

A review of the evidence in support of the SF-36 and General Self-efficacy scale to evaluate a vocational rehabilitation service

Emma Boger, Paula Kersten

Abstract

The purpose of this paper is to examine the evidence for using the General self-efficacy scale and physical functioning scale of the SF-36 to evaluate a vocational rehabilitation service to 1) inform researchers, clinicians and commissioners about the properties of the measures in use and 2) make recommendations for the future evaluation of vocational rehabilitation. **Methods:** Electronic databases, United Kingdom government websites, generic internet search engines and hand searches of reference lists were reviewed for relevant articles. Abstracts were selected against inclusion criteria and relevant articles retrieved for appraisal.

Findings: 19 articles were retrieved and reviewed for inclusion criteria. 7 articles met the inclusion criteria and contributed to the review of the evidence. The use of the SF-36 and GSES in vocational rehabilitation settings appears to be based on evidence drawn from studies with methodological flaws. No evidence could be found in support of the sensitivity and responsiveness of the GSES.

Discussion and implications for practice: The use of outcome measures that have questionable reliability and validity, and an absence of evidence with regard to responsiveness and sensitivity in vocational rehabilitation populations, contributes to an inability to evaluate the effectiveness of vocational rehabilitation services. Further enquiry into how vocational rehabilitation operates, would help to clarify the nature and range of specific activities to be targeted and aid the selection of existing appropriate measures or the development of new measures that reflect the conceptual premises that support vocational rehabilitation.

Background

For seventy years the United Kingdom (UK) has provided a welfare state for vulnerable individuals requiring financial support. In 2009, the number of claimants unable to work due to a health condition was 2.7million (ONS, 2009). This represented an approximate cost to the UK economy of £100billion/year, equating to more than the entire National Health Service (NHS) budget in 2008 (Bell et al., 2008). The average duration of claims in 2002 was nine years, compared to three years in 1985

(DWP, 2002). Evidence suggests that the longer an individual is out of meaningful occupation, the higher their likelihood of experiencing depression, feelings of inadequacy, additional illness and a reduced quality of life (Wadell and Burton, 2006).

To tackle this growing issue in England and Wales, a joint NHS and Department of Work and Pensions (DWP) vocational rehabilitation intervention for recipients of health benefits was initiated. This intervention was known as the Condition Management Programme (CMP) and was administered in 26 regions. One of these regions was Dorset, a medium sized county in the south of England, with a population of around 710,000, 6% of which were in receipt of a health related benefit in 2009 (Council, 2009).

CMP aimed to progress people towards work by reducing fears regarding health and work and enabling claimants to become 'expert' in managing their condition. Clients were allocated a practitioner to support them on a 1:1 basis with individualised psychoeducational interventions. Within

Key words: health-related benefits, vocational rehabilitation, SF-36, general self-efficacy, return to work, condition management.

Dorset CMP, the health professionals consisted of Physiotherapists, Occupational Therapists, and Mental and Adult Health Nurses. Clients were assigned to practitioners, independent of practitioner experience or background as a generic approach to therapy was advocated. However, practitioners were autonomous in exercising 'clinical judgement' to negotiate and agree a plan of action with clients based on a detailed biopsychosocial assessment. Group sessions lead by qualified health practitioners and support workers were also offered to clients. Attendance was not compulsory. CMP was intended for people receiving a health-related benefit for any health condition, however, the majority of referrals could be categorised into three conditions; Mental Health, Musculoskeletal and Cardiovascular.

CMP funding was reliant on the

Author details

Emma Boger PGDip, BSc (hons), RGN, (Post-graduate researcher, Faculty of Health Sciences, University of Southampton, United Kingdom)

Paula Kersten PhD, MSc, BSc (Associate Professor, School of Rehabilitation & Occupation Studies, University of Auckland, New Zealand)

effectiveness and value for money it demonstrated, therefore the use of appropriate outcome measures to aid evaluation was vital to the services longevity. Typically, CMP was funded for a finite time period per client (approximately 13 weeks). The outcome measures used must therefore be able to demonstrate changes in this period. The outcome measures must also be responsive to relevant change due to the intervention, and not change that would occur regardless either due to time, or variation in health condition.

Arguably, the desired outcome following vocational rehabilitation is return to work or work-related activity (Black, 2008, BSRM, 2000). However, this may be an unrealistic expectation for clients who have been out of work for substantial time periods, and within the context of a downward economic climate. Current opinion within vocational rehabilitation suggests it may be more suitable to measure improvements in well-being and confidence, such as self-efficacy, that impact upon a client's readiness to work (Ford and Plowright, 2009, King et al., 2007). CMP in Dorset used the Short-Form 36 item Health Survey Questionnaire (SF-36) (Ware et al., 1993) and the General Self Efficacy Scale (GSES) (Schwarzer and Jerusalem, 1995) as the key measures to report outcomes to commissioners. The service did not routinely follow-up clients; data were collected pre and post intervention.

According to the World Health

Organisation–International Classification of Function, Disability and Health (WHO-ICF) model, the domain of impairment relates to disability and is deduced by the observation of behaviour and the capability to perform specific tasks (WHO, 2001). The CMP seeks to assess this by asking clients to complete a self-report measure (SF-36) pre and post participation. The SF-36 (Ware et al., 1993) is a generic health status measure that measures physical, emotional and social health well-being and is widely used by different health care settings (Bowling, 2007). The SF-36 has 36 items, each having a yes/no response on a hierarchical scale of severity. Each element is scored on a 0-100 scale, with higher scores representing better function or fewer problems.

In addition to physical function, Dorset CMP evaluated changes to self-efficacy, in relation to the WHO-ICF domain of participation, qualified as 'the meaning attributed to the behaviour by the individual' (WHO, 2001). Self-efficacy is described as the level of confidence in one's own ability to perform a specific task or situation (Bandura, 1997). According to Bandura, those with a strong sense of self-efficacy view challenging problems as tasks to be mastered and have a stronger sense of commitment to their interests and activities. Self-efficacy has been positively associated with return to work in a range of health conditions (Booth and James, 2008, Collins et al., 2000, Repetto, 2004, Shirom et al., 2008, Waghorn et al., 2007, Westaby and

Braithwaite, 2003). A key aim of CMP is to improve self-management of health, with an overall aim of promoting return to economic activity. CMP practitioners in Dorset felt this aim is in part met by raising self-efficacy in individuals. The GSES was therefore introduced to appraise this concept, pre and post participation in Dorset CMP, in January 2010.

The GSES is a self-report measure of perceived general self-efficacy, developed for use in individuals with generic health conditions (Schwarzer and Jerusalem, 1995). The measure examines beliefs in one's capability to handle new and difficult tasks in a variety of different domains. It consists of 10 items that are rated on a 4-point scale with the anchors 'not at all true' and 'exactly true'. Higher scores on this measure indicate higher levels of general self-efficacy.

The aim of this study was to investigate the evidence for using the PF scale of the SF-36 and the GSES to appropriately evaluate CMP in Dorset, with regard to their respective WHO-ICF domains of impairment and participation, and discuss the implications for future practice.

Methods

A search of the literature was conducted to examine the evidence supporting the use of the SF-36 and GSES in the clinical context of vocational rehabilitation. The search strategy was focused using the PICO framework (CEBM, 2010) to derive specific search terms (table 1).

Population	Key terms	Intervention	Outcome
(<65 years)- Health, long-term health conditions, Cardiovascular, Musculo-skeletal, Mental Health, vocational rehabilitation	Generalized self-efficacy scale, GSE,GSES, Schwarzer and Jerusalem (authors,	Not applicable	Self-efficacy, GSE, GSES, self-care, self-awareness, self- management
AND	AND		SF-36, generic health measure, physical functioning
Use of the GSES with these populations	WHO-ICF participation SF-36, Ware (author), generic health assessment tools, physical functioning		Validity, reliability, sensitivity, responsiveness, clinical utility
AND	AND		
Use of the SF-36 (PF) with these populations	WHO-ICF impairment		

The following electronic databases were searched; CINAHL; PSYCHinfo and Medline. The databases were selected due to comprehensive cover of journals relating to medical and health related journals. Article reference lists, website of UK government health department, generic internet search engines, and stroke-specific organisations were also searched. Dissertations and conference abstracts were excluded however searches for publications by dissertation or conference abstract authors were conducted. Search terms were chosen to represent the following properties of outcome measure quality; reliability, validity, sensitivity and responsiveness (Valderas et al., 2008).

No time limitations were set, since older literature on measurement properties is still relevant. Searches were conducted in April 2010 to specifically find evidence of those properties with vocational rehabilitation populations. This is crucial since a measure's reliability and validity are on-going properties, dependent upon the context and population with which it is used (DeVellis, 2003, Streiner and Norman, 2008). For example, assessment of self-efficacy in a traumatic brain injury may be different to a vocational rehabilitation population since the issues faced by both populations will have similarities and differences.

Generic internet search engines were also searched. Studies were excluded if they investigated postal or proxy reliability and validity unless this was how they were used in vocational rehabilitation. All relevant full text articles

were assessed using an appropriate standard critical appraisal tool (SPH, 2010).

Each abstract was screened to eliminate articles that were not relevant, based on the following inclusion criteria:

(1) the study was published in English; (2) the article study populations were specifically vocational rehabilitation, and (3) was published between January 1990 and June 2011, to examine the current and most relevant evidence for practice.

Findings and implications for practice

Search results

A total of 103 articles were identified relating to the GSES and 73 articles relating to the physical functioning scale of the SF-36. Of those, 33 abstracts were identified as potentially relevant studies and were screened (143 duplicate records were excluded). 19 articles were retrieved and reviewed for inclusion criteria (studies were excluded because they did not meet the detailed criteria or if they reported on an earlier phase of the same study). A total of seven articles met the inclusion criteria and contributed to the review of the evidence.

SF-36 Physical Functioning scale

Gatchel et al (Gatchel et al., 1998) examined the validity, responsiveness and clinical utility of the SF-36 in work-disabled patients (n=286) with chronic back pain who underwent a tertiary rehabilitation programme, a similar intervention to CMP. Patients completed the SF-36 pre and post participation. To assess clinical utility and content va-

lidity staff were asked to comment on pre-programme SF-36 scores and what they might mean, in relation to any management plans. Pre and post scores were analysed using a paired t-test and found to be significant. Scores were then compared with those from other battery items to see which measure was most responsive and to assess convergent validity. The PF scale of the SF-36 was found to significantly correlate with the other back pain measures.

However, the results must be viewed with caution since inappropriate statistical tests were used for the analysis. The SF-36 generates non-parametric data with arbitrary values assigned to response labels, therefore the t-test and Pearson Product-Moment Correlation (both parametric tests) are unsuitable (Sim and Wright, 2000). The interpretation of data in this manner does not discriminate between the various attributes of the participants in the study, nor the degree of correlation between the response and the underlying assigned construct. The use of non-parametric tests such as chi square and Mann-Witney U would be more fitting to avoid imprecise conclusions (Streiner and Norman, 2008).

In addition, details on the assessment of content validity are negligible. Information on the staff involved, how many, their experience and their conclusions are necessary to establish if this process was rigorous. The authors concluded that the study provided evidence of construct validity for the PF scale, though the statistical analyses and methodology used to determine this were unsuitable. The discussion reported that staff found no relationship between SF-

36 scores and completion and progression through the programme, yet the authors suggest the SF-36 has clinical utility with this population, which appears discordant. Davidson and Keating (Davidson and Keating, 2002) examined the use of the PF scale for reliability and responsiveness in the assessment of low back pain (LBP) by physiotherapists. The sample (n=106) came from a range of settings and areas to ensure it was representative of different treatment modes and patient socioeconomic backgrounds. The SF-36 PF scale and four other commonly used measures for the assessment of LBP were completed in a random sequence at commencement of treatment and then mailed to participants along with a 7-level global change scale, six weeks later. A reminder and replacement set were mailed to non-responders after three weeks, resulting in an adequate response rate of 51% (Openheim, 1992). Using the global scale (ranging from 1="completely gone," to 7="much worse."), participants were classified into one of two groups- 'improved' (points 1-3) or 'unchanged' (points 4-6). A paired t-test was performed on the responses of a sub-group (n=16) who responded as 'a little better' which the authors pose demonstrated the grouping was a valid assumption.

Several problems exist with this approach. Firstly, treatment of the global scale as parametric data is misleading since the values given to intervals on the scale bear no relation to the numerical label they are assigned. Consequently the t-test is not an appropriate test and therefore invalid assumptions are drawn from this analysis. Secondly, it was not specified which group those who responded as 'much worse'

would fit, since neither group is relevant. Thirdly, grouping of seven possible responses into two options is arguably a crude assumption and undermines the attempt to examine the clinical context of change.

Reliability was examined through Intraclass Correlation Coefficient (ICC) calculations for each measure and was reported as greater than 0.8 for the SF-36 PF scale, indicating a high degree of reliability for this measure in this population (Fitzpatrick et al., 1998). Responsiveness was examined using standardised response means (SRMs) and receiver operating characteristic (ROC) curves. For the SF-36 PF scale the SRM=0.44, which is considered moderate and the ROC=0.74, which is considered acceptable (Cohen, 1997). The results remain questionable due to the grouping of respondents into two categories prior to analysis, and therefore do not provide a dependable indication of the scales responsiveness.

Evidence for the sensitivity and responsiveness of the PF scale in a vocational rehabilitation population was absent from the searches. Harwood and Ebrahim (Harwood and Ebrahim, 2000) compared the sensitivity of the PF scale, with the London Handicap scale (Harwood and Ebrahim, 1995), Nottingham activities of daily living scale (Nouri and Lincoln, 1987) and an untested self-report global scale to measure clinically meaningful change, in patients pre and post-surgery for Hip replacement. 123 patients waiting for hip surgery participated in the study with an initial response rate of 63% (n=81), reducing to 58% (n=75) 3 months post-surgery, and 48% (n=58) 6-12 months post-surgery. It is not pos-

sible to specify if any important differences existed regarding those who dropped out, since none were provided by the authors. The mean age of respondents was 72 years, so with regard to applying the findings to CMP is unrepresentative. The response rates are reasonable, though it is important to consider that participants least pleased with their results may be least likely to respond to questionnaires (Openheim, 2000). It is important to consider possible bias with participants who are grateful for treatment being reluctant to declare poor outcomes (Marshall et al., 2006, Sitzia and Wood, 1997).

Responsiveness was analysed by calculating effect sizes from mean and standard deviation scores. Effect size statistics do not account for variation between respondents at baseline and follow-up. The standard deviations at follow-up were very different from baseline, indicating possible variation therefore use of the standardised response mean statistic would be more appropriate (de Vet et al., 2011). The effect sizes for the SF-36 PF scale at three and 6-12 months post-surgery were high at 1.13 and 1.33 respectively. Considering an effect size of 0.5-0.8 is considered significant (Cohen, 1997), this raises possible questions regarding the integrity of the results.

However, change to PF is highly likely to improve following this kind of intervention. Initial scores on the PF scales were skewed towards the worse end of the scale, which provides ample scope for measuring improvement. Had participants deteriorated or stayed the same, a 'floor effect' is likely to have been observed. Questions exist regarding the sensitivity of the SF-36 PF scale for use in CMP, where PF changes are likely to be

less obvious and a 'floor effect' observed. Within CMP, changes to PF are most likely to occur indirectly following changes to mood, confidence and well-being, which can influence activity levels. The 'floor effect' of the SF-36 has been described in various health conditions (Busija et al., 2008, Kersten et al., 1999, Lai et al., 2003), and constitutes an important consideration for vocational rehabilitation.

A more meaningful analysis of the data could incorporate Item Response Theory to examine the relationship between the SF-36 items level of difficulty and the participants' level of ability. McHorney et al (McHorney et al., 1997) examined Likert-scale scoring of the SF-36 PF scale using a Rasch analysis (Bond and Fox, 2001), in terms of the scales' internal validity and responsiveness over a two year period. The study used a comprehensive sample of complete data sets (n= 3445).

Relative precision estimates were used to estimate how much more or less precise the Rasch method was in relation to the standard (Likert-style scale), using standard error calculations. Rasch analysis demonstrated an uneven distribution of items across the proposed hierarchy of health, with item grouping in the middle range of PF ability. This infers that score change may be influenced by baseline scores and therefore should be considered when interpreting changes in health on an individual basis. The authors concluded the Likert-style scoring method was adequate when comparing groups however, the SF-36 is not used in this context by CMP so is a further consideration for the evaluation of such services.

The General Self Efficacy Scale

(GSES)

The validity of the GSES in Arthritis patients, which represent a core client group to CMP, had been examined (Barlow et al., 1996). The authors recruited patients from out-patient clinics and from a charity. Following translation from German into English, the GSES was administered via telephone, to 37 patients to investigate face validity. Minor changes were made as a result of these comments to reflect the phraseology and social conventions implicit in language. The authors did not validate these alterations nor re-test the alterations sufficiently to ensure the trans-adaptation across language and culture was rigorous (Hoegh and Hoegh, 2009). The scale was then administered to a further 16 individuals. The authors concluded that following 'no confusion' from the second sample (n=16), the GSES translation had face validity.

Several limitations exist in this approach. Firstly, the authors recruited individuals who were less than 50 years of age, which is arguably not representative of patients with Arthritis. Secondly whilst telephone administration can be time and cost effective, it does not allow the researcher to be sure that the patient is completing it unaided, or to observe difficulties with reading and comprehending the scale, which is vital to examine face validity. Thirdly, there are no details regarding the analysis of the results and how face validity was concluded. The authors do not give details on how they chose the sample for the second round, or why the size was so much smaller which increases the chance of their findings occurring by chance. More recently, techniques such as Cognitive Interviewing (Willis, 2005), pro-

vide a robust methodology to ensure that PROMs have acceptable face and content validity (Carbone et al., 2002, Christodoulou et al., 2008, Wu and McSweeney, 2004).

Chiu and Tsang (Chiu and Tsang, 2004) examined the reliability and validity of the GSES with individuals with Schizophrenia in a Hong Kong population. Content validity was appropriately investigated by an eight expert panel and deemed to be good to excellent. Participants were recruited from three different hospitals. The paper omits details on how they were recruited, over what time period and if the hospital populations differed demographically, all of which potentially affect results. Participants completed the questionnaire with 'assistance' from staff. Whilst potentially necessary in this client group, the paper does not clarify what assistance was provided or if staff or participants were aware completion of the scale was for research purposes, introducing potential bias. The paper does not state if ethical approval was required or granted, creating doubts concerning the robustness of methodology.

The scale was re-administered two weeks later to examine stability over time (test-retest reliability). Appropriate statistical tests, the intra-class correlation coefficient model and Cronbach's α coefficient were used to explore test-retest reliability and internal consistency, respectively. Test-retest reliability was 0.85 and internal consistency 0.92. Both are high scores indicating good reliability and internal consistency (Streiner and Norman, 2008). Details were omitted regarding the clinically appropriate time elapse between tests to ensure potential measured change was not incidental. Overall, this paper

contained scant details regarding methodology and implications on practice were difficult to contextualise due to the absence of any conclusions by the authors.

Scherbaum et al (Scherbaum et al., 2006) explored Item Response Theory to examine construct validity and reliability of the GSES and two other measures of self-efficacy in Psychology students (n=606). Random data sets of responses were also generated using a computer programme to conduct Modified Parallel Analysis (MPA). Internal consistency for the GSES was 0.61, indicating an acceptable level of reliability (de Vet et al., 2011). MPA revealed the first factor of the GSES accounted for 35% of the variance, indicating good item stability (Streiner and Norman, 2008). Cronbach's α were significant for correlations between the other measures of self-efficacy, signifying good reliability.

The Graded Response Model (Samejima, 1969) was adopted to explore the relationship between item difficulty and participant ability. Results indicated that items on the GSES adequately discriminated between individuals with similar but differing levels of self-efficacy. Graphic depiction of test information functions and standard errors of measurement indicated that the GSES performs best in individuals with average or below average levels of self-efficacy.

This study poses some limitations when considering this measure for vocational rehabilitation. The sample was compiled of healthy individuals, attending University with 62% in part-time employment, so is arguably unrepresentative of typical clients. Participants were given credits towards their degree,

possibly indicating an inherently higher level of self-efficacy in participants. Measures were administered in the same order, therefore a response pattern bias cannot be excluded. Overall, the paper indicates the GSES to possess construct validity with this sample.

With regard to responsiveness and sensitivity of the GSES, the author has been unable to identify papers examining these concepts through electronic database searches, nor the scales' website and user guide. Contact with the GSES authors has been, to date, unyielding. This measure is used to explain behaviour and relationships between traits and states, so perhaps it is unsurprising the literature focuses on validity and reliability. Responsiveness and sensitivity detail would provide useful information concerning the GSES's ability to detect clinically meaningful change. As searches have proved unsuccessful in identifying examples where changes to self-efficacy have been measured with the GSES, it poses questions about its' current use with vocational rehabilitation.

When measuring an outcome it is important to define the expected change or the goal of treatment, and then define the data needed to measure the phenomenon or construct (Laver-Fawcett, 2007). Within the context of vocational rehabilitation, problems exist with implementing this approach since CMP is aimed at generic health conditions, consequently anticipated change varies considerably. The problem is further illustrated by a lack of consensus regarding the most appropriate outcome measure to use (Ford and Plowright, 2009).

For a measure to have clinical utility from a CMP, or other vocational

rehabilitation settings, it must be brief and easy to complete, simple to score, and have no serious floor or ceiling effects in a generic clinical population. Additionally, it must not hinder the relationship between client and practitioner. The author routinely experiences negative comments from clients following completion of the SF-36. For example, a client with Multiple Sclerosis found the PF scale 'totally irrelevant and ridiculous'. Similar criticisms have been reported in the literature (Gatchel et al., 1998, de Groot et al., 2006, Kersten et al., 1999).

CMP adopted the SF-36 since it was user friendly, recognised and a seemingly extensively validated generic measure. It was used historically by other CMPs, so implementation in Dorset was unquestioned. However, it was designed to detect health changes in healthy populations, not individuals with health problems (Ware et al., 1993) consequently use without further validation in CMP is questionable. Arguably, this measure was hastily adopted without careful scrutiny.

CMP introduced the GSES as practitioners began to question the limitations of the SF-36. Its' implementation reflected CMPs' belief, that health related quality of life is considered dependent upon the interpretation and perceptions of the individual. However, this concept is difficult to demonstrate to service commissioners who regard return to work as the optimum outcome and who place less value on latent variables, such as self-efficacy. The paucity of evidence regarding the responsiveness and sensitivity of the GSES, notwithstanding the SF-36 in vocational rehabilitation populations, infer that current use to detect changes

to self-efficacy within vocational rehabilitation maybe unfounded.

Funding for the CMP service finished in April 2011, following questions regarding the cost-effectiveness of the intervention. A study examining the cost-effectiveness, concluded few conclusions could be made regarding the efficacy and effectiveness of interventions to aid those on health benefits back to work, due to a lack of credible evidence (Hayday et al., 2008). A crucial limitation of CMP was that evaluation was not co-ordinated nationally and consensus of the theory underpinning CMP was not established prior to implementation, therefore, arguably it was not an evidence based intervention. Being clear of the theoretical premise of an intervention is a crucial step to effective evaluation (MRC, 2000). Therefore appropriate criteria to indicate change and potential outcomes were debatably absent from CMP.

Conclusions

There is a need to develop a robust measure appropriate for use in a CMP setting for those with generic health problems. The current use of the SF-36 and the GSES by the service is flawed with both remaining un-validated measures for use with this population. Rigorous testing of both measures for validity, reliability, responsiveness and sensitivity with this population is required to establish clinical utility. Commissioners need to provide guidance to practitioners with regard to the expected outcomes of commissioned services to aid selection or development of appropriate outcome measures. Until such clarification, researchers and clinicians should, where possible, select outcome measures

with reliability and validity data in the population to be tested in the vocational rehabilitation intervention. The selection of outcome measures developed with involvement from the target population is also advocated. This ensures that what is meaningful to the patient is more likely to be captured appropriately, thus enhancing content validity (Lasch et al., 2010).

In the meantime, researchers must support clinicians by conducting further work to examine the concept and theoretical premises of vocational rehabilitation and develop appropriate measures if required.

References

- BANDURA, A. 1997. Self-efficacy. The exercise of control, New York, Freeman Press.
- BARLOW, J. H., WILLIAMS, B. & WRIGHT, C. 1996. The Generalized Self-Efficacy Scale in people with arthritis. *Arthritis Care And Research: The Official Journal Of The Arthritis Health Professions Association*, 9, 189-196.
- BELL, M., KOSSYKH, M., RIDGE, N. & WOOLLEY, D. 2008. An empirical analysis of the effect of health and economic growth in the UK: Research Report. London: Health and Safety Executive.
- BLACK, C. 2008. Working for a healthier tomorrow. In: PENSIONS, D. O. W. A. (ed.). London: HM Stationary Office.
- BOND, T. & FOX, C. 2001. Basic Principles of the Rasch Model. In: BOND, T. & FOX, C. (eds.) *Applying the Rasch Model: fundamental measurement in the human sciences*. New Jersey, USA: Lawrence Erlbaum Associates, Inc.
- BOOTH, D. & JAMES, R. 2008. A literature review of self-efficacy and effective jobsearch. *Journal of Occupational Psychology, Employment and Disability*, 10, 27-42.
- BOWLING, A. 2007. Health status assessment. In: AYERS, S., BAUM, A., MCMANUS, C., NEWMAN, S., WALLSTON, K., WEINMAN, J. & WEST, R. (eds.) *Cambridge Handbook of Psychology, Health and Medicine*. Cambridge, UK: Cambridge University Press.
- BSRM 2000. Vocational rehabilitation – the way forward. London: British Society of Rehabilitation Medicine.
- BUSIJA, L., OSBORNE, R., NILSDOTTER, A., BUCHBINDER, R. & ROOS, E. 2008. Magnitude and meaningfulness of change in SF-36 scores in four types of orthopedic surgery. *Health and Quality of Life*

Outcomes, 55.
CARBONE, E., CAMPBELL, M. & HONESS-MORREALE, L. 2002. Use of cognitive interview techniques in the development of nutrition surveys and interactive nutrition messages for low-income populations. *Journal of the American Dietetic Association*, 102, 690-696.

CEBM. 2010. Asking Focused Questions [Online]. Oxford: Centre for Evidenced Based Medicine, University of Oxford. Available: <http://www.cebm.net/EBMtools/> [Accessed 2nd November 2010 2010].

CHIU, F. & TSANG, H. 2004. Validation of the Chinese general self-efficacy scale among individuals with schizophrenia in Hong Kong. *Int J Rehabil Res*, 27, 159-61.

CHRISTODOULOU, C., JUNGHAENEL, D. U., DEWALT, D. A., ROTHROCK, N. & STONE, A. A. 2008. Cognitive interviewing in the evaluation of fatigue items: results from the patient-reported outcomes measurement information system (PROMIS). *Qual Life Res*, 17, 1239-46.

COHEN, J. 1997. *Statistical power for the behavioural sciences*, New York, Academic Press.

COLLINS, M., MOWBRAY, C. & BYBEE, D. 2000. Characteristics predicting successful outcomes of participants with severe mental illness in supported education. *Psychiatr Serv*, 51, 774-80.

COUNCIL, D. C. 2009. Key facts on population [Online]. Dorchester: Dorset County Council. Available: <http://www.dorsetforyou.com/344862>.

DAVIDSON, M. & KEATING, J. 2002. A Comparison of Five Low Back Disability Questionnaires: Reliability and Responsiveness. *Physical Therapy*, 82, 8-24.

DE GROOT, V., BECKERMAN, H., UITDEHAAG, B., DE VET, H., LANKHORST, G., POLMAN, C. & LM, B. 2006. The usefulness of evaluative outcome measures in

patients with multiple sclerosis. *Brain*, 129, 2648-2659.

DE VET, H., TERWEE, C., MOKKINK, L. & KNOL, D. 2011. *Measurement in Medicine*, Cambridge, University Press.

DEVELLIS, R. 2003. *Scale development. Theory and Applications*, Thousand Oaks, Sage Publications.

DWP 2002. *Pathways to work: helping people into employment*. Green paper. London: Department of Work and Pensions.

FITZPATRICK, R., DAVEY, C., BUXTON, M. & JONES, D. 1998. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*, 2, i-iv, 1-74.

FORD, F. & PLOWRIGHT, C. 2009. Realistic evaluation of the impact and outcomes of the Condition Management Pilots. London: Department of Health.

GATCHEL, R., POLATIN, P., MAYER, T., ROBINSON, R. & DERSH, J. 1998. Use of the SF-36 Health Status Survey with a Chronically Disabled Back Pain Population: Strengths and Limitations. *Journal of Occupational Rehabilitation*, 08, 237-246.

HARWOOD, R. & EBRAHIM, S. 1995. *Manual of the London Handicap Scale*, Nottingham, University of Nottingham.

HARWOOD, R. & EBRAHIM, S. 2000. A comparison of the responsiveness of the Nottingham extended activities of daily living scale, London handicap scale and SF-36. *Disabil Rehabil*, 22, 786-93.

HAYDAY, S., RICK, J., CARROLL, C., JAGGER, N. & HILLGAE, J. 2008. Review of the effectiveness and cost effectiveness of interventions, strategies and policies to help recipients of Incapacity Benefit return to employment (paid and unpaid): Research Document no 8. London: Institute of Employment Studies.

HOEGH, M. & HOEGH, S. 2009. Trans-adapting outcome measures in rehabilitation: Cross-cultural issues.

Neuropsychol Rehabil, 19, 955-70.

KERSTEN, P., MULLEE, M., SMITH, J., MCLELLAN, L. & GEORGE, S. 1999. Generic health status measures are unsuitable for measuring health status in severely disabled People. *Clinical Rehabilitation*, 219.

KING, R., LLOYD, C. & MEEHAN, T. (eds.) 2007. *Handbook of Psychosocial Rehabilitation*, Oxford: Blackwell Publishing Ltd.

LAI, S., PERERA, S., DUNCAN, P. & BODE, R. 2003. Physical and Social Functioning After Stroke: Comparison of the Stroke Impact Scale and Short Form-36. *Journal of the American Heart Association*, 34, 488-493.

LASCH, K., MARQUIS, P., VIGNEUX, M., ABETZ, L., ARNOULD, B., BAYLISS, M., CRAWFORD, B. & ROSA, K. 2010. PRO development: rigorous qualitative research as the crucial foundation. *Qual Life Res*, 19, 1087-96.

LAVIER-FAWCET, A. 2007. *Principles of Assessment and Outcome Measurement for Occupational Therapists and Physiotherapists: Theory, Skills and Application*, Chichester, Wiley Blackwell.

MARSHALL, S., HAYWOOD, K. & FITZPATRICK, R. 2006. Impact of patient-reported outcome measures on routine practice: a structured review. *Journal of Evaluation in Clinical Practice*, 12, 559-568.

MCHORNEY, C., HALEY, S. & WARE, J., JR. 1997. Evaluation of the MOS SF-36 Physical Functioning Scale (PF-10): II. Comparison of relative precision using Likert and Rasch scoring methods. *J Clin Epidemiol*, 50, 451-61.

MRC 2000. *A framework for development and evaluation of RCTs for complex interventions to improve health*. London: Medical Research Council.

NOURI, F. & LINCOLN, N. 1987. An extended activities of daily living scale for stroke patients. *Clinical*

Rehabilitation, 1, 301-305.

ONS 2009. Summary report on labour market statistics. London: Office of National Statistics.

OPPENHEIM, A. 2000. Questionnaire Design, Interviewing and Attitude Measurement, London, Pinter.

REPETTO, S. 2004. The role of self-efficacy, self-esteem and depression in job search efforts for women in welfare to work programs. PhD Thesis, University of Adelphi.

SAMEJIMA, F. 1969. Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*.

SCHERBAUM, C., COHEN-CHARASH, Y. & KERN, M. 2006. Measuring General Self-Efficacy: A Comparison of Three Measures Using Item Response Theory. *Educational and Psychological Measurement*, 66, 1047-1063.

SCHWARZER, R. & JERUSALEM, M. 1995. Generalized Self-Efficacy scale. In: WEINMAN, J., WRIGHT, S. & JOHNSTON, M. (eds.) *Measures in health psychology: A user's portfolio. Causal and control beliefs*. Windsor: NFER-Nelson.

SHIROM, A., VINOKUR, A. & PRICE, R. 2008. Self efficacy as a moderator of the effects of job search workshops on reemployment; a field study. *Journal of Applied Social Psychology*, 75, 853-863.

SIM, J. & WRIGHT, C. 2000. *Research in Health Care- Concepts, Designs and Methods*, Cheltenham, Stanley Thornes.

SITZIA, J. & WOOD, N. 1997. Patient satisfaction: A review of issues and concepts. *Social Science and Medicine*, 45, 1829-1843.

SPH. 2010. *Solutions for Public Health- CASP Tools* [Online]. London: Dept. of Health. Available: <http://www.sph.nhs.uk/what-we-do/public-health-workforce/resources/critical-appraisals-skills-programme> [Accessed November 23rd 2010].

STREINER, D. & NORMAN, G. 2008. *Health Measurement Scales, a practical guide to their development and use.*, Oxford, University Press.

VALDERAS, J., FERRER, M., MENDIVIL, J., GARIN, O., RAJMIL, L., HERDMAN, M. & ALONSO, J. 2008. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. *Value Health*, 11, 700-8.

WADELL, G. & BURTON, A. 2006. *Is work good for your health and well-being? An independent review of the scientific evidence for the Department of Work and Pensions*. London: The Stationary Office.

WAGHORN, G., CHANT, D. & KING, R. 2007. Work-related subjective experiences, work-related self-efficacy, and career learning among people with psychiatric disabilities. *American Journal of Psychiatric Rehabilitation*, 10.

WARE, J., SNOW, K., KOSINSKI, M. & GANDEK, B. 1993. *SF-36 Health Survey: Manual and Interpretation Guide*. In: INSTITUTE, T. H. (ed.). Boston: New England Medical Center.

WESTABY, J. & BRAITHWAITE, K. 2003. Specific factors underlying reemployment self-efficacy: Comparing control belief and motivational reason methods for the recently unemployed. *Journal of Applied Behavioral Science*, 39, 415-437.

WHO. 2001. *World Health Organization. International Classification of Functioning, Disability and Health* [Online]. Geneva: World Health Organisation. Available: <http://www.who.int/classifications/icf/en> [Accessed January 2011].

WILLIS, G. 2005. *Cognitive interviewing: A tool for improving questionnaire design*, Thousand Oaks, Sage.

WU, H.-S. & MCSWEENEY, M. 2004. Assessing fatigue in persons with cancer. *Cancer*, 101, 1685-1695.