

Evaluating Counselling Outcome

Why is it necessary? How can it be done?¹

Robert Manthei

Abstract

In the current competitive funding climate, there is a pressing need for counselling outcome research that demonstrates the effectiveness of current practices. However, a dearth of such research in New Zealand remains in spite of this need. Several reasons are offered for counsellors' and agencies' resistance to conducting outcome research. Various ways of gathering such data are discussed in terms of their simplicity of use, effectiveness, and manageability. Two scales in particular are recommended for use: PCOMS and PSYCHLOPS.

Keywords: counselling, outcome research, PCOMS, PSYCHLOPS

The pressing need for counselling outcome data

Over the last several years, the funding climate for counselling has hardened, especially in relation to the willingness of outside funders to continue supporting counselling agencies financially without clear evidence that such counselling makes a positive difference in the lives of clients. More recently, the Minister of Social Development, Anne Tolley (2015), confirmed this trend when she said that future contracts with MSD would be “built around positive results and evidence of what is working.” This message is not new to counsellors. In 2013, Crocket cited the need in New Zealand for counselling outcome research that demonstrated the effectiveness of current practices. While I agree with Crocket that there is a need (which I would now call “urgent”) for counsellors, especially agencies and individuals applying for external funding, to demonstrate that what they do with their clients is effective, there seems to be resistance to doing such evaluation. This resistance can take many forms:

- 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 (Granello, Granello, & Lee, 1999; Sitza & Wood, 1997);
- the fact that there are few inducements to collecting such data.

In an attempt to reduce this reluctance to evaluate outcomes, the focus of this paper is on the “Why” and the “How” of demonstrating counselling outcomes in ways that are simple, effective, and manageable.

Why research counselling effectiveness?

As the cost of counselling rises, third-party funders are requiring recipients of their money to demonstrate how it has been used and how effective it has been in improving the lives of clients. In effect, both the funding bodies (and the clients) are demanding greater accountability and expecting services that accord with up-to-date research and “evidence-based practice” (Goss & Rose, 2002; Granello et al., 1999; Sexton, Whiston, Bleuer, & Walz, 1997)—that is, the “conscientious, explicit and judicious use of current best evidence in making decisions about the care of individuals” (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996, p. 71). It is no longer good enough for counsellors and counselling agencies merely to espouse the “art of counselling” position in explaining their practices (Sexton et al., 1997), especially in the face of ethical requirements to offer demonstrated best practices and to do the client no harm (see, for example, the codes of ethics of the ACA [2014], BACP [2013], and NZAC [2014]). John Murphy, a professor of psychology and counselling, has said:

Sometimes, I'll hear people say, “I just know intuitively how my client is doing.” That scares me.... Without denying the part that art and experience play in this profession, I think it would be arrogant for someone to rely solely on their judgment—and not some sort of testing—to determine whether counseling is effective. (as cited in Morkides, 2009)

There are other reasons why demonstrating counselling effectiveness can be useful (see, for example, Goss & Rose, 2002; Granello et al., 1999; Sexton et al., 1997), such as exploring why counselling works, how it works, or discovering previously unknown relationships among the variables involved.² Other reasons for researching counselling outcomes include:

- Accountability : to provide evidence of effectiveness for policymakers, employers, funders, the public, and clients;
- Industry standards: a growing expectation that services must/should produce quality outcomes;
- Self-reflection: to share and review your work with others (e.g., in supervision, with colleagues);
- Personal benefits: as a challenge; out of professional interest or pride; for reasons of career enhancement;
- State of mind: “outcome mindedness” is much like counselling: both involve testing, curiosity, and “what if?” thinking;
- Discovering new knowledge and “best” practices.

As far back as 1982, Attkisson and Zwick recommended four key criteria when selecting procedures for assessing counselling outcomes: procedures need to be brief, low cost, easy to administer, and generate results that are easy to summarise.

Brevity seems to be a key criterion if an outcome measure is to be widely used. If it is too long and seen to be too complicated, counsellors may not use even a well-researched and validated scale. So, how brief can brief be? While it is accepted that longer measures are more valid than shorter ones, “the answer to the question regarding when a measure is too long is simple: When clinicians won’t use it” (Duncan & Reese, 2013, p. 135). In more practical terms, I would say that anything that takes longer to complete than two to four minutes is too long.

This article looks at four methods that meet these criteria, and discusses the benefits and limitations of each. All are client self-report methods, are relatively easy to administer, and involve reasonably non-complicated procedures for analysing and interpreting the data. As for relying on clients for the data (in fact, client self-report data is the most frequently used type of data in therapy research), Professor Murphy asks:

Who better to ask about the effectiveness of treatment than the client himself? The consumer should be the primary measure, if not the sole measure. (as cited in Morkides, 2009)

Client satisfaction surveys

This is perhaps the most frequently used method of assessing *quality of service*. However, the results are often, incorrectly, presented as measures of “outcome effectiveness.”

Satisfaction is defined as “all inquiries into the extent to which services gratify the client’s wants, wishes, or desires for treatment” (Lebow, 1983, p. 212). Satisfaction surveys that ask clients to evaluate their experience with the agency and their counsellor are relatively low-cost, widely used, and usually easy to administer and score. However, there are so many variations available that none has yet to be adopted as a “go to” instrument. Also, because surveys are usually administered after a client finishes counselling, there is often a large attrition rate, especially with mailed questionnaires (40–50%) (Lebow, 1983). It is simply difficult to get people to complete them. In addition, there are often no reliability and validity data on the survey instrument used. In general, however, satisfaction surveys usually produce high rates of satisfaction (80%+) with less than 10% of respondents expressing any dissatisfaction (Lebow, 1983).

This latter finding has led some researchers to discount the value of such high positive rates (Sitza & Wood, 1997) and to accept them only as an indication that nothing extremely bad occurred, not as an assurance of quality service delivery (p. 1840; see also Williams, Coyle, & Healy, 1998). Another criticism of satisfaction surveys is that clients’ beliefs and values about an experience cannot be represented by general expressions of satisfaction, and therefore, “many satisfaction surveys provide only an illusion of consumerism producing results which tend only to endorse the status quo” (Williams, 1994, p. 509).

Nevertheless, use of satisfaction surveys became widespread as the “clients as consumers” movement gained acceptance (Fontana, Ford, & Rosenheck, 2003; Sitza & Wood, 1997; Williams, 1994). They have been commonly used for a long time now across the wide spectrum of service provision organisations, including medical and mental health, a trend Lebow (1983) noted more than 30 years ago.

Satisfaction with services may not be the same as the actual quality of the services delivered, as research with library users, for example, has found (see HERNON & Altman, 2010). The same may be true for counselling and other mental health services where the clients’ expressed satisfaction with the service received is not always correlated with ratings of their symptoms (or problems), or the correlation between the two is only modest (Fontana et al., 2003). Paul West, an American counsellor educator, is quoted as saying:

You have your client satisfaction surveys...but we don't know if they are just evidence of the strength of the client-counselor relationship. A client may say you're the best thing since soft butter, but it might not be evidence of the effectiveness of counseling. (as cited in Morkides, 2009)

Some researchers, e.g. Lebow (1982), have reported that the correlation of satisfaction with *global outcome* is high, but is less so when more specific outcome indicators are used. These conflicting findings suggest that “significant but low correlations between satisfaction and success indicate that satisfaction ratings cannot replace success ratings or other outcome indicators in assessments of quality of care” (Edwards, Yarvis, Mueller, & Langsley, 1978, p. 1), and that satisfaction with treatment may be assessed as an indication of quality care separate from clinical outcome (Fontana et al., 2003).

Use of a standardised psychological or mental health instrument

Typically, these sorts of measures are positivist in their approach, use standardised questionnaires consisting of pre-set and psychometrically determined items, and report validity and reliability information. The resulting data, called nomothetic data, allow researchers to compare groups of respondents and to generalise the results to larger groups (Sales & Alves, 2012). Choosing an instrument involves, at the very least, considering the type of data desired (e.g., behavioural, emotional, cognitive, social interactional, etc.); the cost of the instrument; and the time required to administer, score, and analyse the results (Granello et al., 1999). Lambert and Vermeersch (2013) listed 13 criteria they considered to be essential to choosing the right measure, a process that would be beyond the capability and interest of most agencies and practising counsellors.

Below are some typically used measures (used mostly in research studies, I believe), although in my opinion they do not always meet the previously stated criteria of being brief, low cost, and easy to administer, with the results being easy to summarise. The reasons they are not widely used in counselling practices are not known, but probably have to do with their perceived complexity, the lack of incentive for counsellors to collect outcome data, the fear of what such data might reveal about the quality of the service(s) provided, and a lack of expertise or confidence in using such instruments. Thus, to paraphrase Duncan and Reese’s (2013, p. 135) view of this situation: “When is a measure too complicated? When clinicians won’t use it.”

The measures listed below are not offered as the “best” of their type. Rather, they are intended to give readers some examples of scales in current use (internationally) that rely on client self-report and are said to be brief and easy to administer. For those who would like further information about any one of them, a starting reference is provided. The extent to which any of these are used by New Zealand counsellors or agencies is not documented, although I believe that:

- most could be acquired for minimal/affordable cost; and
- they are recognised as “proven” evaluation tools.

In spite of these advantages, however, I believe most of them would be considered too long, even though they are touted as relatively brief (taking between two and seven minutes to complete); too complicated; too “psychological/medical model” in their orientation and presentation; and too difficult to analyse without some statistical skills.

Finally, they tend either to measure a specifically diagnosed problem (e.g., BDI-II, depression), or generic health or wellbeing. An exception is Outcome Questionnaire-45 (OQ-45; Lambert et al., 2004) which introduces the concept of using a measure that not only assesses counselling outcome but also can be used to *influence the course of counselling as it progresses*. This key point is discussed more fully in relation to PCOMS (see below), which uses the data as a tool of therapy by providing session-by-session feedback to clients, and to monitor counselling outcome.

The Beck Depression Inventory

The Beck Depression Inventory (Beck, 1972) is a 21-item multiple-choice self-report instrument that has been widely used for measuring a specific condition: depression.

Core Outcome Measure

The Core Outcome Measure (Evans et al., 2002) is used pre- and post-counselling to assess overall wellbeing, current problems, life functioning, and risk. It uses 34 client-rated items to arrive at a measure of global distress.

The Outcome Questionnaire-45

The OQ-45 (Lambert et al., 2004) is a 45-item self-report questionnaire that can be used for repeated measurements during the course of counselling. This use represents a shift from merely measuring change in counselling to incorporating the process of assessment into counselling itself (Lambert, 2010). The OQ-45 takes seven minutes to complete and can be scored and analysed electronically. It assesses three areas: symptom distress, interpersonal relations, and social role, as well as providing a total score of wellbeing.

Sample items: Rated on a five-point scale, from “Never” to “Always”

- i. I get along with others
- ii. I tire quickly
- iii. I feel no interest in things
- iv. I feel stressed at work/school

The Quality of Life Scale

The Quality of Life Scale (Heinrichs, Hanlon, & Carpenter, 1984) is a client self-report scale that uses 32 items to assess global wellbeing. It is designed for clients 17 years and over.

SF Health Surveys

SF Health Surveys (SF Health Surveys, n.d.) are said to be the most frequently reported measure of health and wellbeing in the literature (Ashworth et al., 2004). Ratings are client self-reports and the measure is designed for people 18 years and older. The survey is available in 36-, 12-, and 8-item versions; all three measure eight health domains.

General Well-being Schedule

The General Well-being Schedule (Dupuy, 1977) is an index of general wellbeing consisting of 18 items assessing life satisfaction and level of psychological distress. Scoring is based on a scale ranging from 0 to 110, with lower scores indicating more severe distress. The three levels of distress are sectioned accordingly: 0 to 60 reflects “severe distress;” 61 to 72 “moderate distress;” and 73 to 110 “positive well-being.” The Schedule collects client self-reported information and is self-administered by clients.

Sample items: The Schedule “contains questions about how you feel and how things have been going with you. For each question, mark (X) the answer which best applies to you.”

- | |
|--|
| <ol style="list-style-type: none">1. How have you been feeling in general? (DURING THE PAST MONTH)<ol style="list-style-type: none">1) In excellent spirits; 2) In very good spirits; 3) In good spirits mostly; 4) I have been up and down in spirits a lot; 5) In low spirits mostly; 6) In very low spirits.
2. Have you been bothered by nervousness or your “nerves”? (DURING THE PAST MONTH)<ol style="list-style-type: none">1) Extremely so—to the point where I could not work or take care of things; 2) Very much so; 3) Quite a bit; 4) Some—enough to bother me; 5) A little; 6) Not at all. |
|--|

Partners for Change Outcome Management System (PCOMS)³

PCOMS (Duncan, 2012) is another example of a nomothetic measure that allows for comparisons across groups of clients. It is discussed separately because of its international popularity and widespread use among counsellors in many countries. It is also becoming popular in New Zealand, with several agencies using it: Relationship Services Aotearoa (RA) (unfortunately, PCOMS was never fully implemented before

Outcome Rating Scale (ORS)

Looking back over the last week, including today, help us understand how you have been feeling by rating how well you have been doing in the following areas of your life, where marks to the left represent low levels and marks to the right indicate high levels. *If you are filling out this form for another person, please fill out according to how you think he or she is doing.*

Individually
(Personal well-being)

|-----|

Interpersonally
(Family, close relationships)

|-----|

Socially
(Work, school, friendships)

|-----|

Overall
(General sense of well-being)

|-----|

Institute for the Study of Therapeutic Change
<http://heartandsoulofchange.com/measures/>
 © 2000, Scott D. Miller and Barry L. Duncan

Figure 1: Outcome Rating Scale (ORS)

RA's closure in 2015); WellElder (Wellington) (for details on its use, see Manthei & Nourse, 2012); Methodist Mission Southern (Dunedin); and Wesley Community Action (Wellington). There may be others that I am unaware of.

PCOMS was previously known (and still is to many) as CDOI—Client Directed, Outcomes Informed. The measure contains two 4-item scales: the Outcome Rating Scale (ORS) and the Session Rating Scale (SRS). Although based on the Outcome Questionnaire (Lambert, 2010), it is shorter, takes less time to complete, is best when used as an aid to on-going counselling on a session-by-session basis (a procedure called FIT, or Feedback-Informed Treatment), and uses only a single question to assess each of the underlying four factors in the OQ. If you are confused about all of the acronyms involved or want background on them, see Miller (2014).

Session Rating Scale (SRS V.3.0)		
Please rate today's session by placing a mark on the line nearest to the description that best fits your experience.		
I did not feel heard, understood, and respected.	Relationship ┆-----┆	I felt heard, understood, and respected.
We did not work on or talk about what I wanted to work on or talk about.	Goals and Topics ┆-----┆	We worked on and talked about what I wanted to work on and talk about.
The therapist's approach is not a good fit for me.	Approach or Method ┆-----┆	The therapist's approach is a good fit for me.
There was something missing in the session today.	Overall ┆-----┆	Overall, today's session was right for me.
Institute for the Study of Therapeutic Change http://heartandsoulofchange.com/measures/ © 2002, Scott D. Miller, Barry L. Duncan, & Lynn Johnson		

Figure 2: Session Rating Scale (SRS V.3.0)

The first of the two scales, the ORS, is administered to clients at the beginning of a counselling session and is meant to assess change since the previous session (Duncan, 2014, provides a good description of how it is used in practice). It is scored using a 10-cm, visual analogue scale. Clients place a mark at the point on the scale at which they judge themselves to be. The four lengths are then measured and totalled to get an overall score (min = 0; max = 40). The four questions in the ORS are shown in Figure 1.

The second scale, the SRS (see Figure 2), uses a similar format and scoring. It is administered at the end of each session and assesses four aspects of the therapeutic alliance: the relational bond, the goals and topics of the session, the approach or method used, and the client's overall impression of the session (Duncan, 2012).

Because the SRS assesses the tone or climate of the actual sessions, rather than the client's perceptions of how he or she is functioning, only the ORS is considered further in this article.

Studies that have reported ORS scores have used either clinical and/or non-clinical samples, the distinction being:

Clinical samples include people who sought or were referred for psychological help. Their ORS scores are expected to be less than 25 at the beginning of counselling (Miller & Bargmann, 2012).

Non-clinical samples consist of people from the general population who are not presumed to need or desire psychological help. Their scores are expected to be 25 or higher. "A score of 25, the clinical cut off, differentiates those who are experiencing enough distress to be in a helping relationship from those who are not" (Duncan & Miller, 2014). A score of 25 or more is thus the dividing line between the clinical and non-clinical populations (Miller & Bargmann, 2012; Duncan & Miller, 2014; Reese, Toland, Slone, & Norsworthy, 2010). The cut-off points for adolescents (28) and children (32) have been found to be higher than for adults (Green & Latchford, 2012).

A further distinction has been made in the data reported for clinical samples between scores for those clients who received feedback on their ORS ratings during their counselling (as the scale is meant to be used; see Duncan, 2014) and those who only completed the scale before and after their counselling and therefore received no feedback during counselling. This group of clients were simply given what is called "treatment as usual" (TAU).

"Feedback" clients take part in a transparent, session-by-session discussion between counsellor and client about the client's ratings of his/her problems (ORS score) and the relationship with the counsellor (SRS score). This discussion is used to influence the course of counselling. This feature of PCOMS is important because research has clearly demonstrated that "feedback" clients make greater gains in counselling than TAU clients (see Tables 4 and 5 for details).

"Treatment-as-usual" clients are involved in no such session-by-session assessment or discussion, and counselling is delivered as usual.

Key information about the ORS includes:

1. The ORS is not tied to any particular theoretical orientation (Anker, Duncan, & Sparks, 2009).
2. Since there is a high correlation among the four items, the ORS can be used as a global measure by summing up the four item scores to get a total ORS score (Miller, Duncan, Brown, Sparks, & Claud, 2003).

3. There is a high client completion rate (Miller, Duncan, Sorrell, & Brown, 2005) and the ORS usually takes only a minute or two to complete.
4. The ORS is strongly correlated with other mental health measures, e.g., OQ-45 and the Depression Anxiety Stress Scale (Bringinghurst, Watson, Miller, & Duncan, 2006; Campbell & Hemsley, 2009; Duncan, 2014; Green & Latchford, 2012; Miller et al., 2003).
5. The average intake pre-counselling score in outpatient mental health settings (where people are seeking counselling help) is 18–19 (Miller & Bargmann, 2012).
6. Between 25 and 33% of adults (18+ years) from a clinical population (that is, those seeking or referred for counselling) score above 25 at intake (the clinical cut-off point) (Miller & Bargmann, 2012). For example, this rate was reported to be 27% (in Reese, Duncan, Bohanske, Owen, & Mina, 2014), 28.4% (in Reese, Norsworthy, & Rowlands, 2009), and 25% (in Miller, Duncan, Brown, Sorrell, & Chalk, 2006), all within the suggested range.
7. People scoring above 25 at intake are at a heightened risk for deterioration (Miller & Bargmann, 2012; Miller et al., 2005). This is thought to be due to the client having been mandated (ordered) to receive counselling, wanting help with a particular problem, already being a highly functioning person who wants to grow or actualise further, or having misunderstood the scale in some way (Miller & Bargmann, 2012).

Miller and Duncan (2004, as cited in Anker et al., 2009) recommended the following guidelines for describing change in counselling as measured by the ORS (also described in Reese et al., 2010):

Clinically Significant Change: beginning counselling with a score below the clinical cut-off of 25 points on the four scales and improving at least 5 points to a score of at least 25 or more.

Reliable Change: a gain of 5 points, but not exceeding the clinical cut-off at the end of counselling.

No Change: failing to gain (or lose) 5 points during counselling.

Deterioration/Worse Off: a drop of 5 or more points from pre- to post-counselling.

Findings relating to the utility of the ORS

The utility of the ORS has been assessed in a number of studies internationally. In Tables 1 and 2 below, scores are compared for those clients who received feedback on

Table 1. Proportions of clients benefitting from counselling using feedback

	Change =	Clinical	Reliable	None	Worse
Manthei & Nourse, 2012:	n=281, NZ		68.9%	16%	10.9%
Anker et al., 2009:	n=206, Nor	40.8	9.7	13.6	1.9
Reese et al., 2009: (2 studies)	n=50, USA n=45, USA		80 66.7	16 28.9	4 4.4
Reese et al., 2010:	n=54, USA	48.1	16.7	27.8	7.4
Cooper et al., 2013:	n=288, Ire	88.7		7.4	3.9
Schuman et al., 2014:	n=137, USA	28.4	20.4	36.5	14.6
Miller et al., 2006:	n=3612, USA		47	45	8

Note: Some of the above studies have removed those clients who scored 25 or above at intake from further analysis.

Table 2. Proportions of clients benefitting from counselling who did not receive feedback during the counselling

	Change =	Clinical	Reliable	None	Worse
Anker et al., 2009:	n=204, Nor	10.8%	11.8%	23.5%	3.9%
Reese et al., 2009: (2 studies)	n=24, USA n=29, USA		54.2 41.4	33.3 55.2	12.5 3.4
Reese et al., 2010:	n=38, USA	29.6	14.8	11.1	3.7
Schuman et al., 2014:	n=126, USA	11.9	23.8	49.2	11.9

their ORS ratings while counselling was taking place (see Duncan, 2014) and clients who completed the scale before and after their counselling but received no feedback during counselling.

Are there national or cultural differences in scores on the ORS and SRS?

Differences have been found in two Dutch studies using the ORS. In the first (Hafhenscheid, Duncan, & Miller, 2010), a reliable change score was found to be 8, not 5 as reported by Miller and Bargmann (2012). In the second (Janse, Boezen-Hilberdink, van Dijk, Verbraak, & Hutschemaeckers, 2014), a cut-off point of 24—compared to the USA cut-off point of 25—was found, and the reliable change score was established as 9, not 5. These two studies indicate that there may be important cultural or national differences on the ORS and SRS across countries.

I was unable to find any other attempts to standardise the ORS in additional countries. It seems to be assumed that the USA baseline scores can be readily used across countries. As evidence of this assumption, the ORS and SRS have been translated into a number of different languages—Dutch, Finnish, French, German, Greek, Hebrew, Norwegian, Slovak, Spanish, and Swedish—and, with the exception of the Netherlands, are presumably being used in those countries without country-specific standardised scores (Hafhenscheid et al., 2010). A Māori version of the two scales has been published, called the Kaupapa Outcome Rating Scale (Drury, 2007). Data from the use of this te reo version have not been published to my knowledge.

Are there racial differences in ORS scores?

Very few racially specific data were reported. An exception was Reese et al. (2014) who reported pre- and post-scores for a clinical sample of outpatient clients in the USA, including Hispanic, Afro-American, Native American, Asian American, Euro-American, and “Other” clients. There were no significant differences in pre- or post-ORS scores among the racial groups.

Are there gender differences in pre- and post-ORS scores?

No gender differences were highlighted in a USA study (Reese et al., 2014), nor in separate studies based in the Netherlands (Janse et al., 2014), Norway (Anker, Owen, Duncan, & Sparks, 2010), and New Zealand (Manthei & Nourse, 2012). However, two other studies from the USA reported significant differences in pre-ORS scores, but the results were in conflict, and no useful explanations or reasons were offered (the differences were described as “perplexing”) (Bringhurst et al., 2006).

What is the typical pre-counselling ORS score for the general population?

Typical pre-counselling scores in the general population, based on non-clinical samples, are shown in Table 3.

What is the difference in pre- and post-counselling scores between clients who received feedback on their ORS scores as part of their counselling and those who did not?

Several studies have shown that the use of the ORS as it was designed—to give session-by-session feedback to clients on their progress and their relationship with the counsellor—results in significantly better outcomes (compare the average gain scores in Table 4 with that in Table 5) (Miller et al., 2006; Reese et al., 2009, 2010; Schuman, Slone, Reese, & Duncan, 2014).

Table 3. Typical pre-counselling ORS score for the general population (non-clinical samples)

	Pre ORS	
From a Dutch sample (n=116) (Janse et al., 2014):	m = 29.6	sd = 6.6
From a USA sample (n=86) (Miller et al., 2003):	m = 27.9	sd = 6.
From a USA sample (n= 98) (Bringhurst et al., 2006):	m = 29.9	sd = 7.5

Table 4. Pre- and post-counselling ORS scores for clinical samples receiving feedback with their counselling

	Pre	Post	Gain
Adult means, all countries: Range:	20.2 (8.6) (18–23)	27.4 (8.1) (23–32)	7.2
Student means, all countries: Range:	21.1 (8.1)	29.9 (7.4) (18.5–25)	8.8 (25–33)

Table 5. Pre- and post-counselling ORS scores for clients who did not receive feedback on their scores as part of their counselling

	Pre	Post	Gain
Adults' means: Range:	20.6 (8.1) (18–24)	26.1 (8.7) (22–30)	5.5

Table 4 provides a summary of data from 14 studies across five different countries, including New Zealand; Table 5 summaries data from eight studies in two countries.

Are there different pre-counselling mean ORS scores for different problems/diagnoses?

Although there are some data addressing this question, they are not sufficient to enable definitive statements to be made. For example, the average pre-ORS scores for clients with drug and alcohol problems were reported to be higher than those of the general mental health population (24.1 vs 19.6) (Janse et al., 2014; Miller et al., 2003). However, this was not substantiated by two American studies (Reese et al., 2014; Schuman et al., 2014), nor by an Australian study (Quirk, Miller, Duncan, & Owen, 2013).

In regard to other diagnoses, based on a large sample (n=5186 adults) and clients in 12 diagnostic categories (e.g., substance abuse, mood disorders, anxiety, schizophrenia, bipolar), no significant differences in outcome were found (Reese et al., 2014).

Overall suitability of the ORS for use in New Zealand

Because the available research literature is sparse, the figures reported in published studies should be used with caution, especially if the scale is used in New Zealand where it has not been standardised.

There are several other considerations, as discussed below:

1. There is often a substantial shrinkage of subjects from intake to completion of counselling. This can be as high as 50% or more (Anker et al., 2009, 2010; Cooper, Stewart, Sparks, & Bunting, 2013; Reese et al., 2009, 2014). There is a question about what happens to these clients in terms of why they initially sought counselling. In contrast, New Zealand data on the elderly (Manthei & Nourse, 2012) showed that less than 1% of clients were “no shows,” and only 4% missed appointments during the course of their counselling. These markedly low figures were thought to be indicative of a highly motivated and conscientious older population of clients. The high rates are consistent with those reported in the literature about the elderly (Gum et al., 2006; Kuruvilla, Fenwick, Haque, & Vassilas, 2006).
2. One-session clients can form as much as 20% of the total client load (Reese et al., 2014), but they are often excluded from the ORS analyses (Anker et al., 2009; Miller et al., 2006; Reese et al., 2014). This is done in spite of the fact that Miller et al. (2006) discovered that as many as 80% of one-session clients were found to have made positive change when they were followed up. These numbers are supported by New Zealand findings. For example, Manthei (1996, 2012) found that in two different agencies the number of one-session clients was 15% and 23% of the total, respectively. It was also found that when followed up, 62% of single-session clients said their situations were improved (Manthei, 1996). Why exclude these clients from analysis when one-session therapy is so common here and in other countries?
3. Clients in clinical samples who score 25+ at intake are often excluded from analysis (Miller et al., 2006; Reese et al., 2010, 2014), yet they have chosen to seek counselling for reasons of their own. Why exclude them if their reasons for seeking counselling are sensible and legitimate?

Nevertheless, there appears to be general agreement about some aspects of the scale:

1. The ORS scale can be used with confidence to measure therapeutic change, both under the recommended “feedback” condition (FIT) and without use of “feedback.” If feedback is used, greater gains in counselling can be expected. In addition, testing for a “retrospective shift” could also maximise gain (Manthei, 1997). Briefly, retrospective pre-testing involves having clients rate themselves again after their post-counselling rating, but as they now perceive themselves to have been before starting their counselling. This “looking back to how they were before counselling” rating can then be used to assess more accurately the effectiveness of training, or in this case, counselling (see Hoogstraten, 1982). In most cases, clients would rate themselves worse, or in greater distress, retrospectively than they did at pre-testing (usually due to greater insight achieved during counselling), which would result in greater improvement/gain (the difference between retrospective pre-counselling rating and post-counselling rating).
2. Average intake scores on the ORS for those seeking or being referred for counselling are usually 18 to 22. Typical scores for the general population seem to be 27 to 29.
3. About 25 to 33% of intake clients will score 25 or more on the ORS. This does not invalidate them from benefitting from counselling, but it might affect how they are treated and how their counselling might progress.
4. There may be cultural or national differences in ORS pre- and post-counselling scores. Not enough research evidence is available to confirm this. In the meantime, the scale continues to be used in many different countries as though it were culturally neutral.
5. Evidence for differences in outcome for gender and across different diagnostic categories is not clear.
6. The clinical cut-off point—originally identified as 25 in the USA—might be different for countries other than the USA (e.g., 24 for the Netherlands) and for different age groups (e.g., 32 for children and 28 for adolescents).

At the finish of counselling, it would be reasonable to expect at least 50% of clients to register gains of 5 points or more.

Typically, use of the ORS has revealed that less than 10% of clients deteriorate during counselling. This rate accords with other research.

Summary of the advantages and disadvantages of PCOMS

Advantages

- It is quick to complete, but measuring the point on the line at which the client's mark appears could be clumsy (it could be modified to make it simpler: a 10-cm-long line could be treated as a 10-point scale).
- There is no fee for individual practitioners using PCOMS, and agencies and individuals can register at <http://scott-d-miller-ph-d.myshopify.com/collections/performance-metrics/products/performance-metrics-for-the-ors-and-srs>. Anyone interested in using the forms should register and also find out what other support services might be available, such as help with data analysis and interpretation.
- Both client and counsellor can complete the ORS. This allows two estimates of success and comparisons to be made—bearing in mind that it is common for counsellors to rate success “higher” than clients. This is not surprising given that clients and their counsellors disagree about many aspects of their shared counselling experience (Manthei, 2007).
- The CDOI (as the PCOMS is known to many) is already being used in several agencies in New Zealand and there is a wealth of international data with which to compare local results.

Disadvantages

- One obvious shortcoming is that it measures only general or global wellbeing. It does not assess the specific problems/complaints that brought the client to counselling.
- There may be counsellor resistance to having their performance rated. However, this would be a difficulty with almost any measure used. On the other hand there is probably little or no client resistance to its use, as long as they understand its purpose.

Patient-Generated Outcome Measures (PGOMS)

These are also referred to as PROMS (Patient-Reported Outcome Measures). In particular, this article looks at PSYCHLOPS (Psychological Outcome Profiles).

Patient-generated outcome evaluations (e.g., PGOMS, PROMS) are becoming more accepted in both mental health and medical settings (Patel, Veenstra, & Patrick, 2003; Sales & Alves, 2012). The measures are constructivist in that they provide idiographic, or individualised, perspectives on client-specified problems. Such measures tend to be client-friendly (after all, the client chooses the problems to be

worked on); specific, rather than global, evaluations of a client's wellbeing; flexible; and helpful to clarify the process of counselling (Sales & Alves, 2012). However, there are advantages to both nomothetic and idiographic approaches and some researchers have devised hybrid evaluation measures that use both.

One such measure is PSYCHLOPS, developed to assess counselling outcomes in terms of client-specified concerns (Ashworth et al., 2004), and then validated against the CORE-OM (Ashworth et al., 2005) and the HADS (Hospital Anxiety Depression Scale; Ashworth, Evans, & Clement, 2009). In both of those studies, the scale was reported to be comparable to improvement scores from the other two scales, but, importantly, it was found to be "a more sensitive measure of change than existing outcome measures" (Ashworth, 2007, p.201). Test-retest reliability was reported to be "satisfactory" by Evans, Ashworth, and Peters (2010). The scale has also been validated using an Icelandic clinical population (Heioinsson, Kristjandothir, Olanson, & Siguronsson, 2013).

PSYCHLOPS evaluates three domains (see Figure 3): problems (idiographic data; two questions); functioning (idiographic data; one question); and overall wellbeing (nomothetic data; one question). The post-counselling form includes a fifth question assessing outcome or improvement using a five-point scale from "much better" to "much worse" (Ashworth et al., 2009). The scoring has changed over several years, but the most current version adds up the scores for the first four questions for a maximum total of 20. If, however, clients name only one problem instead of the requested two, the problem score is doubled so the total possible score remains 20 (Evans et al., 2010).

The first two questions ask clients to describe in their own words two current problems and rate how much and for how long the problems have affected them. Originally designed as a pre-/post-counselling measure, recent usage now incorporates a during-counselling version so that mid (or during) therapy scores can be obtained. These additional scores did not change outcome effect size, nor did the introduction of additional problem(s) affect estimated change (Czachowski, Seed, Schofield, & Ashworth, 2011). In practice, the pre-therapy score is compared with subsequent scores (during therapy and post-therapy). The difference is then called the "change score."

An individualised measure such as PSYCHLOPS has the advantage of assessing the effect of counselling on client-defined problems or concerns, not just generic items assessing aspects of functioning or wellbeing. This targeted approach to problem improvement could be of considerable interest to third-party funders such as the Ministries of Social Development, Justice, Corrections, etc., which often fund specific programmes designed to deal with specific client behaviours.

Figure 3: An example of the PSYCHLOPS evaluation. For copies of the format currently in use, see www.psychlops.org

1 a Choose the problem that troubles you most. *(Please write it below.)*

b How much has it affected you over the last week?
(Please tick one box below.)

0 1 2 3 4 5

Not at all affected Severely affected

c How long ago were you first concerned about this problem?
(Please tick one box below.)

Under one month
 Between one and three months
 Over three months but under one year
 One to five years
 Over five years

2 a Choose another problem that troubles you. *(Please write it below.)*

b How much has it affected you over the last week?
(Please tick one box below.)

0 1 2 3 4 5

Not at all affected Severely affected

c How long ago were you first concerned about this problem?
(Please tick one box below.)

Under one month
 Between one and three months
 Over three months but under one year
 One to five years
 Over five years

3 a Choose one thing that is hard to do because of your problem (or problems).
(Please write it below.)

b How hard has it been to do this thing over the last week?
(Please tick one box below.)

0 1 2 3 4 5

Not at all hard Very hard

continued on following page

4	How have you felt in yourself this last week? (<i>Please tick one box below.</i>)							
	0	1	2	3	4	5		
	Very good	<input type="radio"/>	Very bad					
5	<i>(For use with post-counselling administration)</i> Now that you have finished your course of therapy, how would you describe yourself?							
	Much better	A little bit better			About the same			
	1	2			3			
		A little worse		Much worse				
		4		5				

Summary of the advantages and disadvantages of PSYCHLOPS

Advantages

- It is relatively quick and easy to complete, although the client must be willing to engage in the process (this shouldn't be a problem).
- Although problem specification is up to the client, actual ratings could be completed by both client and counsellor. This would generate two estimates of success and the resulting scores could be compared.
- There is a growing amount of international data on the use of PSYCHLOPS (particularly in England and Europe) with which local results could be compared.
- Its main strength is that it measures improvement in "target" problems specifically nominated by the client.
- Because it actively involves the client, it can be used as a direct aid or tool in counselling, with ratings being referred to and adjusted session by session (in much the same way that PCOMS uses ORS and SRS scores in session-by-session feedback).

Disadvantages

- There is a cost for using it, but I suspect it is minimal, probably less than the cost of PCOMS.
- There may be some counsellor resistance to its use, although there should be little or no client resistance.

Conclusion

It seems that either PCOMS or PSYCHLOPS, or a combination of the two, would provide the best assessment choices. As Ashworth, the developer of PSYCHLOPS,

summarised (2007, p.202): “[PSYCHLOPS] adds another dimension to outcome measurement, and not as an alternative to conventional standardised measures.” Thus, PCOMS would provide a useful overall measure of general progress, while PSYCHLOPS would give external funders progress data specific to client aims in particular programmes. Of course, any counsellor or agency intending to initiate a counselling evaluation procedure using either (or both) of these scales will need to read about them in more detail before determining how they best fit with their own practice or their agency’s structure and procedures.

Endnotes

1. Paper presented at the NZAC Conference, Auckland, July 17–18, 2015.
2. Effectiveness studies relate to demonstrating outcome effects under normal, or natural, field conditions that are often beyond the investigator’s control. They generate practice-based evidence and address the question “Does it work in routine practice?” Efficacy studies, on the other hand, refer to laboratory-based delivery where use of random assignment, control groups, and manualised delivery of counselling are utilised. They lead to evidence-based practice and address the question “Does it work?” The former approach is a more realistic, real-world, test of effect, whereas the latter is often not replicable nor generalisable because such controlled conditions cannot be achieved in the field. (BACP, n.d.; Granello et al., 1999).
3. In New Zealand, Exess Connectivity has the licensing rights for PCOMS. It is available in both English and te reo (see <http://www.exess.co.nz>, or Drury (2007), for the Kaupapa Outcome Rating Scale).

References

- ACA (2014). *Code of ethics*. <https://www.counseling.org>
- Anker, M. G., Duncan, B. L., & Sparks, J. A. (2009). Using client feedback to improve couple therapy outcomes: A randomized clinical trial in a naturalistic setting. *Journal of Consulting and Clinical Psychology, 77*(4), 693–704.
- Anker, M., Owen, J., Duncan, B., & Sparks, J. (2010). The alliance in couple therapy: Partner influence, early change, and alliance patterns in a naturalistic sample. *Journal of Consulting and Clinical Psychology, 78*(5), 635–645.
- Ashworth, M. (2007). PSYCHLOPS: A psychometric outcome measure that is finding a niche. *Counselling and Psychotherapy Research, 7*(4), 201–202.
- Ashworth, M., Evans, C., & Clement, S. (2009). Measuring psychological outcomes after

- cognitive behaviour therapy in primary care: A comparison between a new patient-generated measure “PSYCHLOPS” (Psychological Outcome Profiles) and “HADS” (Hospital Anxiety and Depression Scale). *Journal of Mental Health*, 18(2), 169–177.
- Ashworth, M., Robinson, S., Godfrey, E., Parmentier, K., Shepherd, M., Christey, J., ... Matthews, V. (2005). The experiences of therapists using a new client-centred psychometric instrument, PSYCHLOPS (Psychological Outcome Profiles). *Counselling and Psychotherapeutic Research*, 5(1), 37–42.
- 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.
- Attkisson, C. C., & Zwick, R. (1982). The Client Satisfaction Questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Program Planning*, 5, 233–237.
- BACP (2013). *Ethical framework for good practice in counselling*. <http://bacp.co.uk>
- BACP (n.d.). *Effectiveness of counselling*. www.bacp.co.uk/research/resources
- Beck, A. T. (1972). *Depression: Causes and treatment*. Philadelphia, PA: University of Pennsylvania Press.
- Bringhurst, D. L., Watson, C. W., Miller, S. D., & Duncan, B. L. (2006). The reliability and validity of the Outcome Rating Scale: A replication study of a brief clinical measure. *Journal of Brief Therapy*, 5, 23–30.
- Campbell, A., & Hemsley, S. (2009). Outcome Rating Scale and Session Rating Scale in psychological practice: Clinical utility of ultra-brief measures. *Clinical Psychologist*, 13(1), 1–9.
- Cooper, M., Stewart, D., Sparks, J., & Bunting, L. (2013). School-based counselling using systematic feedback: A cohort study evaluating outcomes and predictors of change. *Psychotherapy Research*, 23(4), 474–488.
- Crocket, A. (2013). Articulating counselling in the context of registration and New Right policy. *New Zealand Journal of Counselling*, 33(2), 31–45.
- Czachowski, S., Seed, P., Schofield, P., & Ashworth, M. (2011). Measuring psychological change during Cognitive Behavioural Therapy in primary care: A Polish study using ‘PSYCHLOPS’ (Psychological Outcome Profiles). *PLoS ONE*, 6(12), e27378. (Online journal)
- Drury, N. (2007). The Kaupapa Outcome Rating Scale. *New Zealand Journal of Counselling*, 27(1), 21–32.
- Duncan, B. (2012). The Partners for Change Outcome Management System (PCOMS): The heart and soul of change project. *Canadian Psychology*, 53(2), 93–104.
- Duncan, B. L. (2014). *On becoming a better therapist: Evidence-based practice one client at a time* (2nd ed.). Washington, DC: American Psychological Association.
- Duncan, B. L., & Miller, S. D. (2014). “When I’m good, I’m very good, but when I’m bad I’m

- better”: A new mantra for psychotherapists. Retrieved October 26, 2014 from [www.psychotherapy.net/article/therapy-effectiveness#section-the-outcome-rating-scale-\(ors\)](http://www.psychotherapy.net/article/therapy-effectiveness#section-the-outcome-rating-scale-(ors))
- Duncan, B. L., & Reese, R. J. (2013). Clinical and scientific considerations in progress monitoring: When is a measure too long? *Canadian Psychology, 54*(2), 135–137.
- Dupuy, H. J. (1977). The General Well-being Schedule. In I. McDowell & C. Newell (Eds.), *Measuring health: A guide to rating scales and questionnaire* (2nd ed.) (pp.206–213). New York, NY: Oxford University Press.
- Edwards, D. W., Yarvis R. M., Mueller, D. P., & Langsley, D. G. (1978). Does patient satisfaction correlate with success? *Hospital & Community Psychiatry, 29*(3), 188–190.
- Evans, C., Ashworth, M., & Peters, M. (2010). Are problems prevalent and stable in non-clinical populations? Problems and test–retest stability of a patient-generated measure, PSYCHLOPS (Psychological Outcome Profiles), in a non-clinical student sample. *British Journal of Guidance and Counselling, 38*(4), 431–439.
- 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.
- Fontana, A., Ford, J. D., & Rosenheck, R. (2003). A multivariate model of patients’ satisfaction with treatment for posttraumatic stress disorder. *Journal of Traumatic Stress, 16*(1), 93–106.
- Goss, S., & Rose, S. (2002) Evidence-based practice: A guide for counsellors and psychotherapists. *New Zealand Journal of Counselling, 23*(2), 67–76.
- Granello, D. H., Granello, P. F., & Lee, F. (1999). Measuring treatment outcomes and client satisfaction in a partial hospitalization program. *Journal of Behavioral Health Services & Research, 26*(1), 50–63.
- Green, D., & Latchford, G. (2012). *Maximising the benefits of psychotherapy: A practice-based approach*. Chichester, UK: John Wiley & Sons.
- Gum, A. M., Areán, P. A., Hunkler, E., Tang, L., Katon, W., Hitchcock, P., ... Unützer, J. (2006). Depression treatment preferences in older primary care patients. *The Gerontologist, 46*(1), 14–22.
- Hafhenscheid, A., Duncan, B. L., & Miller, S. D. (2010). The Outcome and Session Rating Scale: A cross-cultural examination of the psychometric properties of the Dutch translation. *Journal of Brief Therapy 7*(1&2), 1–12.
- Heinrichs, D. W., Hanlon, T. E., & Carpenter, W. T. (1984). The Quality of Life Scale: An instrument for rating the schizophrenic deficit syndrome. *Schizophrenia Bulletin, 10*(3), 388–398.
- Heoinsson, H., Kristjandothir, H., Olanson, D. P., & Siguronnsson, J. F. (2013). A validation and replication study of the patient-generated measure PSYCHLOPS on an Icelandic clinical population. *European Journal of Psychological Assessment, 29*(2), 89–95.
- Hernon, P., & Altman, E. (2010). *Assessing service quality: Satisfying the expectations of library customers* (2nd ed.). Chicago, IL: American Library Association.

- Hoogstraten, J. (1982). The retrospective pre-test in an educational training context. *Journal of Experimental Education*, 50, 200–204.
- Janse, P., Boezen-Hilberdink, L., van Dijk, M. K., Verbraak, M. J. P. M., & Hutschemaekers, G. J. M. (2014). Measuring feedback from clients: The psychometric properties of the Dutch Outcome Rating Scale and Session Rating Scale. *European Journal of Psychological Assessment*, 30(2), 86–92.
- Kuruville, T., Fenwick, C. D., Haque, M. S., & Vassilas, C. A. (2006). Elderly depressed patients: What are their views on treatment options? *Aging & Mental Health*, 10(2), 204–206.
- Lambert, M. J. (2010). *Prevention of failure: The use of measuring, monitoring, and feedback in clinical practice*. Washington, DC: American Psychological Association.
- 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.
- Lambert, M. J., & Vermeersch, D. A. (2013). Psychological assessment in treatment. In D. F. Geisinger, B. A. Bracken, J. F. Carlson, J.-I. Hansen, N. R. Kuncel, S. P. Riese, & M. C. Rodriguez (Eds.), *APA handbook of testing and assessment in psychology, vol 2: Testing and assessment in counseling psychology*, pp. 213–229. Washington, DC: American Psychological Association.
- Lebow, J. (1982). Pragmatic decisions in the evolution of consumer satisfaction with mental health treatment. *Evaluation and Program Planning*, 5, 349–356.
- Lebow, J. (1983). Research assessing consumer satisfaction with mental health treatment: A review of findings. *Evaluation and Program Planning*, 6, 211–236.
- Manthei, R. J. (1996). A follow-up study of clients who fail to begin counselling or terminate after one session. *International Journal for the Advancement of Counselling*, 18(2), 115–128.
- Manthei, R. J. (1997). The response-shift bias in a counsellor education programme. *British Journal of Guidance & Counselling*, 25(2), 229–237.
- Manthei, R. J. (2007). Client–counsellor agreement on what happens in counselling. *British Journal of Guidance & Counselling*, 35(3), 261–281.
- Manthei, R. (2012). Counselling effectiveness at a city counselling centre. *New Zealand Journal of Counselling*, 32(1), 37–55.
- Manthei, R., & Nourse, R. (2012). Evaluation of a counselling service for the elderly. *New Zealand Journal of Counselling*, 32(2), 29–53.
- Miller, S. D. (2014). What's in an acronym? CDOI, FIT, PCOMS, ORS, SRS...all BS? <http://www.scottdmiller.com/feedback-informed-treatment-fit/whats-acronym-cdoi-fit-pcoms-ors-srs-bs/>
- Miller, S. D., & Bargmann, S. (2012). The Outcome Rating Scale (ORS) and the Session Rating Scale (SRS). *Integrating Science and Practice*, 2(2), 28–31.
- Miller, S. D., Duncan, B. L., Brown, J., Sorrell, R., & Chalk, M. B. (2006). Using formal client

- feedback to improve retention and outcome: Making on-going, real-time assessment feasible. *Journal of Brief Therapy*, 5(1), 5–22.
- 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.
- Miller, S. D., Duncan, B. L., Sorrell, R., & Brown, G. S. (2005). The Partners for Change Outcome Management System. *Journal of Clinical Psychology*, 61, 199–208.
- Morkides, C. (2009, July 14). Measuring counseling success. *Counseling Today*. Retrieved from <http://ct.counseling.org/2009/07/measuring-counselor-success/>
- NZAC (2104). *Code of ethics*. http://www.nzac.org.nz/code_of_ethics.cfm
- Patel, K. K., Veenstra, D. L., & Patrick, D. L. (2003). A review of selected patient-generated outcome measures and their application in clinical trials. *Value in Health*, 6(5), 595–603.
- Quirk, K., Miller, S., Duncan, B., & Owen, J. (2013). Group session rating scale: Preliminary psychometrics in substance abuse group interventions. *Counselling and Psychotherapy Research*, 3, 194–200.
- Reese, R. J., Duncan, B., Bohanske, R., Owen, J., & Minami, T. (2014). Benchmarking outcomes in a public behavioral health setting: Feedback as a quality improvement strategy. *Journal of Consulting and Clinical Psychology*, 82(4), 731–742.
- Reese, R. J., Norsworthy, L. A., & Rowlands, S. (2009). Does a continuous feedback model improve psychotherapy outcomes? *Psychotherapy Theory, Research, Practice, Training* 46, 418–431.
- Reese, R. J., Toland, M. D., Slone, N. C., & Norsworthy, L. A. (2010). Effect of client feedback on couple psychotherapy outcomes. *Psychotherapy Theory, Research, Practice, Training*, 47(4), 616–630.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence-based medicine: What it is and what it is not. *British Medical Journal*, 312, 71–72.
- Sales, C. M. D., & Alves, P. C. G. (2012). Individualized patient-progress systems: Why we need to move towards a personalized evaluation of psychological treatments. *Canadian Psychology*, 53(2), 115–121.
- Schuman, D. L., Slone, N. C., Reese, R. J., & Duncan, B. (2014). Efficacy of client feedback in group psychotherapy with soldiers referred for substance abuse treatment. *Psychotherapy Research*. Retrieved October 25, 2014 from <http://dx.doi.org/10.1080/10503307.2014.900875>
- Sexton, T. L., Whiston, S. C., Bleuer, J. C., & Walz, G. R. (1997). *Integrating outcome research into counseling practice and training*. Alexandria, VA: American Counseling Association.
- SF Health Surveys, at <https://www.optum.com/optum-outcomes/what-we-do/health-surveys.html>
- Sitza, J., & Wood, N. (1997). Patient satisfaction: A review of issues and concepts. *Social Science Medicine*, 45(12), 1829–1843.

- Tolley, A., (2015, 4 June). Strategy to improve results in social services. www.beehive.govt.nz/release/strategy-improve-results-social-services
- Williams, B. (1994). Patient satisfaction: A valid concept? *Social Science Medicine*, 38(4), 506–516.
- Williams, B., Coyle, J., & Healy, D. (1998). The meaning of patient satisfaction: An explanation of high reported levels. *Social Science Medicine*, 47(9), 1351–1359.