Fidelity in complex behaviour change interventions: a standardised approach to evaluate intervention integrity.

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-003555
Article Type:	Research
Date Submitted by the Author:	12-Jul-2013
Complete List of Authors:	Mars, Thomas; Blizard Institute Barts and The London School of Medicine and Dentistry, Department of Primary Care and Public Health Ellard, David; Warwick Medical School, Clinical trials Unit Carnes, Dawn; Blizard Institute, Barts and The London School of Medicine and Dentistry, Centre for Primary Care and Public Health Homer, Kate; Blizard Institute, Barts and The London School of Medicine and Dentistry, Centre for Primary Care and Public Health Underwood, Martin; Warwick Medical School, Division of Health Sciences Taylor, Stephanie; Blizard Institute, Barts and The London School of Medicine and Dentistry, Centre for Primary Care and Public Health
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Patient-centred medicine
Keywords:	Complex Interventions, Fidelity, Treatment Integrity

.

Title: Fidelity in complex behaviour change interventions: a standardised approach to evaluate intervention integrity.

Authors:

1)* Mr. Tom Mars, Researcher, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB. t.s.mars@qmul.ac.uk

2) Dr. David Ellard, Senior Research Fellow, Clinical Trials Unit (T0.10), Warwick Medical school, University of Warwick, Coventry. CV4 7AL. D.R.Ellard@warwick.ac.uk.

3) Dr. Dawn Carnes, Senior Researcher and COPERS Study Manager, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB. d.carnes@qmul.ac.uk.

4) Ms. Kate Homer, Senior Researcher, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB. k.homer@qmul.ac.uk.

5) Professor Martin Underwood, Professor of Primary Care Research, Head of Division, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry. CV4 7AL. M.Underwood@warwick.ac.uk.

6) Professor Stephanie J.C. Taylor, Professor in Public Health and Primary Care, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB. s.j.c.taylor@qmul.ac.uk.

*corresponding author

Keywords: Complex interventions, Fidelity, Treatment Integrity

Word count: 3325

Abstract

Objectives: The aim of this study was to a) demonstrate the development and testing of tools and procedures designed to monitor and assess the integrity of a complex intervention for chronic pain (COPERS self-management course); and b) make recommendations based on our experiences.

Design: Fidelity assessment of a two arm randomised controlled trial intervention, assessing adherence and competence of the facilitators delivering the intervention.

Setting: The intervention was delivered in the community in two centres in the UK: one inner city and one a mix of rural and urban locations.

Participants: 403 people with chronic musculoskeletal pain were enrolled in the intervention arm and 300 attended the self-management course. Thirty lay and healthcare professionals were trained and 24 delivered the courses (two per course). We ran 31 courses for up to 16 people per course all were audio recorded.

Interventions: The course was run over three and a half days; facilitators delivered a semi-structured manualised course.

Outcomes: We designed three measures to evaluate fidelity assessing adherence to the manual, competence and overall impression.

Results: We evaluated a random sample of four components from each course (n=122). The evaluation forms were reliable and had good face validity. There were high levels of adherence in the delivery, overall adherence was 2 (maximum 2 range, 1.67-2.00), facilitator competence exhibited more variability, overall competence was 1.5 (maximum 2, range, 1.25-2.00). Overall impression was 3 (maximum 4, range, 2.00-3.00).

Conclusions: Monitoring and assessing adherence and competence at the point of intervention delivery can be realised most efficiently by embedding the principles of fidelity measurement within the design stage of complex interventions and the training and assessment of those delivering the intervention. More work is necessary to ensure more robust systems of fidelity evaluation accompany the growth of complex interventions.

Trial Registration: ISRCTN No: ISRCTN24426731

Keywords: Complex interventions, Fidelity, Treatment Integrity

Article summary

Article focus

We outline the procedures for assessing the intervention integrity of a self-management intervention for chronic musculoskeletal pain.

Key messages

The robust evaluation of intervention integrity is dependent on the *a priori* formulation of criteria based on the theoretical underpinnings, aims and content of the intervention.

Adherence and competence criteria should inform the training for those delivering the intervention and incorporated into programme manuals and supporting materials.

The evaluation of intervention integrity requires assessors who are trained and have a sophisticated understanding of the intervention.

Strengths and limitations of this study

To our knowledge our work is the most systematic and rigorous evaluation of the intervention integrity of a complex behavior change intervention to date.

The lack of valid and reliable measures of adherence and competence make the assessment of their impact on outcomes difficult.

Background

Tackling the challenges posed by chronic illness requires initiatives focussed on changing individual behaviour.[1] This has resulted in the proliferation of interventions of increasing complexity**.** Complex interventions have multiple interacting components and are recognised in Medical Research Council (MRC) guidance as having varied and challenging issues in their design, evaluation and implementation. [2] This guidance recognises that intervention fidelity is under-evaluated. Intervention fidelity is defined as the use of methodological strategies to monitor and enhance the reliability (i.e. the consistency) and validity (i.e. the appropriateness) of behavioural programmes.[3]

The construct of 'intervention fidelity' originated from concerns about the 'treatment integrity' of psychotherapeutic interventions expressed in the 1980s and 90s.[4-6] The monitoring, measurement and assessment of intervention fidelity is important as it has been demonstrated that fidelity is a mediator of study outcomes.[7-10] The analysis of intervention fidelity can provide explanations of research findings [5,11] for example where interventions lack impact, this may reflect implementation failure rather than genuine ineffectiveness.[2]

In the last twenty years, the notion of intervention fidelity has become increasingly differentiated and multi-layered.[12-14] There is an emerging science of intervention fidelity, relating to complex interventions presenting researchers with a number of conceptual, methodological and operational challenges.[15-19] There is an on-going debate about how core elements of fidelity are defined and measured [7,16,20] and a recognition of the need for reliable fidelity measurement instruments.[16,21]

There is little consensus about the key elements that contribute to intervention fidelity, possibly because it is a multidimensional construct.[12] Recent work has identified 5 domains of fidelity: study design, training, intervention delivery, intervention receipt by participants and intervention enactment, defined as the extent to which participants apply the skills learned.[13, 22]

In this paper we focus on the domain of intervention delivery or integrity, defined as the monitoring and assessment of behaviours at the point of intervention delivery. Intervention integrity is often considered to be

the heart of fidelity.[18] The effectiveness of complex interventions may be dependent on the 'skills' of those delivering them.[19] 'Skills' can be characterised by separate but related constructs of adherence and competence. Adherence is defined as: the extent to which a person delivers the essential content, delivery strategies and theories prescribed by the intervention designers and avoids activities proscribed by them. Competence refers to the level of 'skill' demonstrated by those delivering an intervention and may include the ability to respond appropriately to a wide variety of contextual cues. Competence is less likely to be assessed than adherence.[19] This may be a reflection of the on-going debate surrounding the definition of competence and 'skill',[6] the methodological difficulties surrounding the monitoring and measurement of competence,[23] and the significant expenditure of time and resource required collect and analyse competence data.[6]

The association between levels of adherence and levels of competence is unclear,[11, 25] and the impact of varying levels of adherence on outcomes is unresolved. Some studies have concluded that high levels of adherence may reflect a lack of flexibility and compromise outcomes,[25] however others have concluded that high levels of adherence are associated with improved outcomes.[24, 27] This suggests that the relationship between outcomes and adherence is not linear, and that flexibility and deviation from predefined protocols may result in lower levels of adherence but produce optimal results.

It has been argued that that the significant resource costs of maintaining a high level of vigilance in treatment fidelity are more than outweighed by the scientific, economic and stakeholder consequences of disseminating inadequately tested interventions or of implementing potentially effective programmes poorly.[3,13,28] Recent evidence suggests that the assessment of intervention fidelity is not being conducted widely or systematically.[1,13,18]

The aim of this study was to a) demonstrate the development and testing of tools and procedures designed to monitor and assess the integrity of a complex intervention for chronic pain (COPERS self-management course); and b) make recommendations based on our experiences.

Methods

The COPERS study

The **CO**ping with persistent **P**ain, **E**ffectiveness **R**esearch into **S**elf-management (COPERS) is a complex behaviour change intervention. It is a self-management course aimed at enabling participants living with long-term musculoskeletal pain to improve the quality of their lives. COPERS is a three day course run for groups of between 8 and16 people. Specifically trained facilitators, one a healthcare professional and one a lay facilitator with experience of living with long term pain conduct the groups. We tested the course's effectiveness in randomised controlled trial (RCT) (ISRCTN 24426731). As part of the trial we developed, tested and implemented a methodology to assess the intervention integrity of the COPERS course as it was delivered to trial participants. In this paper we describe how we assessed fidelity, the challenges we encountered measuring integrity, competence and adherence. We discuss these and provide recommendations based on our experience to help inform others undertaking fidelity assessment of complex interventions.

Data collection

All the 32 COPERS courses were audio recorded with the consent of participants and these recordings were used to assess and evaluate intervention integrity.

Developing the intervention integrity measures

After piloting but prior to delivery of the COPERS trial we identified 7 of 24 course components considered to be the most likely to effect participant behaviour change. These components focussed on participant education and theoretically driven behaviour change techniques and strategies in contrast to other components that encouraged social interaction, relaxation and postural awareness. Intervention integrity was assessed via our audio-recordings of the components listed in Table 1.

Table 1. Components evaluated

Component	Theoretical bases	Component Description
Component 2: (Day 1) Pain information	Acceptance and commitment theory (ACT).[29-31]	Participants watched a DVD aimed at educating them about chronic pain and introducing them, through facilitated discussion, to the notion of acceptance of their pain.
Component 3: (Day 1) Acceptance	Acceptance and commitment theory (ACT.[29-31]	Participants were asked to consider a scenario about an uninvited/unwanted guest as a metaphor for their pain.
Component 5: (Day 1) The pain cycle	Fear avoidance model.[32]	Groups were introduced to the pain cycle and the varied and individual emotions and behaviours that may perpetuate that cycle.
Component 9: (Day 2) Identifying problems, goal setting and action planning	Cognitive behavioural therapy (CBT) and theories of reasoned action/behaviour.[33,34]	Groups were introduced to strategies to enable them to systematically identify problems, brainstorm creative solutions, set goals and devise strategies to escape the pain cycle.
Component 10: (Day 2) Barriers to change-unhelpful thinking	CBT and Rational Emotive Therapy.[33,35]	Groups were encouraged to consider that reflexive, automatic thinking patterns my prevent individuals from achieving their goals.
Component 11: (Day 2) Barriers to change-reframing negatives to positives	Acceptance and commitment theory (ACT) [29-31] CBT and change management principles). [33]	Participants were asked to consider what they were able to do rather than what they were unable to do.
Component 12: (Day 2) Attention control and distraction	Attention control and distraction techniques.[36]	Participants were introduced to techniques that might enable them to focus their minds away from thoughts about pain.

A review of the existing literature indicated that few trials reported information on, or assessed intervention integrity. We used the monitoring and assessment tools from three trials to inform the development of our measures.[19,37,38] The learning outcomes outlined in the COPERS facilitator training course manual helped us to design a provisional set of criteria to measure:

i) “Adherence”, a component specific measure was designed to assess the delivery of key elements as described in the COPERS facilitators’ manual.

ii) “Competence”, a generic competence measure was designed to determine the extent to which the facilitators created an environment in which participants could share their experiences and learn new skills.

iii) “Overall impression”, this measure was designed to reflect the extent to which the aims and objectives of the component were achieved and how the material was received by the group.

We tested a variety of scoring systems for adherence, competence and “overall impression” including: dichotomous response categories, Likert and numeric scales, frequentist and occurrence/non-occurrence methods. We tested inter-rater and intra-rater reliability and assessment efficiency. The research team revised and amended the evaluation forms after piloting.

Adherence measurement

The adherence evaluation form consisted of items that reflected the occurrence or non-occurrence of an event. Component specific items, relating to the key elements prescribed in the COPERS facilitator’s manual, formed the basis of the assessment, We assessed ‘Yes’, element occurred / was delivered (scored two points), ‘No’, element did not occur / was not delivered (zero points) and ‘Unsure’ (one point).

The number of adherence items varied between the different course components (Table 2). To ensure that all scores from the components were standardised to a consistent scale we summed the ‘raw scores’ for each component and divided them by the total number of items for that component. For example: component two ‘Pain Information’ had six adherence items with a maximum ‘raw’ score of twelve (6x2), the total aggregate six item score for this component was divided by six. Thus a maximum (100%) score was two and a minimum score zero.

Competence measurement

The competence evaluation form was generic, it consisted of items related to: the extent to which the facilitators introduced the aims/rationale of each component, the success or failure of the facilitators to generate group discussion and individual disclosure, whether the facilitators’ consolidated and summarised participant learning at the end of each component and/or linked learning to other components in the COPERS

course. Assessment was given as 'Yes'/ demonstrated' (scored two points), 'No'/ not demonstrated (scored zero points), and 'Unsure' (scored one point). The scores were also standardised by dividing the maximum 'raw' score of 8 by the number of items (i.e. four) thus represented by a maximum of two and a minimum score of zero.

Overall impression rating

We used a generic overall general impression scale ranging from one to four, anchored at one: 'did not go well' and four: 'excellent'.

Selection of components to be evaluated

We used a random sampling grid to select four of the seven selected components on each course. Evaluators listened to each recorded component in its entirety and rated adherence, competence and overall impression using a specially designed evaluation form that enabled evaluators to provide supportive quotes and/or comments to justify their ratings.

A number of components could not be analysed due to equipment failure, facilitators omitting to turn recording equipment on, incomplete recording or poor sound quality; evaluators were instructed to substitute that component with the next available selected component from that course.

Members of the COPERS research team (DE, TM, KH) evaluated / assessed the audio recordings. To minimise bias team members evaluated courses they had not been involved in delivering.

Inter-rater/intra-rater reliability

Ten percent of assessed component recordings were tested for inter and intra-rater reliability. A third party (DC) reviewed the evaluation forms and selected a purposive 10 per cent sample of evaluations that reflected a range of scores. These were used to assess reliability of the scoring methods. A period of at least two weeks between first and second evaluations was adopted for the intra-rater reliability testing. We assessed reliability using percentage agreement for each item rated on the evaluation forms.

Results

Thirty one COPERS courses were delivered. We assessed 122 COPERS components, totaling approximately 71 hours. Due to missing recordings 2 courses were assessed on three rather than four components. A summary of the number of components sampled and evaluated is shown in Table 2.

Table 2. Number of items scored for each component evaluated

Component	Number of components evaluated	Adherence: items evaluated	Competence: items evaluated	Overall Impression: items evaluated
2: Pain information	16	6	4	1
3: Acceptance	17	3	4	1
5: Pain cycle	20	6	4	1
9: Goal setting	19	8	4	1
10: Unhelpful thoughts	18	6	4	1
11: Reframing	28	5	4	1
12: Attention control	14	6	4	1

The overall adherence, competence and impression scores are shown in Table 3. As the scores were not normally distributed the median and 25th and 75th percentiles are presented.

Data analysis

Adherence

Overall COPERS courses achieved the maximum course delivery adherence score (Median 2.00), however there were some component score variations (Table 3). The lowest levels of adherence were observed for component 10: Unhelpful Thinking (Median 1.67, percentile 1.67-2.00), and component 2: Pain Information (Median 1.75, 1.42-2.00).

Competence

Competence scores exhibited higher levels of variability than the adherence scores (Table 3). The overall course delivery competence score was a median of 1.5 (25%-75% percentile1.25-2.00). The highest level of competence was for component 5: Pain cycle (median 1.88, 1.50-2.00) and the lowest for component 12: Attention Control (median 1.13, 1.00-1.63).

Overall impression scores

The median overall impression score for all courses was 3 (maximum 4, 2.00-3.00). There was some component score variability (Table 3). Component 12: Attention Control had an overall impression score of 2.00, reflecting the low facilitator competence scores for this component. Component 11: Reframing had a similarly low overall impression score of 2 (2.00-3.25) although it was delivered with the maximum score for adherence (Median 2 ,1.60-2.00) and good levels of competence (Median, 1.63, 1.25-2.00).

Table 3. Overall adherence competence and impression scores

	Overall adherence		Overall competence		Overall impression	
Component	**Median scores**	25%-75% Percentile	**Median scores**	25%-75% Percentile	**Median scores**	25%-75% Percentile
2: Pain information	**1.75**	1.42-2.00	**1.75**	1.25-2.00	**3.00**	3.00-3.00
3: Acceptance	**2.00**	1.83-2.00	**1.50**	1.00-2.00	**3.00**	2.50-3.00
5: Pain cycle	**2.00**	2.00-2.00	**1.88**	1.50-2.00	**3.00**	3.00-4.00
9: Goal setting	**2.00**	2.00-2.00	**1.50**	1.00-2.00	**3.00**	2.00-3.00
10: Unhelpful thinking	**1.67**	1.67-2.00	**1.50**	1.00-1.81	**3.00**	2.00-3.00
11: Reframing	**2.00**	1.60-2.00	**1.63**	1.25-2.00	**2.00**	2.00-3.25
12: Attention control	**2.00**	1.67-2.00	**1.13**	1.00-1.63	**2.00**	1.75-3.00
Overall course score	**2.00**	1.67-2.00	**1.50**	1.25-2.00	**3.00**	2.00-3.00

Inter-rater/intra-rater reliability

Percentage agreement scores measured inter-rater reliability. Fifteen COPERS components were used to measure inter-rater reliability, they comprised 95 adherence item scores, 71 competence item scores and 15 overall impression scores. Inter-rater agreement was 80% for adherence items, 67% for competence items and 53.5% for overall impression scores.

Intra-rater reliability was measured using assessments from 16 COPERS components comprising 94 adherence item scores, 64 competence item scores and 16 overall impression scores intra-rater reliability 91% for adherence items, 75.7% for competence items and 69% for overall impression scores.

Discussion

The aim of this study was to develop a methodology to assess the level of intervention integrity achieved during the delivery of the COPERS self-management course in a randomized controlled trial setting. To our knowledge this is the most systematic and rigorous published evaluation of the intervention integrity of a complex, theory based behavior change intervention to date. Overall the results suggest that the COPERS course was delivered competently and as intended. We describe the opportunities, challenges, achievements and limitations of this work and discuss these in the context of the emerging science of fidelity assessment with regard to intervention integrity and make recommendations based on our experience which may assist other trialists evaluating complex interventions.

Our work supports the suggestion that effective adherence in complex interventions may involve more than the delivery of prescribed ‘surface’ content but also adherence to essential but non-content related ‘core’ theoretical/structural elements.[14] For example, component10: ‘Unhelpful Thinking’ in the COPERS programme illustrates the challenges in defining adherence in complex interventions. This component was intended to help participants recognise and change patterns of automatic negative and self-limiting thoughts. The course manual outlined the informational content of this component, the structure, sequence, timing and

mode of delivery of the various elements to be used by the facilitators. To deliver this component as prescribed a high level of adherence to both content and structure of the session was required. Component 10 had a relatively low adherence score which was primarily caused by the facilitators' difficulty in maintaining the complex structure of the tasks involved in this component rather than a failure to deliver the prescribed content. High levels of adherence to protocols may be associated with a mechanistic, inflexible or unresponsive delivery style and therefore associated with low levels of competence.[6] Conversely, sometimes facilitator 'failure' to deliver the component content as prescribed i.e. low adherence was directly related to low levels of competence. Parts of the course were designed to promote group participation, but if poorly sequenced or timed they often resulted in a didactic/mechanistic delivery style that inhibited rather than encouraged group disclosure and discussion.

Seemingly low levels of adherence however may not necessarily be associated with poor intervention delivery. For example some facilitators deviated from instructions (and were by definition non-adherent) but these deviations can be re-interpreted positively as the facilitators altered the delivery in response to individual or group intervention receipt. Some of our facilitators subtly changed delivery from the prescribed content in the manual but they still achieved the component's overall aims and objectives. This may be a demonstration of high levels of facilitator competence despite being rated as non-adherent.[14] There is, as yet, little empirical work that demonstrates the conditions that may influence adaptation or reinvention or whether, and in what circumstances, these deviations from prescribed protocol may enhance outcomes or decrease effectiveness.[16]

The monitoring and assessment of competence within the COPERS study illustrated the difficulties associated with its measurement. Recent work has identified competence as a complex construct that includes: the ability to establish collaborative relationships and form alliances with participants,[39] through the use of responsive tailoring of programme content,[39] the pacing of delivery,[40] and the use of positive verbal and non-verbal behaviours.[42]

The findings from the COPERS study support those who consider that levels of competence are more sensitive to contextual factors than adherence.[6] The greater variability in our competence scores, compared to those for adherence reflect, in part, the diversity of facilitator skills required to deliver the COPERS programme and the recognised practical and methodological difficulties in measuring what may seem to be a subjective concept.[6,19,23]

Our work supports the hypothesis that competence as a multidimensional construct. Effective intervention delivery may be influenced and moderated by many factors such as: positive or negative individuals and or groups, individual intervention receipt, component content, facilitator and co-facilitator coherence or incoherence, issues related to the use of computer hardware and software, the venue, the distribution of hand-outs, use of flip charts, the co-ordination and organisation of group activities, feedback and time management. Experience also influences competence and we noted that our facilitators appeared to improve with each course they conducted. Our ratings might also reflect the inexperience of the facilitators who were delivering a new initiative.

The overall impression measure was in part designed to reflect some of the 'non-facilitator determined' factors not evaluated by the adherence and competence measures. This subjective measure assessed the extent to which the component achieved its specific aims and was consistent with the goals of the wider programme. The overall impression measure proved to be challenging to use and the data difficult to interpret. Evaluators found it relatively straightforward to assess a component as either 'Excellent' or 'Did not go well', but the intermediate scores were less reliable.

Limitations

We used audio recordings to evaluate the components but it is doubtful if sound recordings alone can capture the subtleties of facilitator competence involving non-verbal behaviours, the dynamics of both facilitators and individual and group interactions. The adherence measures were designed to assess the fundamental requirements of course delivery, however the use of a generic competence measure may not have reflected the range of skills required to deliver the various course components. The absence of standardised definitions

and the lack of valid and reliable measures of adherence and competence made assessments of the impact on outcomes difficult.[19]

Lessons learned

Our experience of assessing fidelity enabled us to gain valuable insights which may be of use to others evaluating the fidelity of complex interventions, these are summarised in Table 4.

Table 4. Insights / key messages on the application of a standardised approach to evaluate intervention integrity.

i) Evaluation of interventions is dependent on the *a priori* formulation of adherence and competence criteria based on the theoretical underpinnings, aims and content of the intervention.
ii) Adherence and competence criteria should be considered during the intervention design, inform the training for those delivering the intervention and should be incorporated into programme manuals and supporting materials.
iii) Evaluation of intervention integrity requires a sophisticated understanding of the intervention. Comprehensive fidelity assessor/evaluator training is essential.
iv) Evaluation of competence optimally requires data from multiple sources such as audio and video recordings, self-report and independent observation.
v) The comprehensive evaluation of competence requires the creation of measures that are sensitive to the complexity of the construct and take into account the intervention specific contextual variables that influence it.
vi) Levels of intervention integrity may vary over time. To ensure a valid assessment of intervention integrity it should be assessed systematically throughout the delivery phase of a trial.

Conclusions

We are confident that the COPERS intervention was delivered with high levels of adherence good levels of competence and that the programme aims were largely achieved and therefore we anticipate that our outcome data will not be influenced by poor intervention delivery. In this paper we presented a method for assessing

adherence and competence and demonstrated its use in a large pragmatic RCT but we agree with the MRC that more work is necessary to ensure that the growth of complex interventions is accompanied by more robust systems of evaluation.

Author affiliations

[1,3,4,6.] Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB.

[2.] Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry. CV4 7AL.

[6.] Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry. CV4 7AL.

Contributors MU, DC and ST conceived the original idea for the COPERS study. DE wrote the fidelity protocol with input from DC, KH, MU, TM and ST. DC, DE,KH and TM evaluated intervention integrity. TM wrote the first draft of this paper, DC, DE.KH, MU, ST contributed to each successive draft of the manuscript.

Funding This paper presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10189). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NUHR or the Department of Health. This project benefitted from facilities funded through the Birmingham Science City Translational Medicine Clinical Research and Infrastructure Trials platform, with support from Advantage West Midlands.

Competing interests None

Ethics approval Ethics approval was given by Cambridge Ethics Committee 11/EE/046.

Trial registration: ISRCTN24426731

Provenance and peer review Not commissioned; internally peer reviewed.

References

1. Michie S, Fixsen D, Grimshaw JM, et al. Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implementation Science* 2009,4:40.

2. Medical Research Council: Developing and Evaluating Complex Interventions: New Guidance. London. 2008.

3. Bellg AJ, Borrelli B, Resnick B, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology* 2004,23(5):443-451.

4. Yeaton WH, Sechrest L. Critical dimensions in the choice and maintenance of successful treatments: strength, integrity, and effectiveness. *Journal of consulting and clinical psychology* 1981,49(2):156-167.

5. Moncher FJ, Prinz R. Treatment Fidelity in Outcome Studies. *Clinical Psychology Review* 1991, 11:247-266.

6. Waltz J, Addis ME, Koerner K, Jacobson NS. Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of consulting and clinical psychology* 1993,61(4):620-630.

7. Mihalic SF. The Importance of Implementation Fidelity. Violence Prevention Initiative. Boulder, Colorado; 2002.

8. Elliot D, Mihalic S. Issues in disseminating and replicating effective prevention programs. *Prevention Science* 2004,5:47-53.

9. Noel P. The impact of therapeutic case management on participation in adolescent substance abuse treatment. *American Journal of Alcohol Abuse* 2006,13:311-327.

10. Thomas RE, Baker P, Lorenzetti D. Family-based programmes for preventing smoking by children and adolescents. Cochrane Systematic Review CD004493. 2007.

11. Perepletchikova F, Kazdin A. Treatment Integrity and Therapeutic Change: Issues and Research Recommendations. *Clinical Psycholological Practice* 2005,12:365-383.

12. Steckler A, Linnan L, eds. Process Evalaution for Public Health interventions and Research. San Francisco: Wiley (Jossey-Bass); 2002.

13. Borrelli B, Sepinwall D, Ernst D, et al. A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of consulting and clinical psychology* 2005,73(5):852-860.

14. Durlak JA, Du Pre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychology* 2008,41(3-4):327-350.

15. Hulscher ME, Laurant MG, Grol RP. Process evaluation on quality improvement interventions. *Quality Safety and Health Care* 2003,12(1):40-46.

16. Dusenbury L, Brannigan R, Falco M,et al. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Res* 2003,18(2):237-256.

17. Carroll C, Patterson M, Wood S, et al. A conceptual framework for implementation fidelity. *Implement Science* 2007,2:40.

18. Gearing RE, El-Bassel N, Ghesquiere A, et al. Major ingredients of fidelity: a review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Rev* 2011,31(1):79-88.

19. Cross WF, West JC. Examining implementer fidelity: Conceptualizing and measuring adherence and competence. *J Child Serv* 2011,6(1):18-33.

20. Mihalic SF, Argamaso S. Implementing the Life Skills training drug prevention program: factors related to implementation fidelity. *Implementation Science* 2008,3(5).

21. Parham LD, Cohn ES, E, Spitzer S, et al. Development of a fidelity measure for research on the effectiveness of the Ayres Sensory Integration intervention. *Am J Occup Ther* 2011, 65(2):133-142.

22. Borrelli B. The Assessment, Monitoring, and Enhancement of Treatment Fidelity In Public Health Clinical Trials. *J Public Health Dent* 2011,71(s1):S52-S63.

23. Stiles WB, Honos-Webb L, Surko M. Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice* 1998, 5(4):438-458.

24. Barber JP, Mercer D, Krakauer I, et al. Development of an adherence/competence rating scale for individual drug counselling. *Drug Alcohol Depend* 1996,43(3):125-132.

25. Castonguay LG, Goldfried M, Wiser SRP, et al. Predicting the effect of cognitive therapy for depression: a study of unique and common factors. *Journal of consulting and clinical psychology* 1996,64(3):497-504.

26. Huey SJ Jr, Henggeler SW, Melton GB, et al.Mechanisms of change in multisystemic therapy: reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of*

consulting and clinical psychology 2000,68(3):451-467.

27. Barber JP, Crits-Christoph P, Luborsky L. Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *Journal of consulting and clinical psychology* 1996, 64(3):619-622.

28. Henggeler SW, Rowland MD, Pickrel SG, et al. Investigating family-based alternatives to institution-based mental health services for youth: lessons learned from the pilot study of a randomized field trial. *J Clin Child Psychol* 1997,26(3):226-233.

29. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: An Experiential Approach to Behaviour Change: Guilford Press; 2004.

30. Hayes SC, Smith SX. Get Out of Your Mind and into Your Life: The New Acceptance and Commitment Therapy. New Harbinger Publications; 2005.

31. McCracken LM, Vowles KE, Eccleston C: Acceptance of chronic pain: componment analysis and a revised assessment method. *Pain* 2004,107(1-2):159-166.

32. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain* 2000,85(3):317-332.

33. Beck JS. Cognitive Therapy. New York: The Guildford Press; 1995.

34. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Pearson Education;1986.

35. Ellis A, Dryden W.The Practice of Rational Emotive Behavior therapy. 2nd ed: Springer Publishing; 2007.

36. Morley S, Shapiro DA, Biggs J.Developing a treatment manual for attention management in chronic pain. *Cognitive Behaviour Therapy* 2004,33(1):1-11.

37. Carroll KM, Nich C, Rounasville BJ. Use of observer and therapist ratings to monitor delivery of coping skills treatment for cocaine abusers: utility of therapist session checklists. *Psychotherapy Research* 1998,8:307-320.

38. Carroll KM, Nich C, Sifry RL, et al. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence* 2000,57(3):225-238.

39. Creed TA, Kendall PC. Therapist alliance-building behavior within a cognitive-behavioral treatment for anxiety in youth. *Journal of consulting and clinical psychology* 2005,73(3):498-505.

40. Davidson K, Scott J, Schmidt U, et al. Therapist competence and clinical outcome in the Prevention of Parasuicide by Manual Assisted Cognitive Behaviour Therapy trial: the POPMACT study. *Psychol Med* 2004,34(5):855-863.

41. Lochman JE, Boxmeyer C, Powell N, et al. Dissemination of the Coping Power program: importance of intensity of counselor training. *Journal of consulting and clinical psychology* 2009,77(3):397-409.

42. Crowe TP, Grenyer BFS. Is therapist alliance or whole group cohesion more influential in group psychology outcomes? *Clinical Psychology and Psychotherapeutics* 2008,15:239-246.

Fidelity in complex behaviour change interventions: a standardised approach to evaluate intervention integrity.

Journal:	*BMJ Open*
Manuscript ID:	bmjopen-2013-003555.R1
Article Type:	Research
Date Submitted by the Author:	03-Oct-2013
Complete List of Authors:	Mars, Thomas; Blizard Institute Barts and The London School of Medicine and Dentistry, Department of Primary Care and Public Health Ellard, David; Warwick Medical School, Clinical trials Unit Carnes, Dawn; Blizard Institute, Barts and The London School of Medicine and Dentistry, Centre for Primary Care and Public Health Homer, Kate; Blizard Institute, Barts and The London School of Medicine and Dentistry, Centre for Primary Care and Public Health Underwood, Martin; Warwick Medical School, Division of Health Sciences Taylor, Stephanie; Blizard Institute, Barts and The London School of Medicine and Dentistry, Centre for Primary Care and Public Health
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Patient-centred medicine
Keywords:	Complex Interventions, Fidelity, Treatment Integrity

.

Title: Fidelity in complex behaviour change interventions: a standardised approach to evaluate intervention integrity.

Authors:

1)* Mr. Tom Mars, Researcher, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB. t.s.mars@qmul.ac.uk

2) Dr. David Ellard, Senior Research Fellow, Clinical Trials Unit (T0.10), Warwick Medical school, University of Warwick, Coventry. CV4 7AL. D.R.Ellard@warwick.ac.uk.

3) Dr. Dawn Carnes, Senior Researcher and COPERS Study Manager, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB. d.carnes@qmul.ac.uk.

4) Ms. Kate Homer, Senior Researcher, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB. k.homer@qmul.ac.uk.

5) Professor Martin Underwood, Professor of Primary Care Research, Head of Division, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry. CV4 7AL. M.Underwood@warwick.ac.uk.

6) Professor Stephanie J.C. Taylor, Professor in Public Health and Primary Care, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB. s.j.c.taylor@qmul.ac.uk.

*corresponding author

Keywords: Complex interventions, Fidelity, Treatment Integrity

Word count: 3777

Abstract

Objectives: The aim of this study was to a) demonstrate the development and testing of tools and procedures designed to monitor and assess the integrity of a complex intervention for chronic pain (COPERS self-management course); and b) make recommendations based on our experiences.

Design: Fidelity assessment of a two arm randomised controlled trial intervention, assessing adherence and competence of the facilitators delivering the intervention.

Setting: The intervention was delivered in the community in two centres in the UK: one inner city and one a mix of rural and urban locations.

Participants: 403 people with chronic musculoskeletal pain were enrolled in the intervention arm and 300 attended the self-management course. Thirty lay and healthcare professionals were trained and 24 delivered the courses (two per course). We ran 31 courses for up to 16 people per course all were audio recorded.

Interventions: The course was run over three and a half days; facilitators delivered a semi-structured manualised course.

Outcomes: We designed three measures to evaluate fidelity assessing adherence to the manual, competence and overall impression.

Results: We evaluated a random sample of four components from each course (n=122). The evaluation forms were reliable and had good face validity. There were high levels of adherence in the delivery, overall adherence was two (maximum 2, IQR 1.67-2.00), facilitator competence exhibited more variability, overall competence was 1.5 (maximum 2, IQR 1.25-2.00). Overall impression was three (maximum 4, IQR 2.00-3.00).

Conclusions: Monitoring and assessing adherence and competence at the point of intervention delivery can be realised most efficiently by embedding the principles of fidelity measurement within the design stage of complex interventions and the training and assessment of those delivering the intervention. More work is necessary to ensure more robust systems of fidelity evaluation accompany the growth of complex interventions.

Trial Registration: ISRCTN No: ISRCTN24426731

Keywords: Complex interventions, Fidelity, Treatment Integrity

Article summary

Article focus

We outline the procedures for assessing the intervention integrity of a self-management intervention for chronic musculoskeletal pain.

Key messages

The robust evaluation of intervention integrity is dependent on the *a priori* formulation of criteria based on the theoretical underpinnings, aims and content of the intervention.

Adherence and competence criteria should inform the training for those delivering the intervention and incorporated into programme manuals and supporting materials.

The evaluation of intervention integrity requires assessors who are trained and have a sophisticated understanding of the intervention.

Strengths and limitations of this study

To our knowledge our work is the most systematic and rigorous evaluation of the intervention integrity of a complex behavior change intervention to date.

The lack of valid and reliable measures of adherence and competence make the assessment of their impact on outcomes difficult.

Background

Tackling the challenges posed by chronic illness requires initiatives focussed on changing individual behaviour.[1] This has resulted in the proliferation of interventions of increasing complexity**.** Complex interventions have multiple interacting components and are recognised in Medical Research Council (MRC) guidance as having varied and challenging issues in their design, evaluation and implementation. [2] This guidance recognises that intervention fidelity is under-evaluated. Intervention fidelity is defined as the use of methodological strategies to monitor and enhance the reliability (i.e. the consistency) and validity (i.e. the appropriateness) of behavioural programmes.[3]

The construct of 'intervention fidelity' originated from concerns about the 'treatment integrity' of psychotherapeutic interventions expressed in the 1980s and 90s.[4,5,7] The monitoring, measurement and assessment of intervention fidelity is important as it has been demonstrated that fidelity is a mediator of study outcomes.[8-11] The analysis of intervention fidelity can provide explanations of research findings [5,12] for example where interventions lack impact, this may reflect implementation failure rather than genuine ineffectiveness.[2] The assessment of intervention fidelity is significant in the maintenance of both internal and external validity. Internal validity may be compromised by 'Type III errors' [6] that arise from the evaluation of a program that has been inadequately implemented. External validity may be improved by rigorous fidelity assessment that facilitates treatment replication across studies and assists the evaluation and development of treatments in applied settings.

In the last twenty years, the notion of intervention fidelity has become increasingly differentiated and multi-layered.[13-15] There is an emerging science of intervention fidelity, relating to complex interventions presenting researchers with a number of conceptual, methodological and operational challenges.[16-20] There is an on-going debate about how core elements of fidelity are defined and measured [8,17,21] and a recognition of the need for reliable fidelity measurement instruments.[17,22]

There is little consensus about the key elements that contribute to intervention fidelity, possibly because it is a multidimensional construct.[13] Recent work has identified five domains of fidelity: study design, training, intervention delivery, intervention receipt by participants and intervention enactment, defined as the extent to which participants apply the skills learned.[14, 23]

In this paper we focus on the domain of intervention delivery or integrity, defined as the monitoring and assessment of behaviours at the point of intervention delivery. Intervention integrity is often considered to be the heart of fidelity.[19] The effectiveness of complex interventions may be dependent on the 'skills' of those delivering them.[20] 'Skills' can be characterised by separate but related constructs of adherence and competence. Adherence is defined as: the extent to which a person delivers the essential content, delivery strategies and theories prescribed by the intervention designers and avoids activities proscribed by them. Competence refers to the level of 'skill' demonstrated by those delivering an intervention and may include the ability to respond appropriately to a wide variety of contextual cues. Competence is less likely to be assessed than adherence.[19] This may be a reflection of the on-going debate surrounding the definition of competence and 'skill',[7] the methodological difficulties surrounding the monitoring and measurement of competence,[24] and the significant expenditure of time and resource required collect and analyse competence data.[7]

The association between levels of adherence and levels of competence is unclear,[12, 26] and the impact of varying levels of adherence on outcomes is unresolved. Some studies have concluded that high levels of adherence may reflect a lack of flexibility and compromise outcomes,[26] however others have concluded that high levels of adherence are associated with improved outcomes.[25, 28] This suggests that the relationship between outcomes and adherence is not linear, and that flexibility and deviation from predefined protocols may result in lower levels of adherence but produce optimal results.

It has been argued that that the significant resource costs of maintaining a high level of vigilance in treatment fidelity are more than outweighed by the scientific, economic and stakeholder consequences of disseminating inadequately tested interventions or of implementing potentially effective programmes poorly.[3,14,29] Recent

evidence suggests that the assessment of intervention fidelity is not being conducted widely or systematically.[1,14,19]

The aim of this study was to a) demonstrate the development and testing of tools and procedures designed to monitor and assess the intervention integrity of a complex intervention for chronic pain (COPERS self-management course); and b) make recommendations based on our experiences.

Methods

The COPERS study

The **CO**ping with persistent **P**ain, **E**ffectiveness **R**esearch into **S**elf-management (COPERS) is a complex behaviour change intervention. It is a self-management course aimed at enabling participants living with long-term musculoskeletal pain to improve the quality of their lives. COPERS is a three day course run for groups of between eight and sixteen people. Specifically trained facilitators, one a healthcare professional and one a lay facilitator with experience of living with long term pain conduct the groups. We tested the course's effectiveness in randomised controlled trial (RCT) (ISRCTN 24426731). As part of the trial we developed, tested and implemented a methodology to assess the intervention integrity of the COPERS course as it was delivered to trial participants. In this paper we describe how we assessed fidelity, the challenges we encountered measuring integrity, competence and adherence. We discuss these and provide recommendations based on our experience to help inform others undertaking fidelity assessment of complex interventions.

Data collection

All the 32 COPERS courses were audio recorded with the consent of participants and these recordings were used to assess and evaluate intervention integrity.

Developing the intervention integrity measures

After piloting, but prior to delivery of the trial, we identified seven of 24 course components that were based on key cognitive behavioural elements relating to the theoretical foundations of the COPERS intervention, and which we considered to be the most likely to effect participant behaviour change. These components focussed on participant education and theoretically driven behaviour change techniques and strategies in contrast to other components that encouraged social interaction, relaxation and postural awareness. Intervention integrity was assessed via our audio-recordings of the components listed in Table 1.

Table 1. Components evaluated

Component	Theoretical bases	Component Description
Component 2: (Day 1) Pain information	Acceptance and commitment theory (ACT).[30-32]	Participants watched a DVD aimed at educating them about chronic pain and introducing them, through facilitated discussion, to the notion of acceptance of their pain.
Component 3: (Day 1) Acceptance	Acceptance and commitment theory (ACT.[30-32]	Participants were asked to consider a scenario about an uninvited/unwanted guest as a metaphor for their pain.
Component 5: (Day 1) The pain cycle	Fear avoidance model.[33]	Groups were introduced to the pain cycle and the varied and individual emotions and behaviours that may perpetuate that cycle.
Component 9: (Day 2) Identifying problems, goal setting and action planning	Cognitive behavioural therapy (CBT) and theories of reasoned action/behaviour.[34,35]	Groups were introduced to strategies to enable them to systematically identify problems, brainstorm creative solutions, set goals and devise strategies to escape the pain cycle.
Component 10: (Day 2) Barriers to change-unhelpful thinking	CBT and Rational Emotive Therapy.[34,36]	Groups were encouraged to consider that reflexive, automatic thinking patterns my prevent individuals from achieving their goals.
Component 11: (Day 2) Barriers to change-reframing negatives to positives	Acceptance and commitment theory (ACT) [30-32] CBT and change management principles). [34]	Participants were asked to consider what they were able to do rather than what they were unable to do.
Component 12: (Day 2) Attention control and distraction	Attention control and distraction techniques.[37]	Participants were introduced to techniques that might enable them to focus their minds away from thoughts about pain.

A review of the existing literature indicated that few trials reported information on, or assessed intervention integrity. We used the monitoring and assessment tools from three trials to inform the development of our measures.[20,38,39] The learning outcomes outlined in the COPERS facilitator training course manual helped us to design a provisional set of criteria to measure:

i) 'Adherence', a component specific measure was designed to assess the delivery of key elements as described in the COPERS facilitators' manual.

ii) 'Competence', a generic competence measure was designed to determine the extent to which the facilitators created an environment in which participants could share their experiences and learn new skills.

iii) 'Overall impression', this measure was designed to reflect the extent to which the aims and objectives of the component were achieved and how the material was received by the group.

We tested a variety of scoring systems for adherence, competence and "overall impression" including: dichotomous response categories, Likert and numeric scales, frequentist and occurrence/non-occurrence methods. We tested inter-rater and intra-rater reliability and assessment efficiency. The research team revised and amended the evaluation forms after piloting.

Adherence measurement

The adherence evaluation form consisted of items that reflected the occurrence or non-occurrence of an event. Component specific items, relating to the key elements prescribed in the COPERS facilitator's manual, formed the basis of the assessment, We assessed 'Yes', element occurred / was delivered (scored two points), 'No', element did not occur / was not delivered (zero points) and 'Unsure' (one point).

The number of adherence items varied between the different course components (Table 2). To ensure that all scores from the components were standardised to a consistent scale we summed the 'raw scores' for each component and divided them by the total number of items for that component. For example: component two 'Pain Information' had six adherence items with a maximum 'raw' score of twelve (6x2), the total aggregate six

item score for this component was divided by six. Thus a maximum (100%) score was two and a minimum score zero.

Competence measurement

The competence evaluation form was generic, it consisted of items related to: the extent to which the facilitators introduced the aims/rationale of each component, the success or failure of the facilitators to generate group discussion and individual disclosure, whether the facilitators' consolidated and summarised participant learning at the end of each component and/or linked learning to other components in the COPERS course. Assessment was given as 'Yes'/ demonstrated' (scored two points), 'No'/ not demonstrated (scored zero points), and 'Unsure' (scored one point). The scores were also standardised by dividing the maximum 'raw' score of eight by the number of items (i.e. four) thus represented by a maximum of two and a minimum score of zero.

Overall impression rating

We used a generic overall general impression scale ranging from one to four, anchored at one: 'did not go well' and four: 'excellent'.

Selection of components to be evaluated

We used a random sampling grid to select four of the seven selected components on each course. Evaluators listened to each recorded component in its entirety and rated adherence, competence and overall impression using a specially designed evaluation form that enabled evaluators to provide supportive quotes and/or comments to justify their ratings.

A number of components could not be analysed due to equipment failure, facilitators omitting to turn recording equipment on, incomplete recording or poor sound quality; evaluators were instructed to substitute that component with the next available selected component from that course.

Members of the COPERS research team (DE, TM, KH) evaluated / assessed the audio recordings. To minimise bias team members evaluated courses they had not been involved in delivering.

Inter-rater/intra-rater reliability

Ten percent of assessed component recordings totaling seventy one hours intervention time were tested for inter and intra-rater reliability. A third party (DC) reviewed the evaluation forms and selected a purposive 10 per cent sample of evaluations that reflected high and low adherence and competence ratings. These were used to assess reliability of the scoring methods. A period of at least two weeks between first and second evaluations was adopted for the intra-rater reliability testing. We assessed reliability using percentage agreement for each item rated on the evaluation forms.

Results

Thirty one COPERS courses were delivered and components from every course were evaluated. We assessed 122 COPERS components. Due to missing recordings 2 courses were assessed on three rather than four components. A summary of the number of components sampled and evaluated is shown in Table 2.

Table 2. Number of items scored for each component evaluated

Component	Number of components evaluated	Adherence: items evaluated	Competence: items evaluated	Overall Impression: items evaluated
2: Pain information	16	6	4	1
3: Acceptance	17	3	4	1
5: Pain cycle	20	6	4	1
9: Goal setting	19	8	4	1
10: Unhelpful thoughts	18	6	4	1
11: Reframing	28	5	4	1
12: Attention control	14	6	4	1

The overall adherence, competence and impression scores are shown in Table 3. As the scores were not normally distributed the median and the Interquartile Range (IQR) is presented.

Data analysis

Adherence

Overall COPERS courses achieved the maximum course delivery adherence score (Median 2.00), however there were some component score variations (Table 3). The lowest levels of adherence were observed for component 10: Unhelpful Thinking (Median 1.67, IQR 1.67-2.00), and component 2: Pain Information (Median 1.75, IQR 1.42-2.00).

Competence

Competence scores exhibited higher levels of variability than the adherence scores (Table 3). The overall course delivery competence score was a median of 1.5 (IQR 1.25-2.00). The highest level of competence was for component 5: Pain cycle (median 1.88, IQR 1.50-2.00) and the lowest for component 12: Attention Control (median 1.13, IQR 1.00-1.63).

Overall impression scores

The median overall impression score for all courses was 3 (maximum 4, IQR 2.00-3.00). There was some component score variability (Table 3). Component 12: Attention Control had an overall impression score of two, reflecting the low facilitator competence scores for this component. Component 11: Reframing had a similarly low overall impression score of two (IQR 2.00-3.25) although it was delivered with the maximum score for adherence (Median 2 , IQR 1.60-2.00) and good levels of competence (Median, 1.63, IQR 1.25-2.00).

Table 3. Overall adherence competence and impression scores

Component	Overall adherence		Overall competence		Overall impression	
	Median scores	IQR	Median scores	IQR	Median scores	IQR
2: Pain information	**1.75**	1.42-2.00	**1.75**	1.25-2.00	**3.00**	3.00-3.00
3: Acceptance	**2.00**	1.83-2.00	**1.50**	1.00-2.00	**3.00**	2.50-3.00
5: Pain cycle	**2.00**	2.00-2.00	**1.88**	1.50-2.00	**3.00**	3.00-4.00
9: Goal setting	**2.00**	2.00-2.00	**1.50**	1.00-2.00	**3.00**	2.00-3.00
10: Unhelpful thinking	**1.67**	1.67-2.00	**1.50**	1.00-1.81	**3.00**	2.00-3.00
11: Reframing	**2.00**	1.60-2.00	**1.63**	1.25-2.00	**2.00**	2.00-3.25
12: Attention control	**2.00**	1.67-2.00	**1.13**	1.00-1.63	**2.00**	1.75-3.00
Overall course score	**2.00**	1.67-2.00	**1.50**	1.25-2.00	**3.00**	2.00-3.00

Inter-rater/intra-rater reliability

Percentage agreement scores measured inter-rater reliability. 15 COPERS components were used to measure inter-rater reliability, they comprised 95 adherence item scores, 71 competence item scores and 15 overall impression scores. Inter-rater agreement was 80% for adherence items, 67% for competence items and 53.5% for overall impression scores.

Intra-rater reliability was measured using assessments from 16 COPERS components comprising 94 adherence item scores, 64 competence item scores and 16 overall impression scores intra-rater reliability 91% for adherence items, 75.7% for competence items and 69% for overall impression scores.

Discussion

The aim of this study was to develop a methodology to assess the level of intervention integrity achieved during the delivery of the COPERS self-management course in a randomized controlled trial setting. To our knowledge this is the most systematic and rigorous published evaluation of the intervention integrity of a complex, theory based behavior change intervention to date. Overall the results suggest that the COPERS course was delivered competently and as intended. We describe the opportunities, challenges, achievements and limitations of this work and discuss these in the context of the emerging science of fidelity assessment with regard to intervention integrity and make recommendations based on our experience which may assist other trialists evaluating complex interventions.

Our work supports the suggestion that effective adherence in complex interventions may involve more than the delivery of prescribed 'surface' content but also adherence to essential but non-content related 'core' theoretical/structural elements.[15] For example, component10: 'Unhelpful Thinking' in the COPERS programme illustrates the challenges in defining adherence in complex interventions. This component was intended to help participants recognise and change patterns of automatic negative and self-limiting thoughts. The course manual outlined the informational content of this component, the structure, sequence, timing and mode of delivery of the various elements to be used by the facilitators. To deliver this component as prescribed a high level of adherence to both content and structure of the session was required. Component 10 had a relatively low adherence score which was primarily caused by the facilitators' difficulty in maintaining the complex structure of the tasks involved in this component rather than a failure to deliver the prescribed content. High levels of adherence to protocols may be associated with a mechanistic, inflexible or unresponsive delivery style and therefore associated with low levels of competence.[7] Conversely, sometimes facilitator 'failure' to deliver the component content as prescribed i.e. low adherence was directly related to low levels of competence. Parts of the course were designed to promote group participation, but if poorly sequenced or timed they often resulted in a didactic/mechanistic delivery style that inhibited rather than encouraged group disclosure and discussion.

Seemingly low levels of adherence however may not necessarily be associated with poor intervention delivery. For example some facilitators deviated from instructions (and were by definition non-adherent) but these deviations can be re-interpreted positively as the facilitators altered the delivery in response to individual or group intervention receipt. Some of our facilitators subtly changed delivery from the prescribed content in the manual but they still achieved the component's overall aims and objectives. This may be a demonstration of high levels of facilitator competence despite being rated as non-adherent.[15] There is, as yet, little empirical work that demonstrates the conditions that may influence adaptation or reinvention or whether, and in what circumstances, these deviations from prescribed protocol may enhance outcomes or decrease effectiveness.[17]

The monitoring and assessment of competence within the COPERS study illustrated the difficulties associated with its measurement. Recent work has identified competence as a complex construct that includes: the ability to establish collaborative relationships and form alliances with participants,[40] through the use of responsive tailoring of programme content,[40] the pacing of delivery,[41] and the use of positive verbal and non-verbal behaviours.[43]

The findings from the COPERS study support those who consider that levels of competence are more sensitive to contextual factors than adherence.[7] The greater variability in our competence scores, compared to those for adherence reflect, in part, the diversity of facilitator skills required to deliver the COPERS programme and the recognised practical and methodological difficulties in measuring what may seem to be a subjective concept.[7,20,24]

Our work supports the hypothesis that competence as a multidimensional construct. Effective intervention delivery may be influenced and moderated by many factors such as: positive or negative individuals and or groups, individual intervention receipt, component content, facilitator and co-facilitator coherence or incoherence, issues related to the use of computer hardware and software, the venue, the distribution of hand-

outs, use of flip charts, the co-ordination and organisation of group activities, feedback and time management. Experience also influences competence and we noted that our facilitators appeared to improve with each course they conducted. Our ratings might also reflect the inexperience of the facilitators who were delivering a new initiative.

The overall impression measure was in part designed to reflect some of the 'non-facilitator determined' factors not evaluated by the adherence and competence measures. This subjective measure assessed the extent to which the component achieved its specific aims and was consistent with the goals of the wider programme. The overall impression measure proved to be challenging to use and the data difficult to interpret. Evaluators found it relatively straightforward to assess a component as either 'Excellent' or 'Did not go well', but the intermediate scores were less reliable.

Limitations

Within the emerging science of fidelity assessment there is a recognition of the need for reliable measurement instruments. [17,22] The varying levels of inter and intra rater reliability found in our work reflect the conceptual and methodological difficulties of measuring interventionist behaviours at the point of program delivery. We consider that our adherence, competence and overall impression measures are developmental and that in the future the use of triangulated data from multiple sources and more differentiated, contextually sensitive measures specifically designed for complex interventions may prove to be of great value. We used audio recordings to evaluate the components but it is doubtful if sound recordings alone can capture the subtleties of facilitator competence involving non-verbal behaviours, the dynamics of both facilitators and individual and group interactions. Although the assessment of adherence and competence was carried out by evaluators not directly involved in the delivery of each assessed component the overall evaluation of the COPERS intervention was conducted by members of the study team and this may have led to bias. The adherence measures were designed to assess the fundamental requirements of course delivery, however the use of a generic competence measure may not have reflected the range of skills required to deliver the various course

components. The absence of standardised definitions and the lack of valid and reliable measures of adherence and competence made assessments of the impact on outcomes difficult.[20]

Lessons learned

Our experience of assessing fidelity enabled us to gain valuable insights which may be of use to others evaluating the fidelity of complex interventions, these are summarised in Table 4.

Table 4. Insights / key messages on the application of a standardised approach to evaluate intervention integrity.

i) Evaluation of interventions is dependent on the *a priori* formulation of adherence and competence criteria based on the theoretical underpinnings, aims and content of the intervention.
ii) Adherence and competence criteria should be considered during the intervention design, inform the training for those delivering the intervention and should be incorporated into program manuals and supporting materials.
iii) Evaluation of intervention integrity requires a sophisticated understanding of the intervention. Comprehensive and cost-effective fidelity assessor/evaluator training can be provided alongside trainee interventionists within course delivery training programs.
iv) Evaluation of competence optimally requires data from multiple sources such as audio and video recordings, self-report and independent observation.
v) The comprehensive evaluation of competence requires the creation of measures that are sensitive to the complexity of the construct and take into account the intervention specific contextual variables that influence it.
vi) Levels of intervention integrity may vary over time. To ensure a valid assessment of intervention integrity it should be assessed systematically throughout the delivery phase of a trial.

Conclusions

We are confident that the COPERS intervention was delivered with high levels of adherence good levels of competence and that the programme aims were largely achieved and therefore we anticipate that our outcome

data will not be influenced by poor intervention delivery. In this paper we presented a method for assessing adherence and competence and demonstrated its use in a large pragmatic RCT but we agree with the MRC that more work is necessary to ensure that the growth of complex interventions is accompanied by more robust systems of evaluation.

Author affiliations

[1,3,4,6.] Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB.

[2.] Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry. CV4 7AL.

[6.] Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry. CV4 7AL.

Contributors MU, DC and ST conceived the original idea for the COPERS study. DE wrote the fidelity protocol with input from DC, KH, MU, TM and ST. DC, DE,KH and TM evaluated intervention integrity. TM wrote the first draft of this paper, DC, DE.KH, MU, ST contributed to each successive draft of the manuscript.

Funding This paper presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10189). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. This project benefitted from facilities funded through the Birmingham Science City Translational Medicine Clinical Research and Infrastructure Trials platform, with support from Advantage West Midlands.

Competing interests None

Ethics approval Ethics approval was given by Cambridge Ethics Committee 11/EE/046.

Trial registration: ISRCTN24426731

Provenance and peer review Not commissioned; internally peer reviewed.

References

1. Michie S, Fixsen D, Grimshaw JM, et al. Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implementation Science* 2009,4:40.

2. Medical Research Council: Developing and Evaluating Complex Interventions: New Guidance. London. 2008.

3. Bellg AJ, Borrelli B, Resnick B, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology* 2004,23(5):443-451.

4. Yeaton WH, Sechrest L. Critical dimensions in the choice and maintenance of successful treatments: strength, integrity, and effectiveness. *Journal of consulting and clinical psychology* 1981,49(2):156-167.

5. Moncher FJ, Prinz R. Treatment Fidelity in Outcome Studies. *Clinical Psychology Review* 1991, 11:247-266.

6. Basch CE, Sliepcevich EM, Gold RS, Duncan DF, Kolbe LJ. Avoiding Type Three Errors in Health Education Program Calculations: A Case Study. *Health Education Quarterly* 1985,12:315-31.

7. Waltz J, Addis ME, Koerner K, Jacobson NS. Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of consulting and clinical psychology* 1993,61(4):620-630.

8. Mihalic SF. The Importance of Implementation Fidelity. Violence Prevention Initiative. Boulder,

Colorado; 2002.

9. Elliot D, Mihalic S. Issues in disseminating and replicating effective prevention programs. *Prevention Science* 2004,5:47-53.

10. Noel P. The impact of therapeutic case management on participation in adolescent substance abuse treatment. *American Journal of Alcohol Abuse* 2006,13:311-327.

11. Thomas RE, Baker P, Lorenzetti D. Family-based programmes for preventing smoking by children and adolescents. Cochrane Systematic Review CD004493. 2007.

12. Perepletchikova F, Kazdin A. Treatment Integrity and Therapeutic Change: Issues and Research Recommendations. *Clinical Psycholological Practice* 2005,12:365-383.

13. Steckler A, Linnan L, eds. Process Evalaution for Public Health interventions and Research. San Francisco: Wiley (Jossey-Bass); 2002.

14. Borrelli B, Sepinwall D, Ernst D, et al. A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of consulting and clinical psychology* 2005,73(5):852-860.

15. Durlak JA, Du Pre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychology* 2008,41(3-4):327-350.

16. Hulscher ME, Laurant MG, Grol RP. Process evaluation on quality improvement interventions. *Quality Safety and Health Care* 2003,12(1):40-46.

17. Dusenbury L, Brannigan R, Falco M,et al. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Res* 2003,18(2):237-256.

18. Carroll C, Patterson M, Wood S, et al. A conceptual framework for implementation fidelity. *Implement Science* 2007,2:40.

19. Gearing RE, El-Bassel N, Ghesquiere A, et al. Major ingredients of fidelity: a review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Rev* 2011,31(1):79-88.

20. Cross WF, West JC. Examining implementer fidelity: Conceptualizing and measuring adherence and competence. *J Child Serv* 2011,6(1):18-33.

21. Mihalic SF, Argamaso S. Implementing the Life Skills training drug prevention program: factors related to implementation fidelity. *Implementation Science* 2008,3(5).

22. Parham LD, Cohn ES, E, Spitzer S, et al. Development of a fidelity measure for research on the effectiveness of the Ayres Sensory Integration intervention. *Am J Occup Ther* 2011, 65(2):133-142.

23. Borrelli B. The Assessment, Monitoring, and Enhancement of Treatment Fidelity In Public Health Clinical Trials. *J Public Health Dent* 2011,71(s1):S52-S63.

24. Stiles WB, Honos-Webb L, Surko M. Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice* 1998, 5(4):438-458.

25. Barber JP, Mercer D, Krakauer I, et al. Development of an adherence/competence rating scale for individual drug counselling. *Drug Alcohol Depend* 1996,43(3):125-132.

26. Castonguay LG, Goldfried M, Wiser SRP, et al. Predicting the effect of cognitive therapy for depression: a study of unique and common factors. *Journal of consulting and clinical psychology* 1996,64(3):497-504.

27. Huey SJ Jr, Henggeler SW, Melton GB, et al.Mechanisms of change in multisystemic therapy: reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of consulting and clinical psychology* 2000,68(3):451-467.

28. Barber JP, Crits-Christoph P, Luborsky L. Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *Journal of consulting and clinical psychology* 1996, 64(3):619-622.

29. Henggeler SW, Rowland MD, Pickrel SG, et al. Investigating family-based alternatives to institution-based mental health services for youth: lessons learned from the pilot study of a randomized field trial. *J Clin Child Psychol* 1997,26(3):226-233.

30. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: An Experiential Approach to Behaviour Change: Guilford Press; 2004.

31. Hayes SC, Smith SX. Get Out of Your Mind and into Your Life: The New Acceptance and Commitment Therapy. New Harbinger Publications; 2005.

32. McCracken LM, Vowles KE, Eccleston C: Acceptance of chronic pain: componment analysis and a revised assessment method. *Pain* 2004,107(1-2):159-166.

33. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain* 2000,85(3):317-332.

34. Beck JS. Cognitive Therapy. New York: The Guildford Press; 1995.

35. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Pearson Education;1986.

36. Ellis A, Dryden W.The Practice of Rational Emotive Behavior therapy. 2nd ed: Springer Publishing; 2007.

37. Morley S, Shapiro DA, Biggs J.Developing a treatment manual for attention management in chronic pain. *Cognitive Behaviour Therapy* 2004,33(1):1-11.

38. Carroll KM, Nich C, Rounasville BJ. Use of observer and therapist ratings to monitor delivery of coping skills treatment for cocaine abusers: utility of therapist session checklists. *Psychotherapy Research* 1998,8:307-320.

39. Carroll KM, Nich C, Sifry RL, et al. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence* 2000,57(3):225-238.

40. Creed TA, Kendall PC. Therapist alliance-building behavior within a cognitive-behavioral treatment for anxiety in youth. *Journal of consulting and clinical psychology* 2005,73(3):498-505.

41. Davidson K, Scott J, Schmidt U, et al. Therapist competence and clinical outcome in the Prevention of Parasuicide by Manual Assisted Cognitive Behaviour Therapy trial: the POPMACT study. *Psychol Med* 2004,34(5):855-863.

42. Lochman JE, Boxmeyer C, Powell N, et al. Dissemination of the Coping Power program: importance of intensity of counselor training. *Journal of consulting and clinical psychology* 2009,77(3):397-409.

43. Crowe TP, Grenyer BFS. Is therapist alliance or whole group cohesion more influential in group psychology outcomes? *Clinical Psychology and Psychotherapeutics* 2008,15:239-246.

.

Title: Fidelity in complex behaviour change interventions: a standardised approach to evaluate intervention integrity.

Authors:

1)* Mr. Tom Mars, Researcher, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB. t.s.mars@qmul.ac.uk

2) Dr. David Ellard, Senior Research Fellow, Clinical Trials Unit (T0.10), Warwick Medical school, University of Warwick, Coventry. CV4 7AL. D.R.Ellard@warwick.ac.uk.

3) Dr. Dawn Carnes, Senior Researcher and COPERS Study Manager, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB. d.carnes@qmul.ac.uk.

4) Ms. Kate Homer, Senior Researcher, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB. k.homer@qmul.ac.uk.

5) Professor Martin Underwood, Professor of Primary Care Research, Head of Division, Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry. CV4 7AL. M.Underwood@warwick.ac.uk.

6) Professor Stephanie J.C. Taylor, Professor in Public Health and Primary Care, Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB. s.j.c.taylor@qmul.ac.uk.

*corresponding author

Keywords: Complex interventions, Fidelity, Treatment Integrity

Word count: 3777

Abstract

Objectives: The aim of this study was to a) demonstrate the development and testing of tools and procedures designed to monitor and assess the integrity of a complex intervention for chronic pain (COPERS self-management course); and b) make recommendations based on our experiences.

Design: Fidelity assessment of a two arm randomised controlled trial intervention, assessing adherence and competence of the facilitators delivering the intervention.

Setting: The intervention was delivered in the community in two centres in the UK: one inner city and one a mix of rural and urban locations.

Participants: 403 people with chronic musculoskeletal pain were enrolled in the intervention arm and 300 attended the self-management course. Thirty lay and healthcare professionals were trained and 24 delivered the courses (two per course). We ran 31 courses for up to 16 people per course all were audio recorded.

Interventions: The course was run over three and a half days; facilitators delivered a semi-structured manualised course.

Outcomes: We designed three measures to evaluate fidelity assessing adherence to the manual, competence and overall impression.

Results: We evaluated a random sample of four components from each course (n=122). The evaluation forms were reliable and had good face validity. There were high levels of adherence in the delivery, overall adherence was two (maximum 2, IQR 1.67-2.00), facilitator competence exhibited more variability, overall competence was 1.5 (maximum 2, IQR 1.25-2.00). Overall impression was three (maximum 4, IQR 2.00-3.00).

Conclusions: Monitoring and assessing adherence and competence at the point of intervention delivery can be realised most efficiently by embedding the principles of fidelity measurement within the design stage of complex interventions and the training and assessment of those delivering the intervention. More work is necessary to ensure more robust systems of fidelity evaluation accompany the growth of complex interventions.

Trial Registration: ISRCTN No: ISRCTN24426731

Keywords: Complex interventions, Fidelity, Treatment Integrity

Article summary

Article focus

We outline the procedures for assessing the intervention integrity of a self-management intervention for chronic musculoskeletal pain.

Key messages

The robust evaluation of intervention integrity is dependent on the *a priori* formulation of criteria based on the theoretical underpinnings, aims and content of the intervention.

Adherence and competence criteria should inform the training for those delivering the intervention and incorporated into programme manuals and supporting materials.

The evaluation of intervention integrity requires assessors who are trained and have a sophisticated understanding of the intervention.

Strengths and limitations of this study

To our knowledge our work is the most systematic and rigorous evaluation of the intervention integrity of a complex behavior change intervention to date.

The lack of valid and reliable measures of adherence and competence make the assessment of their impact on outcomes difficult.

Background

Tackling the challenges posed by chronic illness requires initiatives focussed on changing individual behaviour.[1] This has resulted in the proliferation of interventions of increasing complexity**.** Complex interventions have multiple interacting components and are recognised in Medical Research Council (MRC) guidance as having varied and challenging issues in their design, evaluation and implementation. [2] This guidance recognises that intervention fidelity is under-evaluated. Intervention fidelity is defined as the use of methodological strategies to monitor and enhance the reliability (i.e. the consistency) and validity (i.e. the appropriateness) of behavioural programmes.[3]

The construct of 'intervention fidelity' originated from concerns about the 'treatment integrity' of psychotherapeutic interventions expressed in the 1980s and 90s.[4,5,7] The monitoring, measurement and assessment of intervention fidelity is important as it has been demonstrated that fidelity is a mediator of study outcomes.[8-11] The analysis of intervention fidelity can provide explanations of research findings [5,12] for example where interventions lack impact, this may reflect implementation failure rather than genuine ineffectiveness.[2] The assessment of intervention fidelity is significant in the maintenance of both internal and external validity. Internal validity may be compromised by 'Type III errors' [6] that arise from the evaluation of a program that has been inadequately implemented. External validity may be improved by rigorous fidelity assessment that facilitates treatment replication across studies and assists the evaluation and development of treatments in applied settings.

In the last twenty years, the notion of intervention fidelity has become increasingly differentiated and multi-layered.[13-15] There is an emerging science of intervention fidelity, relating to complex interventions presenting researchers with a number of conceptual, methodological and operational challenges.[16-20] There is an on-going debate about how core elements of fidelity are defined and measured [8,17,21] and a recognition of the need for reliable fidelity measurement instruments.[17,22]

There is little consensus about the key elements that contribute to intervention fidelity, possibly because it is a multidimensional construct.[13] Recent work has identified five domains of fidelity: study design, training, intervention delivery, intervention receipt by participants and intervention enactment, defined as the extent to which participants apply the skills learned.[14, 23]

In this paper we focus on the domain of intervention delivery or integrity, defined as the monitoring and assessment of behaviours at the point of intervention delivery. Intervention integrity is often considered to be the heart of fidelity.[19] The effectiveness of complex interventions may be dependent on the 'skills' of those delivering them.[20] 'Skills' can be characterised by separate but related constructs of adherence and competence. Adherence is defined as: the extent to which a person delivers the essential content, delivery strategies and theories prescribed by the intervention designers and avoids activities proscribed by them. Competence refers to the level of 'skill' demonstrated by those delivering an intervention and may include the ability to respond appropriately to a wide variety of contextual cues. Competence is less likely to be assessed than adherence.[19] This may be a reflection of the on-going debate surrounding the definition of competence and 'skill',[7] the methodological difficulties surrounding the monitoring and measurement of competence,[24] and the significant expenditure of time and resource required collect and analyse competence data.[7]

The association between levels of adherence and levels of competence is unclear,[12, 26] and the impact of varying levels of adherence on outcomes is unresolved. Some studies have concluded that high levels of adherence may reflect a lack of flexibility and compromise outcomes,[26] however others have concluded that high levels of adherence are associated with improved outcomes.[25, 28] This suggests that the relationship between outcomes and adherence is not linear, and that flexibility and deviation from predefined protocols may result in lower levels of adherence but produce optimal results.

It has been argued that that the significant resource costs of maintaining a high level of vigilance in treatment fidelity are more than outweighed by the scientific, economic and stakeholder consequences of disseminating inadequately tested interventions or of implementing potentially effective programmes poorly.[3,14,29] Recent

evidence suggests that the assessment of intervention fidelity is not being conducted widely or systematically.[1,14,19]

The aim of this study was to a) demonstrate the development and testing of tools and procedures designed to monitor and assess the intervention integrity of a complex intervention for chronic pain (COPERS self-management course); and b) make recommendations based on our experiences.

Methods

The COPERS study

The **CO**ping with persistent **P**ain, **E**ffectiveness **R**esearch into **S**elf-management (COPERS) is a complex behaviour change intervention. It is a self-management course aimed at enabling participants living with long-term musculoskeletal pain to improve the quality of their lives. COPERS is a three day course run for groups of between eight and sixteen people. Specifically trained facilitators, one a healthcare professional and one a lay facilitator with experience of living with long term pain conduct the groups. We tested the course's effectiveness in randomised controlled trial (RCT) (ISRCTN 24426731). As part of the trial we developed, tested and implemented a methodology to assess the intervention integrity of the COPERS course as it was delivered to trial participants. In this paper we describe how we assessed fidelity, the challenges we encountered measuring integrity, competence and adherence. We discuss these and provide recommendations based on our experience to help inform others undertaking fidelity assessment of complex interventions.

Data collection

All the 32 COPERS courses were audio recorded with the consent of participants and these recordings were used to assess and evaluate intervention integrity.

Developing the intervention integrity measures

After piloting, but prior to delivery of the trial, we identified seven of 24 course components that were based on key cognitive behavioural elements relating to the theoretical foundations of the COPERS intervention, and which we considered to be the most likely to effect participant behaviour change. These components focussed on participant education and theoretically driven behaviour change techniques and strategies in contrast to other components that encouraged social interaction, relaxation and postural awareness. Intervention integrity was assessed via our audio-recordings of the components listed in Table 1.

Table 1. Components evaluated

Component	Theoretical bases	Component Description
Component 2: (Day 1) Pain information	Acceptance and commitment theory (ACT).[30-32]	Participants watched a DVD aimed at educating them about chronic pain and introducing them, through facilitated discussion, to the notion of acceptance of their pain.
Component 3: (Day 1) Acceptance	Acceptance and commitment theory (ACT.[30-32]	Participants were asked to consider a scenario about an uninvited/unwanted guest as a metaphor for their pain.
Component 5: (Day 1) The pain cycle	Fear avoidance model.[33]	Groups were introduced to the pain cycle and the varied and individual emotions and behaviours that may perpetuate that cycle.
Component 9: (Day 2) Identifying problems, goal setting and action planning	Cognitive behavioural therapy (CBT) and theories of reasoned action/behaviour.[34,35]	Groups were introduced to strategies to enable them to systematically identify problems, brainstorm creative solutions, set goals and devise strategies to escape the pain cycle.
Component 10: (Day 2) Barriers to change-unhelpful thinking	CBT and Rational Emotive Therapy.[34,36]	Groups were encouraged to consider that reflexive, automatic thinking patterns my prevent individuals from achieving their goals.
Component 11: (Day 2) Barriers to change-reframing negatives to positives	Acceptance and commitment theory (ACT) [30-32] CBT and change management principles). [34]	Participants were asked to consider what they were able to do rather than what they were unable to do.
Component 12: (Day 2) Attention control and distraction	Attention control and distraction techniques.[37]	Participants were introduced to techniques that might enable them to focus their minds away from thoughts about pain.

A review of the existing literature indicated that few trials reported information on, or assessed intervention integrity. We used the monitoring and assessment tools from three trials to inform the development of our measures.[20,38,39] The learning outcomes outlined in the COPERS facilitator training course manual helped us to design a provisional set of criteria to measure:

i) 'Adherence', a component specific measure was designed to assess the delivery of key elements as described in the COPERS facilitators' manual.

ii) 'Competence', a generic competence measure was designed to determine the extent to which the facilitators created an environment in which participants could share their experiences and learn new skills.

iii) 'Overall impression', this measure was designed to reflect the extent to which the aims and objectives of the component were achieved and how the material was received by the group.

We tested a variety of scoring systems for adherence, competence and "overall impression" including: dichotomous response categories, Likert and numeric scales, frequentist and occurrence/non-occurrence methods. We tested inter-rater and intra-rater reliability and assessment efficiency. The research team revised and amended the evaluation forms after piloting.

Adherence measurement

The adherence evaluation form consisted of items that reflected the occurrence or non-occurrence of an event. Component specific items, relating to the key elements prescribed in the COPERS facilitator's manual, formed the basis of the assessment, We assessed 'Yes', element occurred / was delivered (scored two points), 'No', element did not occur / was not delivered (zero points) and 'Unsure' (one point).

The number of adherence items varied between the different course components (Table 2). To ensure that all scores from the components were standardised to a consistent scale we summed the 'raw scores' for each component and divided them by the total number of items for that component. For example: component two 'Pain Information' had six adherence items with a maximum 'raw' score of twelve (6x2), the total aggregate six

item score for this component was divided by six. Thus a maximum (100%) score was two and a minimum score zero.

Competence measurement

The competence evaluation form was generic, it consisted of items related to: the extent to which the facilitators introduced the aims/rationale of each component, the success or failure of the facilitators to generate group discussion and individual disclosure, whether the facilitators' consolidated and summarised participant learning at the end of each component and/or linked learning to other components in the COPERS course. Assessment was given as 'Yes'/ demonstrated' (scored two points), 'No'/ not demonstrated (scored zero points), and 'Unsure' (scored one point). The scores were also standardised by dividing the maximum 'raw' score of eight by the number of items (i.e. four) thus represented by a maximum of two and a minimum score of zero.

Overall impression rating

We used a generic overall general impression scale ranging from one to four, anchored at one: 'did not go well' and four: 'excellent'.

Selection of components to be evaluated

We used a random sampling grid to select four of the seven selected components on each course. Evaluators listened to each recorded component in its entirety and rated adherence, competence and overall impression using a specially designed evaluation form that enabled evaluators to provide supportive quotes and/or comments to justify their ratings.

A number of components could not be analysed due to equipment failure, facilitators omitting to turn recording equipment on, incomplete recording or poor sound quality; evaluators were instructed to substitute that component with the next available selected component from that course.

Members of the COPERS research team (DE, TM, KH) evaluated / assessed the audio recordings. To minimise bias team members evaluated courses they had not been involved in delivering.

Inter-rater/intra-rater reliability

Ten percent of assessed component recordings totaling seventy one hours intervention time were tested for inter and intra-rater reliability. A third party (DC) reviewed the evaluation forms and selected a purposive 10 per cent sample of evaluations that reflected high and low adherence and competence ratings. These were used to assess reliability of the scoring methods. A period of at least two weeks between first and second evaluations was adopted for the intra-rater reliability testing. We assessed reliability using percentage agreement for each item rated on the evaluation forms.

Results

Thirty one COPERS courses were delivered and components from every course were evaluated. We assessed 122 COPERS components. Due to missing recordings 2 courses were assessed on three rather than four components. A summary of the number of components sampled and evaluated is shown in Table 2.

Table 2. Number of items scored for each component evaluated

Component	Number of components evaluated	Adherence: items evaluated	Competence: items evaluated	Overall Impression: items evaluated
2: Pain information	16	6	4	1
3: Acceptance	17	3	4	1
5: Pain cycle	20	6	4	1
9: Goal setting	19	8	4	1
10: Unhelpful thoughts	18	6	4	1
11: Reframing	28	5	4	1
12: Attention control	14	6	4	1

The overall adherence, competence and impression scores are shown in Table 3. As the scores were not normally distributed the median and the Interquartile Range (IQR) is presented.

Data analysis

Adherence

Overall COPERS courses achieved the maximum course delivery adherence score (Median 2.00), however there were some component score variations (Table 3). The lowest levels of adherence were observed for component 10: Unhelpful Thinking (Median 1.67, IQR 1.67-2.00), and component 2: Pain Information (Median 1.75, IQR 1.42-2.00).

Competence

Competence scores exhibited higher levels of variability than the adherence scores (Table 3). The overall course delivery competence score was a median of 1.5 (IQR 1.25-2.00). The highest level of competence was for component 5: Pain cycle (median 1.88, IQR 1.50-2.00) and the lowest for component 12: Attention Control (median 1.13, IQR 1.00-1.63).

Overall impression scores

The median overall impression score for all courses was 3 (maximum 4, IQR 2.00-3.00). There was some component score variability (Table 3). Component 12: Attention Control had an overall impression score of two, reflecting the low facilitator competence scores for this component. Component 11: Reframing had a similarly low overall impression score of two (IQR 2.00-3.25) although it was delivered with the maximum score for adherence (Median 2 , IQR 1.60-2.00) and good levels of competence (Median, 1.63, IQR 1.25-2.00).

Table 3. Overall adherence competence and impression scores

Component	Overall adherence		Overall competence		Overall impression	
	Median scores	IQR	Median scores	IQR	Median scores	IQR
2: Pain information	**1.75**	1.42-2.00	**1.75**	1.25-2.00	**3.00**	3.00-3.00
3: Acceptance	**2.00**	1.83-2.00	**1.50**	1.00-2.00	**3.00**	2.50-3.00
5: Pain cycle	**2.00**	2.00-2.00	**1.88**	1.50-2.00	**3.00**	3.00-4.00
9: Goal setting	**2.00**	2.00-2.00	**1.50**	1.00-2.00	**3.00**	2.00-3.00
10: Unhelpful thinking	**1.67**	1.67-2.00	**1.50**	1.00-1.81	**3.00**	2.00-3.00
11: Reframing	**2.00**	1.60-2.00	**1.63**	1.25-2.00	**2.00**	2.00-3.25
12: Attention control	**2.00**	1.67-2.00	**1.13**	1.00-1.63	**2.00**	1.75-3.00
Overall course score	**2.00**	1.67-2.00	**1.50**	1.25-2.00	**3.00**	2.00-3.00

Inter-rater/intra-rater reliability

Percentage agreement scores measured inter-rater reliability. 15 COPERS components were used to measure inter-rater reliability, they comprised 95 adherence item scores, 71 competence item scores and 15 overall impression scores. Inter-rater agreement was 80% for adherence items, 67% for competence items and 53.5% for overall impression scores.

Intra-rater reliability was measured using assessments from 16 COPERS components comprising 94 adherence item scores, 64 competence item scores and 16 overall impression scores intra-rater reliability 91% for adherence items, 75.7% for competence items and 69% for overall impression scores.

Discussion

The aim of this study was to develop a methodology to assess the level of intervention integrity achieved during the delivery of the COPERS self-management course in a randomized controlled trial setting. To our knowledge this is the most systematic and rigorous published evaluation of the intervention integrity of a complex, theory based behavior change intervention to date. Overall the results suggest that the COPERS course was delivered competently and as intended. We describe the opportunities, challenges, achievements and limitations of this work and discuss these in the context of the emerging science of fidelity assessment with regard to intervention integrity and make recommendations based on our experience which may assist other trialists evaluating complex interventions.

Our work supports the suggestion that effective adherence in complex interventions may involve more than the delivery of prescribed 'surface' content but also adherence to essential but non-content related 'core' theoretical/structural elements.[15] For example, component10: 'Unhelpful Thinking' in the COPERS programme illustrates the challenges in defining adherence in complex interventions. This component was intended to help participants recognise and change patterns of automatic negative and self-limiting thoughts. The course manual outlined the informational content of this component, the structure, sequence, timing and mode of delivery of the various elements to be used by the facilitators. To deliver this component as prescribed a high level of adherence to both content and structure of the session was required. Component 10 had a relatively low adherence score which was primarily caused by the facilitators' difficulty in maintaining the complex structure of the tasks involved in this component rather than a failure to deliver the prescribed content. High levels of adherence to protocols may be associated with a mechanistic, inflexible or unresponsive delivery style and therefore associated with low levels of competence.[7] Conversely, sometimes facilitator 'failure' to deliver the component content as prescribed i.e. low adherence was directly related to low levels of competence. Parts of the course were designed to promote group participation, but if poorly sequenced or timed they often resulted in a didactic/mechanistic delivery style that inhibited rather than encouraged group disclosure and discussion.

Seemingly low levels of adherence however may not necessarily be associated with poor intervention delivery. For example some facilitators deviated from instructions (and were by definition non-adherent) but these deviations can be re-interpreted positively as the facilitators altered the delivery in response to individual or group intervention receipt. Some of our facilitators subtly changed delivery from the prescribed content in the manual but they still achieved the component's overall aims and objectives. This may be a demonstration of high levels of facilitator competence despite being rated as non-adherent.[15] There is, as yet, little empirical work that demonstrates the conditions that may influence adaptation or reinvention or whether, and in what circumstances, these deviations from prescribed protocol may enhance outcomes or decrease effectiveness.[17]

The monitoring and assessment of competence within the COPERS study illustrated the difficulties associated with its measurement. Recent work has identified competence as a complex construct that includes: the ability to establish collaborative relationships and form alliances with participants,[40] through the use of responsive tailoring of programme content,[40] the pacing of delivery,[41] and the use of positive verbal and non-verbal behaviours.[43]

The findings from the COPERS study support those who consider that levels of competence are more sensitive to contextual factors than adherence.[7] The greater variability in our competence scores, compared to those for adherence reflect, in part, the diversity of facilitator skills required to deliver the COPERS programme and the recognised practical and methodological difficulties in measuring what may seem to be a subjective concept.[7,20,24]

Our work supports the hypothesis that competence as a multidimensional construct. Effective intervention delivery may be influenced and moderated by many factors such as: positive or negative individuals and or groups, individual intervention receipt, component content, facilitator and co-facilitator coherence or incoherence, issues related to the use of computer hardware and software, the venue, the distribution of hand-

outs, use of flip charts, the co-ordination and organisation of group activities, feedback and time management. Experience also influences competence and we noted that our facilitators appeared to improve with each course they conducted. Our ratings might also reflect the inexperience of the facilitators who were delivering a new initiative.

The overall impression measure was in part designed to reflect some of the 'non-facilitator determined' factors not evaluated by the adherence and competence measures. This subjective measure assessed the extent to which the component achieved its specific aims and was consistent with the goals of the wider programme. The overall impression measure proved to be challenging to use and the data difficult to interpret. Evaluators found it relatively straightforward to assess a component as either 'Excellent' or 'Did not go well', but the intermediate scores were less reliable.

Limitations

Within the emerging science of fidelity assessment there is a recognition of the need for reliable measurement instruments. [17,22] The varying levels of inter and intra rater reliability found in our work reflect the conceptual and methodological difficulties of measuring interventionist behaviours at the point of program delivery. We consider that our adherence, competence and overall impression measures are developmental and that in the future the use of triangulated data from multiple sources and more differentiated, contextually sensitive measures specifically designed for complex interventions may prove to be of great value. We used audio recordings to evaluate the components but it is doubtful if sound recordings alone can capture the subtleties of facilitator competence involving non-verbal behaviours, the dynamics of both facilitators and individual and group interactions. Although the assessment of adherence and competence was carried out by evaluators not directly involved in the delivery of each assessed component the overall evaluation of the COPERS intervention was conducted by members of the study team and this may have led to bias. The adherence measures were designed to assess the fundamental requirements of course delivery, however the use of a generic competence measure may not have reflected the range of skills required to deliver the various course

components. The absence of standardised definitions and the lack of valid and reliable measures of adherence and competence made assessments of the impact on outcomes difficult.[20]

Lessons learned

Our experience of assessing fidelity enabled us to gain valuable insights which may be of use to others evaluating the fidelity of complex interventions, these are summarised in Table 4.

Table 4. Insights / key messages on the application of a standardised approach to evaluate intervention integrity.

i) Evaluation of interventions is dependent on the *a priori* formulation of adherence and competence criteria based on the theoretical underpinnings, aims and content of the intervention.
ii) Adherence and competence criteria should be considered during the intervention design, inform the training for those delivering the intervention and should be incorporated into program manuals and supporting materials.
iii) Evaluation of intervention integrity requires a sophisticated understanding of the intervention. Comprehensive and cost-effective fidelity assessor/evaluator training can be provided alongside trainee interventionists within course delivery training programs.
iv) Evaluation of competence optimally requires data from multiple sources such as audio and video recordings, self-report and independent observation.
v) The comprehensive evaluation of competence requires the creation of measures that are sensitive to the complexity of the construct and take into account the intervention specific contextual variables that influence it.
vi) Levels of intervention integrity may vary over time. To ensure a valid assessment of intervention integrity it should be assessed systematically throughout the delivery phase of a trial.

Conclusions

We are confident that the COPERS intervention was delivered with high levels of adherence good levels of competence and that the programme aims were largely achieved and therefore we anticipate that our outcome

data will not be influenced by poor intervention delivery. In this paper we presented a method for assessing adherence and competence and demonstrated its use in a large pragmatic RCT but we agree with the MRC that more work is necessary to ensure that the growth of complex interventions is accompanied by more robust systems of evaluation.

Author affiliations

[1,3,4,6.] Centre for Primary Care and Public Health, Blizard Institute, Barts and The London School of Medicine and Dentistry, Yvonne Carter Building, 58 Turner Street, London. E1 2AB.

[2.] Clinical Trials Unit, Warwick Medical School, University of Warwick, Coventry. CV4 7AL.

[6.] Division of Health Sciences, Warwick Medical School, University of Warwick, Coventry. CV4 7AL.

Contributors MU, DC and ST conceived the original idea for the COPERS study. DE wrote the fidelity protocol with input from DC, KH, MU, TM and ST. DC, DE,KH and TM evaluated intervention integrity. TM wrote the first draft of this paper, DC, DE.KH, MU, ST contributed to each successive draft of the manuscript.

Funding This paper presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research scheme (RP-PG-0707-10189). The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. This project benefitted from facilities funded through the Birmingham Science City Translational Medicine Clinical Research and Infrastructure Trials platform, with support from Advantage West Midlands.

Competing interests None

Ethics approval Ethics approval was given by Cambridge Ethics Committee 11/EE/046.

Trial registration: ISRCTN24426731

Provenance and peer review Not commissioned; internally peer reviewed.

References

1. Michie S, Fixsen D, Grimshaw JM, et al. Specifying and reporting complex behaviour change interventions: the need for a scientific method. *Implementation Science* 2009,4:40.

2. Medical Research Council: Developing and Evaluating Complex Interventions: New Guidance. London. 2008.

3. Bellg AJ, Borrelli B, Resnick B, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology* 2004,23(5):443-451.

4. Yeaton WH, Sechrest L. Critical dimensions in the choice and maintenance of successful treatments: strength, integrity, and effectiveness. *Journal of consulting and clinical psychology* 1981,49(2):156-167.

5. Moncher FJ, Prinz R. Treatment Fidelity in Outcome Studies. *Clinical Psychology Review* 1991, 11:247-266.

6. Basch CE, Sliepcevich EM, Gold RS, Duncan DF, Kolbe LJ. Avoiding Type Three Errors in Health Education Program Calculations: A Case Study. *Health Education Quarterly* 1985,12:315-31.

7. Waltz J, Addis ME, Koerner K, Jacobson NS. Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of consulting and clinical psychology* 1993,61(4):620-630.

8. Mihalic SF. The Importance of Implementation Fidelity. Violence Prevention Initiative. Boulder,

Colorado; 2002.

9. Elliot D, Mihalic S. Issues in disseminating and replicating effective prevention programs. *Prevention Science* 2004,5:47-53.

10. Noel P. The impact of therapeutic case management on participation in adolescent substance abuse treatment. *American Journal of Alcohol Abuse* 2006,13:311-327.

11. Thomas RE, Baker P, Lorenzetti D. Family-based programmes for preventing smoking by children and adolescents. Cochrane Systematic Review CD004493. 2007.

12. Perepletchikova F, Kazdin A. Treatment Integrity and Therapeutic Change: Issues and Research Recommendations. *Clinical Psycholological Practice* 2005,12:365-383.

13. Steckler A, Linnan L, eds. Process Evalaution for Public Health interventions and Research. San Francisco: Wiley (Jossey-Bass); 2002.

14. Borrelli B, Sepinwall D, Ernst D, et al. A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of consulting and clinical psychology* 2005,73(5):852-860.

15. Durlak JA, Du Pre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *Am J Community Psychology* 2008,41(3-4):327-350.

16. Hulscher ME, Laurant MG, Grol RP. Process evaluation on quality improvement interventions. *Quality Safety and Health Care* 2003,12(1):40-46.

17. Dusenbury L, Brannigan R, Falco M,et al. A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health Education Res* 2003,18(2):237-256.

18. Carroll C, Patterson M, Wood S, et al. A conceptual framework for implementation fidelity. *Implement Science* 2007,2:40.

19. Gearing RE, El-Bassel N, Ghesquiere A, et al. Major ingredients of fidelity: a review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Rev* 2011,31(1):79-88.

20. Cross WF, West JC. Examining implementer fidelity: Conceptualizing and measuring adherence and competence. *J Child Serv* 2011,6(1):18-33.

21. Mihalic SF, Argamaso S. Implementing the Life Skills training drug prevention program: factors related to implementation fidelity. *Implementation Science* 2008,3(5).

22. Parham LD, Cohn ES, E, Spitzer S, et al. Development of a fidelity measure for research on the effectiveness of the Ayres Sensory Integration intervention. *Am J Occup Ther* 2011, 65(2):133-142.

23. Borrelli B. The Assessment, Monitoring, and Enhancement of Treatment Fidelity In Public Health Clinical Trials. *J Public Health Dent* 2011,71(s1):S52-S63.

24. Stiles WB, Honos-Webb L, Surko M. Responsiveness in psychotherapy. *Clinical Psychology: Science and Practice* 1998, 5(4):438-458.

25. Barber JP, Mercer D, Krakauer I, et al. Development of an adherence/competence rating scale for individual drug counselling. *Drug Alcohol Depend* 1996,43(3):125-132.

26. Castonguay LG, Goldfried M, Wiser SRP, et al. Predicting the effect of cognitive therapy for depression: a study of unique and common factors. *Journal of consulting and clinical psychology* 1996,64(3):497-504.

27. Huey SJ Jr, Henggeler SW, Melton GB, et al.Mechanisms of change in multisystemic therapy: reducing delinquent behavior through therapist adherence and improved family and peer functioning. *Journal of consulting and clinical psychology* 2000,68(3):451-467.

28. Barber JP, Crits-Christoph P, Luborsky L. Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. *Journal of consulting and clinical psychology* 1996, 64(3):619-622.

29. Henggeler SW, Rowland MD, Pickrel SG, et al. Investigating family-based alternatives to institution-based mental health services for youth: lessons learned from the pilot study of a randomized field trial. *J Clin Child Psychol* 1997,26(3):226-233.

30. Hayes SC, Strosahl KD, Wilson KG. Acceptance and Commitment Therapy: An Experiential Approach to Behaviour Change: Guilford Press; 2004.

31. Hayes SC, Smith SX. Get Out of Your Mind and into Your Life: The New Acceptance and Commitment Therapy. New Harbinger Publications; 2005.

32. McCracken LM, Vowles KE, Eccleston C: Acceptance of chronic pain: componment analysis and a revised assessment method. *Pain* 2004,107(1-2):159-166.

33. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain* 2000,85(3):317-332.

34. Beck JS. Cognitive Therapy. New York: The Guildford Press; 1995.

35. Bandura A. Social Foundations of Thought and Action: A Social Cognitive Theory. Pearson Education;1986.

36. Ellis A, Dryden W.The Practice of Rational Emotive Behavior therapy. 2nd ed: Springer Publishing; 2007.

37. Morley S, Shapiro DA, Biggs J.Developing a treatment manual for attention management in chronic pain. *Cognitive Behaviour Therapy* 2004,33(1):1-11.

38. Carroll KM, Nich C, Rounasville BJ. Use of observer and therapist ratings to monitor delivery of coping skills treatment for cocaine abusers: utility of therapist session checklists. *Psychotherapy Research* 1998,8:307-320.

39. Carroll KM, Nich C, Sifry RL, et al. A general system for evaluating therapist adherence and competence in psychotherapy research in the addictions. *Drug and Alcohol Dependence* 2000,57(3):225-238.

40. Creed TA, Kendall PC. Therapist alliance-building behavior within a cognitive-behavioral treatment for anxiety in youth. *Journal of consulting and clinical psychology* 2005,73(3):498-505.

41. Davidson K, Scott J, Schmidt U, et al. Therapist competence and clinical outcome in the Prevention of Parasuicide by Manual Assisted Cognitive Behaviour Therapy trial: the POPMACT study. *Psychol Med* 2004,34(5):855-863.

42. Lochman JE, Boxmeyer C, Powell N, et al. Dissemination of the Coping Power program: importance of intensity of counselor training. *Journal of consulting and clinical psychology* 2009,77(3):397-409.

43. Crowe TP, Grenyer BFS. Is therapist alliance or whole group cohesion more influential in group psychology outcomes? *Clinical Psychology and Psychotherapeutics* 2008,15:239-246.