More than just a measure

Exploring clients' experiences of using a standardised self-report questionnaire to evaluate counselling outcomes

Brian Rodgers

Abstract

The past few decades have seen an increasing need for counsellors to evidence their practice using client self-report outcome questionnaires. However, little research has been undertaken on the experience of clients completing such questionnaires. This qualitative study investigated participants' experiences of completing a standardised outcome measure before and after therapy. Seventeen clients attending a community counselling service in Scotland, United Kingdom, were interviewed about their experiences of completing the measure, and the usefulness of the measure for discerning change from before to after therapy. Thematic analysis was undertaken to identify helpful and unhelpful aspects. Participant experiences were categorised into two beneficial themes ("A 'ruler' to measure against" and "A 'checklist' of common problems") and three hindering or problematic themes ("Restrictive," "Debilitating," and "Misaligned"). Implications for practice are discussed, including the possibility of utilising outcome measures as a valuable adjunct to therapy.

Keywords: counselling, outcome measurement, self-report, clients' experience

Over the past few decades there has been steadily increasing pressure for counsellors and counselling services to provide evidence of the effectiveness of their work (Rowland & Goss, 2000). Indeed, Manthei (2015) has stated that there is now an urgent need for counsellors and agencies in New Zealand to "demonstrate that what they do with their clients is effective" (p. 60). While many counsellors and services have routinely sought feedback from clients, this growing pressure to formally evaluate the outcomes of practice has led to an increasing need to utilise standardised outcome measures. In particular, client- or patient-reported outcome measures (PROMs) are now seen as an important component in the endeavour to improve treatment quality (Dawson, Doll, Fitzpatrick, Jenkinson, & Carr, 2010). Typically, these measures have undergone significant psychometric testing and statistical analysis to demonstrate their reliability and validity for assessing therapeutic outcomes (see, for example, Evans et al., 2002; Lyne, Barrett, Evans, & Barkham, 2006 for analysis of the CORE-OM).

Though New Zealand does not currently have a standardised client self-report measure as part of the Ministry of Health's national mental health and addiction information collection programme (PRIMHD), there is clear evidence from overseas that this is likely to happen (Trauer, 2010). For example, in the United Kingdom, the Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001) and the Generalised Anxiety Disorder assessment (GAD-7; Spitzer, Kroenke, Williams, & Lōwe, 2006) are required to be completed as part of the National Health Service's Increasing Access to Psychological Therapies programme (IAPT; National Collaborating Centre for Mental Health, 2019) while in the United States, the Outcome Questionnaire (OQ45; Lambert et al., 2004) and its derivatives form the basis of a number of clinical management systems used by large managed-care organisations (Brown, Burlingame, Lambert, Jones, & Vaccaro, 2001). In Australia, the Behavior and Symptom Identification Scale (BASIS-R; Eisen, Normand, Belanger, Spiro, & Esch, 2004) and the Kessler 10 (Kessler et al., 2003) form part of the Mental Health National Outcomes and Casemix Collection (2003).

Despite the availability of these measures, and the consistent calls for the routine collection of client-reported outcomes data, practitioners have typically been reluctant to integrate them routinely into their practice (National Mental Health Information Development Expert Advisory Panel, 2013). Manthei (2015) identified a number of potential reasons for this resistance, including:

- the feeling that it is an invasion of the private, professional relationship between counsellor and client;
- fear of being found to be ineffective;
- the belief that the outcome of counselling is not quantifiable;
- the assumption that the process of undertaking such an evaluation is necessarily costly, too complex, and time consuming. (pp. 60–61)

As a practitioner myself, I can relate to most of these points. Indeed, the results reported in the present study arise from a research project intended to develop a qualitative, visual approach to outcomes measurement that was more aligned to the therapeutic process, less intrusive and “technical,” and inherently subjective rather than attempting to be objective (Rodgers, 2010). The incorporation of a quantitative outcome measure in the study was only undertaken to allow a comparison with standard measurement practice. It was hence quite a surprise to me when participants in the study began reporting the significance to them of completing the quantitative questionnaire. This led to a curiosity to explore in more depth the clients’ experiences of completing standardised outcome measures.

A number of studies have been undertaken that have attempted to gather feedback from a client perspective on the use of self-report measures in mental health settings, particularly in Australia as part of the development of the National Outcomes and Casemix Collection (NOCC). These have identified numerous beneficial as well as hindering or problematic aspects of outcome measurement from this client perspective. For example, in a comprehensive report commissioned by the Australian government, Stedman, Yellowlees, Mellsop, Clarke, and Drake (1997) field tested three client-completed outcome measures, including the BASIS-32 (Eisen, Dill, & Grob, 1994), to assess their applicability, acceptability, and practicality for routine measurement of outcomes in mental health services. Specifically, the study asked 183 participants about the understandability of language used, the relevance and importance of the questions asked, and whether or not the measure was useful for assessing treatment progress. Participants completed the outcome measures at three time points: at initial assessment, after 24 hours, and then after approximately three months. After each completion, participants were asked to rate the utility of measures using a four-point scale. In addition, 20 participants engaged in three focus groups to explore their experience in more detail (Stedman et al., 2000). The study identified a number of problems with the outcome measures from clients’ perspectives, including that some questionnaire items were too restrictive, that some of the language was difficult to comprehend and/or open to misinterpretation, and that some items were perceived as irrelevant. Additionally, participants reported that some questionnaires were too general and not comprehensive enough to be useful, that completing questionnaires may be distressing, and that assistance would be required to complete some questionnaires.

Similarly, Graham et al. (2001) reported on a consultation with mental health consumers about future directions for outcomes self-assessments in Victoria, Australia. Specifically, the study comprised 10 focus groups with a total of 58 participants, who were asked about the content areas (or domains) that should be covered in any self-assessment measure, the process of outcome measurement in general, and the suitability of the BASIS-32 (Eisen et al., 1994) outcome measure in particular. Though participants' views were diverse, a number of key themes were reported, including that people tended to consider the ratings obtained from standard questionnaires to be overly simplistic; that items on any one instrument were too limited to reflect everyone's individuality; and that language used can be judgemental, ambiguous, and not user-friendly.

Additionally, Guthrie, McIntosh, Callaly, Trauer, and Coombs (2008) consulted with 50 clients using a structured questionnaire to explore their lived experience of using the BASIS-32 (Eisen et al., 1994) in "real world" clinical practice. Participants were asked about the process of completing the measure, how the measure was used by the clinician, whether or not they felt completing the measure led to better care, if there were any questions that they were afraid to answer, and if they had any suggestions for how the measure could be improved or used more effectively to help clients. Participants reported that fewer than half of the clinicians (42%) had explained what the measure would be used for, and only 45% of the clinicians had discussed the clients' responses with them. Concerns were expressed about possible adverse consequences of answering questions on topics such as sexual activity, suicidal feelings, disturbing or unreal beliefs, and illegal drug use. Additionally, the BASIS-32 (Eisen et al., 1994) was seen as lacking in areas around family and social relationships, personality and day-to-day life, coping in society, and learning how to live better with illness.

Contrasting with these concerns, a number of perceived benefits of using outcome measures have been identified in the literature. For example, in the same study by Guthrie et al. (2008) above, the majority of clients felt that completing measures had helped their clinician to understand them better (76%) and that it had led to better care (66%). Similarly, Graham et al. (2001) reported that clients found that completing measures was helpful in providing feedback on their progress and areas to work on, and for maintaining a record of change over time. Additionally, they found that doing so could empower clients to state how they experienced their illness, and could facilitate a more grounded dialogue with their case manager. Stedman et al. (1997) highlighted a number of additional benefits perceived by clients, including increased self-awareness of their mental health, improving the structure of clinical sessions, and providing a less threatening means of communicating dissatisfaction with a service.

These studies have also highlighted the importance to clients of the way measures are used. For example, Graham et al. (2001) found that clients wanted to know whether or not measures would be used to improve the therapy they received, how anonymity and confidentiality would be ensured, and how any possible negative impact of completing the measures would be handled. In the study by Guthrie et al. (2008), clients reported that they wanted clinicians to take more notice of their responses and discuss questions with them, and that clinicians should explain what outcome measures indicate and the purpose of completing them.

These findings align with those of Black et al. (2009), who found that clients generally welcomed the opportunity to complete outcome questionnaires as they saw this as valuing their input into the treatment process. Similarly to Manthei (2015), Black et al. identified that the biggest barrier to the use of routine outcome measurement was not the clients, but the resistance of clinicians to integrate outcome measures into their clinical practice.

In combination, these studies provide valuable insights into clients' reflections on using self-report outcome measures. However, the studies were not typically designed to capture clients' immediate experience of completing an outcome questionnaire in a clinical setting. Further, the studies did not inquire about clients' experiences of using outcome measures both before and after therapy. The present study was able to do just this, by inquiring into clients' experiences of completing a standardised outcome measure immediately after they had completed it, as well as exploring their experience of using the measure to reflect on any changes from before to after therapy.

Method

The results presented here are derived from data collected during a larger research project which utilised a multi-method approach to explore participants' experiences of assessing the outcomes of therapy (Rodgers, 2010). The study combined both qualitative and quantitative methods modelled on a concurrent nested design (Hanson, Creswell, Plano Clark, Petska, & Creswell, 2005). Here the quantitative and qualitative data were collected at the same time, but priority was given to the qualitative data (i.e. "QUAL + quan"; Hanson et al., 2005, p. 227). The results presented in this article are drawn exclusively from the qualitative portion of the data.

Participants

The study was undertaken within a community counselling service in Scotland, UK. Over a 15-month recruitment period, 78 people consented to be contacted about the study. Of these, 43 people attended a pre-therapy interview with the researcher. Seven participants did not receive any counselling, eight withdrew from the study after completing their counselling, six could not be contacted by the researcher on completion of their counselling, and three failed to attend the arranged post-counselling interview. A further two participants had not finished their counselling by the end of the data collection phase of the research project and were withdrawn from the study. This left a total of 17 participants who completed the study (nine female, eight male) with ages between 24 and 66 (mean 43.36). Presenting problems consisted of depression (five), anger management (four), relationship and family problems (three), anxiety/stress (one), bereavement (one), dependency (one), and unspecified emotional/personal problems (two). The number of sessions attended ranged from one through to 33 with a mean of 16.25.

Instrument

The CORE-OM is a 34-item standardised outcome measure which assesses subjective wellbeing (4 items), commonly experienced problems or symptoms (12 items), life/social functioning (12 items), and risk to self and to others (6 items; CORE System Group, 1998). Respondents are asked to rate each item using a 5-point Likert scale: "Not at all," "Only Occasionally," "Sometimes," "Often," and "Most or all the time." The majority of items are phrased such that a response of "Not at all" equates to 0, while "Most or all the time" equates to 4; however, 8 items are phrased positively (e.g. "I have achieved the things I wanted to") where "Not at all" equates to 4 and "Most or all the time" equates to 0.

The measure is scored by calculating the mean of all items and multiplying by 10 to produce a "clinical score" from 0 to 40 (Barkham, Mellor-Clark, Connell, & Cahill, 2006). Clinical scores from 0 to 5 indicate "healthy," 6 to 9 as "low level," 10 to 14 as "mild level," 15 to 19 as "moderate level," 20 to 24 as "moderate to severe level," and 25 to 40 as "severe level." Using this scoring schema, a value of 10 indicates the clinical cut-off allowing measurement of clinically significant change (Jacobson & Truax, 1991).

Initial analysis demonstrated that the CORE-OM has adequate internal reliability and test-retest stability (Evans et al., 2002). However, more recent analysis revealed a poor fit when attempting to validate the domains of wellbeing, problems, and functioning (Lyne et al., 2006). This has led to the recommendation to utilise an overall clinical score rather than individual domain scores (Barkham et al., 2006).

Data collection

Pre-therapy interviews were conducted with participants prior to their first counselling session. They were given the outcome questionnaire and asked to read and complete this with respect to how they had been over the last week. Participants were encouraged to ask any questions or seek clarification on any items, and informed that they could leave any item blank if they so desired. Once completed, the researcher asked participants about their experience of using the outcome measure, and whether they found any aspect of completing the measure valuable or difficult/problematic.

After the end of their counselling, participants were contacted to arrange a post-therapy interview. They were again asked to complete the outcome questionnaire and about their experience of completing the measure. To assist a participant's reflection on change, the scores for the questionnaire were calculated by the researcher during the interview, and plotted on a graph against their pre-therapy scores and the CORE-OM clinical cut-off scores. The participant was then shown their pre-therapy questionnaire along with the results graph and asked to reflect on any changes they saw.

Analysis

Interviews were digitally recorded then transcribed and analysed with the assistance of the Atlas TI qualitative data analysis package. The analysis process followed the general outline of thematic analyses detailed by Braun and Clarke (2006). First the researcher became immersed in the data by reading the transcripts and listening to the digital recordings. Next initial codes were generated, entailing a process of both summarising the data and explicating implicit meaning (Barker, Pistrang,

& Elliott, 2002). Themes were then identified both inductively (i.e. bottom up) and deductively (i.e. top down). The inductive approach entailed sticking closely to the words of the participants, such that the themes were strongly linked to the data, similar to the processes undertaken in Grounded Theory analysis (Glaser & Strauss, 1967). In comparison, the deductive approach utilised the researcher's theoretical understanding developed through the literature review and previous knowledge to inform the construction of themes. Thus, themes and codes were labelled using a combination of the participants' own words, metaphors, and the researcher's theoretical sensitivity. Themes were reviewed using both psychological reflection and constant comparison until a coherent and integrated structural representation was achieved (Barker et al., 2002).

Ethical considerations

As the study was undertaken in a "real world" clinical setting with a selection of potentially vulnerable participants, a number of ethical considerations needed to be attended to. The project was vetted and approved by both the university's ethics committee and the counselling service's board. Further, the study was designed to conform to the British Association for Counselling and Psychotherapy's *Ethical Guidelines for Researching Counselling and Psychotherapy* (Bond, 2004). All interviews included a debriefing stage at the end to check on participants' psychological wellbeing, and to remind them of the further support that was available if they needed it. A process of addressing consent was engaged in whereby participants had multiple opportunities to withdraw from the project, as well as identifying the level of detail they were willing to have used in any research publications (e.g., full transcripts in the form of case studies; brief, anonymised excerpts only; no details to be published). Participant contact details were securely stored separately from the data, which was itself securely stored and backed up on encrypted computer storage.

Results

Participants' experiences of using the outcome measure were categorised into two beneficial themes ("A 'ruler' to measure against" and "A 'checklist' of common problems") and three hindering or problematic themes ("Restrictive," "Debilitating," and "Misaligned"). Each of these themes was constructed from a number of codes illustrating various dimensions of the theme (see Table 1). These dimensions are indicated in the description of each theme and accompanied by representative quotes from participants (participant number in round brackets). Within these quotes, the interviewer's words are enclosed in square brackets "[…]" while implicit references (e.g., to the CORE-OM) are enclosed in curly brackets "{…}".

Table 1. Themes and codes of clients' experiences of using the CORE-OM

Beneficial		*Hindering or problematic*	
Theme	**Codes**	**Theme**	**Codes**
A "ruler" to measure against	• Consistent • Confirming • Affirmation	Restrictive	• Impersonal • Inflexible • Irrelevant
		Debilitating	• Complex • Challenging • Disempowering
A "checklist" of common problems	• Normalising • Perspective • Symbolisation • Focusing	Misaligned	• Miscommunication • Deception • Misrepresentation

A "ruler" to measure against

The outcome questionnaire was experienced by participants as somewhat like a "ruler" which they could use to measure themselves against. The static nature of the questionnaire gave participants a sense of consistency and solidity. Knowing that the questions had not changed, but that their answers had, gave participants a "solid" sense that things were different. Similarly, participants saw the outcome measure as an objective confirmation of their own sense that things had changed. Here the measure gave people a chance to see definitive change from an external perspective, that it was not just their own subjective view. Participants also saw the change in their scores relative to the "average" person as significant. Here the clinical cut-off score seemed to have real meaning for them, with any movement towards or beyond this seen as an affirmation of their psychological wellbeing.

Example 1

> *It's confirmed in writing that I am on the road to a better life, or recovery. I can say that I feel better in myself…but sometimes I worry that I kid myself on and that I'm thinking "am I saying this to hopefully make myself feel better" or is it actually happening, do I feel it? So when you reflect back to how I was answering questions then to how I'm answering questions now, it is absolutely totally clear that there has been a progression and things are getting better… (032)*

Example 2

> *I remember last time being secretly really delighted that I'd come down so much… not to say there is a "normal" that we should all be based on these answers, but it was just nice to think that I wasn't in a danger zone I suppose, or in a place where my mum would worry about me, or in a high risk area I suppose. So now to see myself completely under it, and generally in every area that bit better… (019)*

A "checklist" of common problems

Participants reported using the outcome questionnaire as a checklist of their problems which they could go through and compare themselves against. Seeing their problems written down on a "standard" questionnaire helped people realise that what they were struggling with was quite normal, that other people must feel the same sort of things. Similarly, the process of going down the list of items and realising they did not score highly on some things helped to put people's problems into perspective. The items on the questionnaire also assisted people to symbolise and "give voice" to their experience, as opposed to feeling confused or unable to articulate what was going on for them. Participants also reported that the specific questions gave them an opportunity for focusing inwards to "check in" with themselves, to actually stop and take the time to consider if something was a problem or not.

Example 3

> *Some of the questions about anxiety just made me realise that it is quite a normal thing, that people do suffer from it…that some of the things that do apply to you, you realise that other people do suffer from that…Sometimes you think no one else*

> *has ever gone through this, or no one else thinks the same as me, or maybe I'm just really awful, or people can just cope better…Sometimes when you read things like that {CORE} you realise that it is [normal]. (006)*

Example 4

> *"I have thought it would be better if I were dead"—"Not at all"! I've never thought it was better if I were dead. So in some way it gave you a wee bit of a strength to work on, even if you are feeling totally "oh my God, I'm deflated, I don't have one more ounce of energy left in me, do I need to fill in this questionnaire"…actually, "no, it isn't better if I were dead"… (032)*

Restrictive

Though participants generally found completing the outcome questionnaire useful, some also found it to be too restrictive, that it was impersonal, inflexible, and did not feel relevant to them. People reported that there was "nothing new" for them in the questionnaire, that it was like filling in a "pop" survey in a detached way that felt artificial and impersonal to them. Others reported feeling that the questionnaire was too regimented and inflexible. The fixed set of questions and limited set of possible responses felt like they were being artificially forced to fit into some sort of box. The questionnaire was also experienced as being very "broad spectrum" and too general, such that it was irrelevant to the problems a person was experiencing.

Example 5

> *You're restricted, you're really restricted here {CORE} right…With that {CORE} you've not got choice, you've got to study and go that's 1,2,3,4. [Like being forced into...] Yes. Regimented. [Regimented. To fit yourself into something which...] Into some box… (008)*

Example 6

> *The questionnaire is general, it's 34 questions, and you're speaking to all different kinds of folk that have all different problems…I understand that you can't really capture someone's personality…[but] some of the questions weren't really relevant to me… (006)*

Debilitating

Some participants also reported feeling debilitated when completing the questionnaire. A few felt they had to go through quite a complex process to decide which anchor point on the scale of the questionnaire best represented their experience, for example that the difference between "sometimes" and "only occasionally" was not entirely obvious. For others, the task of completing the questionnaire was quite a challenging ordeal. The questionnaire required a person's cognitive and rational engagement when it was this very ability to engage cognitively and rationally which was impaired. Participants also reported feeling disempowered by the questionnaire. The "technical" nature of the scoring of the item totals meant that it needed to be given over to an "expert," effectively giving the power to someone else to interpret and make meaning for the person.

Example 7

> *It's again that difference between "not at all" is never, "only occasionally" it could happen once or twice for 5 minutes, "sometimes" you feel like it's happening 50% of the time perhaps, so if I said "sometimes I've felt terribly alone and isolated" that's potentially 50% of the time, whereas if I "occasionally feel terribly alone and isolated" in the course of a week that might be two occasions where for 15 minutes I suddenly think "God, what am I doing?"…which could be quite different to there {pre CORE} where it could have been an ongoing thing. It's only one point away… (019)*

Example 8

> *That's like work. Maybe it's me being lazy but it's like work for somebody whose thinking is the problem…That's taking that which is ill…it's like asking me to do a marathon with a broken leg. You know what I mean. Because that's what you're dealing with. You're dealing with the mind…and it's the mind that's ill. So that's an ordeal, or it can be for somebody that's got anxiety. (008)*

Misaligned

It was also apparent that on a number of occasions, participants responded to the questionnaire in a way which did not align with the intended design of the measure. For example, in some instances there was accidental miscommunication when participants unintentionally scored items in the opposite way to that intended by the questionnaire designer. At other times, there was a more active deception by participants in terms of covering something up. Here participants were cautious about how the questionnaire might be interpreted and so they intentionally changed their responses to questions. It was also apparent that for some participants there was a misrepresentation of the changes from before to after therapy. Here participants reported that their interpretation of questionnaire items had changed substantially over the duration of therapy, such that an actual positive change for the participant was inversely represented as a negative change by the questionnaire score. In other instances where the questionnaire scores indicated that no or minimal change had occurred, participants felt this was not an accurate representation of how different things were for them, and that it missed subtle shifts "between the lines."

Example 9

> *"I have been disturbed by unwanted thoughts and feelings"—at the time {pre therapy} I put "sometimes" but I was actually a little bit surprised that I only put sometimes. I don't think I was maybe being as honest. "I have made plans to end my life"—at the time I put "not at all", again I don't think I was being honest because, not knowing what would happen when I came here…I was still afraid of talking about something like that at the time. I kind of get the feeling that I was probably a bit too scared to write it down... (043)*

Example 10

> *["I have felt criticised by other people" is sort of middle of the road {sometimes}—does that feel like that's a problem to you?] No, criticism used to really annoy me...sometimes the people in my life are quite critical, some of them are just quite opinionated, and I've always been given "you should do this" and "you should do that"…my ex-boyfriend was very critical…and my mother is very critical because she is bi-polar and so she has always criticised me since I was wee…[It's interesting because on the one beforehand {pre CORE}, "I have felt criticised by other people" is actually less, it's "only occasionally"] I don't think I realised it as much...I didn't really realise I was being criticised. I just thought I was doing a lot of things wrong… (006)*

Discussion

The results of the study offer some interesting insights into the experience of clients using a standardised questionnaire to assess the outcomes of their therapy, some which align with practitioner concerns as highlighted by Manthei (2015), while others challenge these.

Similar to Stedman et al. (1997) and Graham et al. (2001), the present study found that some participants experienced the questionnaire as restrictive, simplistic, impersonal, and irrelevant (see examples 5 and 6 above). Likewise, the present study aligns with the findings of Stedman et al. that some questionnaire items can be difficult to comprehend and open to misinterpretation. Further, the study highlights the complexity of the task for some people. Not only do the questions themselves require interpretation, but also the possible response options. Here it can be seen that people are being asked to undertake quite a complex aggregating process of trying to "add up" experiences over the last week and then equate them to relatively vague labels on the questionnaire such as "sometimes" (see example 7 above).

The findings of the current study also highlight the degree to which some people can experience difficulty in completing outcome questionnaires when they are psychologically unwell, and that this can have a debilitating impact on them (Stedman et al., 1997). An example of this was the participant who reported that it was like being asked to run a marathon with a broken leg (see example 8 above). Here the act of having to engage cognitively and rationally in the task was experienced as an ordeal, especially when the problems they were attending therapy for were affecting their ability to engage cognitively and rationally. Likewise, the findings of the current study concur with those of Guthrie et al. (2008) that the lack of knowledge about how responses will be used, especially around items to do with more extreme psychological disturbance such as suicidality (see example 9 above), have the potential to leave clients feeling disempowered.

However, despite these hindering and problematic experiences, the majority of participants reported that they felt they benefited from the use of the outcome measure. Similar to the findings of Graham et al. (2001), people seemed to really value the opportunity to keep track of their progress over time, and that having a pre-therapy measure allowed them to see more clearly the extent of any change (see example 1 above). A number of clients in the study reported that they had forgotten how things were before their therapy began or thought that not much had changed until seeing their pre-therapy scores. Additionally, people valued the comparison to the "average" person and seemed to easily understand the concept of a clinical cut-off (see example 2 above).

A significant finding of the present study that has not been highlighted in previous research in this area is the direct therapeutic value that some people seemed to experience from completing the outcome questionnaire. Normalising, gaining perspective, facilitating accurate symbolisation, and focusing are common therapeutic interventions. It is hence intriguing to hear that some clients were able to make use of the standardised questionnaire in this way (see examples 3 and 4 above). In particular, the client's reaction to the item "I have thought it would be better if I were dead" (see example 4 above) demonstrates a powerful, positive impact of engaging with the questionnaire. Here the outcome measure can be seen acting as a therapeutic intervention in its own right.

What is clear from this study, and from other previous research, is that the endeavour of outcome measurement is neither inherently hindering nor inherently beneficial. Rather, it seems that *how* outcome measures are used is of primary significance. There are clearly opportunities for both helpful and hindering experiences to occur for clients when completing outcome questionnaires.

Implications for practice

In terms of the potentially hindering experiences for clients, a number of suggestions are presented which could assist practitioners to mitigate these effects. For example, client concerns about standardised questionnaires being overly restrictive and impersonal could be countered by introducing the option of a personalised client-generated measure such as PSYCHLOPS (Ashworth et al., 2004) or the Personal Questionnaire (Elliott et al., 2015). Such measures offer clients a flexible, individualised approach to outcome measurement that targets the specific problems they are wanting to work on (Manthei, 2015). This approach also has the potential to reduce the likelihood of clients feeling disempowered, as the construction of individualised measures is inherently more collaborative than using a standardised questionnaire (Manthei, 2015).

To counter the impact of the "cognitive burden" for some clients of completing a standardised questionnaire when they are already feeling debilitated, counsellors could offer to read through a questionnaire for the client, or provide a simpler visual outcome measure such as the Outcome Rating Scale (ORS; Miller, Duncan, Brown, Sparks, & Claud, 2003). In regard to client concerns about the consequences of responding to questions involving more extreme psychological disturbance, or risk items such as suicidality and self-harm, counsellors could invite a discussion around the reason for the items, and what the implications might be.

In terms of the potentially helpful experiences for clients, the findings of the study can be seen to encourage practitioners to routinely utilise outcome measures as an integral part of their practice rather than seeing this as an unwelcome imposition or irrelevant inconvenience. For example, counsellors could use outcome measures as a "touchstone" to help remind clients of where they have come from, and how far they have progressed. Here a questionnaire can be seen as a snapshot of someone's thinking, allowing this to be viewed and reviewed at a later stage to reveal patterns of change.

The finding from the present study of the direct therapeutic value that some participants experienced from completing the questionnaire suggests the possibility of counsellors using such measures as a therapeutic tool to facilitate focusing, symbolisation, and gaining perspective. Here there would also seem to be an opportunity to collaboratively engage with clients in using an outcome measure as part of a check-in process to facilitate a direction for a session, by seeing what "arises" for clients in response to each of the questionnaire items, or by checking in on any changes from last time to identify what needs to be worked on.

Interestingly, this point raises a question about the trend towards the use of briefer versions of standardised measures in clinical settings. While briefer measures may be quicker to complete and hence seen to be less intrusive to the therapeutic process, it could be argued that they also present fewer "items" to check in with, which potentially limits their therapeutic potential. From this perspective, ultra-brief measures such as the ORS (Miller et al., 2003) which only has four items (individual wellbeing, interpersonal wellbeing, social wellbeing, and overall wellbeing) could be seen to be of limited value. Similarly, the use of client-generated measures in isolation potentially limits the opportunity for clients to "check in" against a list of commonly experienced problems.

Implications for policy and future research

As can be seen from the above discussion, no single approach to outcome measurement is likely to mitigate all hindering aspects or maximise all beneficial aspects. This finding is of significance in relation to the movement towards the adoption of a defined set of national outcomes data (e.g. the NOCC in Australia). The present study suggests that any move towards mandating the use of measures needs to not only consider what measures are used, but potentially more importantly *how* any measures will be used by practitioners.

The results of the study also reinforce concerns about a simplistic approach to the quantification of the outcomes of counselling. Within the limited sample of collected data, some of the calculated results obtained from the questionnaire were clearly in “error” from a purely psychometric perspective, and would have returned spurious outcomes data. This reinforces the concern about using outcome questionnaires as a basis for performance evaluation of clinicians. Equally, it is clear that the clients in this study valued the opportunity to give feedback, and saw the questionnaire as an opportunity for their voice to be heard. Again, it would appear that outcome questionnaires are neither inherently “good” nor “bad” in this regard. Rather, it would seem that care needs to be taken to understand the nuances of the process of outcome data collection and how this is conducted.

To further understand these nuances, additional research is required. For example, comparative studies exploring clients’ experiences of using different formats of outcome questionnaire would seem valuable. Here comparisons could be made between short versus long versions of standardised outcome questionnaires, as well as between Likert scales (as used on the CORE-OM) and Visual Analogue Scales (as used on the ORS). These could also be compared to using patient-generated measures such as PSYCHLOPS and the Personal Questionnaire, either in combination or standalone. A key feature of this research would be to conduct studies in practice-based settings in order to understand how clients experience and make use of the various formats in real-world applications.

Additionally, research to explore various ways of collecting and offering information back to clients would be beneficial. For example, the present study used paper questionnaires with a simple hand-drawn chart of pre/post therapy scores. Alternative methods such as using a computer tablet for questionnaire completion with automatic graphing of scores could offer clients different benefits and introduce different hindrances. Additionally, using outcome measures throughout therapeutic processes rather than just at the beginning and end may well alter the perceived usefulness of the questionnaires. Further investigation of the routine use of measures on a weekly basis may reveal significantly different results from those of the present study.

Limitations

In interpreting the above results, it is important to acknowledge a number of limitations of the study.

Firstly, the selected outcome measure, the 34-item CORE-OM (CORE System Group, 1998), was chosen as a standardised questionnaire primarily for research purposes. Though the measure was originally developed as a practice-based questionnaire, numerous shorter versions of the instrument are now available which would typically be used in a clinical setting. However, as the intention was to compare the questionnaire to a newly developed visual outcome method, the full version of the CORE-OM was used in the study as it had greater research validity.

Secondly, the outcome measure was utilised during a research interview conducted by a researcher rather than in a therapy session with a counsellor. This could have significantly altered the dynamics in relation to power, trust, honesty, transparency, etc. for participants, as well as the perceived importance and time given to filling in the questionnaire compared to completing an outcome measure with a counsellor in a counselling session.

Thirdly, the study was conducted as part of a dedicated research project run in parallel to the service's usual practices. This resulted in numerous challenges with the recruitment and retention of participants, such that less than a quarter of those who were initially contacted about the research actually completed the study. Hence the results are drawn from a self-selected sample of clients who had an interest in being part of the research project.

These limitations indicate that caution needs to be taken when inferring the relevance of the findings to using outcome measures in clinical practice. Further research would seem to be warranted here to explore clients' experiences in contexts more representative of usual practice.

Care also needs to be taken when drawing inferences from results obtained in a UK context for application in an Aotearoa New Zealand context, especially from a cultural perspective. The participant sample for the study were uniformly white and Scottish, with the researcher being of similar white European descent. Hence there are likely to be shared cultural norms and assumptions embedded in the findings and, as such, care needs to be taken when implying relevance to other cultural contexts, especially for white clinicians working with indigenous clients (Trauer & Nagel, 2012). Particularly within an Aotearoa New Zealand context, there is danger of re-enacting colonising practices by imposing a Western/Pākehā approach to outcomes measurement (Kingi & Durie, 2000). Here there is potential to conduct additional studies using measures and methods more consistent with te ao Māori, such as Hua Oranga (Kingi & Durie, 2000), and the Kaupapa Outcome Rating Scale (KORS; Drury, 2007).

Conclusions

The findings of the study invite practitioners and agencies to reflect on their stance in relation to using outcome measures in their practice. Counter to the points identified by Manthei (2015) above, the process of outcome measurement would seem to have the potential to offer a valuable adjunct to the therapeutic process. Rather than being just a data-gathering exercise, the findings indicate the importance of *how* questionnaires are used. With the increasing likelihood of the routine evaluation of outcomes being mandated in clinical settings, the findings from this study invite practitioners to reflect on how they might sensitively incorporate outcome measurement into their practice, such that it becomes another tool that clients can utilise in their recovery and growth processes.

References

Ashworth, M., Shepherd, M., Christey, J., Matthews, V., Wright, K., Parmentier, H., … Godfrey, E. (2004). A client-generated psychometric instrument: The development of "PSYCHLOPS". *Counselling and Psychotherapy Research, 4*(2), 27–31. https://doi.org/10.1080/14733140412331383913

Barker, C., Pistrang, N., & Elliott, R. (2002). *Research methods in clinical psychology: An introduction for students and practitioners* (2nd ed). Chichester, England: John Wiley & Sons. https://doi.org/10.1002/0470013435

Barkham, M., Mellor-Clark, J., Connell, J., & Cahill, J. (2006). A core approach to practice-based evidence: A brief history of the origins and applications of the CORE-OM and CORE system. *Counselling and Psychotherapy*

Research, 6(1), 3–15. https://doi.org/10.1080/14733140600581218

Black, J., Lewis, T., McIntosh, P., Callaly, T., Coombs, T., Hunter, A., & Moore, L. (2009). It's not that bad: The views of consumers and carers about routine outcome measurement in mental health. *Australian Health Review*, *33*(1), 93–99. https://doi.org/10.1071/AH090093

Bond, T. (2004). *Ethical guidelines for researching counselling and psychotherapy.* Rugby, England: BACP.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*, 77–101. https://doi.org/10.1191/1478088706qp063oa

Brown, G. S., Burlingame, G. M., Lambert, M. J., Jones, E. B., & Vaccaro, J. (2001). Pushing the quality envelope: A new outcomes management system. *Psychiatric Services*, *52*(7), 925–934. https://doi.org/10.1176/appi.ps.52.7.925

CORE System Group (1998). *The CORE System (Information Management) Handbook*. Leeds, England: CORE System Group.

Dawson, J., Doll, H., Fitzpatrick, R., Jenkinson, C., & Carr, A. J. (2010). The routine use of patient reported outcome measures in healthcare settings. *British Medical Journal, 340*. https://doi.org/10.1136/bmj.c186

Drury, N. (2007). The Kaupapa Outcome Rating Scale. *New Zealand Journal of Counselling, 27*(1), 21–32.

Eisen, S. V., Dill, D. L., & Grob, M. C. (1994). Reliability and validity of a brief patient-report instrument for psychiatric outcome evaluation. *Hospital and Community Psychiatry, 45*, 242–247. https://doi.org/10.1176/ps.45.3.242

Eisen, S. V., Normand, S., Belanger, A. J., Spiro, A., & Esch, D. (2004). The revised Behavior and Symptom Identification Scale (BASIS-R): Reliability and validity. *Medical Care, 42*(12), 1230–1241. https://doi.org/10.1097/00005650-200412000-00010

Elliott, R., Wagner, J., Sales, C., Rodgers, B., Alves, P., & Café, M. J. (2015). Psychometrics of the Personal Questionnaire: A client-generated outcome measure. *Psychological Assessment*, *28*(3), 263–278. https://doi.org/10.1037/pas0000174

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, *180*(1), 51–60. https://doi.org/10.1192/bjp.180.1.51

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. New York, NY: Aldine de Gruyter.

Graham, C., Coombs, T., Buckingham, W., Eagar, K., Trauer, T., & Callaly, T. (2001). *The Victorian mental health outcomes measurement strategy: Consumer perspectives on future directions for outcome self assessment*. Victoria, Australia: Department of Human Services.

Guthrie, D., McIntosh, M., Callaly, T., Trauer, T., & Coombs, T. (2008). Consumer attitudes towards the use of routine outcome measures in a public mental health service: A consumer-driven study. *International Journal of Mental Health Nursing, 17*, 92–97. https://doi.org/10.1111/j.1447-0349.2008.00516.x

Hanson, W. E., Creswell, J. W., Plano Clark, V. L., Petska, K. S., & Creswell, J. D. (2005). Mixed methods research designs in counseling psychology. *Journal of Counseling Psychology, 52*(2), 224–235. https://doi.org/10.1037/0022-0167.52.2.224

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*(1), 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., … Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry, 60*(2), 184–189. https://doi.org/10.1001/archpsyc.60.2.184

Kingi, K. R., & Durie, M. H. (2000). *"Hua oranga": A Mäori measure of mental health outcome*. Report prepared for the Ministry of Health. Palmerston North, New Zealand: Massey University.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *General Internal Medicine, 16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., & Burlington, G. M. (2004). *Administration and scoring manual for the Outcome Questionnaire-45.* Salt Lake City, UT: OQ Measures.

Lyne, K. J., Barrett, P., Evans, C., & Barkham, M. (2006). Dimensions of variation on the CORE-OM. *British Journal of Clinical Psychology*, *45*, 185–203. https://doi.org/10.1348/014466505X39106

Manthei, R. (2015). Evaluating counselling outcome: Why is it necessary? How can it be done? *New Zealand Journal of Counselling, 35*(1), 60–85.

Mental Health National Outcomes and Casemix Collection. (2003). *Overview of clinician-rated and consumer self-report measures*, ver 1.50. Canberra, Australia: Department of Health and Ageing.

Miller, S. D., Duncan, B. L., Brown, J., Sparks, J. A., & Claud, D. A. (2003). The Outcome Rating Scale: A preliminary study of reliability, validity, and feasibility of a brief visual analogue measure. *Journal of Brief Therapy, 2*(2), 91–100.

National Collaborating Centre for Mental Health. (2019). *The improving access to psychological therapies manual.* https://www.england.nhs.uk/wp-content/uploads/2019/02/improving-access-to-psychological-therapies-manual.pdf

National Mental Health Information Development Expert Advisory Panel. (2013). *Mental health national outcomes and casemix collection: NOCC Strategic Directions 2014–2024*. Canberra, Australia.

Rodgers, B. (2010). *Life space mapping: Developing a visual method for investigating the outcomes of counselling and psychotherapy from the client's frame of reference* (Doctoral dissertation, University of Abertay, Dundee, Scotland). https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.650271

Rowland, N., & Goss, S. (2000). *Evidence-based counselling and psychological therapies.* London, England: Brunner-Routledge.

Spitzer, R. L., Kroenke, K., Williams, J. B., & Lōwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *JAMA Internal Medicine, 166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Stedman, T., Yellowlees, P., Drake, S., Chant, D., Clarke, R., & Chapple, B. (2000). The perceived utility of six selected measures of consumer outcomes proposed for routine use in Australian mental health services. *Australian and New Zealand Journal of Psychiatry, 34*, 842–849. https://doi.org/10.1080/j.1440-1614.2000.00803.x

Stedman, T., Yellowlees, P., Mellsop, G., Clarke, R., & Drake, S. (1997). *Measuring consumer outcomes in mental health.* Canberra, Australia: Department of Health and Family Services.

Trauer, T. (2010). *Outcome measurement in mental health: Theory and practice*. New York, NY: Cambridge University Press. https://doi.org/10.1017/CBO9780511760686

Trauer, T., & Nagel, T. (2012). Outcome measurement in adult Indigenous mental health consumers. *Advances in Mental Health*, *11*(1), 98–108. https://doi.org/10.5172/jamh.2012.11.1.106