The development and validation of a shortened version of the Eating Disorder Examination Questionnaire (EDE-Q)

Nicole Gideon

D.Clin.Psy. thesis (Volume 1), 2014

University College London

UCL Doctorate in Clinical Psychology

Thesis declaration form

I confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature:

Name: Nicole Gideon

Date: 19.06.2014

Overview

Volume 1 of this thesis describes the development and initial validation of a shortened version of the Eating Disorder Examination Questionnaire (EDE-Q) for use as a brief and sessional outcome measure in clinical practice. There are three parts to Volume 1.

Part 1 is a literature review on the benefits of providing continuous progress feedback in mental health treatment with a particular focus on different clinical settings. The review identified 14 studies in couple therapy, university counselling, psychiatric outpatient and inpatient services. The findings are discussed with consideration of the studies' quality ratings. However, due to methodological shortcomings and small number of studies it is difficult to draw firm conclusions.

Part 2 describes the multi-method development of the EDE-QS, a shortened version of the widely used EDE-Q, integrating principal component analysis, expert consultation and Rasch modelling. A subsequent online survey completed by people with and without eating disorders examined the new questionnaire's reliability, validity and sensitivity.

Part 3 is a critical appraisal of the research process with a particular focus on the methodological challenges experienced during the scale development process. The implications of using a symptom-specific assessment questionnaire for people with eating disorders are also discussed.

Table of Contents

List of Tables

Part 1: Literature Review

Part 2: Empirical Paper

List of Figures

Part 1: Literature Review

Part 2: Empirical Paper

Acknowledgements

I would like to acknowledge and thank my supervisors Lucy Serpell and Nick Hawkes for their ongoing support throughout this project. Lucy's practical advice and relaxed approach helped me complete this study despite the research and personal challenges I was faced with during my course. Nick helped me to access a large amount of EDE-Q data in his NHS service and provided me with a valuable and exciting placement experience.

Rob Saunders also deserves specific mentioning as he helped me to grapple with the mysteries of Rasch analysis. He was always approachable and extremely helpful.

I am particularly grateful for Jon Mond's help and support during this project. He quickly became an essential member of the research team and offered thoughtful and timely advice, despite the fact that he contributed from Australia.

I would also like to thank B-eat for their efficient help in recruiting participants for the online study and Kate Tchanturia for providing archival EDE-Q data.

My greatest thanks and humble apologies go to my daughter Mia and my partner Neil. I would not have been able to continue with the work and course demands if it was not for their continuous motivation, understanding and emotional support. Thank you both so much.

Part 1: Literature Review

Does feedback on progress improve service users' outcomes in psychotherapy?

Abstract

Aims: This review explores and evaluates the benefits of sessional progress monitoring and feedback on clients' outcomes in mental health treatment across clinical settings and outcome measurement systems.

Method: Fourteen studies were included and reviewed after searching PsychInfo, Embase, PubMed and Medline. The findings were presented for couple therapy, university counselling, psychiatric outpatient and inpatient services. The studies' quality was assessed and scored using the Downs and Black (1998) quality appraisal tool.

Results: Overall, the review showed that continuous progress feedback can improve clients' psychological outcomes, consistent with previous research. It is less clear whether the benefits apply to all clinical settings, as the findings were less conclusive for studies with more severely impaired client groups.

Conclusions: More high quality studies are needed to draw firm conclusions on the observed benefits. It is also essential to investigate the mechanism of change so that robust studies can be designed and evaluated.

Introduction

In recent years routine outcome measure collection in mental health services has gained increased attention. The British government actively promotes the use of outcome measurement and evaluation with the aim to improve people's quality of care (Department of Health, 2012). An outcome measures compendium was published that includes questionnaires covering a broad range of psychological difficulties to guide services and practitioners (National Institute for Mental Health in England, 2008). Through their application in routine practice, specific information can be collected repeatedly without the clinician having to engage in assessment and frequent re-assessment during limited therapeutic time (Lambert & Shimokawa, 2011). Several studies have demonstrated that clinicians are overly optimistic when they judge their patients therapeutic gains and are not able to accurately identify those that are not benefitting from therapy and deteriorating (Hatfield, McCullough, Frantz & Krieger, 2010; Lambert, 2012). Standardised questionnaires can therefore provide more objective feedback than a therapist's clinical intuition. These data are valuable to examine the gains made in mental health treatment overall but, more specifically, can also be used to inform therapists on clients' progress on a sessional basis.

Continuous progress feedback can assist in identifying any problems that may have been missed at the time of referral or highlight emerging difficulties that can subsequently be addressed and prioritised in treatment (Boyce & Browne, 2013; Valderas et al., 2008). This may influence a therapist's plan of action and impact on decision-making with regards to the future course of the intervention (Lambert & Shimokawa, 2011). It may also contribute to improved client- therapist communication. Valderas et al.

(2008) claim that a transparent approach of gaining a shared understanding of the presenting problems could improve adherence to treatment and patient satisfaction. More recently, research studies have examined the impact of reviewing sessional progress information in treatment on clients' outcomes and have found beneficial effects (Lambert, 2013). More specifically, this effect appears to be stronger and more consistent when selecting those people who do not respond to treatment as expected (Shimokawa, Lambert & Smart, 2010).

Despite the government's expectations and research highlighting the benefits of collection and feedback of routine outcome measurement, many clinicians do not recognise its value and clinical utility (Garland, Kruse & Aarons, 2003). Further, it was shown that less than 40% of psychologists (Hatfield & Ogles, 2004) and only a minority of psychiatrists (Gilbody, House & Sheldon, 2002) collect outcomes routinely as part of their clinical practice. Jensen-Doss and Hawley (2010) found that therapists need to perceive the feedback as useful, valid and as adding knowledge beyond their own clinical observations. As the majority of clinicians appear sceptical and are concerned about the time-burden involved in fully engaging in continuous progress monitoring (Garland et al., 2003), it is important to establish and evaluate the generalisabilty of any benefits. This is the aim of the current literature review.

Previous reviews

General health care services

There have been several reviews of outcome monitoring in health care settings, which provided mixed results.

Marshall, Haywood and Fitzpatrick (2006) reviewed the use of patient reported outcome measures across health settings and found that it specifically improved clinicians' detection of mental health difficulties and their diagnostic abilities. Although studies were included that focussed on general health, people with mental health difficulties showed greater improvement. On the contrary, Boyce and Browne's (2013) systematic review found limited support for improved outcomes when progress information was supplied to health care professionals. Only one of 16 reviewed studies demonstrated a statistically significant effect in favour of the feedback group. They however, acknowledged that the results were based on poor quality studies, which prevented them from drawing firm conclusions. Valderas et al.'s (2008) review in general health care identified 15 of 23 studies that showed benefits for the outcome monitoring and feedback group. However, methodological weakness of the reviewed research trials limited the perceived reliability of these results. Carlier et al. (2012) searched the literature up to 2009 and only included physical and mental health studies that had a randomised controlled design. Forty-five of the 52 included studies collected outcome measures about a patient's mental health status. They found that using outcome measures aided professionals in faster decision making with regards to a person's diagnosis and treatment adjustment. It also helped to improve the communication between the clinician and the patient. Ten studies were conducted in a specialist mental health setting. Of those 78% found a positive impact of using outcome measures on their mental health.

The majority of the included studies in these reviews obtained health status data as opposed to progress data. This means that outcome measures were often used as a one-off screening tool. Further, the included studies

were not exclusive to mental health services or compared measures of mental health for people who did not receive treatment or therapy for mental health problems specifically.

Mental health services

Knaup, Koesters, Schoefer, Becker and Puschner (2009) reviewed 12 studies that were conducted in specialist mental health care settings, conducted a meta-analysis and found that feedback of outcomes improved psychotherapy outcomes. More importantly, feedback was more effective if it was provided to patients as well as clinicians, if it was delivered at least twice and included information on treatment progress as opposed to status. They included studies up to March 2008 and found a statistically significant difference between feedback and no feedback groups; the effect size, however, was very small. Further statistical analyses did not support any lasting effects of its benefits. Although conducted within a mental health setting, in this review only half of the included studies shared outcome measure feedback on a weekly basis and three studies provided feedback only once throughout the course of therapy.

Shimokawa et al. (2010) conducted a meta-analytic and mega-analytic review of a specific progress monitoring system, the Outcome Questionnaire Psychotherapy Quality Management System (OQ system; Lambert & Shimokawa, 2011). This system generates computerised progress monitoring data based on the Outcome Questionnaire (OQ-45), which enquires about mental health symptoms, interpersonal problems and social functioning (Lambert et al., 2004). It provides clinicians with an expected trajectory and can therefore highlight those individuals who appear off-track. Those identified as off-track complete further questionnaires to obtain information to

guide clinicians' decision-making on how to adjust their treatment. These additional measures are referred to as Clinical Support Tools (CST; Lambert, 2012). Shimokawa et al. (2010) aggregated data from six of their studies and found that using all types of progress feedback (i.e. to therapists only, to clients and therapists, to therapist in addition to using CST) was beneficial for all clients. This was even more pronounced for those who had been identified as being "not on track" (NOT) and were therefore deteriorating. Feedback to therapists also reduced treatment failures. A limitation of this review is that all but one of the included studies were conducted with patients from a university population, which may have over-represented people with relatively mild clinical symptoms (Simon, Lambert, Harris, Busath & Vazqeuz, 2012). It is further problematic to generalise benefits of continuous outcome monitoring as the review was limited to one specific system, the OQ- system, and one specific outcome measure, the OQ-45. In addition, as the studies using the OQ-system have primarily been carried out by Lambert and others, it is likely that this may have introduced a researcher allegiance bias (Luborsky et al., 1999).

Lambert and Shimokawa's (2011) meta-analysis inspected two different outcome management systems and included nine studies. They compared outcomes of the OQ-system with the Partners for Change Outcome Management System (PCOMS; Miller, Duncan, Sorrell & Brown, 2005). PCOMS consists of two brief four-item rating scales, the Outcome Rating Scale (ORS), which enquires about mental health functioning, and the Session Rating Scale (SRS), which prompts the client to reflect on the therapeutic relationship (Miller & Duncan, 2004). Lambert and Shimokawa (2011) found that both systems improved patients' progress in the feedback conditions. It

showed slightly stronger improvement for those people who were identified as NOT clients and whose therapists applied additional strategies by using CSTs. Although this review emphasised the benefits made by the intervention groups, it is still unclear whether these findings generalise to other settings, systems and outcome measures. It also raises the question whether feedback on continuous progress monitoring is of additional benefit for people who are on a positive trajectory as opposed to being NOT (Lambert, 2013).

Goodman, McKay and DePhilippis (2013) reviewed the literature of outcome research in mental health more recently and found that feedback has positive effects on clients' progress. Although this was a more recent review, they did not use a systematic approach and no specific inclusion criteria for the reviewed studies. They also did not appraise the quality of the included studies and gave them equal weighting. They further specifically focussed on substance use treatments.

Aims of this review

Since Knaup et al.'s review in 2009 there has not been a systematic literature review that focussed on outcome measure feedback in mental health treatment that was inclusive of all outcome management systems and measures. With regards to the scepticism of clinical practitioners about the benefits of progress monitoring, it is of particular interest to investigate whether the observed benefits reported by Shimokawa et al. (2010) and Lambert and Shimokawa (2011) generalise to more severe and diverse clinical populations, who may present to general or inpatient mental health services. The current review will therefore include studies across all clinical populations and present their results separately. It is hoped that this may

provide individual practitioners and services with more conclusive evidence with regards to the benefits of setting up and implementing outcome management systems in different clinical settings.

This review therefore seeks to update the currently available evidence whilst assessing and considering the included studies' quality ratings.

To summarise, the current review aims:

1) to expand the existing evidence base by systematically searching for articles published after Knaup et al.'s review in 2009
2) to include a range of progress feedback systems and measures applied in mental health treatment
3) to use more defined inclusion and exclusion criteria of studies (i.e. include studies with continuous treatment progress feedback only)
4) to assess the quality of the individual studies using a quality appraisal tool
5) to evaluate the effectiveness of outcome feedback in diverse clinical populations and settings

Methods

Search strategy

In December 2013 PsychInfo, Embase, PubMed and Medline were systematically searched using the Ovid platform. To identify relevant articles, keywords were entered in each of the database as listed in Table 1 and combined through the 'OR' and 'AND' command.

The search was limited to include studies between 2008 and December 2013, those published in English and those from peer-reviewed journals. A

google scholar search was also performed, using the search terms "Routine outcome monitoring and feedback psychotherapy".

Table 1*: Database search terms*

Database	Keyterm 1	Keyterm 2	Keyterm 3
PsycInfo	Feedback.mp OR explode: feedback OR knowledge of results	Outcome*.mp OR explode: Psychotherapeutic Outcomes OR explode: Treatment Outcomes	Mental health.mp OR explode: mental health
Embase	Feedback.mp OR feedback system OR explode: feedback system	Outcome*.mp OR explode: treatment outcome OR explode: outcomes research	Mental health.mp OR explode: mental health
Medline	Feedback.mp OR explode: feedback	Outcome*.mp OR "Outcome and Process Assessment (Health Care)" OR Treatment Outcome OR "Outcome Assessment (Health Care)"	Mental health.mp OR explode: mental health OR explode: mental disorders
Combined search: Embase, Medline, PsycInfo	Feedback.mp AND Outcome*.mp AND Mental health.mp AND Psychotherapy.mp AND routine.mp AND monitoring.mp		

Inclusion/exclusion criteria

Studies were included for review if they fulfilled the following criteria:

Participants

- Any age
- Clinical population seeking mental health treatment for any mental health difficulty, including substance misuse
- Any severity of mental health problems

Types of intervention

- Any talking therapy/psychotherapy approach to address a mental health problem, including substance misuse

Outcomes

- Studies that include at least one standardised measure of psychological or psychosocial functioning

Study design

1) collected progress feedback continuously (i.e. sessionally or weekly)
2) used standardised tools of psychological functioning as the primary outcome measure
3) provided progress feedback of psychological or psychosocial functioning to therapists and/or their clients
4) compared experimental groups of giving feedback with no feedback
5) randomised as well as non-randomised controlled studies

A total of 1,151 articles were identified through an initial search (see *Figure 1*). Their titles and abstracts were screened and 31 were included for full text review using the inclusion/exclusion criteria outlined above. Of these, 14 studies were retained for the current review.

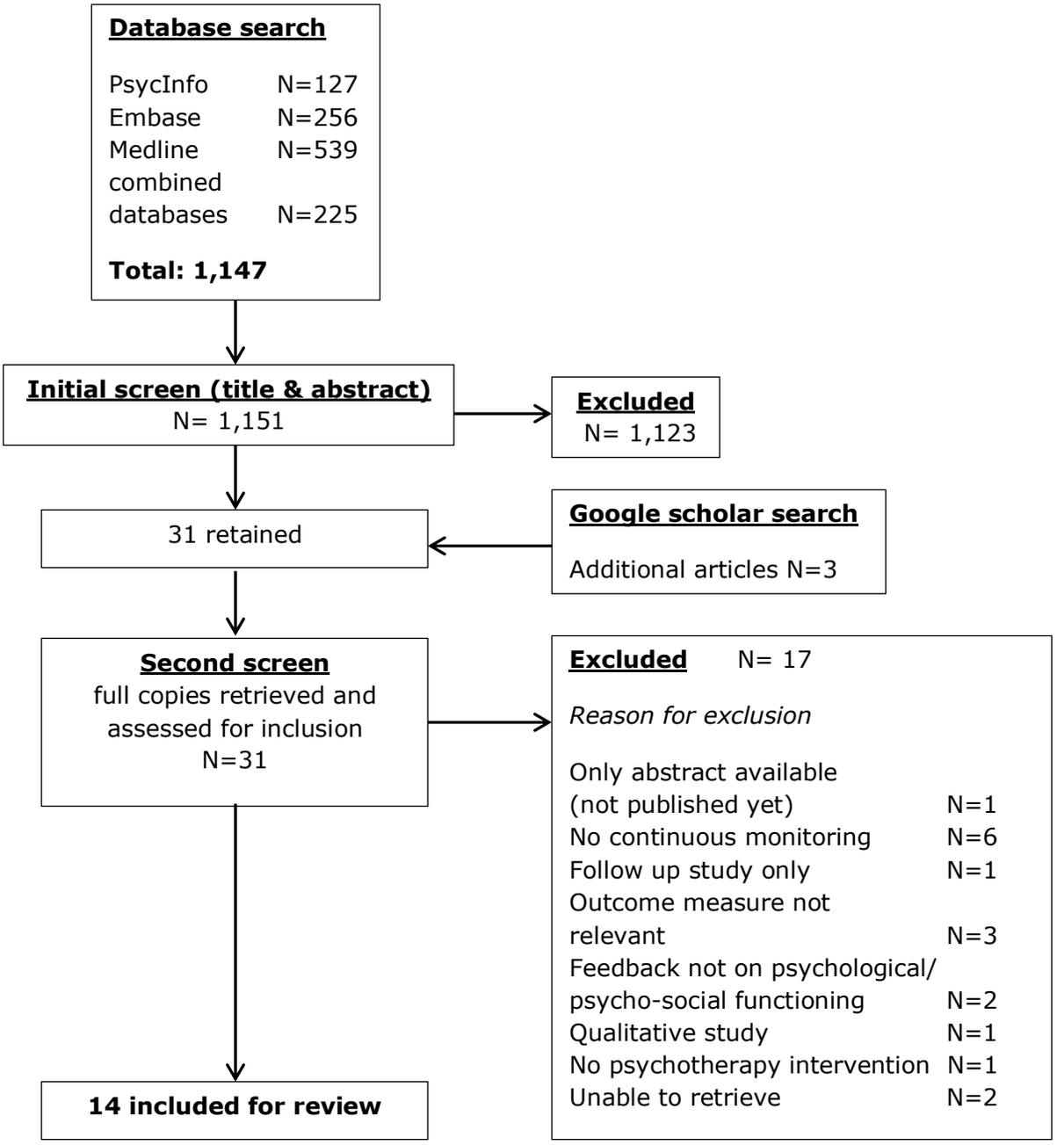

Figure 1: **Flowchart of search strategy**

Quality assessment method

Downs and Black's (1998) critical appraisal tool was used to assess the quality of the included studies (see Appendix A for the complete checklist and scoring). It has shown to have high internal consistency, good test-retest and inter-rater reliability (Downs & Black, 1998). It has also been recommended

to be used for the quality assessment of studies by West et al. (2002), who reviewed several appraisal tools. The checklist consists of 27 questions and it is possible to attain scores ranging from zero to 28 (Question 5 can achieve a score of 2). Only randomised controlled trials can achieve the maximum score.

One question was replaced to account for the nature of the review's research and one was adapted. Question 15 in the original appraisal tool enquires whether the authors attempted to blind researchers measuring the main outcomes of the intervention. As this is not relevant for studies that measure and discuss progress feedback in the therapeutic sessions, it was replaced with "Were therapists experienced professionals with regular caseloads", which had been included in a systematic review by Cahill, Barkham and Stiles (2010). The scoring to Question 27 was adapted and scored as met if the authors had conducted a power calculation and included a sample that was sufficiently powered to detect a significant effect.

Modelled on Hooper, Jutai, Strong & Russell-Minda's (2008) review, the studies were placed into quality rating categories based on the obtained scores (see Table 2).

Table 2*: Scores and quality ratings of Downs and Black's (1998) appraisal tool*

Score	**Quality Rating**
0-14	Poor
15-19	Fair
20-25	Good
26-28	Excellent

Results

Fifteen studies were included for review. Information on their characteristics is presented in Table 3.

Study setting

The majority of studies were conducted in the USA (8), two in Norway, two in Germany, one in Ireland and one in Sweden. The sample size ranged from 43 to 3,919 participants.

In order to examine the outcomes in different clinical populations, the studies were grouped into types of clinical settings and are presented accordingly. Most studies (5) were conducted in psychiatric outpatient or community services. Four took place in university counselling services and three in inpatient mental health services. Only two were carried out in couple therapy services.

Study population

The female-male ratio varied across studies, ranging from 44 to 100% of women. The majority of studies (12) had 50% or more females in their patient sample. Only one study included participants who were under 17 years of age, with a mean age of 15. For the other studies, patients' mean age varied between 20 to 41 years.

Patients presented with mood disorders and anxiety problems in six studies. Additional problems included personality disorder, substance use, adjustment problems, eating disorders, schizophrenia, somatoform disorders and relationship problems. Relationship difficulties were exclusively addressed in two studies. Only one study was included that treated people for substance misuse problems specifically and one that treated people with

eating disorders. Five of the reviewed studies did not specify the mental health difficulties their patients presented with.

Mental health treatment and duration

Participants received a wide range of mental health treatments, including elements of CBT, psychodynamic and systemic approaches, as well as integrative interventions of these. These were delivered by a range of health professionals, such as psychologists, psychiatrists, social workers, trainees and others. Most participants' treatments lasted on average between four and eight sessions or weeks. Three studies provided treatment for up to 13 sessions/weeks on average and only two studies delivered longer treatments, namely 16.5 and 18 sessions/weeks on average. One study did not specify the treatment length of their clients.

Study design

All but two studies employed randomised designs. Only Anker, Duncan and Sparks (2009) conducted a follow-up assessment at six months. The remaining studies took the last outcome measure at the point of discharge or at a pre-defined time point during treatment.

Table 3*: Characteristics of the included studies*

Author(s)	Country	Design; Intervention/ control group N	unit of randomisation	N	Female %	Age range (m)	Mental health difficulty	Treatment; mean sessions/weeks	Therapists (N)
couples therapy									
Anker et al. (2009)	Norway	RCT; 1/1	patient	410 (205 couples)	50	20-71 (37.8)	Relationship difficulties	eclectic (SF, narrative, CBT, humanistic & systemic); 4.6	Psychologists, SW, Psychiatric Nurse (10)
Reese et al. (2010)	USA	RCT; 1/1	patient	92 (46 couples)	50	19-56 (30.2)	Relationship & individual distress	systemic (SF, narrative/postmodern, and strategic therapy); 5.91	practicum students (13)
psychiatric outpatients & community services									
Bickman et al. (2011)	USA	RCT; 1/1 (28 sites)	service site	340	48.8	11-18 (14.8)	Unclear	various (CBT, integrative, behavioural, systemic, play therapy); 16.5	Unclear (144)
By Rise et al. (2012)	Norway	RCT; 1/1	patient	75	62.7	18-70 (25)	Unclear	psychotherapy, CBT, pharmacotherapy; 6	76% psychologists & 24% psychiatric nurses (25)
Crits-Christoph et al. (2012)	USA	2 phases, consecutive clients; 1/1 (3 sites)	none	116	44	Adults (39.6)	Alcohol or drug use	counselling for substance misuse; 12 or 6 months	Unclear (38)
Hansson et al. (2013)	Sweden	RCT; 1/1 (2 sites)	patient	374	73	Adults (39)	Mood & anxiety disorder, personality disorder, other	psychotherapy; 18.2	psychiatrists, mental health nurses & assistants, CP, SW, physiotherapists, OT (56)
Simon et al. (2012)	USA	randomised block design; 1/1	patient	370	64.2	Adults (36.1)	Mood disorder & anxiety, substance use	CBT, interpersonal, humanistic; 6.6	licensed psychologists (4), licensed SW (2)
university counselling services									
Murphy et al. (2012)	Ireland	RCT; 1/1	patient	110	58.2	18- 59 (23.8)	Anxiety, depression, relationship problems, other	constructivist, CBT, psychodynamic, integrative; 3.7	post-graduates in counselling psychology, psychotherapy, social work or family therapy (8)

Reese et al. (2009a) *study 1*	USA	randomised block design; 1/1	patient	74	71.6	18-27 (20.2)	Unclear	CBT, systemic, SF, or integrated/eclectic; 6	Master's level practitioners (5) & practicum students (5)
Reese et al. (2009a) *study 2*	USA	RCT; 1/1	clinicians	74	68.9	18-69 (33)	Unclear	CBT, systemic, SF, or integrated/eclectic; 6.9	practicum students (17)
Reese et al. (2009b)	USA	partially randomised; 1/1 (several sites)	clinicians	95	71.6	15-69 (34.1)	Unclear	psychotherapy; unclear, over 1 academic year	master's level trainees (28) and supervisors (9)
Slade et al. (2008)	USA	randomised quasi-experimental (incl. archival sample); 6/6/2	patient	3,919	62.2	17-58 (23.6)	Mood disorder, adjustment, anxiety, eating disorder, other	CBT, psychodynamic/interpersonal, humanistic/existential, behavioural, other; ranged from 8 to 12.1 for NOT clients	psychologists (28) & doctoral trainees (46)
inpatient services									
Probst et al. (2013)	Germany	RCT; 1/1	patient	43	55.8	Adults (41.2)	Depression, somatoform disorders, anxiety	unclear; 5.9	unclear (17)
Puschner et al. (2009)	Germany	RCT; 1/1	clinicians	294	47.3	Adults (41.2)	Affective disorders, schizophrenia, other	unclear, MDT appointments; 8	psychiatry residents (30), specialist registrars (8), psychotherapists (5), other (1)
Simon et al. (2013)	USA	randomised block design; 1/1	patient	133	100	17-54 (25.5)	Eating disorders	cognitive and/or dialectic behavioural, integrative (client-centred, interpersonal, systemic); 12.6	licensed psychologists (6), licensed SW (7), marriage and family therapists (3)

RCT: randomised controlled trial
SF: solution-focussed
CBT: cognitive-behavioural therapy
CP: clinical psychologists
SW: social worker
OP: occupational therapist
MDT: multi-disciplinary team

Methods of progress monitoring

The feedback systems and outcome measures that were used are summarised in Table 4. All of the studies obtained sessional or weekly progress information, which was completed by the clients. One study also collected outcome information from the therapists and carers. Five studies supplied progress feedback to therapists only but in three of these, therapists were either encouraged to share this information with their clients or they could freely decide whether they wanted to do this. The feedback always included a progress chart on the previous and current outcome scores. Additional features included information on an expected trajectory, feedback messages and colour-coded warning signals. The majority of studies highlighted to clinicians if a patient was at risk of treatment failure or not progressing as expected. Five of these studies provided therapists with specific treatment guidance based on additionally collected data from the CST.

The ORS and/or SRS were used as the feedback measure of choice in six of the included studies. Half of the studies administered the OQ-45 in its original or translated form. Only one study used a different feedback questionnaire.

Therapists in nine studies received training on how to use and interpret the results of the feedback measures.

Study outcomes and quality ratings

The obtained results and quality ratings of the studies are summarised in Table 5.

Overall, 11 of the included studies reported more positive outcomes when continuous progress feedback was provided; however, two of these studies did not reach statistical significance and two analysed data for people

who were identified as NOT only. Three studies did not report any observed benefits of adding progress monitoring and feedback to mental health interventions. There was a mixed range of study quality across clinical settings. None of the included studies obtained a quality rating of "excellent". However, five were rated as "good", and the remaining "fair" or "poor".

In the following, the reviewed studies are briefly presented individually and results are summarised in view of their quality ratings and their study setting.

Table 4*: Summary of progress feedback system*

Authors	Feedback measure	risk clients alert	CST	Completed by	Feedback to	Timing of feedback	Features of Feedback	therapist training time; content
couples therapy (outpatient)								
Anker et al. (2009)	ORS & SRS	yes	no	client	therapist & client	immediate	total score, progress chart, expected treatment response table	8 hr & 9 hr follow-up; instructed to follow manual
Reese et al. (2010)	ORS & SRS	no	no	client	therapist & client	immediate	progress chart	
psychiatric outpatients & community services								
Bickman et al. (2011)	SFSS; others	yes	no	client, therapist, carer	therapist	delayed (med: 9 days)	progress chart	unclear how many hours; regularly scheduled (at least monthly) group teleconferences
By Rise et al. (2012)	ORS & SRS	no	no	client	therapist & client	immediate	progress chart	2 days (12 hrs)
Crits-Christoph (2012)	OQ-45 adjusted	yes	yes	client	therapist	immediate	progress chart, drug & alcohol use scores, colour-coded progress information	hours unclear; orientation and trained in interpretation of feedback reports
Hansson et al. (2013)	OQ-45 Swedish version	yes	no	client	therapist & client	unclear - assumed delayed	therapist: total score, subscale scores and progress chart; patients: progress chart	hrs unclear; group training sessions also individual support
Simon et al. (2012)	OQ-45	yes	yes	client	therapist & client	delayed (unknown)	colour-coded progress information; progress report; decision-tree for problem solving (clinician)	
university counselling services								
Murphy et al. (2012)	ORS	yes	no	client	therapist & client	immediate	progress chart and predicted progress	hrs unclear; read chapter & used PCOMS for one year prior to study
Reese et al. (2009a) *study 1*	ORS & SRS	no	no	client	therapist & client	immediate	progress chart	1hr; summary hand out

Reese et al. (2009a) *study 2*	ORS & SRS	no	no	client	therapist & client	immediate	progress chart	1hr; summary hand out
Reese et al. (2009b)	ORS & SRS	no	no	client	therapist & client	immediate	progress chart	
Slade et al. (2008)	OQ-45	yes	yes	client	therapist (encouraged to share with clients)	immediate & delayed	progress chart, feedback message; clinicians only: suicidality alert if applicable	1hr
inpatient services								
Probst et al. (2013)	OQ-45 German version	yes	yes	client	therapist (free to share with clients)	delayed	colour-coded warning signals for NOT patients; feedback reports; recovery curve	
Puschner et al. (2009)	OQ-45 German version	no	no	client	therapist & client	delayed (1 or 2 days)	progress chart; written summary & recommendations; suicidality alert if applicable; clinician only: colour coded change information & change and status summary	
Simon et al. (2013)	OQ-45	yes	yes	client	therapist (encouraged to share with clients)	unknown	colour-coded progress information; progress chart; written message; decision-tree for problem solving	hrs unclear; provided with rationale & positive impact of using system; how to access IT system

CST: Clinical Support Tool
ORS: Outcome Rating Scale
SRS: Session Rating Scale
SFSS: Symptoms and Functioning Severity Scale
OQ-45: Outcome Questionnaire - 45

Table 5: *Summary of study outcomes & quality ratings*

Authors	outcome measure	outcome at discharge	Effect size	quality score	quality rating
couples therapy					
Anker et al. (2009)	ORS, LW (follow up only)	improved outcomes with feedback; fewer at risk with feedback; more people achieved reliable and clin sign change *at 6 mths follow up* ORS: improved outcomes with feedback; LW: no difference	d = .5 *follow up:* ORS: d= .44	23	Good
Reese et al. (2010)	ORS	improved outcomes with feedback; faster improvement; more people achieved reliable and clin sign change	d = .54	17	Fair
psychiatric outpatients & community services					
Bickman et al. (2011)	SFSS	Faster improvement with feedback as rated by youths, clinicians & carers	Youths: .18 Clinicians: .24 Carers: .27	18	Fair
By Rise et al. (2012)	TAS, CSQ, BASIS-32, PAM, SF-12 - MCS, SF-12 - PCS, ORS, SRS, PM, PP	improved motivation for treatment with feedback but no differences for alliance, satisfaction, mental health symptoms, quality of life & patient participation		22	Good
Crits-Christoph et al. (2012)	OQ-45 adjusted	*Subgroup analysis only for NOT clients* improved outcomes with feedback on alcohol use & from becoming NOT on drug use & OQ-45	Alcohol: d=.26 *from point of NOT:* drugs: d=.38	18	Fair
Hansson et al. (2013)	OQ-45	improved outcomes with feedback but n.s. (ITT: p = .06; PPA: p = .08) *subgroup analysis:* no differences	ITT: g = .21 PPA: g = .24	23	Good
Simon et al. (2012)	OQ-45	no differences *subgroup analysis for NOT clients* more improvement with feedback	*NOT*: d=0.12	14	Poor
university counselling services					
Murphy et al. (2012)	ORS	improved outcomes with feedback but n.s.; no differences on reliable change *subgroup analysis for NOT clients* no differences *post hoc analysis* improved outcomes with feedback for people with anxiety	d=.21 *anxiety*: d=0.6	21	Good
Reese et al. (2009a) *study 1*	ORS	improved outcomes with feedback; more people achieved reliable change with feedback	d= .54	16	Fair

Reese et al. (2009a) *study 2*	ORS	improved outcomes with feedback; more people achieved reliable change with feedback *subgroup analysis for NOT clients* improved outcomes with feedback but n. s. (no p-value given)	d= .49 *subgroup:* d = .07	17	Fair
Reese et al. (2009b)	ORS	more improvement with feedback	eta sq= .07	14	Poor
Slade et al. (2008)	OQ-45	more improvement with feedback; no differences between immediate or delayed feedback; no added benefits for also giving feedback to clients (vs therapists only) *subgroup analysis for NOT clients* as above; faster improvement if feedback is immediate; more improvement if CST used and faster improvement if delivered within one week (vs two weeks).		14	Poor
inpatient services					
Probst et al. (2013)	OQ-45 German version	*Subgroup analysis only for NOT clients* more improvement with feedback; fewer reliably deteriorated patients with feedback	d=.54	21	Good
Puschner et al. (2009)	OQ-45 German version	no differences		21	Good
Simon et al. (2013)	OQ-45; BMI	more improvement with feedback and more people achieved clin sign change; more people achieved reliable change in control group; no differences in BMI	d =.3	19	Fair

ORS: Outcome Rating Scale
LW: Locke-Wallace Martial Adjustment Test
TAS: Treatment Alliance Scale
CSQ: Client Satisfaction Questionnaire
BASIS-32: Behaviour and Symptom Identification Scale 32
PAM: Patient Activation Measure
SF-12: Short Form -12v2, a measure of health-related quality of life
MCS: Mental Component Score
PCS: Physical Component Score
SRS: Session Rating Scale
PM: Patient Motivation
PP: Patient Participation
ITT: Intent to treat
PPA: Per protocol analysis
OQ-45: Outcome Questionnaire – 45
NOT: Not on track
CST: Clinical Support Tool
BMI: Body Mass Index

Couples therapy

Two studies were carried out in couple therapy services.

Anker et al. (2009) conducted a randomised study in a naturalistic couple therapy setting. They allocated 205 White Euro-Scandinavian heterosexual couples (410 individuals) to a feedback or control condition. Therapists were allocated 50% of couples from the feedback and the control group. All participants were blind to their allocated group and completed the same outcome measures, the brief ORS scale. Only those in the feedback condition, however, discussed their scores and progress with their therapist. Therapists received at total of 17 hours of training on how to deal with unexpected treatment progress although adherence to this was not monitored. There were no apparent differences between the experimental groups after randomisation. They found that 41.7% of individuals (both in couple, 22.6%) in the control group benefitted from treatment whereas 64.6% of people (both in couple, 50.5%) improved significantly or clinically in the feedback group. They ran a follow-up investigation after six months and found that the beneficial effect remained albeit slightly diminished (39.1% (both in couple, 18.8%) in control vs 66.7% (both in couple, 47.6%) in feedback). Anker et al. (2009) also found that the feedback condition appeared to have a preventative effect in that less people presented as "at risk" clients (54.4% vs 74.5%).

This was a well-designed study, obtaining a rating of "good" on the Downs and Black (1998) appraisal tool. It had strong external validity and a strength of this study was that patients were blinded to their allocated condition. This was the only study that investigated benefits at

six months post-treatment, although they did not comment on or account for the high attrition rate (over 50%) which may have influenced the results. This study added sound evidence of progress feedback improving outcomes in couple therapy.

Reese, Toland, Slone and Norsworthy (2010) also investigated the effect of feedback in a couple therapy setting. Their aim was to replicate Anker et al.'s (2009) study in the USA. They randomised 46 heterosexual couples to 13 graduate trainee couple therapists and found that, after controlling for pre-treatment ORS scores, people in the feedback condition improved on average by 4.44 points more. They obtained a moderate standardised effect size of .54. They further showed that couples in the feedback condition improved more quickly and more couples and individuals achieved reliable and clinically significant change. For those people who had pre-ORS scores below 25 points, 53.3% achieved clinical significance in the feedback condition as compared to 18.2% of couples in the control condition.

Based on the scoring obtained from the Downs and Black (1998) quality appraisal tool, this study achieved a quality rating of "fair". Their recruitment process was unclear and the randomisation process to different experimental groups was neither mentioned nor described. Although the pre-treatment intake ORS scores were mentioned between the two groups, any other patient/couple characteristics were not further investigated. In addition, the interventions were carried out by trainees only, which may make it difficult to generalise the findings to routine care.

To summarise, the results obtained from research carried out in couple therapy settings appear to be promising. Both studies have found improved outcomes in the progress feedback groups with moderate effect sizes. However, it has to be acknowledged that this review identified only two studies, which limits the generalisabilty of their results. Further, one study was of low quality, which may impact on the reliability of the obtained findings.

Psychiatric outpatients and community services

This review identified five studies that were conducted in psychiatric outpatient or community services.

Bickman, Kelley, Breda, de Andrade & Riemer (2011) used a cluster randomised design to assess whether weekly feedback to clinicians from a private service improves the effectiveness of mental health treatment of youths living in community settings. Twenty-eight community youth service sites across the United States participated in this study. Sessional outcome measures were completed by 144 clinicians, 340 youths and 383 carers. There were more black youths in the feedback group but this was controlled for in the analysis. It was unclear what mental health issues the young people were suffering from or what therapeutic training or qualifications were held by the therapists. They found that clinicians, youths and carers reported that youths in the feedback group showed greater improvement. This effect was strongest for those youths whose therapists had viewed their feedback at least once.

The study scored "fair" on the quality appraisal tool as there were several methodological and reporting problems. There was no detailed information on the recruitment procedure of the participants and no attrition rates or figures were mentioned. It was reported that some clinicians only participated with some clients but no further details were given on this. Also, one third of the clinicians in the feedback group did not view any feedback even though this was part of the study protocol. It is unclear what effect this had on the delivery of feedback information.

By Rise, Eriksen, Grimstad and Steinsbekk (2012) were particularly interested in the effect of therapy progress feedback on client satisfaction and therapeutic alliance when using sessional ORS and SRS data. They also collected measures of psychological functioning at baseline and at six weeks after starting treatment, which meant that they met the inclusion criteria for this review. By Rise et al. (2012) did not find any differences in psychological outcome measures between the feedback and control condition. Treatment motivation was the only construct that showed increased scores for people receiving feedback. The authors concluded that their study sample was relatively small (N=75) and that a larger study with a longer follow-up may have shown statistically significant findings. However, this was one of the few studies included in this review that had conducted an a-priori power analysis and recruited the required sample size.

This was a well-designed and reported study and attained a quality rating of "good".

Crits-Christoph et al. (2012) compared the effect of progress monitoring between a feedback and a control group across four sites in the United States, which included a total of 304 patients who sought individual counselling treatment for substance misuse. There were two phases of data collection; the first included 165 clients who completed a modified OQ-45 without their therapist receiving feedback; the second phase gave feedback to therapists on 139 OQ-45 measures. The OQ-45 was adjusted by adding two items enquiring about a person's alcohol and drug use over the past week. Crits-Christoph et al. (2012) carried out analyses on a subgroup sample only, on people who were identified as not on track (NOT) by the OQ-45 (38.8% in feedback and 37.6% in control condition). Comparing baseline measures to the last session, it was found that the NOT patients in the control condition showed only little improvement in their alcohol use over time which was in contrast to the greater improvements for NOT individuals in the feedback condition. There were similar trends with regards to drug use but this difference did not reach statistical significance. The authors also examined differences between the groups from the point that clients went off-track. Once again, the feedback group improved more on measures on psychological functioning and drug use towards the end of treatment. They therefore concluded that the provision of feedback helped the NOT patients to get back on track. There were differences between feedback effects across the sites and the researchers suggested that these may have been due to greater familiarity with the clinical support tools at one particular site.

This study obtained a quality rating of "fair". The biggest drawback of the study was that it did not have a randomised design and the two

phases occurred consecutively. Although this study demonstrated improvements with progress feedback, these were only investigated in a subgroup sample.

Hansson, Rundberg, Oesterling, Oejehagen & Berglund (2013) assessed the differences between therapists and clients receiving progress feedback with a treatment as usual condition in two psychiatric outpatient clinics in Sweden. They had the largest study sample within this clinical setting (N=374) and the study patients attended on average the most therapy sessions (18.2). They found that greater treatment gains were made in the feedback group although the difference only showed a trend towards statistical significance (ITT: p=.06; PPA: p=.08). They further investigated possible benefits of feedback for people who were identified as NOT (27% in feedback group and 28% in control group). This however did not result in any additional benefits.

Overall, this was a well-reported and designed study and received a quality rating of "good".

Simon et al. (2012) wanted to explore whether their previous research findings from university counselling students extended to a more clinical population. Their randomised controlled trial recruited psychiatric patients from a hospital-based outpatient service in the USA and allocated them to a feedback or control group. The length of the intervention period was unclear. It was also not reported how frequently feedback was provided to therapists, although it appeared that the feedback was not immediate. Simon et al. (2012) did not find any differences in OQ-45 outcome measures between on-track clients at post-

treatment. A subgroup analysis, however, found that people who were identified as NOT (*N*=207) improved twice as much in the feedback group when their therapists used clinical support tools (CST), albeit with a small effect size (*d*=.12).

This study obtained a quality rating of "poor". This was mainly based on reporting omissions. Internal validity was potentially compromised as the participants' demographic characteristics, apart from OQ-45 intake scores, were not investigated between groups to rule out any confounding factors. The number of hypotheses to be tested was not clearly stated at the outset of the article and it was therefore unclear whether the conducted analyses were all within the initial study plans.

To summarise, the reviewed studies conducted in psychiatric outpatient and community settings provided mixed results on the outcomes of progress feedback. The studies' quality ratings varied from good to poor. One study found statistical support for improved outcomes; however, it had a fair quality rating and low effect sizes. One study of good quality found a positive trend but this did not reach statistical significance. The other high quality study did not observe differences between the intervention groups. This was consistent with another study, which, however, proved to be of poor quality. Two studies of lower quality found support for improved outcomes for people receiving feedback who were identified as NOT. The findings from this clinical setting are therefore inconclusive.

University counselling services

This review identified four studies that were conducted in university counselling services.

Murphy, Rashleigh and Timulak (2012) designed a randomised controlled study in an Irish university counselling centre, which provided therapy for a college population with the main presenting problems of depression, anxiety and relationships. The ORS was completed in both treatment groups. In the feedback condition, 59 students and their therapists received outcome feedback, whereas in the control group 51 students never reviewed their progress scores with their therapists. They found that both groups had improved ORS scores at the end of therapy but there were no significant differences between them. The effect size of the feedback group was d=.85 and d=.64 in the control. The authors therefore suggested that individuals in the feedback group improved more but the difference was not statistically significant (p=.23), potentially because of the small sample size. A post-hoc power analysis revealed that a sample size of 786 clients would have been needed to have an 80% chance to detect an effect of 0.2. There were also no differences between groups across the categories of reliably changed, deteriorated and no change (based on the reliable change index). The authors further explored differences in treatment progress for those clients who were at risk of treatment failure and did not find a significant difference. Post-hoc analysis revealed that feedback improved outcomes for 60% of clients presenting with anxiety problems, which was in contrast to 17.7% of people in the control group.

This was a well-designed and reported study which achieved a quality rating of "good". It showed high internal validity and the attrition rates at all phases of the study were made explicit. They however did not present information on each group's demographics although they claimed that statistical tests did not show any significant differences between them.

Reese, Norsworthy and Rowlands (2009a) conducted two studies at two different sites on the same university campus that utilised PCOMS to investigate whether feedback would increase treatment gains on ORS scores. They included people who received individual therapy at a university counselling centre and at a community-based graduate training clinic for marriage and family therapy. The two studies differed in that people in the control group at the university counselling centre did not complete any ORS measure whereas those of the graduate training clinic did so before every session but their therapists would not be informed about the scores and would not discuss these during their appointment. Both studies found that people in the feedback group had better post-treatment scores with a medium to large effect size (d=.54 and d=.49). It was also demonstrated that feedback helped more people to achieve reliable change by the end of treatment in both studies. As study 2 had collected sessional ORS scores in the experimental as well as in the control group, the researchers were able to compare the effect of feedback for those people who were identified as not progressing after three sessions. Sharing and discussing progress feedback helped these individuals to make more treatment gains overall, although this difference was not statistically significant. The authors explained that

this may be due to the low sample size. They further showed that people in the feedback condition achieved reliable change more quickly and in addition found that in study 1, people treated by qualified professionals had improved more quickly than those seen by staff in training.

Overall this study showed that all clients benefitted from implementing an outcome measure feedback system. However, Reese et al. (2009a) scored "fair" on the Downs and Black (1998) quality appraisal tool. There were limitations in reporting and external validity was reduced as their recruitment procedure was unclear. Apart from ORS intake scores, potential confounding factors between groups were not investigated. It also appeared that additional data analyses were carried out but it is not clear whether post-hoc procedures were used.

Reese et al. (2009b) investigated the effect of using outcome measure feedback in trainee supervision on clients' progress and trainees' ratings of their supervisory relationship. They used the PCOMS system across several sites. Therapy trainees in the no-feedback condition collected ORS data at each session but did not discuss the content with clients. They demonstrated that trainees whose supervisors structured their sessions and prioritised cased discussions based on PCOMS data had better client outcomes.

Reese et al. (2009b) achieved a quality rating of "poor" due to several methodological shortcomings. The authors explored differences on client outcomes between the control and feedback group at one study site but did not provide any information on the sample size, client demographics or type of therapy. They further showed inadequate

statistical reporting and the randomisation procedures were not consistent across the different study sites.

Slade, Lambert, Harmon, Smart and Bailey (2008) investigated the impact of progress feedback in a large sample of 3,919 individuals. They investigated immediate against delayed OQ-45 feedback and providing information to therapists only against giving it to both therapists and their clients. They compared the immediate feedback group to two archival groups: weekly feedback and treatment as usual. They further carried out several subgroup analyses for people who were identified as NOT. They found that feedback improved people's outcome. There were no added benefits for supplying feedback immediately (vs delaying it) and directly to the client (vs to the therapist only). For people who were identified as being at risk of treatment failure, similar outcomes were observed. The use of CSTs showed an added benefit on treatment outcomes. In addition, they demonstrated that for this group, improvement was quicker if feedback and the use of CST were provided more timely.

This study was rated as "poor" with regards to quality. Although it provided them with a large sample size, a major limitation was the use of an archival group. They also did not provide information on the samples' demographics and apart from baseline scores potential differences between groups were not explored.

To summarise, the reviewed studies in university counselling services provided evidence for the additional benefits of using progress feedback for people seeking help in university counselling centres.

However, the majority of studies were of fair quality. The only study of good quality did not report any differences between feedback and treatment as usual groups. The positive results obtained in university settings therefore have to be interpreted with caution.

Inpatient services

A search of the literature identified three studies that explored the effect of progress feedback in inpatient settings.

Puschner, Schoefer, Knaup and Becker (2009) explored differences on outcomes between feedback on the OQ-45 (German version) and a control group in an in-patient unit at a psychiatric hospital in Germany. They randomised 48 psychiatric professionals to either feedback or control groups to avoid cross-contamination effects between these groups. The sample consisted of 294 patients, who suffered mainly from affective disorders and a third from schizophrenia. The OQ-45 intake scores of 75.5% of the sample were in the dysfunctional range, representing a more severely impaired population than many other studies. Patients were under the care of a multi-disciplinary team and had sessions with several health professionals. Interestingly, only one of the professionals received the client's progress information (psychiatry residents (14), specialist registrars (5) and psychotherapists (2)). No significant differences were found between the groups when comparing intake and discharge outcome measures after an intention to treat and per protocol analysis. Patients received progress feedback directly but a post-study survey revealed that feedback was rarely discussed in their therapy sessions. Clinicians also indicated that they "disliked treatment recommendations" based on the OQ-45 data.

This study was well-conducted and reported and therefore rated as "good" on the Downs and Black (1998) appraisal tool.

Probst et al. (2013) randomised inpatients at two psychosomatic clinics in Germany to an OQ-45 feedback group and to treatment as usual. In contrast to previous studies, all participants regardless of progress trajectory completed the Assessment of Signal Cases (ASC), which is a CST instrument, every week. They investigated the effect of feedback and CST on those patients only who were identified as being NOT (*N*= 43, 17.1%). It was found that people who were NOT showed significant improvement in OQ-45 scores in the feedback condition. This group also had fewer numbers of people who reliably deteriorated.

This study was rated as "good" on the Downs and Black (1998) appraisal tool. There were some limitations with regards to external validity and the representativeness of the included sample.

Simon et al. (2013) carried out a randomised controlled study in an inpatient eating disorders service using the OQ-45 questionnaire. They randomly allocated patients to a feedback group, including CST for the therapist, or treatment as usual. All the therapists at this service participated in the study and were unfamiliar with the OQ-45 system. They showed that the feedback group improved more on OQ-45 scores post-treatment than the control group (*d*=.3). Although they found a trend suggesting that people diagnosed with bulimia nervosa benefitted less than those diagnosed with anorexia nervosa or eating disorders not otherwise specified, this was not statistically significant. Interestingly, more people achieved clinical recovery in the feedback group but more

people achieved reliable change in the control condition. Combining both frequencies, 75% of people in the feedback group reliably improved or recovered compared to 68.3% of people in the control group.

The study was rated as "fair" on the quality assessment tool. The study's sampling procedures were unclear, which may have introduced sampling bias. They did not clearly state the hypotheses to be tested so that it was impossible to determine whether there was a need to apply any post-hoc adjustments (e.g. analysis on BMI and diagnostic differences).

To summarise, two of the three included studies showed evidence that progress monitoring improves patient outcomes. However, one of these focussed on people who were NOT only and it is unclear whether the whole sample benefitted from receiving feedback. On the contrary, one good quality study did not support this finding. Due to the mixed results and limited number of studies, it is impossible to draw firm conclusions regarding the effectiveness of feedback for this client group.

Discussion

This review set out to expand the current evidence base on the benefits of monitoring and providing therapy progress feedback across a range of mental health settings and clinical populations.

In general, this review adds to previous research findings that continuous progress monitoring and feedback can improve patient outcomes (Goodman et al., 2013; Knaup et al., 2009). Of the 15 reviewed studies, nine found beneficial effects of using progress feedback

systems in mental health treatment, whilst a further two appeared to show benefits but the results were not statistically significant. The studies' effect sizes, where reported, exceeded those obtained by Knaup et al.'s (2009) meta-analysis. Of these studies, five also explored outcomes for people who were identified as not-on-track (NOT) and two others investigated this subgroup exclusively. Significant benefits of progress feedback were reported by four of these and one found a non-significant positive trend. However, two studies could not support this finding which is contrary to previous research (Shimokawa et al., 2010). It is important to acknowledge that the studies that found a positive effect were of moderate to low quality, whereas those that found opposing results were of high quality. It therefore needs to be considered that the positive outcomes may have been subject to study biases. The findings of the current review are therefore less conclusive than previous studies for the subgroup of NOT patients (Lambert & Shimokawa, 2011; Shimokawa et al., 2010).

The review's specific aim was to explore the outcomes in different clinical settings. Studies conducted in *couple therapy* settings showed benefits of progress monitoring. However, there were only two studies included for review and only one was of high quality. Although these findings are promising for the field of couple therapy, more studies are needed to strengthen these findings.

In *outpatient psychiatric or community settings*, the evidence was more mixed. Two studies found that outcomes improved with progress feedback whereas another two studies demonstrated the opposite. Both results were demonstrated by studies of similar quality, which means that

the findings should be given equal weighting and therefore remain inconclusive. The effects of progress feedback exclusively for people identified as NOT were investigated by one study. This showed improvement for the feedback group. However, this study's quality was low and more studies are needed to confirm the results.

All studies investigating progress monitoring in *university counselling centres* found that it had a positive effect on outcomes. However, the evidence for this needs to be interpreted with caution as the majority of the included studies were of low quality (3 out of 4). These findings are consistent with Shimokawa et al.'s (2010) meta-analytic and mega-analytic review. However, further subgroup analysis showed inconsistent results with regards to the added benefits of feedback to NOT clients, which is contrary to Shimokawa et al.'s (2010) and Lambert and Shimokawa's (2011) findings. More high quality studies are needed to strengthen the existing evidence base for the effectiveness of progress feedback on client outcomes in university counselling centres.

The studies from *inpatient settings* presented mixed and inconclusive findings. Positive outcomes were achieved in the feedback group by one study and for NOT patients in another. However, the study that did not find evidence of a feedback effect was of high quality and should receive stronger weighting. Boyce and Browne (2013) suggested that people with more severe impairment might benefit more from progress feedback as there may be "more room for improvement". This idea was, however, not supported by the current review. Due to the inconclusive results and the small number of included studies, it remains

unclear whether progress feedback systems are of increased benefits for people with more severe mental health difficulties.

The findings from this review may suggest that not all patients and clinical settings benefit from progress feedback systems. However, this cannot be concluded with certainty as the mechanisms responsible for change are as of yet unclear (Goodman et al., 2013; Knaup et al., 2009). It is impossible to assess whether the studies included in this review have implemented and delivered the progress feedback systems appropriately and successfully. Although the majority of studies (*N*=12) provided feedback to both therapists and clients or encouraged the therapists to discuss progress scores with their clients, it is unclear whether this was done and if so, how it was discussed. Interestingly, one study surveyed patient and clinician feedback following the intervention (Puschner et al., 2009). In this particular study, which was of high quality but did not find positive outcomes of progress feedback, it revealed that patients received information directly but that it was rarely discussed in their therapy sessions. If discussion of progress feedback in treatment was an active ingredient for change, this could have impacted on the intervention's effectiveness. The same post-study survey highlighted the scepticism of some clinicians, who mentioned that they "disliked the treatment recommendations", which is likely to have influenced their therapy adjustments following feedback. Experienced clinicians in particular may have a sceptical stance towards standardised questionnaire data and resist adjusting their practice if the data challenge their clinical and professional intuition (Bickman, 2008), despite a weak correlation between clinical competence and outcomes (Barber, Sharpless,

Klostermann & McCarthy, 2007). Anker et al.'s (2009) study showed that outcomes improved more for clients of less effective therapists, which suggests that they benefitted most from receiving progress information. Reese et al. (2009b) also found that using feedback in supervision improved patient outcomes for trainees. These studies raise the interesting question of how progress information is viewed, appraised and used to make adjustment to therapeutic work. Future research should therefore focus on the drivers and mechanism for change in progress feedback, which may help the quality assessment of future reviews.

It is also striking that all but one study used the PCOMS and OQ-45 management systems to monitor and deliver outcome information. It would be interesting to assess whether other measures may be equally suitable for the use of continuous progress feedback. This further raises the question whether problem-specific questionnaires could be even more suitable or useful in specialist services, which provide treatment for distinct mental health problems, such as psychotic or eating disorder symptoms. Although the OQ-45 has shown to be sufficiently sensitive to change in a more mildly impaired university student sample (Vermeersch et al., 2004), it may not capture change in more severely impaired population with specific difficulties. This could have contributed to the inconclusive results of the more severely impaired populations.

The most obvious difference between the PCOMS and OQ-45 system is their questionnaires' item-length (45 vs 8 items). Despite this, they have both been shown to be similarly effective in enhancing treatment outcomes (Lambert & Shimokawa, 2011). Shorter outcome questionnaires are less burdensome on clients as well as therapists.

Future research should focus on evaluating brief and symptom-specific measures for their sensitivity to change and suitability as a progress monitoring tool. This would increase choice in outcome tools and possibly enhance the perceived usefulness of these for mental health practitioners and services. This would be particularly important as the PCOMS and the OQ-45 questionnaires are not commonly used in UK mental health services.

Clinical implications

Overall, it can be said that this review found some support for the beneficial effects of using progress feedback in couple therapy and university counselling services. It is as of yet inconclusive, whether these benefits extend to people with more severe mental health problems in psychiatric outpatient and inpatient settings.

The implementation and use of routine progress monitoring and feedback can be a costly and time-consuming undertaking (Bickman, 2008). It is therefore essential that further high quality evidence will explore their effectiveness and benefits to clients to justify spending a service's limited resources. Some services may be already collecting sessional data, like the Improving Access to Psychological Therapies (IAPT) services, which may make continuous progress feedback an easier process. However, we currently lack information on the mechanisms involved that have a positive impact on psychotherapy outcomes when using progress feedback systems. Until this is explored further, it is likely that services and clinicians will remain sceptical about their implementation.

Limitations of review

None of the included studies in this reviewed were given a rating of "excellent" on the Downs and Black (1998) critical appraisal tools and less than half attained a score of "good". There are still several methodological limitations that studies have to overcome in order to produce conclusive and robust evidence. Future studies should therefore be more transparent about their recruitment and sampling procedures, carry out a-priory power calculations, provide information on attrition rates over the course of the study, explore confounding factors between experimental groups and report statistical results appropriately (i.e. include actual p-values). It would be of particular interest if future studies collected information on how the feedback of progress data was conducted and perceived by clients. Although this review used a quality assessment tool, it needs to be acknowledged that there are several checklists available to the research community that vary in their content and ratings. This makes comparisons between studies and future reviews difficult (Mallen, Peat & Croft, 2006).

Finally, it needs to be considered that this review did not consult other raters to assess the inter-rater reliability of the obtained quality ratings.

References

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*, 693-704.

Barber, J.P., Sharpless, B.A., Klostermann, S. & McCarthy, K.S. (2007). Assessing intervention competence and its relation to therapy outcome: a selected review derived from the outcome literature. *Professional Psychology: Research and Practice, 38*, 493-500.

Bickman, L. (2008). A measurement feedback system (MFS) is necessary to improve mental health outcomes. *Journal of the American Academy of Child and Adolescent Psychiatry, 47,* 1114-1119.

Bickman, L., Kelley, S.D., Breda, C., de Andrade, A.R. & Riemer, M. (2011). Effects of routine feedback to clinicians on mental health outcomes of youths: results of a randomized trial. *Psychiatric Services, 62*, 1423-1429.

Boyce, M.B. & Browne, J.P. (2013). Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. *Quality of Life Research, 22*, 2265-2278.

By Rise, M., Eriksen, L., Grimstad, H. & Steinsbekk, A. (2012). The short-term effect on alliance and satisfaction of using patient feedback scales in mental health out-patient treatment. A randomised

controlled trial. *BMC Health Services Research, 12* (348), Retrieved from http://www.biomedcentral.com/1472-6963

Cahill, J., Barkham, M. & Stiles, W.B. (2010). Systematic review of practice-based research on psychological therapies in routine clinical settings. *British Journal of Clinical Psychology, 49,* 421-453.

Carlier, I.V.E., Meuldijk, D., Van Vliet, I.M., Van Fenema, E., Van der Wee, N.J.A. & Zitman, F.G. (2012). Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *Journal of Evaluation in Clinical Practice, 18,* 104-110.

Crits-Christoph, P., Ring-Kurtz, S., Hamilton, J.L., Lambert, M.J., Gallop, R., McClure, B., Kulaga, A. & Rotrosen, J. (2012). A preliminary study of the effects of individual patient-level feedback in outpatient substance abuse treatment programs. *Journal of Substance Abuse Treatment, 42*, 301-309.

Department of Health (2012). *No health without mental health: Implementation framework*. London: Department of Health.

Downs, S.H. & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health, 52*, 377-384.

Garland, A.F., Kruse, M. & Aarons, G.A. (2003). Clinicians and outcome measurement: what's the use? *Journal of Behavioral Health Services & Research, 30*, 393-405.

Gilbody, S.M., House, A.O. & Sheldon, T.A. (2002). Psychiatrists in the

UK do not use outcome measures: National survey. *The British Journal of Psychiatry, 180*, 101-103.

Goodman, J.D., McKay, J.R. & DePhilippis, D. (2013). Progress monitoring in mental health and addiction treatment: a means of improving care. *Professional Psychology: Research and Practice, 44*, 231-246.

Hansson, H., Rundberg, J., Oesterling, A., Oejehagen, A. & Berglund, M. (2013). Intervention with feedback using outcome questionnaire 45 (OQ-45) in a Swedish psychiatric outpatient population. A randomized controlled trial. *Nordic Journal of Psychiatry, 67*, 274-281.

Hatfield, D., McCullough, L., Frantz, S.H.B. & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology & Psychotherapy, 17*, 25-32.

Hatfield, D.R. & Ogles, B.M. (2004). The use of outcome measures by psychologists in clinical practice. *Professional Psychology: Research and Practice, 35,* 485-491.

Hooper, P., Jutai, J.W., Strong, G., Russell-Minda, E. (2008). Age-related macular degeneration and low-vision rehabilitation: a systematic review. *Canadian Journal of Ophtalmology, 43*, 180-187.

Jensen-Doss, A. & Hawley, K.M. (2010). Understanding barriers to evidence-based assessment: Clinician attitudes toward standardized assessment tools. *Journal of Clinical Child & Adolescent Psychology, 39*, 885-896.

Knaup, C., Koesters, M., Schoefer, D., Becker, T. & Puschner, B. (2009). Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *The British Journal of Psychiatry, 195*, 15-22.

Lambert, M.J. (2012). Helping clinicians to use and learn from research-based systems: the OQ-Analyst. *Psychotherapy, 49*, 109-114.

Lambert, M.J. (2013). Outcome in psychotherapy: the past and important advances. *Psychotherapy, 50*, 42-51.

Lambert, M.J., Morton, J.J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R.C.,…Burlingame, G.M. (2004). *Administration and scoring manual for the outcome questionnaire, 45*. Salt Lake City, UT: OQ Measures.

Lambert, M.J. & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy, 48*, 72-79.

Luborsky, L., Diguer, L., Seligman, D.A., Rosenthal, R., Krause, E.D, Johnson, S.,…Schweizer, E. (1999). The researcher's own therapy allegiances: a "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6*, 95-106.

Mallen, C., Peat, G. & Croft, P. (2006). Quality assessment of observational studies is not commonplace in systematic reviews. *Journal of Clinical Epidemiology, 59*, 765-769.

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.

Miller, S.D. & Duncan, B.L. (2004). *The outcome and session rating scales: administration and scoring manual.* Chicago: Institute for the Study of Therapeutic Change.

Miller, S.D., Duncan, B.L., Sorrell, R. & Brown, G.S. (2005). The partners for change outcome system. *Journal of Clinical Psychology: In Session, 61*, 199-208.

Murphy, K.P., Rashleigh, C.M. & Timulak, L. (2012). The relationship between progress feedback and therapeutic outcome in student counselling: A randomised controlled trial. *Counselling Psychology Quarterly, 25*, 1-18.

National Institute for Mental Health in England (2008). Mental health compendium. London: NIMHE.

Probst, T., Lambert, M.J., Loew, T.H., Dahlbender, R.W., Goellner, R. & Tritt, K. (2013). Feedback on patient progress and clinical support tools for therapists: improved outcome for patients at risk of treatment failure in psychosomatic in-patient therapy under the conditions of routine practice. *Journal of Psychosomatic Research, 75*, 255-261.

Puschner, B., Schoefer, D., Knaup C. & Becker, T. (2009). Outcome management in in-patient psychiatric care. *Acta Psychiatrica Scandinavica, 120*, 308-319.

Reese, R.J., Norsworthy, L.A. and Rowlands, S.R. (2009*a*). Does a continuous feedback system improve psychotherapy outcome? *Psychotherapy Theory, Research, Practice, Training, 46*, 418-431.

Reese, R.J., Usher, E.L., Bowman, D.C., Norsworthy, L.A., Halstead, J.L., Rowlands, S.R. & Chisholm, R.R. (2009*b*). Using client feedback in psychotherapy training: an analysis of its influence on supervision and counsellor self-efficacy. *Training and Education in Professional Psychology, 3*, 157-168.

Reese, R.J., Toland, M.D., Slone, N.C. & Norsworthy (2010). Effect of client feedback on couple psychotherapy outcomes. *Psychotherapy Theory, Research, Practice, Training, 47*, 616-630.

Shimokawa, K., Lambert, M.J. & Smart, D.W. (2010). Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *Journal of Consulting and Clinical Psychology, 78*, 298-311.

Simon, W., Lambert, M.J., Harris, M.W., Busath, G. & Vazquez, A. (2012). Providing patient progress information and clinical support tools to therapists: effects on patients at risk of treatment failure. *Psychotherapy Research, 22,* 638-647.

Simon, W., Lambert, M.J., Busath, G., Vazquez, A., Berkeljon, A., Hyer, K., … Berrett, M. (2013). Effects of providing patient progress feedback and clinical support tools to psychotherapists in an inpatient eating disorders treatment program: a randomized controlled study. *Psychotherapy Research, 23*, 287-300.

Slade, K., Lambert, M.J., Harmon, S.C., Smart, D.W. & Bailey, R. (2008). Improving psychotherapy outcome: the use of immediate electronic

feedback and revised clinical support tools. *Clinical Psychology and Psychotherapy, 15*, 287-303.

Valderas, J.M., Kotzeva, A., Espallargues, M., Guyatt, G., Ferrans, C.E., Halyard, M.Y., … Alonso, J. (2008). The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Quality of Life Research, 17*, 179-193.

Vermeersch, D.A., Whipple, J.L., Lambert, M.J., Hawkins, E.J., Burchfield, C.M. & Okiishi, J.C. (2004). Outcome questionnaire: is it sensitive to changes in counselling center clients? *Journal of Counseling Psychology, 51*, 38-49.

West, S., King, V., Carey, T.S., Lohr, K.N., McKoy, N., Sutton, S.F. & Lux, L. (2002). Systems to rate the strength of scientific evidence. In *Evidence Report/Technology Assessment No.47 (Prepared by the Research Trial Institute – University of North Carolina Evidence-based Practice Center under Contrast No 290-97-0011). AHRQ Publication No.02-E016.* Rockville, MD: Agency for Healthcare Research and Quality.

Part 2: Empirical Paper

The development and validation of a shortened version of the Eating Disorder Examination Questionnaire (EDE-Q)

Abstract

Aims: The aim of this study was to develop and validate a short version of the Eating Disorder Examination Questionnaire (EDE-Q) for sessional outcome assessment, which is sensitive to clinical change.

Method: A principal component analysis was conducted to determine the factor structure of 489 EDE-Qs completed by individuals with a range of eating disorders. Rasch analysis was carried out on each identified factor. The statistical information and expert ratings (N=10) informed the inclusion/exclusion criteria for each EDE-Q item. The EDE-Q's response scale properties were also investigated using the Rasch model. Data from people with (N=54) and without eating disorders (N=503) were collected through an online survey to assess the reliability, validity and sensitivity of the new measure.

Results: A 12-item short version, the Eating Disorder Examination Questionnaire Short (EDE-QS) was developed. Initial psychometric evaluation showed that the EDE-QS is a reliable, valid and sensitive questionnaire.

Conclusions: The EDE-QS appears suitable for the use as a brief and user-friendly sessional outcome measure.

Introduction

Eating disorders pose a serious challenge to mental health services due to their often chronic trajectory (Steinhausen, 2002) and far-reaching psycho-social and medical implications (Bohn et al., 2008; Doll, Petersen & Stewart-Brown, 2005). People suffering from eating disorders are at an increased risk of premature death (Arcelus, Mitchell, Wales & Nielsen, 2011) and anorexia nervosa has the highest mortality rate for adolescents amongst all other psychiatric disorders (National Institute for Clinical Excellence, 2004). It is therefore crucial to carry out appropriate assessments of people with eating disorders and monitor their progress throughout therapy so that care and treatment can be optimised.

Eating disorders

Eating disorders broadly fall into the categories of anorexia nervosa (AN), bulimia nervosa (BN), binge eating disorder (BED) and other specified feeding or eating disorder (OSFED) as classified by the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5; American Psychiatric Association, 2013). The common factors across eating disorders are the individual's concerns about their weight and shape and persistent eating behaviours that increase physical health risks. Beyond this, it appears that the individual's self-worth is derived almost exclusively from their physical appearance or ability to influence their weight and eating behaviour (Fairburn & Cooper, 1989). It is difficult to determine the exact prevalence rates as results vary depending on the survey and methodology used (Fairburn & Cooper, 1989; Roth & Fonagy, 2005). However, a recent study estimates that within Europe between

0.2 and 0.5% of people are affected by anorexia nervosa and 0.1-0.9% by bulimia nervosa within a 12 month period (Wittchen et al., 2011).

Eating Disorder Examination Questionnaire

The Eating Disorder Examination Questionnaire (EDE-Q) is a self-report measure, which was developed by Fairburn and Beglin (1994). Its items were derived from the Eating Disorder Examination (EDE), which is a structured and well validated eating disorder assessment interview (Cooper, Cooper & Fairburn, 1989; Fairburn & Cooper, 1993). The EDE is considered to be the "gold standard" in the assessment of eating disorder pathology (Berg, Peterson, Frazier & Crow, 2012). However, clinicians need to be trained to deliver and interpret the EDE successfully, which can be strain on a service's resources. In addition, it takes approximately an hour to carry out the assessment (Fairburn, 2008), which again is a pressure on available therapeutic time and makes multiple measurements throughout the course of therapy unlikely. The EDE-Q was developed in an attempt to produce a self-report questionnaire that can approach the "gold standard" whilst making it more widely accessible and less burdensome for clients (Fairburn & Beglin, 1994).

The latest version of the EDE-Q, the EDE-Q 6.0 (Fairburn & Beglin, 2008), consists of 22 scaled items that assess a person's attitudes towards eating, their physical appearance and weight. These are further categorised into four subscales as for the EDE (Cooper et al., 1989): Restraint, Eating Concern, Shape Concern and Weight Concern.
The Restraint subscale includes items referring to specific dietary behaviours of restricting or limiting one's food intake. Eating Concern

refers to difficulties or worries elicited when eating. The Shape and Weight Concern subscales ask questions about shape and weight related worries and their impact on a person's feelings and view of themselves. There are six frequency items that enquire about overeating, days and episodes of binge eating, self-induced vomiting, laxative use and excessive exercise. All questions refer to the past 28 days. The intention of the EDE-Q is to capture the frequency and severity of these problematic behaviours and cognitions and to monitor changes over time, particularly in response to treatment.

Performance compared to EDE

Since its conception, the EDE-Q has received international attention and is used in clinical practices around the globe. Its validity and reliability has been thoroughly examined across a range of eating disorders. It was concluded that there are acceptable levels of correlation between the EDE and EDE-Q subscales in people with BN and AN (Binford, Le Grange & Jellar, 2005; Carter, Aime & Mills, 2000; Fairburn & Beglin, 1994) and the general population (Mond, Hay, Rodgers, Owen & Beumont, 2004b). Although statistically significant, lower levels of agreement were found in obese bariatric surgery patients (Kalarchian, Wilson, Brolin & Bradley, 2000) and people with BED (Wilfley, Schwartz, Spurrell & Fairburn, 1997). Good levels of internal consistency were shown for EDE-Q total score and subscales in BN (Peterson et al. 2007) and in the general population (Mond, Hay, Rodgers, Owen & Beumont, 2004a) as well as good test-retest reliability for the individual subscales in people with BED (Reas, Grilo & Masheb, 2006) and an adult community sample (Luce & Crowther, 1999).

The frequency and behavioural items, however, showed less consistent results across different populations. Similar responses were found between the EDE and the EDE-Q for objective binge eating behaviours in people with BED (Grilo, Masheb & Wilson, 2001; Reas et al., 2006) and obese bariatric surgery patients (Kalarchian et al., 2000). However, another study could not support these findings for people with BED (Wilfley et al., 1997). Results from a general population sample (Fairburn & Beglin, 1994; Mond et al., 2004b) and from people suffering with bulimia nervosa (Carter et al., 2000) also failed to show consistencies for objective binge eating episodes. Differences were found for laxative use among a community sample (Fairburn & Beglin, 1994) and self-induced vomiting in people with bulimia nervosa (Carter et al., 2000). Some studies found higher rates for the EDE-Q (Fairburn & Beglin, 1994), others for the EDE (Carter et al., 2000; Grilo et al., 2001; Mond et al., 2004b). The behavioural items further showed variation when a test-retest analysis was conducted (Berg et al., 2012; Luce & Crowther, 1999; Mond et al., 2004a).

It is, however, not clear whether the EDE or the EDE-Q is more accurate in frequency data collection as there is a lack of objective measurement (Wilfley et al., 1997) and further research is required (Berg et al., 2012).

Factor Structure

Studies investigating the factor structure of the EDE-Q in a number of different populations were unable to support the existing four factor model for the scaled items as suggested by Cooper et al. (1989).

Hrabosky et al. (2008) ran an exploratory factor analysis with data collected from obese bariatric surgery candidates. This resulted in a four-factor model, consisting of 12 items only. The four factors did not overlap substantially with the original subscales. The authors therefore questioned the utility of the original scale for bariatric surgery patients. Peterson et al. (2007) extracted four factors through an exploratory factor analysis in women with bulimic symptoms. Two factors were similar to the original Eating Concern and Restraint subscales, the other two, however, consisted of different items. A post-hoc analysis found support for a three factor model, in which most shape and weight concern items loaded onto one factor.

Confirmatory factor analysis did not support the proposed four factor structure of eating disordered outpatients and healthy individuals in Australia (Allen, Byrne, Lampard, Watson & Fursland, 2011), nor in a community sample of adolescents in the UK (White, Haycraft, Goodwin & Meyer, 2014). Instead, Allen et al. (2011) suggested a brief one factor model, consisting of eight weight and shape concern items. White et al. (2014), however, found support for a three factor model, which combined the shape and weight concern items into one factor. The remaining two factors strongly resembled the original restraint and eating concern subscales.

To summarise, recent research has questioned the validity of the existing subscales of the EDE-Q. These studies have however found no shared consensus of an alternative factor structure, although there seems to be support for combining the weight and shape concern items. It however has to be acknowledged that samples varied with regards to eating disorder pathology, severity and age range.

The case for routine outcome measurement

The government and commissioners increasingly demand that clinicians and services collect and report on patients' outcomes to improve the delivery of mental health care (Department of Health, 2011). This is supported by research evidence that the collection and feedback of routine outcome measures leads to more positive outcomes (Lambert, 2013). Simon et al. (2013) demonstrated that better outcomes were achieved for people receiving inpatient treatment for an eating disorder if their therapists received regular feedback on their progress. Lambert and Shimokawa (2011) suggested that feedback enables therapists to timely re-evaluate and amend the intervention if needed. Valderas et al. (2008) also claim that sharing progress feedback with a client can improve the therapeutic relationship, shared understanding and treatment adherence. It is therefore suggested that the appropriate use of routine outcome data can provide service users with more accurate treatment progress and improve the quality of the intervention.

The National Institute for Mental Health in England, sponsored by the UK Government, published an outcome measures compendium that includes questionnaires suitable for outcome monitoring across a broad range of psychological difficulties to guide services and practitioners (National

Institute for Mental Health in England, 2008). The EDE-Q is currently the solely recommended outcome tool for the assessment and monitoring of eating disorders. Although the EDE-Q has substantially reduced administration time from the original EDE, it is still too lengthy to be used as a brief sessional outcome measure. It also measures symptoms over the last 28 days, which makes it problematic to use for the measurement of change from one week to the next.

Programmes, such as the Improving Access to Psychological Therapies (IAPT), expect therapists to obtain outcome measures during every clinical contact (IAPT, 2008). Recent data revealed that IAPT practitioners achieved a completion rate of 90% (IAPT, 2012). This suggests that brief outcome measures, such as the GAD-7 (Spitzer, Kroenke, Williams & Loewe, 2006), and PHQ-9 (Kroenke, Spitzer & Williams, 2001), facilitate routine outcome measurement and reduce the practical barriers to sessional data collection. It would therefore be desirable to have a similarly short outcome measure for eating disorders that is sufficiently sensitive to capture change over time.

Aim of the study

Lambert and Hawkins (2004) argued that measures selected for progress monitoring need to fulfil the criteria of being valid, reliable and sensitive to change over time. The aim of this study is therefore to develop a short version of the EDE-Q, the Eating Disorder Examination Questionnaire SHORT (EDE-QS), which meets these criteria and lends itself to be used as a user-friendly and sessional treatment outcome measure.

Study overview

The study was carried out in two phases. **Phase 1** refers to the development of the shortened version and includes 1) an exploratory principal component analysis of original EDE-Q data, 2) Rasch analysis on the identified factors, 3) collection of expert opinion, 4) and integration of these methods to inform item selection and deletion. Applying a combination of these approaches will help balance clinical utility with adequate psychometric requirements (Slade, Thornicroft & Glover, 1999).

The use of Rasch analysis was considered to be particularly important for several reasons (Bond & Fox, 2010; Tennant & Conaghan, 2007):

a) To examine the appropriateness of a rating scale

By running Rasch rating scale diagnostics, it can be assessed whether the chosen response categories of a scale are meaningful and informative. The EDE-Q's response categories, for example, refer to the number of days in a month and are spread over seven categories. If respondents make good use of each category in a way that is consistent with the severity of their eating difficulties (i.e. selecting higher response options if more severely impaired), the scale can be shown to have an adequate format. However, if response categories do not appear to make meaningful distinctions between eating disorder severities or are rarely used, the scale may be optimised by changing the number of response categories.

b) To identify misfitting items

The Rasch analysis determines how well an item "fits" the model and measures the construct that it is meant to measure. If an item has little

predictive value and obtains unexpected ratings, it is said to misfit the model. It therefore introduces random variability into the data.

c) To identify redundant items

Rasch analyses can also establish whether an item produces unique information about a person in relation to the construct at hand. If the answer to one item highly depends on or can be predicted by the answer to another item, it could be argued that one of these items is redundant.

d) To establish difficulty estimates

By following the Rasch model, each item's difficulty estimate is calculated. In this context, this means that there are items which will be endorsed by the majority of people, regardless of the severity of their eating disorder. These items will have a low difficulty estimate and could be referred to as "easy" items. Other items may only be endorsed by people with more significant impairments and therefore represent a more "difficult" item. It can thus be concluded that these people have more severe eating difficulties. The difficulty estimate can be used to select a broad range of "easy" and "difficult" items so that the scale is suitable for people with varying degrees of eating disorder severity and can differentiate between them.

Phase 2 describes the psychometric evaluation of the shortened version's reliability and validity through an online survey, including respondents with eating disorders and the general population.

a) Reliability: It is expected to find high *internal consistency*, by calculating Cronbach's alpha to establish the homogeneity of the scale (Cronbach, 1951). It is also predicted that the measure will show

temporal stability as established by a *test-retest* analysis across two different time points.

b) Construct validity: This will be explored by comparing the shortened version to other eating disorder related questionnaires and the original EDE-Q. It is expected that there will be large positive correlations with these measures, demonstrating *convergent validity*. The EDE-QS will also be compared to measures of mental health functioning. It is expected that higher scores on the EDE-QS will correlate with higher scores on anxious and depressive symptoms and lower scores for quality of life ratings. *Divergent validity* will be assessed by examining the correlation between the shortened questionnaire and the Sociability Scale (Cheek & Buss, 1981), a measure of sociability, which taps into a construct that appears to be unrelated to eating disorder pathology (Miller, Schmidt & Vaillancourt, 2008).

The frequency items will be explored separately. It is expected that respondents will report similar estimates for the EDE-QS and the EDE-Q.

c) Sensitivity: By comparing the total EDE-QS scores for people with and without a current eating disorder, the measure's sensitivity for differentiating between these groups will be established.

1. Phase 1: Questionnaire Reduction

1.1 Methods

1.1.1 Participants

EDE-Q data

Existing EDE-Q data collected during April 2008 and January 2013 by three Eating Disorders Services in North and South London were included in the analyses, resulting in a sample size of 489 patients. The minimum recommended sample size to conduct a factor analysis for a questionnaire of the EDE-Q's length by Tabachnick and Fidell (2001) and Comrey and Lee (1992) is 300 cases.

The majority of the sample was female (90.2%) and ranged from 18 to 72 years (*M*=31.5, *SD*=11.5). The sample included outpatient and inpatient admissions across a range of eating disorders. The Global EDE-Q scale ranged from 1.55 to 6 (*M*= 4.11, *SD*=1.2). Probable DSM-5 diagnoses (American Psychiatric Association, 2013) were derived from EDE-Q responses as access to the diagnostic data was not available for all respondents and it was not possible to carry out diagnostic interviews on all participants. Sixteen percent of respondents were therefore identified as probable AN - restrictive, 15% as probable AN – binge/purge subtype, 21% as probable BN, 18% as probable BED and 30% as probable OSFED (see Appendix B for diagnostic criteria employed). Mean Body Mass Indices (BMI) were 14.23 (*SD*=1.7) for AN – restrictive subtype, 14.79 (*SD*=1.5) for AN – binge/purge subtype, 24.83 (*SD*=7.8) for BN, 37.23 (*SD*=13.8) for BED and 27.27 (*SD*=13.6) for OSFED. There

was limited information available on ethnic background of the respondents and it is therefore not reported here.

1.1.2 Measure

EDE-Questionnaire

The original self-administered EDE-Q 6.0 questionnaire consists of 28 items. Five additional items, which were not included in the analysis, enquire about an individual's weight, height, menstrual cycle and whether the contraceptive pill is taken. Items 1-12 and 19-21 are rated on a seven point scale with response options ranging from 'No days' to 'Every day' over the past 28 days. Items 22-28 are rated on a seven point scale with response options ranging from 'not at all' to 'markedly' over the past 28 days. In total, there are 22 scaled items, which are categorised into one of four subscales: *restraint, eating concern, shape concern* and *weight concern*. They are scored by taking the mean value of each subscale. This can range from zero to a maximum score of six. The mean score across all subscales results in the *Global EDE-Q score*.

Items 13-18 elicit open responses to the number of times or days of specific eating behaviours, such as *objective binge eating* (OBE), s*elf-induced vomiting* (SIV), *laxative use* (LAX) or *excessive exercise* (EX), over the last 28 days. These are not included in the subscale scores, but give an indication of symptom severity and aid diagnosis.

1.1.3 Procedures

The research study underwent proportionate review and was approved by a NHS ethics committee (see Appendix C for approval letter).

EDE-Q data

Existing EDE-Q data from three London Eating Disorders Services were collated from paper questionnaires and electronic files and entered into one spreadsheet. There was less than 5% of missing data for each scaled item. These were imputed using the Expectation Maximisation method (EM) as supplied in the Statistical Package for the Social Sciences (SPSS, version 21). Little's MCAR test was not significant ($\chi(741)=754.79$, $p=.35$), which suggests that the data were missing completely at random. There was no greater difference than 0.02 points for individual items between actual and estimated means. The total EDE-Q mean score before imputation was 4.21 (*SD*=1.15) and 3.98 (*SD*=1.11) after.

Expert Survey

An online survey was emailed to the research team's professional contacts. These individuals were expert eating disorder clinicians with a minimum of six years expertise in working with people with eating disorders. They were asked to give their opinion on the importance of EDE-Q items with regards to their ability to indicate clinically significant change in severity of eating disorders. Their task was to categorise each EDE-Q item into "least important", "very important-might be good to include" or "most important – needs to be included".

1.1.4 Statistical Analyses

The Rasch model works on the assumption of unidimensionality, which means that all items in a questionnaire should address one conceptual issue (Tennant & Conaghan, 2007). Although the EDE-Q assesses the

overall construct of eating pathology, previous studies have provided evidence of multi-dimensionality across several diagnostic presentations. Due to the inconsistent results with regards to number of factors and associated items in the literature (e.g., Allen et al., 2011; Hrabosky et al., 2008), an exploratory principal component analysis (PCA) was conducted in a sample that included a good spread of diagnostic categories and eating disorder severity. This was necessary to identify the underlying components or dimensions of the EDE-Q in a general eating disordered population so that Rasch modelling could be carried out separately on each dimension (DeVellis, 2012; Franchignoni et al., 2010).

a. Exploratory PCA

An exploratory PCA was carried out, using oblimin rotation (oblique), without specifying a number of factors as the items were free to correlate. The Statistical Package for Social Sciences (SPSS, version 21) was used. Only scaled EDE-Q items were included in the analyses.

As suggested by Field (2005), for a sample size larger than 250 and the average communality being greater than .6, the Kaiser's criterion can be applied and eigenvalues above 1 would be retained. The scree plot was also visually investigated to determine the number of factors. Items loading above .3 were retained.

b. Rasch Analysis

Winsteps software was used (version Bond&FoxSteps, Bond & Fox, 2010). The polytomous Rasch rating scale model was used because the EDE-Q's response scale is ordinal with seven response options.

Rating Scale Diagnostics

As a first step, the characteristics of the rating scale were examined according to criteria as set out by Linacre (2002), before the overall fit to the model was assessed (Lundstroem & Pesudovs, 2009):

a. At least 10 responses should be present in each response category.
b. There should be a regular distribution of responses across response categories.
c. There should be a consistent increase of average measures with each category.

   Average measures are representative of eating disorder severity. An increase in response category should therefore demonstrate an increase in eating disorder severity, as people with greater eating disorder pathology are expected to endorse higher response options.
d. Step calibrations (or category thresholds) should increase monotonically.

   Distinct steps between thresholds indicate that each category has the highest probability to be endorsed by respondents with a specific severity of eating disorder pathology. Therefore, higher categories should have a greater probability of being selected by respondents with greater severity of eating disorders. Thresholds should be appropriately distanced from one another and increase by at least 1.4 logits but no more than 5 logits (Bond & Fox, 2010), although this value reduces with larger number of

categories (e.g., at least 1.0 logit for a five category scale (Linacre, 2002)).

e. Category outfit mean square values should be less than 2. Greater values are an indication of excessive randomness and noise in the data.

To examine the distribution of response categories, the probability curves for each factor were also inspected visually. The individual curves should show distinct peaks for each category, indicating that each is the most probable response for some part of the eating disorder pathology (Bond & Fox, 2010). Person and item separation indices were also inspected to assess whether collapsing of response categories improved the reliability of persons and items. Bond and Fox (2010) argue that the indices should have values of at least 2.

If infrequent or inconsistent use of response categories was indicated by the rating scale diagnostics, collapsing of categories was considered. The rating scale diagnostics and probability curves of the collapsed models were then compared to the original to identify the optimal number of response categories (Bond & Fox, 2010).

Item fit to model

Mean square infit and outfit values are used to assess each item's performance to the expected overall model. Values between .7 and 1.40 indicate acceptable fit to the model (Bond & Fox, 2010).

The item-measure correlation was also investigated. Values greater than .3 demonstrate that the item is sufficiently correlated to the overall concept or model (Williams et al., 2009).

The poorest fitting items were considered for deletion.

Local dependence

Residual correlations between items within a scale were examined for local dependency, which indicates that responses to one item are dependent on another and therefore imply item redundancy (Tennant & Conaghan, 2007). Residual item correlations that have values greater than .3 of the overall average of all correlations suggest local dependence (das Nair, Moreton & Lincoln, 2011). Where this applied, deletion of one of the dependent items was considered.

Difficulty Estimate

The Rasch model provides a difficulty estimate for each item, which can be considered as the level of eating disorder severity in this context. More positive estimates are most likely to be endorsed by more severely eating disordered individuals, whereas lower estimates (including negative values) are most likely to be endorsed by less severely impaired people. The aim was to assess a wide range of eating disorder pathology and items were considered for inclusion if they were either high or low on the severity estimate. Items that showed a strong overlap of severity (i.e., differences <0.20) were considered for deletion (Greco, Lambert & Baer, 2008).

Rasch analysis can also be used to assess whether all items enquire about one theoretical construct and indicate *unidimensionality* of a scale. This was not assessed separately as the PCA was conducted to identify the dimensions of the EDE-Q. *Differential item functioning (DIF)* was also

not assessed as it may be expected between diagnoses and gender which would not be of concern for the questionnaire's purpose.

f. Expert Survey

The categories were given values from 0-2 and summed up for each EDE-Q item. Their total scores were used to obtain an overall rating of importance for clinical change. These could range from a minimum score of zero to a maximum score of 20. Experts were also invited to provide free responses to any aspects of the EDE-Q.

g. Combination of methods

Information from the exploratory PCA, Rasch Modelling and expert survey was combined to make decisions on the inclusion and exclusion of items (see Appendix D for inclusion/exclusion criteria).

1.2 Results

1.2.1 Expert Survey

Ten experts submitted responses and seven (70%) provided demographic information. Of these, 86% were female with an age range of 33 to 63 years (M= 47, SD = 9.6). All respondents were White, with the majority (86%) being White British. There was an almost even professional division between professional backgrounds, with 57% being Psychiatrists and 43% Clinical Psychologists. Their therapeutic trainings included Franz CP, MClinPsych, MRCPsych, CAT, CBT, IPT, sex therapy, MBBS, MD and Family Therapy Diploma. Most experts (71%) have worked in the field of eating disorders for more than ten years, the remaining between six and ten years. All respondents were currently

working in an outpatient setting, and one third were also based in inpatient, one third in psychological therapies and one third in day programmes services. One respondent worked with children, everyone else with adults with eating disorders. The experts estimated that 50% of their clients presented with AN (*SD*=32.8), 20% with BN (*SD*=17.2), 3% with BED (*SD*=4.08) and 27% with OSFED (*SD*=21.2).

The obtained expert ratings ranged from zero to a highest score of 15. See Table *2* for individual item ratings.

1.2.2 Exploratory principal component analysis

Kaiser-Meyer-Olkin measure of sampling adequacy (KMO) produced a value of .874 which indicates that the analysis provided distinct and reliable factors (Field, 2005). Bartlett's test of sphericity was highly significant (5,289.84; *p<.001*), indicating the data were appropriate for conducting a PCA.

The average communality was .63 and therefore Kaiser's criterion was applied (Field, 2005). This suggested a five factor solution. The scree plot's curve started to tail off after three factors; however, there was another slight drop after five factors and this, in combination with the Kaiser's criterion confirmed the final selection of five factors.

Factor 1, which was labelled *Drive for Thinness,* explained 33.01% of the total variance and included six items. These were included across all of the four original subscales. Factor 2, labelled *Body Dissatisfaction*, added 13.04% of variance, consisted of six items and included four items from the original Shape Concern subscale. Factor 3 explained 6.53% of variance, was labelled *Dietary Restraint* and included four items. Three of these are also found in the original Restraint subscale. Factor 4 added

5.34% of explained variance, was labelled *Guilty Affect* and consisted of two items only. Both items are found in the original Eating Concern scale. The final Factor 5 explained 4.98% of variance, referred to *Overvaluation of shape/weight* and consisted of four items, of which two were part of the original Shape Concern scale. See Table 2 for individual factor loadings.

The component correlation matrix showed that the correlation between factors ranged from .17 to .35, which demonstrated complete separation between factors.

1.2.3 Rasch analysis

As the fourth factor comprised only two items, it was not included in a separate Rasch analysis (Siegert, Jackson, Tennant & Turner-Stokes, 2010; Williams et al., 2009).

Rating scale diagnostics

Rating scale diagnostics and probability curves were examined for each factor. All response categories included more than ten observations and none of the categories had outfit mean squares greater than 2. However, responses across categories were not evenly distributed (e.g., rating scale category 2 held consistently less than 10% of responses) and all had disordered category thresholds, which was also clearly visible from the probability curves (see *Figure 1* for an example*).*

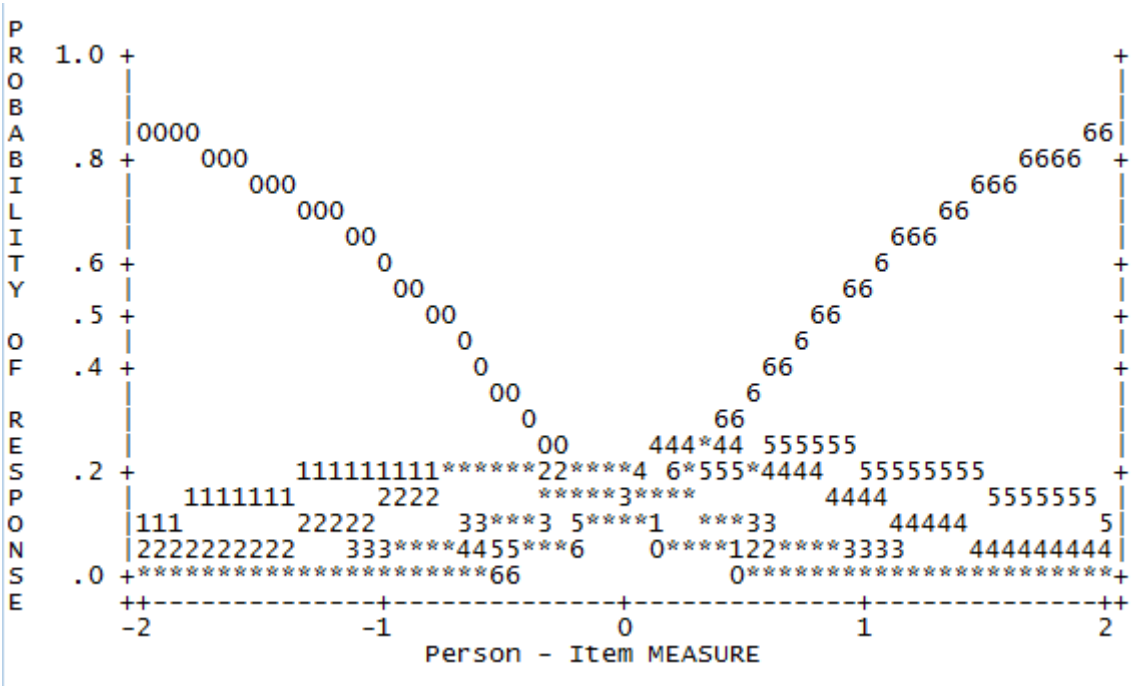

Figure 1: **Response probability curve with original 7-point response options (Factor 5)**

The original rating scale consists of seven response options with values ranging from zero to six (0123456). It was investigated if collapsing of categories resulted in improved rating scale diagnostics. By combining response options 1 and 2, as well as options 3 and 4, and 5 and 6, a four-point response scale was produced, which included values ranging from zero to three (0112233) (see Table 3). This was applied across all factors in order to give a unified response scale for all items, which was considered essential for a user-friendly shortened version.

Rating scale diagnostics demonstrated improved category thresholds and probability curves for the revised four-point response scale (see Table 1).

Table 1: ***Rating scale diagnostics, reliability indices and visual inspection of probability curves for original and collapsed 4-point rating scale***

	Rating scale	Regular response frequency	Step calibrations	Outfit mean square	Person separation	Item separation	Probability curve
FACTOR 1	original	no	disordered	< 2.0	1.34	9.36	0 and 6 peak only
	4-point	improved	disordered	< 2.0	1.25	8.76	0, 1 and 3 peak
FACTOR 2	original	no	disordered	< 2.0	1.19	2.24	0, 4 and 6 peak
	4-point	improved	ordered	< 2.0	0.98	2.64	all peak
FACTOR 3	original	no	disordered	< 2.0	0.99	20.17	0 and 6 peak only
	4-point	improved	ordered	< 2.0	0.84	5.47	all peak
FACTOR 5	original	no	disordered	< 2.0	0.95	7.81	0, 4 and 6 peak
	4-point	improved	ordered	< 2.0	0.69	7.53	all peak

The distribution of response frequencies improved across all factors once a four point response option was used. All but one factor now showed ordered step calibrations, which was further confirmed by the probability curves, showing more distinct peaks. The person separation indices slightly reduced for each factor, which indicates that the items do not separate the respondents as well as they might. However, it was decided to prioritise ordered thresholds over an already low person separation index. The item separation indices also reduced through collapsing the response scale; however, these remained above the threshold of 2.0 as specified by Bond and Fox (2010), and were therefore considered satisfactory.

Although probability curves improved markedly, they still showed respondents' tendency to endorse the extreme points of the questionnaire, namely "no days" and "every day" (see *Figure 2*).

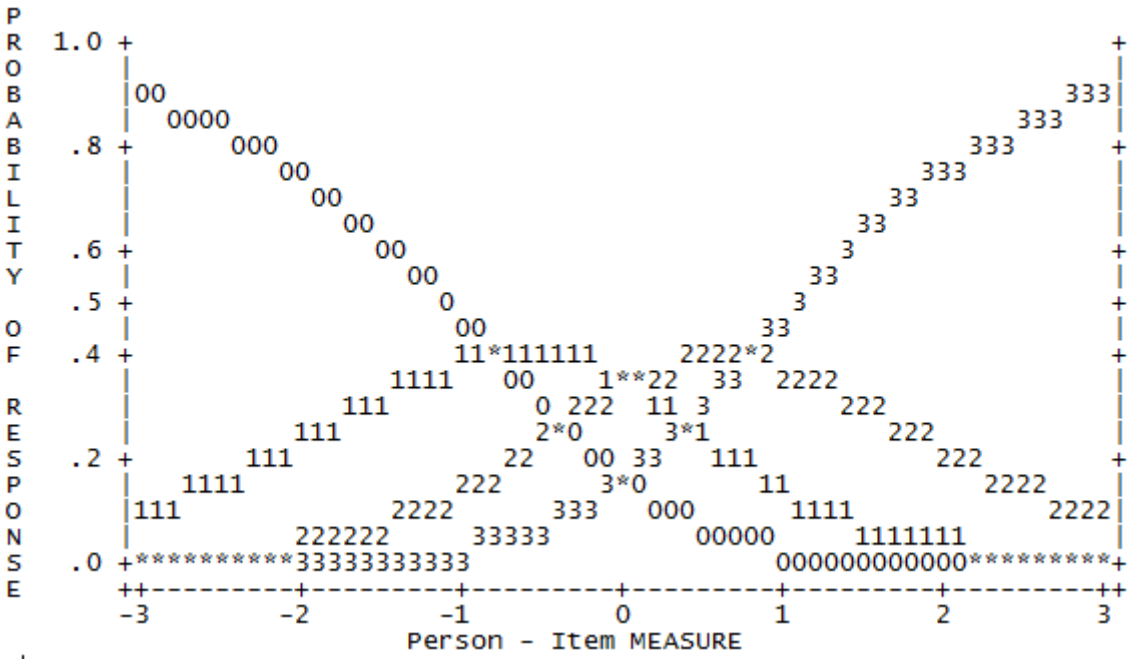

Figure 2: **Response probability curve with collapsed 4-point response options (Factor 5)**

1.2.4 Reduction of EDE-Q items

After collapsing the response scale, Rasch analyses were carried out on each dimension. These results were used in combination with those obtained from the expert survey to inform deletion of items. Table 2 shows the results of the principal component analysis, Rasch model and expert survey combined. The following items were deleted:

Factor 1: Item 24 was identified as misfitting the Rasch model and deleted. Items 7 and 21 highly overlapped with regards to item difficulty or severity. As item 7 had a higher expert rating, it was retained whilst item 21 was deleted. The severity level of item 7 indicated that it was likely to be endorsed by many people with eating disorders. Item 2 was appropriately distanced in terms of severity from item 7 and most likely to be endorsed by respondents with more severe eating problems. It was therefore retained and items 5 and 6 were deleted.

Factor 2: Investigation of local dependence revealed that responses to items 11 and 12, 25 and 26, and 27 and 28 were highly dependent on one another. Items 11 and 28 had lower expert ratings and were therefore deleted. Items 25 and 26 had an almost identical wording; one referring to dissatisfaction with shape, the other to dissatisfaction with weight. Hrabosky et al. (2008) reported similar results in that people awaiting bariatric surgery did not distinguish between concerns about shape and weight. Therefore, instead of deleting one of the dependent items, they were combined to now include reference to satisfaction with "shape and weight". Item 27 had a lower

expert rating than items 12 and 25/26 and was therefore deleted from Factor 2.

Factor 3: Item 10 was identified as misfitting the Rasch model. It was however retained, as it obtained a high rating by experts and was relevant for meeting anorexia nervosa diagnostic criteria. Item 1 obtained the EDE-Q's highest expert rating and was therefore retained. The remaining items, 3 and 4, were removed.

Factor 4: Both items were removed as it was impossible to investigate their psychometric properties using the Rasch model. In addition, they did not obtain a very high rating by experts and were therefore deemed as less important.

Factor 5: Item 22 refers to the importance of one's weight, item 23 to the importance of one's shape. As they showed local dependence, an almost identical level of severity and their content was relevant to diagnostic criteria, it was decided to combine them into a single item, referring to the importance of one's weight and shape. Item 9 approached the threshold of item misfit and was therefore removed and item 8 was retained.

Frequency items: The frequency items were inspected in a similar fashion, investigating expert ratings and diagnostic relevance.
Item 15 had high overlap in content with item 13 and 14. As the latter were rated higher by experts, item 15 was removed. Item 14 refers to a loss of control over eating. This has shown to be a better predictor of eating disorder pathologies than objective binge eating (Latner, Mond, Mackenzie, Haynes & Hay, 2014). In order to have an independent item

on perceived loss of control over eating (Mond, Hall, Bentley, Harrison, Gratwick-Sarl & Lewis, 2014) as well as a measure of objective binge eating, the order of items 13 and 14 was reversed. Respondents are therefore asked about perceived loss of control first, which is followed by a question on objective binge eating episodes.

Items 16 and 17 refer to compensatory behaviours (i.e. taking laxatives and vomiting) and were combined into a single item. To reduce missing responses and increase simplicity of coding, a Likert-scale response format was adopted.

Table 2: ***Summary of PCA, Rasch analysis, expert survey and diagnostic relevance***

	item	**item content**	**PCA**	**RASCH**					**Expert Survey**	**Diagnostic relevance**
			Factor Loading	Difficulty Estimate	S.E.	Infit MNSQ	Outfit MNSQ	Item Total Correl.	Local Item Dependence *item (correl.)*	Expert Rating	
Factor 1	2	Long periods without eating	0.68	1	0.06	0.87	0.83	0.73		5	AN
	5	Empty stomach	0.7	0.13	0.05	0.89	0.79	0.71		5	
	6	Flat stomach	0.47	-0.6	0.06	1.2	1.06	0.6		5	
	7	Preoccupation with food	0.55	-0.35	0.06	0.89	0.84	0.65		11	
	21	Concerned to be seen eating	0.53	-0.14	0.05	1.03	1.11	0.6		5	
	24	Upset to be weighed	0.52	-0.04	0.05	1.22	1.29	0.58		0	
Factor 2	11	Feeling of fatness	0.51	0.16	0.08	1.17	1.01	0.7	12 (0.36)	6	
	12	Desire to lose weight	0.49	0.44	0.08	1.14	1.09	0.73	11 (0.36)	10	
	25	Dissatisfaction with weight	0.84	-0.11	0.09	0.83	0.81	0.72	26 (0.31)	5	
	26	Dissatisfaction with shape	0.8	-0.33	0.09	0.82	0.85	0.68	25 (0.31)	4	
	27	Discomfort seeing body	0.8	-0.14	0.09	0.86	0.95	0.69	28 (0.33)	4	
	28	Discomfort being seen	0.78	-0.02	0.08	1.08	1.16	0.67	27 (0.33)	1	

Factor 3	1	Limit amount of food	0.79	0.07	0.07	0.85	0.83	0.8		15	AN
	3	Exclude foods	0.81	0.37	0.07	0.81	0.78	0.83		10	AN
	4	Dietary rules	0.75	0.31	0.07	1.05	1.02	0.79		10	AN
	10	Fear of weight gain	0.5	-0.75	0.09	1.4	1.41	0.62		12	AN
Factor 4	19	Eating in secret	0.74							5	
	20	Feeling guilty	0.51							5	
Factor 5	8	Preoccupation with shape/weight	0.51	0.91	0.07	1.17	1.11	0.78		8	
	9	Fear of losing control	0.63	0.26	0.07	1.32	1.27	0.71		9	BN
	22	Importance of weight	0.65	-0.59	0.09	0.78	0.77	0.71	23 (0.78)	9	AN/BN
	23	Importance of shape	0.69	-0.58	0.09	0.69	0.65	0.72	22 (0.78)	7	AN/BN
Frequency items	13	times of overeating								10	BED/BN
	14	times of having lost control								11	BED/BN
	15	days of overeating & loss of control								7	BED/BN
	16	times of SIV								14	BN/AN
	17	times of taking LAX								12	BN/AN
	18	times of excessive EX								13	BN/AN

S.E. = Standard Error
MNSQ = Mean square
correl. = correlation
AN = Anorexia nervosa
BN = Bulimia nervosa
BED = Binge eating disorder
SIV = Self-induced vomiting
LAX = Laxative use
EX = excessive exercise

The purpose of the study was to develop a measure that is suitable as a sessional outcome measure, which is likely to be weekly. The response scale was therefore recoded from a 28 day reference to seven days, corresponding with the collapsed categories (see Table 3).

Table 3: ***Change in response scale categories from original to 4-point scale***

EDE-Q	Response Scale	**0**	**1**	**2**	**3**	**4**	**5**	**6**
	Days/mth	0	1-5	6-12	13-15	16-22	23-27	28-31
EDE-QS	Collapsed Response Scale	**0**	**1**	**1**	**2**	**2**	**3**	**3**
	Days/mth	0	1-12		13-22		23-31	
	Days/week	0	1-2		3-5		6-7	

This resulted in a shortened version of the EDE-QS.

1.3 Discussion

The aim of Phase 1 of this study was to develop a shortened version of the EDE-Q (version 6.0) from questionnaire responses of people presenting with a wide range of eating disorders. Through a combination and integration of several methods (i.e. PCA, Rasch modelling and expert opinion) a 12-item questionnaire, the EDE-QS, was produced.

The exploratory PCA produced a five-factor model that did not replicate the original EDE-Q subscales. This is consistent with other studies, which carried out factor analytic methods and arrived at several different factor structures for bariatric surgery candidates (Hrabosky et al., 2008), the general eating disorder population (Aardoom, Dingemans, Slof Op't Landt & Furth, 2012; Allen et al., 2011) and women with bulimic symptoms (Peterson

et al., 2007). There was a significant overlap of this study's Factor 2 and 3, which refer to Body Dissatisfaction and Dietary Restraint, with the factors found in more recent exploratory PCA studies (Peterson et al., 2007; White et al., 2014). The results were however less consistent for the remaining factors. The observed inconsistencies may be due to different eating disorder populations across studies. There were, however, also differences in factor structure when comparing this study's mixed eating disorder population to that of Aardoom et al. (2012).

These findings add to the evidence that there is not sufficient psychometric support for the existing EDE-Q subscales. However, this and other studies have not found a consensus on a more accurate factor structure either.

The conducted Rasch analyses identified problems with the EDE-Q's response scale, which refers to a time frame of 28 days, divided into seven possible response options. Respondents most commonly selected the most extreme response options ("no days" and "every day"). Further, the Rasch thresholds between categories were disordered. This means that the given categories are not selected in a way consistent with the respondents' severity of eating disorder (Bond & Fox, 2010). For example, more severely impaired persons may have selected response option three (13-15 days), whereas only mildly impaired people chose option four (16-22 days). This may have arisen due to differences and difficulties in calculating the exact number of days within one month. Observed differences between EDE-Q and EDE ratings have been thought to be due to problems in retrieval strategies (Mond et al., 2004b). During the EDE interview memory prompts are given by using a calendar, which may enhance recall (Fairburn, 2008). It is unclear which

retrieval strategies are applied during the self-report questionnaire; it is however likely that there is a high cognitive demand which may impact on respondents' accuracy of recall.

Based on Rasch's rating scale diagnostics, the original seven-item response categories were changed to a four-point rating scale, which markedly improved the items' response curves, although people were still most likely to endorse the end points of the scale, namely "no days" or "every day". The reference point for the scale was also changed from 28 days to seven, which is likely to reduce the cognitive demand on people to compute and remember the frequencies of their thoughts, attitudes and behaviours. It would therefore be interesting for a future study to run rating scale diagnostics on EDE-QS data to assess whether more people now endorsed intermediate responses. In addition, referring to the past week helps to obtain more accurate diagnostic criteria of symptoms being present at least once per week, given the questionnaire has been completed sessionally over three consecutive months (American Psychiatric Association, 2013). By providing scaled response options for the frequency items, it is also more likely that missing or unreadable data can be minimised.

The Rasch model and results of the expert survey aided in the selection of items to remove and retain for the shortened version. By combining statistical and expert based methods it was aimed to develop a shortened version that is psychometrically as well as theoretically sound to adequately capture the construct of eating disorders (Coste, Guillemin, Pouchot & Fermanian, 1997). Overall, this resulted in a final scale of 12 items.

2. Phase 2: Psychometric validation of the EDE-QS

2.1 Methods

2.1.1 Participants and Procedure

An email appeal was sent out to all students of a London university providing a link to an online survey. The same link was advertised on the website of a charity supporting current and former sufferers of eating disorders. The link was further emailed to the charity's email distribution list. This resulted in 559 people completing the survey, which consisted of several online questionnaires, chosen to aid in validating the EDE-QS. Respondents were invited to provide their email address so that they could be contacted again a few days later to complete the EDE-QS only. Of 482 people who were contacted again, 335 (69.5%) completed the EDE-QS a second time.

The research study underwent proportionate review and was approved by a NHS ethics committee. All participants were given information about the scope and aims of the study, confidentiality and data protection (see Appendix C for ethical approval letter and participant information). By participating they were able to enter a draw to win one of two £50 vouchers for a store of their choice. Participation in the Retest survey meant that their names were entered twice in the raffle.

2.1.2 Measures

The original *EDE-Q* and the *EDE-QS* was included as previously described. The EDE-Q showed excellent internal consistency in the current sample (α = .96).

Sociability Scale. The five-item Sociability Scale as described by Cheek and Buss (1981) is rated on a five-point Likert scale from zero ("not at all") to four ("extremely"). It asks respondents to rate the extent of which statements, such as "I like to be with people", are typical of them. The current study showed good internal consistency of the scale (α = .81).

Clinical Impairment Assessment (CIA). The CIA is a 16-item measure developed for the purpose of assessing impairments secondary to eating disorders. It enquires about an individual's personal, cognitive and social functioning on a four-point Likert scale. It has shown good psychometric properties and is useful in predicting eating disorder case status (Bohn et al., 2008). The CIA's internal consistency was excellent in this study (α = .96).

Generalised Anxiety Disorder Questionnaire (GAD-7). The GAD-7 is a brief screening instrument for generalised anxiety disorder. Respondents are asked to rate the occurrence of anxiety symptoms on a four-point Likert scale over the past two weeks, ranging from "not at all" to "nearly every day". Spitzer et al. (2006) demonstrated that the GAD-7 had good validity when the results of the questionnaire were compared to independent mental health diagnoses, functional status measures, disability days and health care use. It also showed good validity and reliability in the general population (Loewe et al., 2008). Internal consistency in the current study was very good (α = .92).

Patient Health Questionnaire (PHQ-9). The PHQ-9 is a brief depression scale with a four-point Likert response option and is widely used as a standardised sessional measure in UK mental health services. It assesses depressive symptomatology within the last two weeks, which ranges from "no days" to "every day". It has shown to be a reliable and valid assessment

instrument of depression severity (Kroenke et al., 2001). The PHQ-9 showed strong internal consistency in this study (α = .91).

SCOFF Questionnaire. The SCOFF is a brief five-item screening questionnaire designed to detect eating disorders (Morgan, Reid & Lacey, 1999). Questions about key characteristics of anorexia and bulimia nervosa can be answered with a "yes" or "no" response. Two "yes" responses or more indicate that it is likely that the person may be suffering from an eating disorder. The measure showed good validity in comparison to a clinical interview for eating disorders (Hill, Reid, Morgan & Lacey, 2010). Internal consistency for the current study as measured by Cronbach's alpha was low (α = .64).

Short Evaluation of Eating Disorders (SEED). This questionnaire is a brief eating disorder assessment instrument. It consists of six items (total of 13 questions) from which an anorexia (ANTSI) and bulimia total severity index (BNTSI) can be derived. These range from a score of zero ("no symptoms") to three ("extreme symptoms"). It demonstrated acceptable construct validity and was also able to discriminate between eating disorder cases and non-cases (Bauer, Winn, Schmidt & Kordy, 2005). The SEED demonstrated an acceptable level of internal consistency in the current study (α = .76).

The World Health Organization Quality of Life (WHOQOL)-BREF. The WHOQOL-BREF is a shorter version of the original international quality of life questionnaire (WHOQOL-100). It is a 26 items measure that enquires about four domains: physical health, psychological health, social relationships and environment. It has been evaluated internationally and demonstrated to be a

valid and reliable instrument (Amir et al., 2000). Internal consistency for this study was satisfactory for physical health (α = .77), social relationships (α = .73) and environment (α = .78). It was good for psychological health (α = .88).

2.1.3 Statistical Analysis

All analyses were carried out using SPSS (version 21).

Missing data: Surveys were included if the respondent had completed the EDE-QS. All scaled questionnaires (excluding demographic information) had a forced response to their items so that it was impossible to continue with the survey if any items were skipped. This resulted in no missing data for EDE-Q, EDE-QS, Sociability Scale, CIA and GAD-7. Five participants did not complete the PHQ-9, six the SCOFF, seven the SEED and nine the WHOQOL-BREF as these respondents terminated the survey before completion. For these, missing values were entered into SPSS.

Preliminary Normality Testing

Tests of normality were carried out to determine whether responses to the scales were normally distributed. If this was the case, Pearson's correlation co-efficient was used for correlational analyses. For non-normally distributed data, Spearman's Rho was applied.

The non-parametric Whitney-U test and chi square analyses were used to examine differences between two groups.

Reliability

Internal consistency: Cronbach's alpha coefficient was calculated to assess the homogeneity of the EDE-QS scale.

Kline (1999) suggested that Cronbach's alpha above .8 indicates good

reliability. Bland and Altman (1997) emphasised that measures for clinical applications should have a Cronbach's alpha of at least .9.

Test-retest reliability: An Intra Class Correlation coefficient (ICC) was computed between the overall EDE-QS at two administrations, using a two way random model and type absolute agreement.

Validity

Convergent validity: Tests of convergent validity were carried out. It was hypothesised that positive correlations would be obtained between the EDE-Q, SEED, ANTSI and BNTSI, SCOFF and CIA. It was also expected that there would be positive but possibly weaker correlations with GAD-7, PHQ-9 and WHOQOL-BREF. The analyses were carried out separately for those respondents who reported currently not to be suffering from an eating disorder and those who stated they did. Statistically significant correlations were considered to have a small (+/- 0.1), medium (+/- 0.3) or large (+/- 0.5) effect size (Field, 2005).

As the change in the EDE-QS response scale meant that it was not possible to directly compare the frequencies for the behavioural EDE-Q items (e.g. binge eating) within a one-month period, the kappa statistics was utilised to assess the chance-corrected level of agreement between the measures, i.e. the absence or presence of specific regular behaviours (at least once/week).

Divergent validity: It was expected that there would not be a specific association between the EDE-QS and the Sociability Scale as Miller et al. (2008) have not found a specific association of sociability and eating difficulties in a non-clinical student population.

Sensitivity

An independent samples test was conducted to examine the difference between the EDE-QS scores of people with and without current self-reported eating disorders. It was hypothesised that there would be statistical differences between both groups.

2.2 Results

Participants' characteristics

In total, 559 people completed the online survey. Of these, 54 (9.7%) self-reported that they currently suffered from an eating disorder. It was not possible to verify diagnoses. It was reported that 25 (46.3%) with eating disorders were alerted to the study through the university email appeal and another 25 (46.3%) were recruited through advertisement provided by the eating disorder charity. In contrast, 491 (97.6%) people, who stated that they did not have an eating disorder, heard about the study through their university email.

The demographic information of the participants is presented in Table 4. The sample was further divided into those that stated they were currently suffering from an eating disorder and those that said they were not. Chi square analysis revealed that fewer men reported to be suffering from an eating disorder (χ^2 (1) =11.45, p<.001). There were also differences in levels of education between the groups (χ^2 (7)=39.36, p<.001), in that more people with an eating disorder than expected reported "basic schooling", "higher education" and "vocational and work qualifications" as their highest obtained qualification.

Table 4*: Participants' demographic information*

Participant demographics		**all (N=559)**	**current ED* (N=54)**	**no current ED* (N=503)**
		%	**%**	**%**
Age	18-24	60.6	64.8	60.2
	25-34	31.7	22.2	32.6
	35-44	5.2	7.4	5
	45-54	1.1	1.9	1
	55-64	0.7	1.9	0.6
	65-74	0.5	0	0.6
	missing	0.2	1.9	0
Gender	Male	19.1	1.9	20.9
	Female	80.9	98.1	79.1
Ethnicity	White British	43.8	68.5	41.2
	White Irish	2.1	0	2.4
	White Other	31.8	18.5	33.2
	Black Caribbean	0.4	0	0.4
	Black African	1.1	1.9	1
	Indian	2.7	0	3
	Pakistani	0.4	0	0.4
	Bangladeshi	0.7	0	0.8
	Other Asian	2.3	1.9	2.4
	Mixed - White & Caribbean	0.2	0	0.2
	Mixed - White & African	0.4	0	0.4
	Mixed - White & Asian	1.6	0	1.8
	Mixed - Other	1.6	1.9	1.6
	Chinese	7.5	7.4	7.6
	Any other ethnicity	2.9	0	3.2
	Prefer not to say	0.4	0	0.4
	missing	0.2	0	0.2
Education	Basic schooling	0.4	3.7	0
	Higher education	30.8	48.1	29
	Basic university	58.1	40.7	60
	Vocational & work	0.2	1.9	0
	Higher university	6.4	5.6	6.4
	Professional qualification	2	0	2.2
	Foreign qualification	2	0	2.2
	Other	0.2	0	0.2
Past ED* diagnosis		13.2	66.7	7.6
Current ED*	yes	9.7	100	0
	no	90	0	100
	missing	.4	0	0

*ED = eating disorder

2.2.1 Reliability

Internal consistency: Cronbach's alpha coefficient showed that internal consistency was high (α = .913).

All items correlated with the overall scale with item-total correlations ranging from .43 to .8. Apart from questions 7 and 8, deletion of any items would not result in an improved Cronbach's alpha value. If items 7 and 8 were deleted, Cronbach's alpha would increase to .914, which is not a substantial increase in reliability.

Test-retest reliability: Participants completed the EDE-QS for the second time on average 7.4 days later with a minimum of two and a maximum of 29 days. The mean total scores of the EDE-QS at time 1 was 7.19 (*SD*=6.4) and time 2 was 7.48 (*SD*=6.31). The ICC demonstrated a high degree of temporal stability (*ICC* =.92; p<.001) with a 95% confidence interval from .91 to .94.

2.2.2 Validity

Convergent validity: There were significant and high correlations between the EDE-QS and the EDE-Q as predicted. The EDE-QS showed further strong positive associations between other measures of eating disorder pathology as anticipated, although there was only a medium effect size for the BNTSI of the SEED. As hypothesised, the correlations with measures of general psychological functioning showed slightly weaker, albeit statistically significant, correlations. As expected, there were negative correlations with the WHOQOL-BREF domains, with the exception of the 'environment' domain for people with a current eating disorder. This indicated that higher EDE-QS scores were associated with reduced Quality of Life scores.

Divergent validity: Consistent with the initial hypothesis, there was no apparent association between the EDE-QS and the measure of sociability for people without a current eating disorder. However, contrary to expectations, there was a negative correlation for people who were currently suffering from an eating disorder, indicating that high scores on the EDE-QS are associated with lower sociability.

Table 5 shows all correlations between the EDE-QS and the individual measures.

Table 5: ***Convergent and discriminant validity correlations for EDE-QS***

	EDE-QS (no ED) *r*	**EDE-QS (current ED)** *r*
Convergent		
EDE-Q	.91**	.9**
CIA	.82**	.85**
SCOFF	.6**	.58*
SEED ANTSI	.64**	.55**
SEED BNTSI	.53**	.35*
GAD-7	.4**	.5**
PHQ-9	.51**	.66**
WHOQOL-BREF		
Physical health	-.35**	-.31*
Psychological	-.54**	-.52**
Social relationships	-.28**	-.47**
Environment	-.26**	-.06
Divergent		
Sociability Scale	-.07	-.44**

* p<0.05
** p<0.01
ED=eating disorder

Frequencies results

The chance corrected agreement between the EDE-QS and EDE-Q ratings of presence of at least one behaviour per week for people with self-reported eating disorders was excellent for days of binge eating (*kappa*=.7, *t*=5.3, *p*<0.001), for laxative use or self-induced vomiting (*kappa*=.84, *t*=6.24,

$p<0.001$) and for excessive exercise (*kappa*=.89, $t=6.45$, $p<.001$) (Landis & Koch, 1977).

2.2.3 Sensitivity

People who reported to currently be suffering from an eating disorder (*Mdn*=17.5) scored higher on the EDE-QS than those who indicated not to have an eating disorder *(Mdn*=5.0; *U*=3209.5, $p<.001$).

Completion time

The majority of respondents of the test-retest survey (*N*=276; 82.2%) completed and submitted the EDE-QS within three minutes. It was impossible to obtain an estimate for completion time of the full EDE-Q. However, Fairburn and Beglin (1994) reported that their respondents completed the pen and paper version within 15 minutes.

2.3 Discussion

The second phase of this study provided preliminary results for the psychometric evaluation of the EDE-QS. The shortened questionnaire demonstrated high internal consistency and excellent test-retest reliability. It further showed good convergent validity with the long version, other eating disorder measures and measures for anxiety, depression and aspects of quality of life as hypothesised, both for people with and without an eating disorder. There was a high consistency between the EDE-Q and the EDE-QS with regards to reporting behaviours typical of eating difficulties, i.e. binge eating, self-induced vomiting, laxative use and excessive exercise. The EDE-QS is also sufficiently sensitive to differentiate between people who report to be suffering from an eating disorder and those who do not.

Analyses on the respondents' demographic information showed that men were underrepresented in the group with eating disorders. This is consistent with findings from current research on gender differences in eating disorder prevalence rates (Hilbert, de Zwaan & Braehler, 2012; National Institute for Clinical Excellence, 2004; Striegel-Moore, et al., 2009). People without eating disorders in this study were also more likely to have obtained higher levels of formal education. However, this was to be expected as the majority of people without eating disorders were recruited from a university population.

The high correlations with the original EDE-Q suggests that the most relevant and informative items have been retained in the shortened version. Aardoom et al. (2012) established that the EDE-Q's global score is a valid indicator for a person's level of eating disorder severity. It is therefore likely, that the EDE-QS total score may be similarly sensitive to eating disorder impairment. This however needs to be further investigated.

The EDE-QS also showed large positive associations with other measures of eating disorder pathology, with the exception of the SEED's BN TSI, which was of a medium strength. However, the SEED's initial validation study (Bauer et al., 2005) showed similarly small correlation coefficients (r=.32) when correlated with the Eating Disorder Inventory (Garner, Olmsted & Polivy, 1983), a well-established and comprehensive assessment instrument for measuring symptoms of anorexia and bulimia nervosa. Bohn et al. (2008) observed a correlation coefficient of .89 between the EDE-Q and the CIA. This is comparable to that obtained between the EDE-QS and the CIA in the general population (r=.82) and in people with an eating disorder (r=.85) in this study. As expected, an increase in problematic eating

behaviours and attitudes resulted in an increase of feelings of anxiety and low mood, as these are known to be highly comorbid (Fornari et al., 1992). As hypothesised, negative correlations between the EDE-QS and measures of quality of life were found, which is consistent with other research studies (Jenkins, Hoste, Meyer & Blissett, 2011). The only exception was the domain of 'environment', which did not show a significant negative association with the EDE-QS for people with eating disorders. Environmental factors, such as financial resources and transport, may not necessarily be compromised with an increase in eating difficulties. It is, however, unclear why a significant negative correlation was found in people without an eating disorder.

Although the EDE-Q's internal and temporal reliability has been investigated and well established (Luce & Crowther, 1996; Mond et al., 2004a; Reas et al., 2006), as well as its convergent validity with the EDE subscales (e.g. Binford et al., 2005), to the author's knowledge there are no research studies assessing the EDE-Q global score's convergent validity with other measures of eating disorders (Berg et al., 2012). It is therefore not possible to assess how well the EDE-QS performed in comparison to the EDE-Q.

There was no particular association between the EDE-QS and a measure of sociability in people without an eating disorder, as expected. However, people who identified themselves as having an eating disorder showed a negative correlation between the EDE-QS and the Sociability scale, which was contrary to initial prediction. However, considering that in Miller et al.'s (2008) study, on which the hypothesis was based, participants were recruited from the general population, this may not be surprising. Whilst non-clinical individuals' degree of sociability may not be related to their eating

behaviours and attitudes, people who are suffering from an eating disorder may feel less sociable and have reduced social contacts. Items of the CIA (Bohn et al., 2008) refer to aspects of sociability (e.g. "stopped you going out with others" or "interfered with your relationship with others") and impairments in these are predictive of eating disorder problems. Therefore, a negative correlation, as observed in this study, was perhaps to be expected. It proved difficult to compare the reported frequencies of behaviours characteristic to eating problems, such as binge eating and compensatory behaviours, for the EDE-Q and EDE-QS due to a change in the referenced time frame (past 28 days vs seven days). In addition, the EDE-QS has a Likert scale format for frequency items, which makes it impossible to capture people's exact estimates. However, it was possible to compare the number of people who had engaged in these behaviours regularly, i.e. at least once a week, which corresponds with the diagnostic criteria as set out by the DSM-V (American Psychiatric Association, 2013). Although this is a fairly rough comparison, it indicated high agreement between the EDE-QS and the EDE-Q.

The analyses further revealed that the EDE-QS showed significantly higher total scores for people with eating disorders, which suggests that it may have the potential to distinguish between different levels of eating disorder severity. Further research needs to be carried out to establish whether the EDE-QS is also sensitive enough to capture change in eating disorder severity, which would be of immense clinical importance.

Overall, the findings of the current study suggest that the EDE-QS is an effective, reliable and valid measure for assessing eating disorder pathology.

General Discussion

This study developed and validated a shortened version of the EDE-Q (Fairburn & Beglin, 2008). Through the use of a multi-method approach, a 12-item questionnaire, the EDE-QS, was developed. Overall, the psychometric evaluation of the EDE-QS demonstrated that it is valid and reliable measure.

There are other brief questionnaires which were developed to assess eating disorders, such as the SEED, SCOFF, Eating Disturbance Scale (EDS-5; Rosenvinge et al., 2001) or the Eating Disorder Examination – Screen (EDE-S; Beglin & Fairburn, 1992). However, these questionnaires were either developed as screening instruments or focus specifically on anorexia and bulimia nervosa. The EDE-QS has undergone rigorous and multi-method development resulting in a brief questionnaire which retains sensitivity to people's severity of eating difficulties.

Clinical implications

Due to its brevity (general completion time within three minutes) and revised response categories, the EDE-QS will lend itself to being used as a sessional measure and therefore permit ongoing progress monitoring, which has demonstrated to improve clients' outcomes (Lambert, 2013). However, future research (some of which is currently underway in our group) needs to determine the ability of the EDE-QS to measure change over time. Based on Fairburn, Cooper, Shafran and Wilson's (2008) transdiagnostic protocol for the treatment of eating disorders, positive outcomes are more likely if change in eating behaviours and symptoms occurs within the first six weeks of starting therapy and should therefore be a focus of attention. Continuous outcome monitoring referring to the past week can therefore provide valuable

information to the clinician. Changes or in fact the absence of changes in symptoms may provide useful material for therapeutic discussions and could help the clinician shape the intervention, which researchers view as the key element for the observed benefits of progress monitoring (Boyce & Brown, 2013; Lambert & Shimokawa, 2011).

Anorexia nervosa remains one of the most challenging eating disorders and treatment attempts are often unsuccessful with potentially devastating consequences for the client (Arcelus et al., 2011; Wilson, Grilo & Vitousek, 2007). It is still unclear which types of treatment and more specifically which elements of treatment are most beneficial (Attia, 2011; Bulik, Berkman, Brownley, Sedway & Lohr, 2007). A progress monitoring instrument could be utilised in research studies to identify moments of change in eating attitudes and behaviours and may help to shed some light on the most helpful therapeutic sessions.

Limitations of study

There are several limitations to the current study. There was no ethnicity data available for the archival EDE-Q data, which raises uncertainties about the generalisability of the data. However, the large and diverse sample was a strength of this phase of the study.

The sample size of consulted experts in this study was small, and a convenience sampling method was used, which may have limited generalisability. According to Okoli and Pawlowski (2004), opinions gathered by ten people are sufficient to obtain general agreement. However, as the expert opinion had a high bearing on the inclusion and exclusion of items on

the questionnaire, a larger sample may have been desirable to ensure consensus.

For the psychometric analysis, the sample was divided into people who currently had an eating disorder and those who reported that they did not. This is a crude measure of eating disorder diagnosis and the use of diagnostic assessments or interviews would have been more accurate. However, it was not possible to conduct diagnostic assessments within the remit and time constraints of this thesis. The number of people who identified themselves as having an eating disorder was relatively small (N=54). A post-hoc power analysis, however, revealed that based on the obtained correlation of r=.9, it was large enough to have a 95% chance of correctly rejecting a null hypothesis of r=.7 at the .05 significance level. The minimum required sample size was 32. It would however be desirable to continue the EDE-QS's psychometric evaluation using a larger eating disorder sample.

Another major drawback of the study was the lack of service user involvement in the development of the questionnaire. The initial project proposal set out to obtain service user feedback on the EDE-QS to establish its acceptability and utility of the items and wordings to respondents. Due to time constraints this part of the research project had to be regretfully dropped.

Future research

Future studies should assess the EDE-QS's sensitivity to change. Pre-and post-treatment EDE-Q and EDE-QS data in two eating disorders services are currently being collected. Their analyses will provide helpful information about the new measure's ability to detect change over the course of psychological treatment.

It would be desirable to establish clinically significant change indices or cut-off points to differentiate between non-clinical and clinical impairment in eating disorders. This could guide clinicians with regard to treatment planning and prioritising.

It would have been useful to assess the psychometric validity of the EDE-QS by diagnostic group but the sample size was too small for this. It would therefore be useful if future studies validated its psychometric properties for people with BN, AN, BED and OSFED to provide evidence for its applicability across diagnostic groups.

Future research should also investigate the acceptability and comprehensibility of the EDE-QS amongst service users.

References

Aardoom, J.J., Dingemans, A.E., Slof Op't Landt, M.C.T. & Van Furth, E.F. (2012). Norms and discriminative validity of the eating disorder examination questionnaire (EDE-Q). *Eating Behaviors, 13*, 305-309.

Allen, K.L., Byrne, S.M., Lampard, A., Watson, H. & Fursland, A. (2011). Confirmatory factor analysis of the eating disorder examination-questionnaire (EDE-Q). *Eating Behaviours, 12,* 143-151.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Amir, M., Fleck, M., Herrman, H., Lomanchenkov, A., Lucas, R. & Patrick, D. (2000). Reliability, validity and reproducibility of the WHOQOL-Bref in six countries. *Quality of Life Research, 9*, 320.

Arcelus, J., Mitchell, A.J., Wales, J. & Nielsen, S. (2011). Mortality rates in patients with anorexia nervosa and other eating disorders. A meta-analysis of 36 studies. *Archives of General Psychiatry, 68*, 724-731.

Attia, E. (2011). Anorexia nervosa: current status and future directions. *Annual Review of Medicine, 61*, 425-435.

Bauer, S., Winn, S., Schmidt, U. & Kordy, H. (2005). Construction, scoring and validation of the short evaluation of eating disorders (SEED). *European Eating Disorders Review, 13*, 191-200.

Beglin, S.J. & Fairburn, C. G. (1992). Evaluation of a new instrument for the detection of eating disorders in community samples. *Psychiatry Research, 44*, 191-201.

Berg, K.C., Peterson, C.B., Frazier, P. & Crow, S.J. (2012). Psychometric evaluation of the eating disorder examination and eating disorder examination-questionnaire: A systematic review of the literature. *International Journal of Eating Disorders, 45*, 428-438.

Binford, R.B., Le Grange, D. & Jellar, C.C. (2005). Eating disorders examination versus eating disorders examination-questionnaire in adolescents with full and partial-syndrome bulimia nervosa and anorexia nervosa. *International Journal of Eating Disorders, 37*, 44-49.

Bland, J.M. & Altman, D.G. (1997). Statistics notes: Cronbach's alpha. *British Medical Journal, 314*, 572.

Bohn, K., Doll, H.A., Cooper, Z., O'Connor, M., Palmer, R.L. & Fairburn, C.G. (2008). The measurement of impairment due to eating disorder psychopathology. *Behaviour Research and Therapy, 46,* 1105-1110.

Bond, T.G. & Fox, C.M. (2010*). Applying the rasch model: Fundamental measurement in the human sciences* (2nd ed.). New York: Routledge.

Boyce, M.B. & Browne, J.P. (2013). Does providing feedback on patient-reported outcomes to healthcare professionals result in better outcomes for patients? A systematic review. *Quality of Life Research, 22*, 2265-2278.

Bulik, C.M., Berkman, N.D., Brownley, K.A., Sedway, J.A. & Lohr, K.N. (2007). Anorexia nervosa treatment: A systematic review of randomized controlled trials. *International Journal of Eating Disorders, 40*, 310-320.

Carter, J.C., Aime, A.A. & Mills, J.S. (2000). Assessment of bulimia nervosa: A comparison of interview and self-report questionnaire methods. *International Journal of Eating Disorders, 30,* 187-192.

Cheek, J.M. & Buss, A.H. (1981). Shyness and sociability. *Journal of Personality and Social Psychology, 41*, 330-339.

Comrey, A.L. & Lee, H.B. (1992). *A first course in factor analysis* (2nd edition). Hillsdale, NJ: Erlbaum.

Cooper, Z., Cooper, P.J. & Fairburn, C.G. (1989). The validity of the eating disorder examination and its subscales. *British Journal of Psychiatry, 154*, 807-812.

Coste, J., Guillemin, F., Pouchot, J. & Fermanian, J. (1997). Methodological approaches to shortening composite measurement scales. *Journal of Clinical Epidemiology, 50*, 247-252.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. P*sychometrika, 16,* 297-334.

Das Nair, R., Moreton, B.J. & Lincoln, N.B. (2011). Rasch analysis of the nottingham extended activities of daily living scale. *Journal of Rehabilitation Medicine, 43*, 944-950.

Department of Health (2011). *No health without mental health. Delivering better mental health outcomes for people of all ages*. London: Department of Health.

DeVellis, R.F. (2012). *Scale development: theory and applications* (3rd edition). London: Sage Publications.

Doll, H.A., Petersen, S.E. & Stewart-Brown, S.L. (2005). Eating disorders and emotional and physical well-being: Associations between student self-reports of eating disorders and quality of life as measures by the SF-36. *Quality of Life Research, 14*, 705-717.

Fairburn, C.G. (2008). *Cognitive behavior therapy and eating disorders*. New York: Guilford Press.

Fairburn, C.G. & Beglin, S.J. (1994). Assessment of eating disorder psychopathology: Interview or self-report questionnaire? *International Journal of Eating Disorders, 16*, 625-632.

Fairburn, C.G. & Beglin, S.J. (2008). Eating disorder examination questionnaire (EDE-Q 6.0). In: C.G. Fairburn (Ed.), *Cognitive behavior therapy and eating disorders*. New York: Guilford Press.

Fairburn, C.G. & Cooper, Z. (1989). Eating disorders. In K. Hawton, P.M. Salkovskis, J.W. Kirk & D.M. Clark (Eds.), *Cognitive-behavioural approaches to adult psychiatric disorder: A practical guide*. Oxford: Oxford University Press.

Fairburn, C.G. & Cooper, Z. (1993). The eating disorder examination (12th ed.). In: C.G. Fairburn & G.T. Wilson (Eds.), *Binge eating: Nature, assessment and treatment.* New York: Guilford Press.

Fairburn, C.G., Cooper, Z., Shafran, R. & Wilson, G.T. (2008). Eating disorders: A transdiagnostic protocol. In: D.H. Barlow (Ed.), *Clinical handbook of psychological disorders* (4th ed.). New York: Guildford Press.

Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage Publications Ltd.

Fornari, V., Kaplan, M., Sandberg, D.E., Matthews, M., Skolnick, N. & Katz, J.L. (1992). Depressive and anxiety disorders in anorexia nervosa and bulimia nervosa. *International Journal of Eating Disorders, 12*, 21-29.

Franchignoni, F., Giordano, A., Sartorio, F., Vercelli, S., Pascariello, B. & Ferriero, G. (2010). Suggestions for refinement of the disabilities of the arm, shoulder and hand outcome measure (DASH): A factor analysis and rasch validation study. *Archives of Physical Medicine and Rehabilitation, 91*, 1370-1377.

Garner, D.M., Olmsted, M.P. & Polivy, J. (1983). Development and validation of a multidimensional eating disorder inventory for anorexia nervosa and bulimia. *International Journal of Eating Disorders, 2*, 15-34.

Greco, L.A., Lambert, W. & Baer, R.A. (2008). Psychological inflexibility in childhood and adolescence: Development and evaluation of the avoidance and fusion questionnaire for youth. *Psychological Assessment, 20*, 93-102.

Grilo, C.M., Masheb, R.M. & Wilson, G.T. (2001). A comparison of different methods for assessing the features of eating disorders in patients with binge eating disorder. *Journal of Consulting and Clinical Psychology, 69*, 317-322.

Hilbert, A., de Zwaan, M. & Braehler (2012). How frequent are eating disturbances in the population? Norms of the eating disorder examination questionnaire. *PLoS ONE 7(1):* e29125. doi:10.1371/journal.pone.0029125

Hill, L.S., Reid, F., Morgan, J.F & Lacey, J.H. (2010). SCOFF, the development of an eating disorder screening questionnaire. *International Journal of Eating Disorders, 43*, 344-351.

Hrabosky, J.I., White, M.A., Masheb, R.M., Rothschild, B.S., Burke-Martindale, C.H. & Grilo, C.M. (2008). Psychometric evaluation of the eating disorder examination-questionnaire for bariatric surgery candidates. *Obesity, 16*, 763-769.

Improving Access to Psychological Therapies (2008). *Improving access to psychological therapies (IAPT) outcomes toolkit 2008/9.* London: IAPT.

Improving Access to Psychological Therapies (2012). *IAPT three-year report. The first million patients.* London: IAPT.

Jenkins, P.E., Hoste, R.R., Meyer, C. & Blissett, J.M. (2011). Eating disorders and quality of life: A review of the literature. *Clinical Psychology Review, 31*, 113-121.

Jones, P.W., Harding, G., Berry, P., Wiklund, I., Chen, W.-H. & Kline Leidy, N. (2009). Development and first validation of the COPD assessment test. *European Respiratory Journal, 34*, 648-654.

Kalarchian, M.A., Wilson, G.T., Brolin, R.E. & Bradley, L (2000). Assessment of eating disorders in bariatric surgery candidates: Self-report questionnaire versus interview. *International Journal of Eating Disorders, 28*, 465-469.

Kline, P. (1999). *The handbook of psychological testing* (2nd ed.). London: Routledge.

Kroenke, K., Spitzer, R.L. & Williams, J.B.W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*, 606-613.

Lambert, M.J. (2013). Outcome in psychotherapy: The past and important advances. *Psychotherapy, 50*, 42-51.

Lambert, M.J. & Hawkins, E.J. (2004). Measuring outcome in professional practice: Considerations in selecting and using brief outcome instruments. *Professional Psychology: Research and Practice, 35*, 492-499.

Lambert, M.J. & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy, 48*, 72-79.

Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Latner, J.D., Mond, J.M., Kelly, M.C., Haynes, S.N. & Hay, P.J. (2014). The loss of control over eating scale: Development and psychometric evaluation. *International Journal of Eating Disorders,* doi: 10.1002/eat.22296

Linacre, J.M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*, 85-106.

Loewe, B., Decker, O., Mueller, S., Braehler, E., Schellberg, D., Herzog, W. & Herzberg, P.Y. (2008). Validation and standardisation of the generalised anxiety disorder screener (GAD-7) in the general population. *Medical Care, 46,* 266-274.

Luce, K.H. & Crowther, J.H. (1999). The reliability of the eating disorder examination – self report questionnaire version (EDE-Q). *International Journal of Eating Disorders, 25*, 349-351.

Lundstroem, M. & Pesudovs, K. (2009). Nine-item short-form rasch-scaled revision of the catquest questionnaire. *Journal of Cataract & Refractive Surgery, 35,* 504-513.

Miller, J.L., Schmidt, L.A. & Vaillancourt, T. (2008). Shyness, sociability, and eating problems in a non-clinical sample of female undergraduates. *Eating Behaviors, 9*, 352-359.

Mond, J.M., Hall, A., Bentley, C., Harrison, C., Gratwick-Sarll, K. & Lewis, V. (2014). Eating-disordered behaviour in adolescent boys: Eating disorder examination questionnaire norms. *International Journal of Eating Disorders, 47*, 335-341.

Mond, J.M., Hay, P.J., Rodgers, B., Owen, C. & Beumont, P.J.V. (2004a). Temporal stability of the eating disorder examination questionnaire. *International Journal of Eating Disorders, 36*, 195-203.

Mond, J.M., Hay, P.J., Rodgers, B., Owen, C. & Beumont, P.J.V. (2004b). Validity of the eating disorder examination questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour Research and Therapy, 42,* 551-567.

Morgan, J.F., Reid, F. & Lacey, J.H. (1999). The SCOFF questionnaire: Assessment of a new screening tool for eating disorders. *British Medical Journal, 319*, 1467-1468.

National Institute for Clinical Excellence (2004*). Eating disorders: Core interventions in the treatment and management of anorexia nervosa, bulimia nervosa and related eating disorders*. London: NICE.

National Institute for Mental Health in England (2008). *Mental health compendium*. London: NIMHE.

Okoli, C. & Pawlowski, S.D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management, 42*, 15-29.

Peterson, C.B., Crosby, R.C., Wonderlich, S.A., Joiner, T., Crow, S.J., Mitchell, J.E., … Le Grange, D. (2007). Psychometric properties of the eating disorder examination questionnaire: Factor structure and internal consistency. *International Journal of Eating Disorders, 40,* 386-389.

Reas, D.L., Grilo, C.M. & Masheb, R.M. (2006). Reliability of the eating disorder examination-questionnaire in patients with binge eating disorder. *Behaviour Research and Therapy, 44,* 43-51.

Rosenvinge, J.H., Perry, J.A., Bjorgum, L., Bergersen, T.D., Silvera, D.H. & Holte, A. (2001). A new instrument measuring disturbed eating patterns in community populations: Development and initial validation of a five-item scale (EDS-5). *European Eating Disorders Review, 9,* 123-132.

Roth, A. & Fonagy, P. (2005). *What works for whom? A critical review of psychotherapy research* (2nd ed.*).* London: The Guilford Press.

Siegert, R.J., Jackson, D.M., Tennant, A. & Turner-Stokes, L. (2010). Factor analysis and rasch analysis of the zarit burden interview for acquired brain injury carer research. *Journal of Rehabilitation Medicine, 42,* 302-309.

Simon, W., Lambert, M.J., Busath, G., Vazquez, A., Berkeljon, A., Hyer, K., … Berrett, M. (2013). Effects of providing patient progress feedback and clinical support tools to psychotherapists in an inpatient eating disorders treatment program: A randomised controlled study. *Psychotherapy Research, 23,* 287-300.

Slade, M., Thornicroft, G. & Glover, G. (1999). The feasibility of routine outcome measures in mental health. *Social Psychiatry and Psychiatric Epidemiology, 34*, 243-249.

Spitzer, R.L., Kroenke, K., Williams, J.B.W. & Loewe, B. (2006). A brief measure for assessing generalised anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*, 1092-1097.

Steinhausen, H.C. (2002). Outcome of eating disorders. *Child and Adolescent Psychiatric Clinics of North America, 18*, 225-242.

Striegel-Moore, R.H., Rosselli, F., Perrin, N., DeBar, L., Wilson, G.T., May, A. & Kraemer, H.C. (2009). Gender difference in the prevalence of eating disorder symptoms. *International Journal of Eating Disorders, 42*, 471-474.

Tabachnick, B.G. & Fidell, L.S. (2001). *Using multivariate statistics* (4th edition). Boston: Allyn & Bacon.

Tennant, A. & Conaghan, P.G. (2007). The rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a rasch paper? *Arthritis & Rheumatism (Arthritis Care & Research), 57*, 1358-1362.

Valderas, J.M., Kotzeva, A., Espallargues, M., Guyatt, G., Ferrans, C.E., Halyard, M.Y., … Alonso, J. (2008). The impact of measuring patient-reported outcomes in clinical practice: A systematic review of the literature. *Quality of Life Research, 17*, 179-193.

White, H.J., Haycraft, E., Goodwin, H. & Meyer, C. (2014). Eating disorder examination questionnaire: Factor structure for adolescent girls and boys. *International Journal of Eating Disorders, 47*, 99-104.

Wilfley, D.E., Schwartz, M.B., Spurrell, E.B. & Fairburn, C.G. (1997). Assessing the specific psychopathology of binge eating disorder patients: Interview or self-report? *Behaviour Research and Therapy, 35,* 1151-1159.

Wilson, G.T., Grilo, C.M. & Vitousek (2007). Psychological treatment of eating disorders. *American Psychologist, 62*, 199-216.

Wittchen, H.U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Joensson, B., … Steinhausen, H.-C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010*. European Neuropsychopharmacology, 21*, 655-679.

Williams, R.T., Heinemann, A.W., Bode, R.K., Wilson, C.S., Fann, J.R. & Tate, D.G. (2009). Improving measurement properties of the patient health questionnaire-9 with rating scale analysis. *Rehabilitation Psychology, 54*, 198-203.

Part 3: Critical Appraisal

Introduction

The following is a critical appraisal of the empirical research study that I undertook as part of my clinical psychology training. I will begin by describing how I became interested in undertaking this particular research project. I will then discuss the challenges faced by researchers who set out to develop a short outcome measure, with a particular focus on methodological issues. Further, I will discuss the utility of the new outcome measure, the EDE-QS, in clinical practice with people suffering from eating disorders by drawing on qualitative research and my personal experience of working in an eating disorders service.

Interest in developing a shortened EDE-Q

I was allocated to an Improving Access to Psychological Therapy (IAPT) service for my first placement on the clinical psychology course. Amongst the many people that I saw, was one young woman who presented with a strong urge to restrict her food intake and bulimic behaviours. Besides the usually administered IAPT outcome measures, I asked her to complete the Eating Disorder Examination Questionnaire (EDE-Q), as recommended by IAPT's minimum data set, at the beginning and end of our sessions. I scored her responses and remember the difficulty I had assessing her level of change using her pre and post treatment scores. There was some but not a huge amount of change in her scores. However, as the EDE-Q refers to the past 28 days, during which I had seen her four times, I was left unsure how much an impact our intervention (of 10 sessions) had made on these specific symptoms at the final session.

Working at IAPT I quickly became a proponent of sessional outcome measure collection. Despite initial worries that this may impact too much on my clients' time and nerves, I realised that completion rarely took longer than three minutes. Once both parties were used to it, the questionnaires were completed swiftly before each treatment session. However, even then I regretted that the wealth of information that we were collecting within IAPT was reduced to comparing pre-and post-treatment scores. I feel that I missed a great opportunity of discussing people's progress during treatment based on their questionnaire scores in supervision and/or therapy.

When my external supervisor mentioned the idea of producing a shortened version of the EDE-Q, I was quickly convinced that I had the motivation to develop a more user-friendly questionnaire that could be used in IAPT or eating disorder specialist services for sessional outcome collection.

Challenges in short form development

Development of a short form is a risky undertaking. Substantial effort and time has to be invested to construct and psychometrically validate a shortened version. It is essential that the short form remains sufficiently valid so that it can be accepted and used in clinical practice. However, by shortening a scale, its validity will inevitably be reduced (Smith, McCarthy & Anderson, 2000). Kruyen, Emons and Sijtsma (2014) also warn that fewer items reduce a questionnaire's ability to detect clinically meaningful change on an individual level. It is therefore important to carefully identify items that best represent the construct to be tested so that the result is a useful and sensitive measurement tool.

Validity of existing scale

Smith et al. (2000) consider it imperative to only shorten those questionnaires that have a robust evidence base of their validity. As the EDE-Q has been the subject of frequent psychometric investigation (Berg, Peterson, Frazier & Crow, 2012), I was convinced that its shortening was a safe undertaking. The most recent version is the EDE-Q 6.0 (Fairburn & Beglin, 2008). However, upon closer examination of the literature I noticed that the majority of studies were based on older versions of the EDE-Q, which included 36 (Binford, Le Grange & Jellar, 2005; Carter, Aime & Mills, 2000; Carter, Stewart & Fairburn, 2001; Mond, Hay, Rodgers, Owen & Beumont, 2004; Reas, Grilo & Masheb, 2006), 38 (Kalarchian, Wilson, Brolin & Bradley, 2000), 40 (Engelsen & Laberg, 2001) or 41 (Luce & Crowther, 1999) questionnaire items. Some studies did not clarify which version they had used at all (Allen, Byrne, Lampard, Watson & Fursland, 2011; Grilo, Masheb & Wilson, 2001) and others mentioned that they used the 3rd version, but did not specify the number of items included in this (Wilfley, Schwartz, Spurrell & Fairburn, 1997). It was of concern that the extent of the differences between these versions was unclear. I attempted to investigate this further.

I was able to find only the most recent version online, the EDE-Q 6.0, which was used in this study and contains a total of 28 scaled items. One of my supervisors provided me with an older, 36 item version, which was still used in their eating disorder service. The difference in number of items between these two versions is due to a re-structure of the frequency questions although the content remained mainly the same. However, one frequency item has been dropped completely. This referred to *subjective* binge eating (SBE), which implies the person had a sense of loss of control

during eating but may not have consumed an unusually large amount of food. It is unknown why this item has been dropped from the current version and is potentially controversial as research has shown that SBE can be a better predictor of eating disorder pathologies than objective binge eating (personal correspondence with Jon Mond; Latner, Mond, Mackenzie, Haynes & Hay, 2014). Further changes to the current version include adjustments in wording of almost all scaled items (e.g. the item *"Have you tried to avoid eating any foods which you like in order to influence your shape or weight?"* was reworded to *"Have you tried to exclude from your diet any foods that you like in order to influence your shape or weight (whether or not you have succeeded?"*). Although the two versions are extremely similar, it still remains uncertain whether both are understood and responded to in the same way. Both by searching the literature and contacting the lead author of the EDE-Q Professor Chris Fairburn, I attempted to determine whether any studies have investigated the comparability of the different versions. Chris Fairburn replied that *"the various iterations of the EDE-Q mostly involve subtle changes to the wording - I don't have the details. […] the important thing to know is that they are all compatible with earlier versions."* Although this response implied that Chris Fairburn did not consider the changes over time to be significant, the uncertainty of the current scale's validity still remains. It raises the question of whether the most recent and older versions are equally valid and reliable, as well as whether studies using different versions are comparable.

Deciding on appropriate methods for item reduction

Continuing with the research project, the first hurdle I had to overcome was to decide on the best method(s) of identifying the most useful items in the

existing scale. When I explored the literature, it became apparent that there were many recommendations on initial scale development but information on shortening a scale was limited. The default method used by most researchers appears to be classical test theory and factor analysis (Kruyen et al., 2014). The utility or appropriateness of these techniques, however, often remains unchallenged. There are several problems with factor analytic techniques, which include the following: They require a full data set, which often leads to imputing data; their mathematics are based on linearity but test scores are not; the appropriateness of the response scale remains unquestioned and intervals between Likert-scales are treated as equi-distanced, which is rarely the case; and analyses of different samples rarely provide the same factor structures (Wright, 1995). The latter is believed to be caused by factor analysis' inability to consider that people, in this context with different levels of psychological impairment, will respond in a way that is consistent with that particular item's sensitivity to measure a certain level of psychological distress. Hence, different samples will show a different response pattern.

These statistical shortcomings do not apply to Rasch Modelling, which is based on item response theory. After reading introductory texts, I became excited and fascinated by the theoretical underpinning and practical application. However, it posed the challenge of finding appropriate support as Rasch Modelling went beyond my statistical teaching and the expertise of everyone working in the clinical psychology department. Eventually and luckily I was introduced to PhD student Rob Saunders, who had experience in Rasch Modelling and was willing to support me with my research study. He pointed me towards further reading, gave advice on which software to use as well as giving me a space to discuss the results.

Integration of Statistical Methods

Because Rasch Modelling is still in its infancy or, perhaps more precisely, not yet widely integrated into the psychological sciences, there is no clear consensus on how to employ its techniques, especially in conjunction with factor analytic methods.

Some of the consulted studies carried out Rasch analyses first to assess a scale's unidimensionality and then conducted a confirmatory factor analysis (Cole, Rabin, Smith & Kaufman, 2004). Others carried out an exploratory factor analysis first to identify the scale's dimensions and used Rasch analysis on each factor for further evaluation of the scale (Franchignoni et al., 2010; Greco, Lambert & Baer, 2008). Some used Rasch analysis to assess the scales dimensions and did not include factor analytic methods at all (Jones et al., 2009) and others again used confirmatory factor analysis as well as assessing unidimensionality of each scale using the Rasch approach (Lamoureux et al., 2007). As a novice to the field working in a profession with limited statistical knowledge, it was a challenge to make commitments to using a particular methodology. This was particularly difficult, as none of the studies justified why they had chosen the type and order of the applied methods over others. Due to the inconsistent findings with regards to the subscales of the EDE-Q, the use of a confirmatory factor analysis was ruled out. Being unfamiliar with the Rasch model at this stage of the research process, I decided to start with exploratory principal component analysis. In hindsight, I would have also liked to explore the EDE-Q's dimensions using Rasch analysis.

Although exploratory principal component analysis is often applied in psychological research, I was again met with uncertainty when I had to interpret the statistical output. For example, it was not very clear how many factors should be extracted. Kaiser's criterion suggests including factors that have Eigenvalues of more than 1.0 (Kaiser, 1960). However, as with all cut-offs, the dividing line is fairly arbitrary and the question is whether factors that have Eigenvalues slightly above one should be included, which was the case for the current study. Consulting the scree plot is another subjective way of selecting a number of factors. It can be relatively straight forward if there is a distinct bend in the graph (DeVellis, 2012). However, a more gradually declining curve introduces once again an element of subjectivity, which was definitely the case for my data. I decided on the strict Eigenvalue cut-off of 1.0, which somewhat resonated with the scree plot, so that I could justify my decision-making processes with reference to the literature. However, it is reasonable to consider that another researcher with a different background and different research experience would have decided differently and only included three or four factors. This is likely to have had huge implications for the following Rasch analyses, which treated each factor as a separate dimension for analysis. Fewer factors would have established different dimensions and different items may have been identified as misfitting the Rasch model. This could have resulted in the deletion of a different set of items, which may have produced an alternative version to the current EDE-QS. Already, at this very first step in the research process, my decision-making influenced the fate of the final measure and leaves the study vulnerable to criticism.

Moving on in the research process, it became apparent that there are differences in the way Rasch modelling is applied. Some researchers only used certain aspects, such as assessment of unidimensionality, item severity and item misfit (Cole et al., 2004). Others made full use of it (Lundstroem & Pesudovs, 2009) and investigated the existing rating scale, model fit, item fit, differential item functioning, reliability, local independence of items and undimensionality as recommended by Tennant and Conaghan (2007). I excluded only those techniques that I believed to have been addressed by the exploratory factor analysis, i.e. assessment of unidimensionality, and those that seemed less relevant for the purpose of the current study, i.e. differential item functioning.

It is recommended to combine a number of techniques, specifically statistical and judgement-based methods, when developing a questionnaire. It is however less clear how to integrate these. I placed greater emphasis on statistical results based on the assumption that the EDE-Q had been shown to have high correlations with the "gold standard" of eating disorder assessment, the Eating Disorder Examination Interview, as recommended by Coste, Guillemin, Pouchot and Fermanian (1997). Expert ratings were however given priority if a particularly high score was obtained, indicating strong consensus amongst the experts, or if statistical methods were insufficient to guide further decision-making. It was hoped that in this way, a balance between statistical techniques and clinical importance was achieved.

To say the least, learning about and teaching myself a new statistical method has been anxiety and uncertainty provoking, whilst at the same time providing me with curiosity and excitement of applying an innovative model that appears to better meet the needs of scale development. However, it is

essential to have necessary supports in place to discuss and overcome difficulties and challenges.

By conducting a research study using statistical techniques, one assumes to circumvent the issue of subjective judgement, which is commonly raised as the major criticism of qualitative research (Dixon-Woods, Shaw, Agarwal & Smith, 2004). However, it soon became apparent that I was faced with a plethora of subjective decision-making due to a lack of well-established and agreed procedures and methods. Apart from dealing with uncomfortable feelings of uncertainty as the lead researcher, it raises questions about the assumed objectivity of the research process in scale development. It is unclear whether another researcher would have made the same decisions as me. It is possible that they would have developed a shortened version containing a different set of items. As researchers, we often think that the statistical tests applied to quantitative data are robust and reliable. Perhaps as a profession we should become more aware of the subjective decision-making involved within quantitative research methods.

Service user involvement

This study was conducted within a strict time-limit and I set myself an ambitious task to develop as well as validate a new questionnaire. Unfortunately, I had underestimated how long some parts of the project, such as the NHS ethics application, may take. After a serious delay I had to drop aspects of the research protocol and decided to forego the service user evaluation of the short version. This was to my great regret as I consider service users to be the true experts in this context. Hence, I feel service users should be consulted on the appropriateness of the scale and wording of individual items in future (DeVellis, 2012). Due to the lack of empirical data

on the adaption of the EDE-Q over time, it is unclear whether service users have previously been consulted and involved in the changes to its wording. Service user involvement and feedback has shown to be beneficial in questionnaire development (Chen, Tam, Wong, Law & Chiu, 2005) and its omission now poses a major limitation to this research study.

Service users' perception of eating disorder treatment has been investigated qualitatively and the findings suggest that treatment is highly valued if the underlying issues which led to eating difficulties are addressed and understood (De la Rie, Noordenbos, Donker & van Furth, 2008; Pettersen & Rosenvinge, 2002). It has also been found that people regard positive life events, in particular those that refer to improved social relations, as important steps in the recovery process. Further, self-acceptance and improved emotional expression and management were captured as essential for people who reflected on their recovery process (Federici & Kaplan, 2008; Pettersen & Rosenvinge, 2002). The EDE-Q and EDE-QS, however, focus on eating behaviours, attitudes and symptoms. The questionnaires therefore suggest symptom reduction as the main indicator of improvement or recovery, which is consistent with research on therapists' view of successful treatment (De la Rie et al., 2008). Pettersen and Rosenvinge (2002) found that some service users reported full recovery but explained that their attitudes towards food and body image remained problematic, even after successful treatment. A person's perceived recovery might therefore not necessarily translate into a reduction of eating disorder symptoms. The authors speculated that this may be due to the often ego-syntonic nature of eating pathology. Based on these findings, it would have been useful to also obtain service users' ratings on the importance for inclusion of EDE-Q items in

a short version. The qualitative research suggests that their ratings may have differed from those of the experts consulted in this study. In addition, service users may have felt that items relating to interpersonal and emotional issues are missing in the EDE-QS. Whilst this should be acknowledged, it is important to remember that the purpose of this study was to reduce an existing questionnaire as opposed to developing a new measure.

From my personal experience of working in an eating disorder service I learnt that a person's eating behaviour often fulfilled an essential function in their lives. Some people binged to deal with uncomfortable feelings and others restricted food as a means of emotional avoidance, self-punishment or to communicate distress that otherwise went unnoticed. The main focus of treatment on an inpatient unit for anorexia was to increase food intake and avoid compensatory behaviours because of the severe medical risks associated with low weight. However, it seems essential to provide individuals with alternative coping strategies so that they can deal with those distressing experiences that resulted in disordered eating in the first place in a more helpful way. I observed that several individuals followed their meal plans and gained weight, which would have resulted in some reduction of their score on the EDE-QS and in fact on any other eating disorder questionnaire. Their levels of distress, however, were as high as ever as their means of coping with emotional challenges had been removed. It is therefore essential to consider questionnaire scores in addition to other clinical information about a person, and not view them as a substitute. I also wonder whether an eating disorder specific outcome measure like the EDE-QS, given its focus on symptoms and behaviours, should be used in

conjunction with an additional measure, perhaps a brief quality of life questionnaire.

The EDE-QS's emphasis on behaviours and attitudes further suggests that improvement of the eating disorder is dependent exclusively on individual change. It disregards a person's interpersonal difficulties and social context, which have been suggested as potential etiological factors for developing an eating disorder (Rieger et al., 2010). Changes in these domains can be essential factors contributing to recovery (Federici & Kaplan, 2008). These are, however, not captured in the EDE-QS and practitioners should bear this in mind when providing treatment for eating disorders and evaluating change over time.

To summarise, it would have been extremely useful if service users had been involved in this project. Their views on the importance of items for inclusion in the EDE-QS and on the wording of the final items would have increased the measure's acceptability and ease of comprehension. When using the EDE-QS in clinical practice, it is important to consider that despite the presence of eating disorder symptoms, people may feel better and regard themselves as recovered. The opposite may, however, also occur in that people's symptoms reduce but their levels of distress remain high. In addition, there are several aspects to an individual's life which contribute to recovery that are not captured in the EDE-QS. It is therefore essential that clinicians using the EDE-QS are aware of the limitations mentioned here and that its scores are interpreted in conjunction with a person's subjective account.

Conclusion

Whilst I advocate and strongly support the use of psychological outcome measures, a sceptical stance needs to be retained. Firstly, health professionals need to be aware of the methodological limitations in scale development when choosing questionnaires for their clinical practice. Prior to commencing this study I was certainly not aware of the many subjective decisions I would need to make during the scale reduction process due to a lack of rigorously established guidelines. Researchers and clinicians should also be made aware of the lack of service user involvement and input in this study's scale reduction process.

It is further essential not to lose the person's individual experience within the list of symptoms or questionnaire scores, particularly as the EDE-QS does not include any items on emotional, social or interpersonal difficulties. Clinicians may therefore consider including an additional outcome measure to obtain a more holistic view on an individual's experience in eating disorder services.

References

Allen, K.L., Byrne, S.M., Lampard, A., Watson, H. & Fursland, A. (2011). Confirmatory factor analysis of the eating disorder examination-questionnaire (EDE-Q). *Eating Behaviours, 12,* 143-151.

Berg, K.C., Peterson, C.B., Frazier, P. & Crow, S.J. (2012). Psychometric evaluation of the eating disorder examination and eating disorder examination-questionnaire: A systematic review of the literature. *International Journal of Eating Disorders, 45*, 428-438.

Binford, R.B., Le Grange, D. & Jellar, C.C. (2005). Eating disorders examination versus eating disorders examination-questionnaire in adolescents with full and partial-syndrome bulimia nervosa and anorexia nervosa. *International Journal of Eating Disorders, 37*, 44-49.

Carter, J.C., Aime, A.A. & Mills, J.S. (2000). Assessment of bulimia nervosa: A comparison of interview and self-report questionnaire methods. *International Journal of Eating Disorders, 30,* 187-192.

Carter, J.C., Stewart, D.A. & Fairburn, C.G. (2001). Eating disorder examination questionnaire: Norms for young adolescent girls. *Behaviour Research and Therapy, 39*, 625-632.

Chen, E.Y.H., Tam, D.K.P., Wong, J.W.S., Law, C.W. & Chiu, C.P.Y. (2005). Self-administered instrument to measure the patient's experience of recovery after first-episode psychosis: Development and validation of the psychosis recovery inventory. *Australian and New Zealand Journal of Psychiatry, 39*, 493-499.

Cole, J.C., Rabin, A.S., Smith, T.L. & Kaufman, A.S. (2004). Development and validation of a rasch-derived CES-D short form. *Psychological Assessment, 16,* 360-372.

Coste, J., Guillemin, F., Pouchot, J. & Fermanian, J. (1997). Methodological approaches to shortening composite measurement scales. *Journal of Clinical Epidemiology, 50*, 247-252.

De la Rie, S., Noordenbos, G., Donker, M. & van Furth, E. (2008). The quality of treatment of eating disorders: A comparison of the therapists' and the patients' perspective. *International Journal of Eating Disorders, 41,* 307-317.

DeVellis, R.F. (2012). *Scale development: Theory and applications* (3rd ed.). London: Sage Publications.

Dixon-Woods, M., Shaw, R.L., Agarwal, S. & Smith, J.A. (2004). The problem of appraising qualitative research. *Quality of Safety and Health Care, 13*, 223-225.

Engelsen, B.K. & Laberg, J.K. (2001). A comparison of three questionnaires (EAT-12, EDI, and EDE-Q) for assessment of eating problems in healthy female adolescents. *Nordic Journal of Psychiatry, 55*, 129-135.

Fairburn, C.G. & Beglin, S.J. (2008). Eating disorder examination questionnaire (EDE-Q 6.0). In: C.G. Fairburn (ed.), *Cognitive behavior therapy and eating disorders*. New York: Guilford Press.

Federici, A. & Kaplan, A.S. (2008). The patient's account of relapse and recovery in anorexia nervosa: A qualitative study. *European Eating Disorders Review, 16*, 1-10.

Franchignoni, F., Giordano, A., Sartorio, F., Vercelli, S., Pascariello, B. & Ferriero, G. (2010). Suggestions for refinement of the disabilities of the arm, shoulder and hand outcome measure (DASH): A factor analysis and rasch validation study. *Archives of Physical Medicine and Rehabilitation, 91*, 1370-1377.

Greco, L.A., Lambert, W. & Baer, R.A. (2008). Psychological inflexibility in childhood and adolescence: Development and evaluation of the avoidance and fusion questionnaire for youth. *Psychological Assessment, 20*, 93-102.

Grilo, C.M., Masheb, R.M. & Wilson, G.T. (2001). A comparison of different methods for assessing the features of eating disorders in patients with binge eating disorder. *Journal of Consulting and Clinical Psychology, 69*, 317-322.

Jones, P.W., Harding, G., Berry, P., Wiklund, I., Chen, W.-H. & Kline Leidy, N. (2009). Development and first validation of the COPD assessment test. *European Respiratory Journal, 34*, 648-654.

Kaiser, H.F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141-151.

Kalarchian, M.A., Wilson, G.T., Brolin, R.E. & Bradley, L (2000). Assessment of eating disorders in bariatric surgery candidates: Self-report questionnaire versus interview. *International Journal of Eating Disorders, 28*, 465-469.

Kruyen, P.M., Emons, W.H.M. & Sijtsma, K. (2014). Assessing individual change using short tests and questionnaire. *Applied Psychological Measurement, 38*, 201-216.

Lamoureux, E.L., Pallant, J.F., Pesudovs, K., Rees, G., Hassell, J.B. & Keeffe, J.E. (2007). The impact of vision impairment questionnaire: An assessment of its domain structrure using confirmatory factor analysis and rasch analysis. *Investigative Ophthalmology & Visual Science, 48*, 1001-1006.

Latner, J.D., Mond, J.M., Mackenzie, C.K., Haynes, S.N. & Hay, P.J. (2014). The loss of control over eating scale: Development and psychometric evaluation. *International Journal of Eating Disorders*, doi: 10.1002/eat.22296

Luce, K.H. & Crowther, J.H. (1999). The reliability of the eating disorder examination – self report questionnaire version (EDE-Q). *International Journal of Eating Disorders, 25*, 349-351.

Lundstroem, M. & Pesudovs, K. (2009). Nine-item short-form rasch-scaled revision of the catquest questionnaire. *Journal of Cataract & Refractive Surgery, 35,* 504-513.

Mond, J.M., Hay, P.J., Rodgers, B., Owen, C. & Beumont, P.J.V. (2004). Temporal stability of the eating disorder examination questionnaire. *International Journal of Eating Disorders, 36*, 195-203.

Pettersen, G. & Rosenvinge, J.H. (2002). Improvement and recovery from eating disorders: A patient perspective. *Eating Disorders: The Journal of Treatment & Prevention, 10*, 61-71.

Reas, D.L., Grilo, C.M. & Masheb, R.M. (2006). Reliability of the eating disorder examination-questionnaire in patients with binge eating disorder. *Behaviour Research and Therapy, 44,* 43-51.

Rieger, E., Van Buren, D.J., Bishop, M., Tanofsky-Kraff, M., Welch, R. & Wilfley, D.E. (2010). An eating disorder-specific model of interpersonal psychotherapy (IPT-ED): Causal pathways and treatment implications. *Clinical Psychology Review, 30*, 400-410.

Smith, G.T., McCarthy, D.M. & Anderson, K.G. (2000). On the sins of short-form development. *Psychological Assessment, 12,* 102-111.

Tennant, A. & Conaghan, P.G. (2007). The rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should on look for in a rasch paper? *Arthritis & Rheumatism (Arthritis Care & Research), 57*, 1358-1362.

Wright, B.D. (1996). Comparing rasch measurement and factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal, 3*, 3-24.

Appendix A

Downs and Black quality appraisal questions

Question	**Scoring**
Reporting	
1. Is the hypothesis/aim/objective of the study clearly described?	Yes=1, No=0
2. Are the main outcomes to be measured clearly described in the Introduction or Methods section?	Yes=1, No=0
3. Are the characteristics of the patients included in the study clearly described?	Yes=1, No=0
4. Are the interventions of interest clearly described?	Yes=1, No=0
5. Are the distributions of principal confounders in each group of subjects to be compared clearly described?	Yes=2, Partially=1, No=0
6. Are the main findings of the study clearly described?	Yes=1, No=0
7. Does the study provide estimates of the random variability in the data for the main outcomes?	Yes=1, No=0
8. Have all important adverse events that may be a consequence of the intervention been reported?	Yes=1, No=0

9. Have the characteristics of patients lost to follow-up been described?	Yes=1, No=0
10. Have actual probability values been reported for the main outcomes except where the probability value is less than 0.001?	Yes=1, No=0
External validity	
11. Were the subjects asked to participate in the study representative of the entire population from which they were recruited?	Yes=1, No=0, Unable to determine=0
12. Were the subjects who were prepared to participate representative of the entire population from which they were recruited?	Yes=1, No=0, Unable to determine=0
13. Were staff, places, and facilities where the patients were treated, representative of treatment the majority of patients receive?	Yes=1, No=0, Unable to determine=0
14. Were therapists experienced professionals with regular caseloads? *(replaced by author)*	Yes=1, No=0, Unable to determine=0
Internal validity - bias	
15. Was an attempt made to blind study subjects to the intervention they have received?	Yes=1, No=0, Unable to determine=0
16. If any of the results of the study were based on “data dredging”, was this made clear?	Yes=1, No=0, Unable to determine=0
17. In trials and cohort studies, do the analyses adjust for different lengths of follow-up of patients, or in case-control studies, is the time period between the intervention and outcome the same for cases and controls?	Yes=1, No=0, Unable to determine=0
18. Were the statistical tests used to assess the main outcomes appropriate?	Yes=1, No=0, Unable to determine=0
19. Was compliance with the intervention/s reliable?	Yes=1, No=0, Unable to determine=0

20. Were the main outcome measures used accurate (valid and reliable)?	Yes=1, No=0, Unable to determine=0
Internal validity - confounding (selection bias)	
21. Were the patients in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited from the same population?	Yes=1, No=0, Unable to determine=0
22. Were study subjects in different intervention groups (trials and cohort studies) or were the cases and controls (case-control studies) recruited over the same period of time?	Yes=1, No=0, Unable to determine=0
23. Were study subjects randomised to intervention groups?	Yes=1, No=0, Unable to determine=0
24. Was the randomised intervention assignment concealed from both patients and health care staff until recruitment was complete and irrevocable?	Yes=1, No=0, Unable to determine=0
25. Was there adequate adjustment for confounding in the analyses from which the main findings were drawn?	Yes=1, No=0, Unable to determine=0
26. Were losses of patients to follow-up taken into account?	Yes=1, No=0, Unable to determine=0
Power	
27. Has a power analysis been performed and was the included sample sufficiently powered? *(adapted by author)*	Yes=1, No=0

Appendix B

Diagnostic Criteria derived from EDE-Q 6.0 data

Table 6: ***Diagnostic eating disorder criteria***

Diagnosis	BMI	Compensatory behaviour	Frequency over past 28 days
Anorexia Nervosa- restrictive	< 17.5	No regular binge eating and purging episodes	Less than once/week
Anorexia Nervosa – binge/purge subtype	< 17.5	Binge eating OR purging episodes	At least once/week
Bulimia Nervosa	>18.5	Binge eating AND purging episodes	At least once/week
Binge Eating Disorder	>18.5	Binge eating episodes	At least once/week
Other Specified Feeding or Eating Disorder (OSFED)	Remaining cases		

Appendix C

Ethical approval letter

NHS
Health Research Authority
NRES Committee East of England - Hatfield
Room 002, TEDCO Business Centre
Rolling Mill Road
Jarrow
Tyne and Wear
NE32 3DT

Telephone: 0191 428 3561

07 August 2013

Dr Lucy Serpell
Lecturer
University College London
Research Department of Clinical Educational & Health Psychology
Gower Street
London
WC1E 6BT

Dear Dr Serpell

Study title:	**The development and validation of a shortened version of the Eating-Disorder-Examination -Questionnaire**
REC reference:	**13/EE/0254**
IRAS project ID:	**126942**

Thank you for your letter of 05 August 2013, responding to the Proportionate Review Sub-Committee's request for changes to the documentation for the above study.

The revised documentation has been reviewed and approved by the sub-committee.

We plan to publish your research summary wording for the above study on the NRES website, together with your contact details, unless you expressly withhold permission to do so. Publication will be no earlier than three months from the date of this favourable opinion letter. Should you wish to provide a substitute contact point, require further information, or wish to withhold permission to publish, please contact the Co-ordinator Sarah Grimshaw, nrescommittee.eastofengland-hatfield@nhs.net

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

Ethical review of research sites

The favourable opinion applies to all NHS sites taking part in the study, subject to management permission being obtained from the NHS/HSC R&D office prior to the start of the study (see "Conditions of the favourable opinion" below).

Conditions of the favourable opinion

The favourable opinion is subject to the following conditions being met prior to the start of the study.

Management permission or approval must be obtained from each host organisation prior to the start of the study at the site concerned.

Management permission ("R&D approval") should be sought from all NHS organisations involved in the study in accordance with NHS research governance arrangements.

Guidance on applying for NHS permission for research is available in the Integrated Research Application System or at http://www.rdforum.nhs.uk.

Where a NHS organisation's role in the study is limited to identifying and referring potential participants to research sites ("participant identification centre"), guidance should be sought from the R&D office on the information it requires to give permission for this activity.

For non-NHS sites, site management permission should be obtained in accordance with the procedures of the relevant host organisation.

Sponsors are not required to notify the Committee of approvals from host organisations.

It is the responsibility of the sponsor to ensure that all the conditions are complied with before the start of the study or its initiation at a particular site (as applicable).

You should notify the REC in writing once all conditions have been met (except for site approvals from host organisations) and provide copies of any revised documentation with updated version numbers. The REC will acknowledge receipt and provide a final list of the approved documentation for the study, which can be made available to host organisations to facilitate their permission for the study. Failure to provide the final versions to the REC may cause delay in obtaining permissions.

Approved documents

The documents reviewed and approved by the Committee are:

Document	*Version*	*Date*
Evidence of insurance or indemnity	Gallagher London Policy No B1262FI0103012	
GP/Consultant Information Sheets	1.1	02 April 2013
Investigator CV	Lucy Serpell	02 July 2013
Investigator CV	Nicole Gideon	08 March 2013
Other: Advertisement: Poster Online Study [redacted]	1.1	02 April 2013

Other: Advertisement: Poster Pre-Post Treatment [redacted]	1.1	02 April 2013
Other: Advertisement: Poster Pre-Post Treatment [redacted]	1.1	02 April 2013
Other: Advertisement: Poster Online Study [redacted]	1.1	02 April 2013
Other: Poster Online Study UCL & BEAT	1.1	02 April 2013
Other: Advertisement: Poster Short Online Study BEAT & UCL	1.1	02 April 2013
Other: Data Protection Registration	UCL Registration No Z6364106/2013/04/20	11 April 2013
Participant Consent Form: [redacted]	2.1	31 July 2013
Participant Consent Form: [redacted]	2.2	31 July 2013
Participant Information Sheet: Experts	1.1	02 April 2013
Participant Information Sheet: Online Short	2.1	31 July 2013
Participant Information Sheet: [redacted]	2.4	31 July 2013
Participant Information Sheet: [redacted]	2.3	31 July 2013
Participant Information Sheet: Online	2.2	31 July 2013
Protocol	1	29 January 2013
Questionnaire: Pre-testing Questionnaire	1.0	01 July 2013
Questionnaire: Participant Background Information	1.1	02 April 2013
Questionnaire: EDE-Q 6.0		
Questionnaire: Sociability Scale		
Questionnaire: CIA		
Questionnaire: EAT-26		
Questionnaire: CORE		
Questionnaire: GAD-7		
Questionnaire: PHQ-9		
Questionnaire: SCOFF		
Questionnaire: SEED		
Questionnaire: WHOQOL-BREF		
REC application	IRAS Version 3.5, 126942/479579/1/553	
Referees or other scientific critique report	Sunjeev Kamboj	08 February 2013
Response to Request for Further Information	Nicole Gideon	02 August 2013
Summary/Synopsis	1.0	03 April 2013

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Reporting requirements

The attached document "After ethical review – guidance for researchers" gives detailed guidance on reporting requirements for studies with a favourable opinion, including:

- Notifying substantial amendments
- Adding new sites and investigators
- Notification of serious breaches of the protocol
- Progress and safety reports
- Notifying the end of the study

The NRES website also provides guidance on these topics, which is updated in the light of changes in reporting requirements or procedures.

Feedback

You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.

Further information is available at National Research Ethics Service website > After Review

13/EE/0254	**Please quote this number on all correspondence**

We are pleased to welcome researchers and R & D staff at our NRES committee members' training days – see details at http://www.hra.nhs.uk/hra-training/

With the Committee's best wishes for the success of this project.

Yours sincerely
pp

Mr David Grayson
Chair

Email: nrescommittee.eastofengland-hatfield@nhs.net

Enclosures: *"After ethical review – guidance for researchers" SL-AR2*

Copy to: *Dr Clara Kalu, University College London*
Mrs Angela Williams, Camden & Islington NHS Foundation Trust

Participant information sheet

The development and validation of a shortened version of the Eating Disorder Examination – Questionnaire (EDE-Q)

(student research project)

Please see below for further information and contact details

Investigators: Lucy Serpell (Clinical Psychologist)
Nicholas Hawkes (Clinical Psychologist)
Nicole Gideon (Trainee Clinical Psychologist)
Email: nicole.gideon.11@ucl.ac.uk

Research Department of Clinical, Educational & Health Psychology
University College London
1-19 Torrington Place
London, WC1E 7HB

Details of Study
This study aims to develop a shortened version of the commonly used Eating Disorder Examination - Questionnaire (EDE-Q). The current version of the questionnaire takes a long time to complete and some people find it hard to fill it in. We would like to develop a more user-friendly version that can be completed quickly. We also need to check that the new questionnaire is valid and reliable if we are to draw meaningful conclusions from it.

Validity refers to whether the questionnaire actually measures what it is supposed to. For example, a questionnaire about depression should be measuring aspects of depression rather than another mental health problem such as anxiety. Reliability refers to the extent to which the questionnaire gives the same results each time it is used.
To check the reliability and validity of this new questionnaire, we need to compare it against existing questionnaires. We also need to give it to the same people at two different times.

Why have I been invited?
You have been invited because you are 18 years of age or older, can read and understand a good level of English and you fall into one of the two groups that are being studied:

-You have never had an eating disorder diagnosis
or
- You have/had a diagnosis of an eating disorder, either now or in the past.

If you are currently in treatment for an eating disorder, please discuss participation in the study with your lead clinician.

Do I have to take part?
No, it is up to you to decide. We will outline the study in this information sheet and if you would like to participate we will ask you to give your consent. You are free to withdraw at any time, without giving a reason.

What will happen to me if I take part?
You will be asked to spend between 15 and 20 minutes online, completing questionnaires. Some of the questionnaires ask general questions about mental health difficulties, eating habits and thoughts and perceptions about yourself. This includes completing the long and the new shortened version of the EDE-Q. You can do this at any time that is convenient for you.

In order to test reliability, we would like to contact you again and ask you to complete the shortened EDE-Q a second time 7-14 days after you first completed it. You do not have to do this if you do not want to. However, if you do complete the study twice you will be entered into the raffle twice (see below for details of the raffle), doubling your chances of winning!

What are the possible disadvantages and risks of taking part?
You will be asked some general questions about your mental health, your eating habits and thoughts that you may have about yourself. If you find it upsetting to think about these things then there is a small chance that you may feel upset after doing this study. There are no other disadvantages or risks involved in this study.

What are the possible benefits of taking part?
You will be entered into a raffle, along with other participants in this study, to potentially win one of two £50 vouchers for a shop of your choice. In addition, we hope you will find it a positive experience and the knowledge gained from this study will be of help to people with eating disorders and mental health difficulties in the future.

Will my taking part in the study be kept confidential?
Yes. The guidelines in the Data Protection Act (1998) will be followed, meaning that all information about you will be handled in confidence. An identification code will be allocated to you and the information we collect will be recorded and put into an electronic database using this code rather than your name. This means that your data will be anonymous and therefore it will not be possible for us to withdraw your data after you have submitted your questionnaires. The data will be used for research purposes only.

What if there is a problem?
If answering these questions makes you upset or worried, you can find advice and support from:

Beat: Beating Eating Disorders Website: www.b-eat.co.uk Helpline: 0845 634 1414	or	NHS Direct Website: www.nhsdirect.nhs.uk Helpline: 111 or 0845 4647

Email: help@b-eat.co.uk

What if I feel unhappy about the way I have been treated?
If you feel unhappy about the way you have been treated at any point during this study and would like to make a complaint, please contact Nicole Gideon nicole.gideon.11@ucl.ac.uk. If you are not satisfied with the response, please contact the chief investigator Lucy Serpell l.serpell@ucl.ac.uk.

What will happen to the results of the research study?
The data and results from this study may be published in psychology journals or used in scientific reports. As the data will be anonymised, you will never be identified by name.

Who is organising and funding the research?
This study is organised by Dr Lucy Serpell, Dr Nicholas Hawkes and Nici Gideon and it is sponsored by the University College London.

Who has approved this study?
This study has been approved by a NHS Ethics Committee, the NRES Committee East of England-Hatfield, which has reviewed it in detail.

Thank you for reading this information sheet. Please discuss the information above with others if you wish or email Nicole (Nicole.gideon.11@ucl.ac.uk) if there is anything that is not clear or if you would like more information.

Appendix D

Exclusion and inclusion criteria for EDE-QS based on combined methods

EXCLUSION

1. Identify those items that misfit the Rasch model (as indicated by outfit/infit mean square values) and consider for deletion
2. Identify those pairs of items that exhibit local dependency and delete one of them based on a) expert preference or b) fit to model or c) diagnostic importance
3. Identify those items that fall within a similar range of eating disorder severity and consider deletion of one of them based on a) expert preference or b) fit to model or c) diagnostic importance

INCLUSION

1. Identify those items that have been rated highest by experts
2. Identify those items that are important for assessing diagnostic criteria
3. Identify two items per factor (high and low severity) and discuss rationale for inclusion with regards to a) expert preference or b) fit to model or c) diagnostic importance