simulated and real meetings. Since the normality of the differences was unclear, particularly for autonomy (Shapiro–Wilk's test, $P < .001$), we also performed non-parametric Wilcoxon signed-rank tests to confirm the reliability of the results. The results remained largely consistent across all these analyses, with a statistically significant change observed only in empathy. Our analysis involved multiple statistical tests, which increases the likelihood of finding a significant result by chance. To account for this, we adjusted our results using the widely accepted Benjamini–Hochberg procedure, which controls the false discovery rate (FDR). After this adjustment, the intervention's effect on Empathy was statistically significant only when using a lenient FDR threshold of 20 percent. The result did not meet the more conventional and robust threshold of 5 percent, suggesting this finding should be interpreted with caution as preliminary.

Person-centred, idiographic analysis (Supplementary Figure S2 in OSF) revealed the individual-level variation underlying the aggregate change presented in Table 3. Perhaps most importantly, we could see that the apparent drop in the mean of Evocation is driven by a single participant who scored 4 in their T1 recording and 1 in their T3 recording. This individual also contributed disproportionately negatively to the group-level results in collaboration and purposefulness, where other

Table 3. Means and standard deviations of skill scores at T1 and T3 in real service user meetings ($n = 17$).

	T1		T3		t(df)	P	Cohen's d	95% Confidence interval for Cohen's d	
	M	SD	M	SD				Lower	Upper
Evocation	3.20	0.84	2.60	1.52	t(4) = −0.802	.468	−0.359	−1.247	0.570
Collaboration	3.29	0.92	3.53	0.87	t(16) = 0.775	.450	0.188	−0.295	0.665
Autonomy	3.12	0.33	3.35	0.49	1.725	.104	0.418	−0.084	0.909
Empathy	2.88	0.78	3.53	0.72	2.524	.023*	0.612	0.084	1.124
Purposefulness	3.41	0.71	3.47	1.01	0.194	.848	0.047	−0.429	0.522
Clarity of concerns	3.24	0.90	3.53	1.23	1.429	.172	0.346	−0.831	0.149
Child focus	3.08	0.52	3.00	1.04	t(11) = −0.290	.777	−0.084	−0.649	0.485

*$P < .05$.

participants changed only a little. These findings highlight the importance of evaluating individual-level trajectories in our context.

EDA of SWIM and survey data correspondence

Initial EDA of person-level skills generally aligned with average skill change (steps one and two of analysis, see supplementary material, Exploratory Data Analysis (EDA) section). It yielded the following sizes for each category of interest: 'Improvers' ($n = 9$), 'Decliners' ($n = 4$), and 'Others' (mixed or stagnant skill changes; $n = 9$).

To triangulate the SWIM-based performance groupings, we examined their correspondence with participants' self-reported MI skill use intentions, self-efficacy, and amount of practicing (Supplementary Figure S6 in OSF). The analysis revealed a consistent pattern across all three survey categories. The 'Improvers' group reported the highest mean levels of self-efficacy, intention to use MI skills, and practicing of those skills. The 'Decliners' group reported the lowest mean levels in all three areas, while the 'Others' group consistently scored between the two. The plot presenting the individual data points and visual robustness analysis is presented in Supplementary Figures S7 and S8.

The composite scores were determined unidimensional ($\omega t = 0.77–0.89$; see Supplementary material, Statistical procedures section). Therefore, a Jonckheere–Terpstra test with Benjamini–Hochberg correction was applied to formally characterize the observed ordering. The test indicated statistically significant monotonic trends for all three composite variables: self-efficacy (JT $= 103$, adjusted $P = .002$), practicing (JT $= 165$, adjusted $P = .005$), and intention (JT $= 89$, adjusted $P = .024$). These results quantify the association between SWIM-assessed skill changes and self-reported process measures within this sample. However, because the statistical test was selected post-hoc and conducted on the same data as the observed patterns, the reported *P*-values underestimate true error rates. Hence, findings should be considered exploratory evidence for this relationship.

Discussion

The study examined training outcomes of social workers participating in a MI training tailored to child and family social work. The study has five key findings. First, we observed no statistically significant changes in interaction skills other than empathy. However, the small sample size and wide confidence intervals emphasize the need for cautious interpretation. The observed improvement in empathy may indicate that it is the first skill in which a positive change occurs. This would be in line with

Miller and Moyers (2006) suggestion of the eight stages in learning MI, according to which the progress begins with general openness to MI and then acquiring basic techniques needed to demonstrate empathy.

Secondly, our analyses revealed a strong, statistically significant relationship between observed skill improvement and the psychological drivers of behaviour change: social workers who improved most also reported the highest levels of practice, self-efficacy, and intention to use the skills. This demonstrates the importance of also engaging the motivation of workers. As Antoine Saint-Exupéry said: 'If you want to build a ship, don't drum up the men to gather wood, divide the work, and give orders. Instead, teach them to yearn for the vast and endless sea.' (as cited in Forrester et al. 2021). For training in communication skills, we may need more emphasis on supporting the participants' autonomous motivation (Ryan and Deci 2017). Additionally, attention should be paid to the participants' opportunities to practice and their self-confidence.

Thirdly, this study is one of the first endeavours to rate social workers' practice skills within statutory child and family social work outside the UK. Our study found it was possible to code reliably for elements of social work communication skills with the SWIM coding tool. In the future, a reliable skills coding scheme such as SWIM offers various possibilities for research on effectiveness of social work training and practice. Additionally, coding one's own practice is beneficial for learning new skills and hence recommended by MI trainers (Forrester et al. 2021). Listening and coding may promote self-awareness (e.g. identifying certain manners) and can help to set learning targets such as recognizing change talk. Practitioners may also overestimate how MI-consistent they are (see, e.g. Beckman et al. 2022). Coding practice makes this visible and may increase fidelity to MI.

While participants in this study were not required to code their recordings, they considered recording and listening to their own practice highly beneficial in learning MI (Isokuortti et al. manuscript in preparation). Future training could incorporate SWIM more systematically in the self-practice programme and coaching sessions. Additionally, the future training of trainers should also include training in coding trainees' communication skills.

Fourth, a methodological observation concerns how consistent social work practice skills may be between meetings. An assumption of this and similar studies is that we are sampling some sort of underlying level of practice skill. This is a reasonable assumption, on which much analysis of professional practice and education is based. However, social work conversations involve multiple people, and each one is framed by unique needs and institutional requirements. In such a situation, one might expect a high level of variation for each individual worker in the skills they demonstrate in different meetings, and indeed this has been found when consistency is analysed (Forrester et al. submitted for publication). This

effectively creates a lot of noise, which may make detecting the genuine signal difficult. Put simply, to reliably identify whether training had or had not made a difference, one might need a very large sample.

Fifth, we observe the importance of data visualization and EDA. The triangulation of SWIM ratings with self-reported process variables proved an interesting avenue for future research—we encourage researchers to heed recent calls to explore the space of behaviour change beyond group-level averages (Speelman et al. 2024). Due to EDA, we found correspondence between survey responses and skill improvement, as well as observed that a single participant's T3 service user recording scored significantly worse than their T1 recording. This latter, singular event would not have been observable absent EDA and explains the apparent reduction in group-level evocation scores. This singular case also draws down means in collaboration and purposefulness change scores.

Limitations, strengths, and future research

First, this study was an exploratory feasibility study, and therefore, it did not involve a control group.

Secondly, attrition in our sample was high—only twenty-seven percent of all meetings during recording weeks were recorded. One of the main reasons for not recording a meeting was that five or more participants were present. This may reflect the current trend to organize large network meetings in Finland. In the future, more research is needed to assess how the dynamics in such large meetings promote or hinder skilful conversations between practitioners and families. Another reason for attrition was that practitioners felt asking the service user's permission would harm the relationship. While such concern is relevant in the child protection context, it is also possible that practitioners have served—perhaps unintentionally—as informal gatekeepers due to their own concerns regarding the study (see Mezey et al. 2015). Our sample seems to lack difficult conversations, potentially due to such gatekeeping.

Another explanation for the lack of discussing difficult subjects might be practitioners' tendency to avoid them. During our coding process, we observed that social workers tended not to bring harmful behaviour into the discussion and occasionally ignored the topic even if a service user referred to such a problem. Moreover, this observation highlighted the need to revise the SWIM coding handbook (Moyers et al. 2014; Forrester et al. 2024). While ignoring harmful behaviour lowers the scores for clarity about risk and need, we suggest this omission should also be acknowledged in evocation ratings. The current description expects practitioners to address the change somehow, for example, by providing reasons to change (low on scale) or evoking the service user's own reasons (high on scale), but it does not consider practice where

behaviour change is dismissed even if it should not be. Another area requiring revision was collaboration. In its current form, the description and verbal anchors of collaboration do not cover practice where the practitioner is mainly following the service user because they seem not to know how to balance the conversation.

This brings us to our third limitation. Despite having extensive discussions with the coding team, we achieved suboptimal inter-rater reliability ranging from moderate to good. Still, the reliability is sufficient for drawing the current preliminary conclusions.

Fourth, the small sample size restricts the statistical power to detect changes in interaction skills beyond large effects. This limitation is reflected in the wide confidence intervals, adding uncertainty to the findings. Future research with larger samples is essential to assess the replicability of these findings, particularly in determining whether empathy consistently emerges as the first area of change in similar training contexts, as well as whether similar findings emerge in person-based analyses. Furthermore, a larger sample size may reveal changes in skills beyond empathy.

As for the strengths, this is one of the few non-UK-based studies focusing on training and learning MI in child and family social work. Moreover, the study carefully adapted a research-based MI training package into Finnish context and complemented it with elements informed by behaviour change science. As for methodological strengths, the study applied SWIM, an established rating tool permitting objective and comparable assessment of communication skills.

Future research may similarly benefit from analysing participant subgroups in terms of their response to the intervention and thus create a better understanding what underlies poor performance in decliners or non-improvers. This may ultimately lead to more effective interventions.

Finally, while it was not possible to collect sufficient data to assess the effectiveness of the training comprehensively, a total of 125 real service user meetings were recorded during the data collection periods. As such, this type of observational data is unique in the Finnish context. Social work is a public service delivered in private places, with power asymmetries and conversations about sometimes difficult issues. Recording of direct practice was time-consuming and sensitive requiring significant effort from the research group as well as the participating social workers and their clients. However, we believe it was worthwhile, as direct access to these very private conversations is essential for understanding and improving the services provided to some of the most vulnerable families.

Conclusions

While it is not possible to draw strong conclusions about the training's effectiveness due to the small sample size, the finding of increased

empathy suggests that it may be the first skill to change in novice MI practitioners. Our findings also demonstrate that communication skills vary between practitioners, and hence, it is necessary to pay more attention to communication skills training to ensure the uniform quality of social work practice. Integrating insights from behaviour change science may help social work trainers enhance training outcomes. In particular, we found strong relationships between worker motivation to change and improvements in practice. It seems likely that a focus on enhancing motivation and supporting behaviour change is at least as important if not more important than teaching skills.

Acknowledgements

Supplementary data

Supplementary data is available at *British Journal of Social Work* online.

Conflicts of interest. None declared.

Funding

References

Beckman, M. et al. (2022) 'Correspondence Between Practitioners' Self-assessment and Independent Motivational Interviewing Treatment Integrity Ratings', *Frontiers in Psychology*, 13: 890579. https://doi.org/10.3389/fpsyg.2022.890579

Boyle, S., Vseteckova, J., and Higgins, M. (2019) 'Impact of Motivational Interviewing by Social Workers on Service Users: A Systematic Review', *Research on Social Work Practice*, 29: 863–75. https://doi.org/10.1177/1049731519827377

Crutzen, R., and Peters, G.-J. Y. (2017) 'Scale Quality: Alpha Is an Inadequate Estimate and Factor-Analytic Evidence Is Needed First of All', *Health Psychology Review*, 11: 242–7. https://doi.org/10/gdg43k

Finnish Social Welfare Act (15 §).

Fishbein, M. and Ajzen, I. (2010) *Predicting and Changing Behavior: The Reasoned Action Approach*. New York: Psychology Press.

Franconeri, S. L. et al. (2021) 'The Science of Visual Data Communication: What Works', *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 22: 110–61. https://doi.org/10.1177/15291006211051956

Forrester, D., Westlake, D., and Killian, M. (submitted) 'What Is the Relationship between Social Work Skills in Simulated and Real Interviews?'

Forrester, D. et al. (2019) 'What Is the Relationship between Worker Skills and Outcomes for Families in Child and Family Social Work?', *The British Journal of Social Work*, 49: 2148–67. https://doi.org/10.1093/bjsw/bcy126

Forrester, D. et al. (2018) 'A Randomized Controlled Trial of Training in Motivational Interviewing for Child Protection', *Children and Youth Services Review*, 88: 180–90. https://doi.org/10.1016/j.childyouth.2018.02.014

Forrester, D., Wilkins, D. and Whittaker, C. (2021) *Motivational Interviewing for Working with Children and Families. A Practical Guide for Early Intervention and Child Protection*. London and Philadelphia: Jessica Kingsley Publishers.

Forrester, D., Whittaker, C., and Wilkins, D. (2024) *Social Work and Interviewing Motivationally (SWIM) Coding Handbook, Version 2*. Cardiff: CASCADE Centre, Cardiff University. https://cascadewales.org/wp-content/uploads/sites/3/2021/06/Social-Work-IM-Skills-Coding-Handbook.pdf, accessed 15 March 2024.

Healy, K. (2018) *The Skilled Communicator in Social Work. The Art and Science of Communication in Practice*. London: Macmillan Education, Palgrave.

Heino, M. T. J. et al. (2019) 'Visualisation and Network Analysis of Physical Activity and Its Determinants: Demonstrating Opportunities in Analysing Baseline Associations in the Let's Move It Trial', *Health Psychology and Behavioral Medicine*, 7: 269–89. https://doi.org/10/gjfxzk

Heino, M. T. J. et al. (2021) 'Studying Behaviour Change Mechanisms under Complexity', *Behavioral Sciences*, 11: 77. https://doi.org/10.3390/bs11050077

Isokuortti, N. et al. (in preparation) 'Training to Motivational Interviewing in Child and Family Social Work: A Mixed Methods Process Evaluation'.

Jonckheere, A. R. (1954) 'A Distribution-Free k-Sample Test Against Ordered Alternatives', *Biometrika*, 41: 133–45. https://doi.org/10.2307/2333011

Koo, T., and Li, M. (2016) 'A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research', *Journal of Chiropractic Medicine*, 15: 155–63. https://doi.org/10.1016/j.jcm.2016.02.012

Koprowska, J. (2020) *Communication and Interpersonal Skills in Social Work*, 5th edn. London: Learning Matters.

Laursen, B., and Hoff, E. (2006) 'Person-Centered and Variable-Centered Approaches to Longitudinal Data', *Merrill-Palmer Quarterly*, 52: 377–89. https://www.jstor.org/stable/23096200

Lundahl, B. W. et al. (2010) 'Meta- analysis of Motivational Interviewing: Twenty Five Years of Research', *Research on Social Work Practice*, 20: 137–60. https://doi.org/10.1177/1049731509347850

McDonald, R. P. (1999) *Test Theory: A Unified Treatment*. New Jersey: Lawrence Erlbaum Associates, Inc.

McEachan, R. et al. (2016) 'Meta-Analysis of the Reasoned Action Approach (RAA) to Understanding Health Behaviors', *Annals of Behavioral Medicine: a Publication of the Society of Behavioral Medicine*, 50: 592–612. https://doi.org/10.1007/s12160-016-9798-4

Miller, W. R., and Moyers, T. B. (2006) 'Eight Stages in Learning Motivational Interviewing', *Journal of Teaching in the Addictions*, 5: 3–17. https://doi.org/10.1300/J188v05n01_02

Miller, W. R., and Rollnick, S. (2023) *Motivational Interviewing: Helping People Change and Grow*, 4th edn. New York: Guilford Publications.

Mezey, G. et al. (2015) 'Challenges to Undertaking Randomised Trials with Looked After Children in Social Care Settings', *Trials*, 16: 206. https://doi.org/10.1186/s13063-015-0708-z

Molenaar, P. C. M. (2007) 'Psychological Methodology will Change Profoundly Due to the Necessity to Focus on Intra-individual Variation', *Integrative Psychological & Behavioral Science*, 41: 35–40. https://doi.org/10/cw6xcm

Moyers, T. B., Manuel, J. K., and Ernst, D. (2014) *Motivational Interviewing Treatment Integrity Coding Manual 4.1.* Unpublished Manual.

R Core Team (2025) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reith-Hall, E., and Montgomery, P. (2023) 'Communication Skills Training for Improving the Communicative Abilities of Student Social Workers', *Campbell Systematic Reviews*, 19: e1309. https://doi.org/10.1002/cl2.1309

Revelle, W. (2025) *Psych: Procedures for Personality and Psychological Research*. Evanston, Illinois: Northwestern University (Version 2.5.3). https://CRAN.R-project.org/package=psych

Ryan, R. M., and Deci, E. L. (2017) *Self-determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness*. New York: Guilford Press.

Saqr, M., Vogelsmeier, L. V. D. E., and López-Pernas, S. (2024) 'Capturing Where the Learning Process Takes Place: A Person-specific and Person-centered Primer', *Learning and Individual Differences*, 113: 102492. https://doi.org/10.1016/j.lindif.2024.102492

Shah, A. et al. (2019) 'Partnering with Parents: Reviewing the Evidence for Motivational Interviewing in Child Welfare', *Families in Society: The Journal of Contemporary Social Services*, 100: 52–67. https://doi.org/10.1177/1044389418803455

Smedslund, G. et al. (2011) 'Motivational Interviewing for Substance Abuse', *The Cochrane Database of Systematic Reviews*, 2011: CD008063. https://doi.org/10.1002/14651858.CD008063.pub2

Speelman, C. P. et al. (2024) 'Most Psychological Researchers Assume Their Samples Are Ergodic: Evidence from a Year of Articles in Three Major Journals', *Collabra: Psychology*, 10: 92888. https://doi.org/10.1525/collabra.92888

Terpstra, T. J. (1952) 'The Asymptotic Normality and Consistency of Kendall's Test Against Trend, When Ties Are Present in One Ranking', *Indagationes Mathematicae (Proceedings)*, 55: 327–33.

Trevithick, P. (2012) *Social Work Skills and Knowledge. A Practice Handbook*. Maidenhead: Open University Press.

Tukey, J. W. (1977) *Exploratory Data Analysis* (Vol. 2). Massachusetts: Addison-Wesley.

Whittaker, C. et al. (2016) 'Can We Reliably Measure Social Work Communication Skills? Development of a Scale to Measure Child and Family Social Work Direct Practice', *International Journal of Child and Family Welfare*, 17: 47–63.