School of Medicine

Ysgol Meddygaeth

Developing a Short-form of the Genetic Counselling Outcome Scale

MPhil Cancer and Genetics

Cardiff University School of Medicine

2018

By Peter Grant

C1675275

Contents

Chapter Four

Chapter Five

Chapter Six

Appendix

DECLARATION

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed …… (candidate)

Date ……

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of ……(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed …… (candidate)

Date ……

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed …… (candidate)

Date ……

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ……..…..….. (candidate)

Date ……

Acknowledgements

With thanks to all members of the research team: Marion McAllister [a], Maria Pampaka [b], Katherine Payne [c], Angus Clarke [d]

[a] Centre for Medical Education, School of Medicine, Cardiff University, Cardiff, UK

[b] Departments of Social Statistics (School of Social Science) and Education (School of Environment, Education and Development), The University of Manchester, Manchester, UK

[c] Division of Population Health, Health Services Research and Primary Care, The University of Manchester, Manchester, UK

[d] Division of Cancer & Genetics, School of Medicine, Cardiff University, Cardiff, UK.

To Dr Marion McAllister, my wonder woman, for her enduring patience and guidance.

And to George Basnett and Rob Winter, for providing the fun.

Summary

The aim of this study was to develop a short form of the 24-item Genetic Counselling Outcome Scale (GCOS-24), suitable for use in the clinical setting and in evaluations of genetic counselling and testing services. The study comprised four phases. Phase I: Cognitive interviews were used to explore interpretability of GCOS-24 items and which GCOS-24 items were most valued by the target population. Ten cognitive interviews were conducted with individuals affected by or at risk for a genetic condition, recruited from patient support groups. Phase II: Quantitative analysis of an existing data set of GCOS-24 responses (n = 395), using Classical Test Theory (CTT) methods to identify underlying traits, and Item Response Theory (IRT) methods to examine item discrimination. Phase III: Item Selection. The results from Phases I & II were used to inform the selection of a set of GCOS-24 items. The Rasch rating scale model (Andrich, 1978) was also used to explore functional problems with the seven-point Likert Scale. A six-item questionnaire with a five-point Likert Scale was produced (GCOS-6). In Phase IV the reliability and discriminative ability of the new instrument was tested through a test-retest study. GCOS-6 displays excellent test-retest reliability (0.788) and moderate internal consistency (α = .570). This study represents a potential first step in the development of a measure which could be used in the evaluation of technologies and services used in genetic counselling and testing services.

Tables

Figures

1. Introduction

The goal of this chapter is to present an introduction to the context of this thesis, including an outline of why this research was carried out and why it is of significance in the field of clinical genetics. The research problem will be described, as well as the overall project aim.

1.1 Context of Research

Genetic counselling and associated testing services (hereafter shortened to 'clinical genetics services' (CGS)) is a medical speciality which can offer a number of potential benefits to individuals and families affected by possible genetic conditions. Studies have provided evidence that patients attend CGS seeking information and a supportive relationship, and that the benefits of CGS include relief of uncertainty and feelings of vulnerability, increased self-efficacy, and adaptation to the genetic condition in the family (Bernhardt *et al.,* 2000; MacLeod *et al.,* 2002; McAllister *et al.,* 2008; Payne *et al.,* 2007; Skirton, 2001; Slomp *et al.,* 2017). One stated aim of prenatal genetic counselling, for example, is to assist the patient in making decisions regarding invasive testing (Beulen *et al.*, 2016).

Robust and validated measures of these benefits are needed to provide evidence to service commissioners about the outcomes of investing in existing CGS or future service developments. Evaluations of CGS have traditionally examined outcome variables such as information recall, reproductive intentions and decisions made, and patient satisfaction (Clarke *et al.,* 1996). Measures of process such as waiting times and numbers of patients seen have also been used, as well as the performance characteristics of genetics tests (e.g. sensitivity, specificity and predictive values) (Clarke *et al.,* 1996; Payne *et al.,* 2008). More recently, clinical genetics professionals have contended that the traditional approaches to outcome measurement are not relevant or appropriate, and that insufficient attention has been paid to outcomes relevant to the population of individuals who use CGS (Clarke, 1996; MacLeod, 2002; McAllister *et al.*, 2008; McAllister & Dearing, 2015; Payne *et al.,* 2008). Moreover, many of the measures which have been used to evaluate CGS have not undergone rigorous psychometric validation, often assessed for internal consistency alone (Payne *et al.,* 2008; McAllister & Dearing 2015).

1.2 Patient-Reported Outcome Measures

Patient-reported outcomes measures (PROMs) are questionnaires designed to measure healthcare outcomes directly from the perspective of the patient, and over recent years they have been gaining prominence in healthcare valuation across the world. In the UK, routine use of PROMs in the NHS was recommended by the Department of Health for the purpose of providing data on quality of care (DoH, 2008), and this has since been operationalised for all NHS hernia repairs, varicose vein treatments, and hip and knee replacements in England (Diness *et al.,* 2017; Judge *et al.,* 2012; Nuttall *et al.,* 2013). PROMs are also of increasing importance in US healthcare, with the Department of Health and Human Services (HHS) and the Food and Drug Administration (FDA) now recommending PROMs data should be used to support medical product labelling claims (FDA, 2009). The recognised value of PROMs is further demonstrated by the Patient Reported Outcomes Measurement Information System (PROMIS) initiative, which catalogues validated PROMs for use in evaluating physical, mental and social health in adults and children. It is designed to enhance communication between clinicians and researchers, and is available in many languages. In short, PROMs offer valuable tools for service evaluation and audit of practice.

Standardised and widely-validated PROMs such as the EQ-5D (Brooks, 1996) or SF-6D (Brazier et al., 2002), used for service evaluation across certain branches of healthcare, are not appropriate in the context of CGS because they focus on a restricted number of outcome domains, including the physical health status of the patient. Certain items within the EQ-5D, for example, explore the ability of the respondent to walk about and dress themselves; certain items within the SF-6D assess whether health affects physical functioning. Many genetic conditions can neither be treated nor cured, and, apart from the monitoring or testing for complications of a genetic condition, interventions offered by genetic counselling are not expected to affect physical health status. Although in some cases patient morbidity or mortality may show improvement in the long-term, for example with those who are offered screening or surgery options for hereditary cardiac or cancer syndromes, these changes would not be directly attributable to genetic counselling and testing. Health-Related Quality of Life (HRQoL) is a multi-dimensional concept that includes elements relating to physical, emotional, psychological, and social domains of health. HRQoL outcomes are valued by CGS patients and clinicians (Payne *et al.,* 2007), and HRQoL instruments have been

recommended for use as measures of effect in evaluations of interventions in medical genetics (Stevenson & Carey, 2009).

1.3 Research Problem and Project Aim

The evaluation of CGS requires a robust and valid PROM, capturing relevant outcomes which are valued by CGS patients. This study aims to establish a PROM which would be appropriate for routine use in audit and clinical evaluations of CGS.

2. Literature Review

This chapter presents a description of what genetic counselling and associated testing services (CGS) entails, followed by a critical review of the published research regarding outcome measures in genetic counselling and testing services. The aim of the review was to identify, synthesise and critically appraise the relevant literature, and in doing so to justify why this current research project is necessary and of value in the advancement of healthcare research. Key terms have been defined, and the aims of the project have been refined according to the findings of the review.

2.1 What do we mean by 'Genetic counselling and associated testing'?

As far as medical specialities go, genetic counselling boasts a relatively short history. Since first being titled as such by Sheldon Reed in 1947 (Reed, 1955), it has gone from being an isolated activity to being integrated as a major component of clinical genetics and a legitimate branch of healthcare, and the range and complexity of issues which the service is now expected to encompass has expanded considerably. In the UK, genetic counselling is regarded as an integral part of the genetic testing process, strongly recommended by the NHS in most genetic testing situations (Harding, 2016).

The current gold standard definition for genetic counselling was published in 2006 by the Genetic Counseling Task Force of the National Society of Genetic Counselors (NSGC) in the US (Resta *et al.,* 2006). The study made use of input from the membership, leaders of genetic advocacy groups and genetic professional organisations, and was endorsed by the NSGC Board of Directors. The creation was spurred by the need to maintain common practice following the advent of genomic medicine (Resta *et al.,* 2006), and following the expansion of genetic counselling beyond traditional settings (Bennett *et al.,* 2003; Ciarleglio *et al.,* 2003). It reads as follows:

> 'Genetic counselling is the process of helping people understand and adapt to the medical, psychological and familial implications of genetic contributions to disease. This process integrates the following:

- Interpretation of family and medical histories to assess the chance of disease occurrence or recurrence;
- Education about inheritance, testing, management, prevention, resources and research;
- Counselling to promote informed choices and adaptation' (Resta *et al.,* 2006, p77).

Standardised definitions can help to encourage common practice and ensure that patients receive appropriate medical care. With that said, genetic counselling is not a standardised process, and it should be remembered that definitions may not be representative of all situations. As Matloff (1994) demonstrated in a survey of over 200 genetic counsellors in the US, the content of genetic counselling sessions will vary from counsellor to counsellor and from centre to centre. Similarly the focus of the service will shift depending on the genetic condition at hand, and specific objectives and outcomes will naturally show differences between patients, as shown by Michie *et al.* (1996) who analysed patient expectations, patient concerns, and patient outcomes from 131 genetic counselling consultations, and by Macleod *et al.* (2002) who examined counselees' perceptions of their consultation.

Furthermore, genetic counselling services vary between countries and cultures (Fathzadeh *et al.,* 2008 (Iran); Mohanty & Dias, 2011 (India); Pampols *et al.,* 2016 (Spain); Temtamy & Hussen, 2017 (Egypt)). Ethical, religious, and moral values can be significantly different both intra- and internationally, as can be the standard of healthcare available to patients. As such, the process of genetic counselling will be shaped by the respective clinical, technological, ethical and societal milieux (Fathzadeh *et al.,* 2008 (Iran); Mohanty & Das, 2011 (India); Pampols *et al.,* 2016 (Spain); Temtamy & Hussen, 2017 (Egypt)). Although the NSGC definition may represent the speciality from the perspective of those individuals in the US at that time, certain components may be lacking or of limited relevance in, say, Egypt or India. For instance, the NSGC definition does not mention spiritual beliefs. Whilst this may be of lesser, and arguably diminishing importance in Western societies such as the UK, US, Canada and Australia, in other cultures this could be a significant consideration to address in counselling sessions and as such would be a priority for inclusion in a definition.

In the UK, one of the key features of modern genetic counselling is that the service is centred around the patient and their family members (Hough, 2002; Middleton *et al.,*

2015; Ormond, 2013; Tluczek *et al.,* 2011). A predominantly 'non-directive' approach is taken, meaning that the counsellor does not try to guide the patient towards any particular decision, for example whether to terminate a pregnancy or to have a genomic test. Instead, the counsellor works with the patient to educate and inform, in order to build an understanding of what it means to have a genetic condition in the family and what options are available to them.

> "It involves a person-centred approach where the genetic counsellor helps the patient to incorporate the genetic information into their lives, adjust to it, rationalise it, think through how they want to act on it and rehearse how they wish to explain it to relatives." (Hough, 2002. p51)

Genetic counselling patients may likely have a number of questions and concerns, and may carry considerable emotional distress (Clarke *et al.,* 1996; Duric *et al*., 2003; Hamilton *et al.,* 2009; Nordin *et al.,* 2011). As described by McCarthy-Veach *et al.* (2003) in their genetic counselling practice manual, a patient may come for genetic counselling at one of the most vulnerable moments in their life. Their child may have been diagnosed with a neurodegenerative condition; there may be fear over the potential effects of a hereditary trait; or there may be grief if a genetic condition has resulted in the premature death of a family member. It is therefore essential for genetic counsellors to listen and communicate effectively with their patients, to exhibit sensitivity and compassion, and to provide emotional support where necessary.

Genetic testing is a type of medical test which involves the study of a person's DNA. It usually involves having a sample of blood or tissue taken, and may be carried out to diagnose a genetic condition, to help determine the chances of developing a genetic condition, or to determine whether a person is a carrier of a genetic mutation. In some cases genetic testing can be performed to find out the likelihood of a baby being born with a certain genetic condition. Examples of prenatal testing processes include amniocentesis, whereby cells are extracted from the mother's womb using a needle, chorionic villus sampling, which involves the removal and testing of placental cells, and cell-free fetal DNA screening (also called non-invasive prenatal screening), which detects defects in the fetal DNA that is released by the placenta into the mother's bloodstream during pregnancy. A referral to genetic testing will usually be accompanied by a referral to genetic counselling, allowing individuals to discuss the risks, benefits and limitations of genetic testing with a trained professional.

Although commonly used interchangeably, the terms 'genetics' and 'genomics' are not synonymous. 'Genetics' is the study of heredity, of the genes people inherit and pass down through their family. 'Genomics' refers to the study of all genes within an organism, including their functions and relationships. There is currently debate over whether 'genomic counselling' and 'genomic testing' are becoming ever-more appropriate terms as we transition from single-gene focused genetic counselling and testing to the routine incorporation of genomic medicine (Ormond, 2013). For the purposes of this thesis, the traditional terminology of 'genetic counselling' and 'genetic testing' has been used throughout.

In summary, genetic counselling is the process of helping people understand and adapt to the medical, psychological and familial implications of genetic contributions to disease. Specific objectives and outcomes may vary from patient to patient, and may be influenced by a number of factors such as the condition at hand and geographical location, but current practice recommends a non-directive, patient-centred approach should be taken in order to help build an understanding of what it means to have a genetic condition in the family and what options are available to them. Genetic testing can be used to confirm or rule out a suspected genetic condition or help to determine a person's chance of developing or passing on a genetic disorder.

2.2 Literature Search Methodology

2.2.1 Introduction

The overall aim of this study, as stated on page three, was to establish a PROM which would be appropriate for routine use in audit and clinical evaluations of CGS. However, before jumping into the often arduous and time-consuming task of creating a novel health measurement scale, it is recommended that researchers should first look for existing validated measures (DeVellis, 2011; Streiner & Norman, 2008).

Prior to this project, the only published systematic review of outcome measures in CGS had been carried out by Payne *et al.* (2008), a study which identified 67 validated outcome measures and concluded that no single measure at the time encompassed all aspects of the potential benefits from using a CGS. A more recent review by McAllister & Dearing (2015) identified additional measures, but results were used specifically to analyse outcome domains. Over the last ten years, the speciality of clinical genetics has

seen rapid advances. Existing technologies have improved and novel technologies have appeared, and our collective knowledge about how genetics might influence disease has increased. Within clinical practice, genetic testing is increasingly being performed outside the traditional bounds of CGS and is now moving into other specialities. This process is referred to as 'mainstreaming genetics' and is occurring in the context of cancer predisposition genes (Rahman, 2014), paediatrics (Valente *et al.,* 2008), and neurogenetic testing (Lo *et al.,* 2014). Furthermore, recent economic evaluations in CGS have found the high degree of heterogeneity in outcome measures as being a principal methodological limitation (Djalalov *et al.,* 2011; D'Andrea *et al.,* 2015). The aim of the following literature review was therefore to provide a full, thorough, and current account of validated outcome measures which have been used in CGS. In other words, the question driving the review was:

> 'Is there an existing patient-reported outcome measure which would be appropriate for routine use in audit and clinical evaluations of CGS?'

2.2.2 Search Design

The aim of this search was to identify validated outcome measures which had previously been used in the evaluation of CGS. An outcome measure was defined as: 'any instrument used to measure, evaluate or assess the impact of CGS on the patient'. The reason for only including validated outcome measures was that validation is a requirement of robust evaluations. For the purposes of this review, validation was met if a measure had passed some form of psychometric assessment.

Being an unfunded MPhil project, this review was not intended to be a systematic review; no formal meta-analysis of included articles was conducted and multiple independent reviewers were not used. The scope of the review was limited to published works in English which were available online, either freely or through Cardiff University access. Given time constraints, the period of search and writing was limited to Jun 2017 – Jun 2018.

2.2.3 Search Strategy

An electronic search of The Cardiff University Ovid database from 1940 to present was used as the primary resource, but further databases such as Embase (1980 to present), the NHS Health Economic Evaluations Database (1900 to present), Medline (1966 to present) and the Cochrane database (1900 to present) were also utilised following reference to the systematic review of Payne *et al.* (2008). Search terms included: "genetic"; "genomic"; "counsel(l)ing"; "testing"; "clinical genetics"; "outcomes"; "patient outcomes"; "patient reported outcomes"; "PROM"; "measure"; "survey"; "questionnaire"; "scale"; and terms were again cross-referenced with Payne to check for omissions. Overall, the search strategy was put together through consultation with supervisors MM & KP, and using the existing systematic review of Payne *et al.* (2008).

2.2.4 Selection and Extraction

An initial screen of titles and abstracts was carried out by one reviewer (PG), and articles were rejected if they were clearly not relevant to outcome measures in CGS. If relevance was uncertain, the full text was located and examined. Articles met the inclusion criteria if a validated outcome measure was created or applied for the purpose of evaluating some aspect of CGS. Articles were excluded if they were not written in English, if the outcome measure was not validated, or if the measure was not appropriate for use within routine CGS. For the purpose of this study, the completion of any psychometric test was sufficient to meet the validity criteria.

If a validated outcome measure was identified, a tailored spreadsheet was then used to extract information about the measure. The degree of psychometric validation was noted, as was the purpose of the measure.

2.3 Results

The search methodology identified 151 titles and abstracts which appeared to be relevant and which were chosen for more detailed examination. From these, 86 papers were selected for inclusion in the final review. A total of 82 validated outcome measures were referred to in these 86 studies (Table 2.1).

Table 2.1: Validated outcome measures used in the evaluation of clinical genetics services

Outcome measure	Primary Source(s)	Purpose	Type of measure
Anticipated impact of results	Hailey *et al.* (2000) Lerman *et al.* (1995)	To assess the likelihood of a variety of possible psychological reactions to a positive and negative test result.	Rating scale Genetics specific
Appropriateness of genetic testing delivery	Andrea *et al.* (2018)	To investigate the appropriateness of genetic testing delivery and post-testing healthcare pathways.	Rating scale Genetics specific
Assessment of benefits and risk of breast cancer testing	Hailey *et al.* (2000) Lerman *et al.* (1995)	To assess the perceived benefits and risks of genetic testing.	Rating scale Genetics specific
Audit Tool for Genetic Services	Skirton *et al.* (2005)	To measure outcomes of clinical genetics services.	Rating scale Genetics specific
Beck Depression Inventory (BDI)	Su *et al.* (2009)	To measure the intensity of depression in psychiatrically diagnosed patients and for detecting depression in normal populations.	Rating scale Non-genetics specific
Beliefs About Breast Cancer Genetic Testing	Bowen *et al.* (2002)	To measure specific beliefs about breast cancer genetic testing.	Rating scale Genetics specific
Body Image/Sexuality Scale (BISS)	Lodder *et al.* (2002) Van Oostrum *et al.* (2003)	To assess body image and general sexual functioning	Rating scale Non-Genetics specific
(Breast) Cancer Attitude Inventory (CAI) and Anxiety sub-scale (BCANX)	Berrenberg (1991) Hailey *et al.* (2000)	To assess attitudes towards cancer.	Rating scale Non-genetics (cancer) specific
Breast cancer (hereditary) concern	Stalmeier *et al.* (1999)	To determine concern about breast cancer	Rating scale Genetics specific
Breast Cancer Genetic Counselling Knowledge	Erblich *et al.* (2005)	To assess knowledge of information generally provided during breast cancer genetic counselling.	True/False & Multiple Choice Genetics specific

Outcome measure	Primary Source(s)	Purpose	Type of measure
Questionnaire (BGKQ)			
Breast Cancer Worry / Cancer Worry Scale	Lerman *et al.* (1991) Van Oostrum *et al.* (2003)	To assess dimensions of cancer worry	Rating scale Non-genetics (cancer) specific
Brief Symptom Inventory	Derogatis & Melisaratos (1983)	To assess psychological symptom patterns in normal populations and in psychiatric patients.	Rating scale Non-genetics specific
Cancer Anxiety and Helplessness Scale	Kash *et al.* (1992)	To assess women's general cancer anxiety and sense of helplessness.	Rating scale Non-genetics specific
Center for Epidemiologic Studies Depression-Scale (CES-D) and brief form	Radloff (1977) Ross & Mirowsky (1984)	To measure depressive symptomatology in the general population rather than the assessment for diagnosis at clinical intake and/or evaluation of severity of illness over the course of treatment.	Rating scale Non-genetics specific
Clinical Genetics Satisfaction (CGS) indicator.	Zellerino *et al.* (2009)	To evaluate patient satisfaction with genetic counselling.	Rating scale Genetics specific
Decision Evaluation Scale	Stalmeier *et al.* (2005)	To assess how patients evaluate their medical treatment choice.	Rating scale Non-genetics specific
Decisional Conflict Scale (DCS)	O'Connor (1995)	To measure decisional conflict, which is a state of uncertainty about the course of action to take.	Rating scale Non-genetics specific
Decision making process	Brain *et al.* (2005) Michie *et al.* (1997)	To measure the extent to which women thought or 'agonised' about the decision.	Rating and multiple-choice Genetics specific
Desire to participate in the shared decision making program	Stalmeier *et al.* (1999)	To measure desire to participate in the shared decision making program	Rating scale Genetics specific
Emotional reaction to the program information	Stalmeier *et al.* (1999)	To measure the emotional reaction to information given on the shared decision making program	Rating scale Genetics specific

Outcome measure	Primary Source(s)	Purpose	Type of measure
Evaluation of practical issues and responsibilities.	Otten *et al.* (2016)	To assess experiences with preparing for online counselling (e.g. clarity of the instructions email).	Multiple-choice Genetics specific
Expectations of online counselling	Otten *et al.* (2016)	To assess patients' expectations of online counselling.	Rating scale Genetics specific
Family Environment Scale (FES)	Moos & Moos (1994) Halvorsen (1991)	Designed to measure the social-environmental characteristics of all types of families.	Rating scale Non-genetics specific
Functional Assessment of Cancer Therapy-General (FACT)	Cella *et al.* (1993) Brady *et al.* (1997)	To measure quality of life in patients with cancer. There is also a scale specific to breast cancer.	Rating scale Non-genetics specific
General Health Questionnaire (GHQ)	Goldberg & Williams (1988) Goldberg & Hillier (1979)	To detect those with a diagnosable psychiatric disorder. It looks at two areas: inability to carry out one's normal 'healthy' functions and the appearance of new phenomena of a distressing nature.	Rating scale Non-genetics specific
Genetics Appointment Patient Satisfaction Score (GAPPS)	Westwood *et al.* (2012)	To test whether primary care genetic-led genetics education improves both non-cancer and cancer-referral rates.	Rating scale Genetics specific
Genetic Counselling Outcome Scale (GCOS-24)	McAllister *et al.* (2011b)	To capture empowerment, a construct encompassing many patient outcomes from CGS.	Rating scale Genetics specific
Genetic Counseling Satisfaction Scale (GCSS)	Tercyak *et al.* (2001)	To assess patient satisfaction with the process and content of genetic counselling	Rating scale Genetics specific
Genetic Knowledge Index (GKI)	Furr & Kelly (1999)	To measure level of genetic knowledge, not specific to a genetic disease.	Rating scale Genetics specific
Global Severity Index (GSI) of the Symptom Check List-90 (SCL90)	Derogatis (1983)	The SCL-90R was designed to reflect the psychological symptom patterns of psychiatric and medical patients. To measure the degree to which they suffered from psychological complaints	Rating scale Non-genetics specific
Health Beliefs Model (screening	Kash *et al.* (1992)	To assess perceived susceptibility to disease, severity of disease,	Rating scale

Outcome measure	Primary Source(s)	Purpose	Type of measure
and breast cancer)		benefits of intervention, risks of intervention, and practical obstacles to intervention.	Non-genetics specific
Health Orientation Scale	Woolridge & Murray (1989)	Designed to objectively appraise the psychological implications of identification as a sickle cell gene carrier. Also used to assess the emotional implications of being a carrier of the CF-gene	Rating scale Genetics specific
Hopkins Symptom Checklist (HSCL)	Derogatis *et al.* (1974)	To assess the presence and severity of anxiety and depression symptoms over the previous month. It is a self-report symptom inventory.	Rating scale Non-genetics specific
Hospital Anxiety and Depression Scale (HADS)	Zigmond & Snaith (1983) Van Oostrum *et al.* (2003)	Self-assessment mood scale designed for use in non-psychiatric hospital patients to detect states of depression and anxiety.	Rating scale Non-genetics specific
Illness Perception Questionnaire (IPQ)	Cho *et al.* (2012)	To measure perceived control over risk.	Rating Scale Non-genetics specific
Impact of Event Scale (IES)	Horowitz *et al.* (1979) Van Oostrum *et al.* (2003)	To evaluate current subjective distress for any life event. The wording is not anchored to a specific occurrence but to the particular qualities of conscious experience that encompass all such events.	Rating scale Non-genetics specific
Intention to act upon shared decision making program	Stalmeier *et al.* (1999)	To measure the intention to act upon the shared decision making program	Rating scale Genetics specific
Knowledge About Breast Cancer	Donovan & Tucker (2000) Stager (1993) Vaeth (1993)	To assess women's knowledge of several dimensions of breast cancer.	Rating scale Generic
Knowledge about genetic testing for inherited cancer	Benkendorf *et al.* (1997) Lerman *et al.* (1996)	To assess knowledge of inheritance of breast-ovarian cancer susceptibility and genetic testing.	True/false rating Genetics specific
Knowledge about genetic risk for breast cancer	Donovan & Tucker (2000)	To assess women's knowledge about the hereditary nature of breast cancer and the increased risk of breast and ovarian cancer	Rating scale Genetics specific

Outcome measure	Primary Source(s)	Purpose	Type of measure
		associated with altered BRCA1 or BRCA2 gene.	
Knowledge Scale about Breast (and Ovarian) Cancer and Hereditary	Ondrusek *et al.* (1999)	To test general knowledge about breast cancer and hereditary breast cancer among women at low to moderate risk of hereditary breast cancer.	Rating scale Non-genetics specific
Life Orientation Test (LOT)	Scheier *et al.* (1994) Carver *et al.* (1994)	To measure the level of optimism in one's outlook on life	Rating scale Non-genetics specific
Measure of Counselees' Knowledge of Down Syndrome	Braitman & Antley. (1978)	To measure counselees' knowledge and/or understanding of Down syndrome	Multiple choice Genetics specific
Medical Communication Behaviour System (MCBS)	Wolraich *et al.* (1986)	To assess physician-patient interactions that involve giving distressful information.	Rating scale Genetics specific
Medical Interview Satisfaction Scale - modified (MISS)	Wolf *et al.* (1978)	To assess the patient's perception of a particular care encounter rather than satisfaction with medical care in general	Rating scale Non-genetics specific
Medical Outcomes Short-Form Survey (SF-36 and SF-12)	Ware (1993) Jenkinson *et al.* (1996)	To measure quality of life.	Rating scale Non-genetics specific
Medical Outcomes Study (MOS) Social Support Survey	Sherbourne & Stewart (1991)	To measure the current availability of social support	Rating scale Non-genetics specific
Monitoring Blunting Style Scale (MBSS)	Miller (1987)	To determine information-seeking coping style.	True/false rating Non-genetics specific
Minnesota Multiphasic Personality Inventory (MMPI)	Graham (1987)	To assess general personality profile.	True/false Non-genetics specific
Modified Maternal Serum Screening Knowledge Questionnaire (MSSKQ)	Goel *et al.* (1996)	To assess knowledge about maternal serum screening. Modified to assess knowledge of prenatal testing in general rather than maternal serum screening	Rating scale Genetics specific

Outcome measure	Primary Source(s)	Purpose	Type of measure
modified Tolerance for Ambiguity Scale (TFA)	Geller *et al.* (1993)	To measure ambiguity tolerance as a more general personality trait. Intolerance for ambiguity has been defined as 'the tendency to perceive situations that are novel, complex or insoluble, as sources of threat.'	Rating scale Non-genetics specific
Multidimensional Impact of Cancer Risk Assessment (MICRA)	Cella *et al*. (2002)	To assess concerns and psychosocial issues associated with genetic testing for cancer risk	Rating scale Genetics specific
Openness to Discuss Cancer in the Family Scale (ODCFS)	Mesters *et al.* (1997) Van Oostrum *et al.* (2003)	To assess openness of communication about cancer (and cancer genetic test result) in the nuclear family (partner and children) and the family of origin (parents, siblings).	Rating scale Non-genetics specific
Patient health questionnaire (PHQ-9)	Meiser *et al.* (2013)	To evaluate individuals with a family history of depression.	Rating Scale Non-genetics specific
Patient Satisfaction with Genetic Counselling	Brain *et al.* (2000) Shiloh *et al.* (1990)	To assess patient satisfaction with the genetic counseling process.	Rating scale Genetics specific
Penn State Cancer Genetics Program Survey	Kausmeyer *et al.* (2006)	To assess sources of patient referrals, patient satisfaction and expectations, changes in risk perception and decision making based on knowledge gained from the cancer risk-assessment.	Multiple choice Genetics specific
Perceived-Devaluation-Discrimination-Scale (PDDS)	Meiser *et al.* (2013)	To assess perceived stigma of depression.	Rating Scale Non-genetics specific
Perceived personal control (PPC)	Berkenstadt *et al*. (1999) Otten *et al.* (2016)	To measure PPC.	Rating scale Genetics specific
Perceived Risk of Breast Cancer	Brain *et al.* (1999)	To assess perceived personal risk of developing breast cancer.	Rating scale Generic
Perceptions of the benefits, limitations and risks of genetic	Donovan & Tucker (2000) Hughes *et al.* (1997)	To assess perceptions of the benefits, limitations and risks of genetic testing for breast-ovarian cancer risk.	Rating scale

Outcome measure	Primary Source(s)	Purpose	Type of measure
testing	Audrain *et al.* (1995)		Genetics specific
Pharmacogenetics in Psychiatry follow-up questionnaire (PIP-FQ)	Walden *et al.* (2015)	To examine treatment outcomes in psychiatric care after genetic information was provided to patients.	Rating scale Genetics specific
Profile of Mood State (POMS)	McNair *et al.* (1981)	To measure mood states in psychiatric outpatients and for assessing changes in such patients. It is also used in non-patient populations.	Rating scale Non-genetics specific
Prostate cancer genetic screening survey	Doukas (2004)	To explore what values and expectations influence the intention of men to undergo genetic testing for prostate cancer risk	Rating scale Genetics specific
Psychological Adaptation to Genetic Information Scale (PAGIS)	Read *et al*. (2005)	To measure multiple dimensions of psychological adaptation to genetic information to facilitate evaluation of the efficacy of counseling and supportive interventions and to identify people at risk for coping difficulties.	Rating scale Genetics specific
Psychological Consequences Questionnaire (PCQ)	Cockburn *et al.* (1992)	To assess the psychological consequences of breast mammography on well-being	Rating scale Non-genetics specific
Quality of Care Through the Patients' Eyes (QUOTE)-gene^{CA}	Pieterse *et al.* (2005)	To measure the needs and preferences in genetic counseling for hereditary cancer before their first consultation.	Rating scale Genetics specific
Risk comprehension and subjective knowledge of women in the shared decision making program	Stalmeier *et al.* (1999)	To assess risk comprehension and subjective knowledge of the women in the shared decision making program	Rating scale Genetics specific
Rosenberg Self-Esteem Scale	Rosenberg (1965) Curbow & Somerfield (1991)	Global measure of self-esteem considered to be an indicator of psychological adjustment. This measure was originally developed to measure adolescents' global feelings of self-worth or self-acceptance.	Rating scale Non-genetics specific

Outcome measure	Primary Source(s)	Purpose	Type of measure
Satisfaction with Decision Scale	Brain *et al.* (2005) Holmes-Rovner *et al.* (1996)	To measure satisfaction with a medical decision. Developed in the context of postmenopausal hormone-replacement therapy decisions.	Rating scale Non-genetics specific
Satisfaction with Genetic Counselling Questionnaire	Hilgart *et al.* (2012)	To evaluate the impact of cancer genetic risk-assessment services on patients at risk of familial breast cancer.	Rating scale Genetics specific
Satisfaction with shared decision making program	Stalmeier *et al.* (1999)	To measure the level of satisfaction with the shared decision making program	Rating scale Genetics specific
Self-rating Depression Scale (SDS)	Zung (1965)	To measure, using self-rating and interviewer rating, depressive disorder.	Rating scale Non-genetics specific
Shared decision making program rationale acceptability	Stalmeier *et al.* (1999)	To measure the acceptability of the rationale for the shared decision making program	Rating scale Genetics specific
Short-form Health Survey (SF-12)	Hubalek *et al.* (2016)	To examine long-term psychosocial consequences and counsellees' satisfaction after genetic counselling for breast and ovarian cancer.	Rating Scale Non-genetics specific
Short-form Health Survey (SF-36)	Bowen & Powers (2010)	To measure perceived quality of life.	Rating Scale Non-genetics specific
Spielberger State Trait Anxiety Inventory (STAI) and state scale (STAI-State)	Spielberger et al. (1970) Marteau & Bekker (1992)	To measure anxiety. The STAI differentiates between the temporary condition (state anxiety) and the more general and long-standing condition (trait anxiety). Adapted for use in children.	Rating scale Non-genetics specific
Spiritual Well-Being Scale (SWBS)	Ellison & Smith (1991) Gioiella *et al.* (1998)	To assess personal spiritual meaning and satisfaction.	Rating scale Non-genetics specific
Subjective Quality of Life Profile (SQLP)	Dazord (1995)	To assess subjective quality of life in patients or healthy people and explore the various dimensions of quality of life.	Rating scale Non-genetics specific

Outcome measure	Primary Source(s)	Purpose	Type of measure
Telemedicine Satisfaction Questionnaire (TSQ)	Otten *et al.* (2015)	To measure expected satisfaction with Telemedicine and perceived user satisfaction.	Rating scale Genetics specific
Tennessee Self-Concept Scale	Fitts (1965)	The scale is intended to summarize an individual's feeling of self-worth, the degree to which the self-image is realistic, and whether or not that self-image is a deviant one.	Rating scale Non-genetics specific
Utrechtse Coping List (UCL)	Westbrook (1979)	To evaluate coping strategies such as: active coping, palliative coping, avoiding reactions, social support seeking, depressive-regressive coping, expression of emotions or anger and comforting ideas.	Rating scale Non-genetics specific
Worry Interference Scale (WIS)	Trask (2001)	To assess the degree to which thoughts about breast cancer are perceived as interfering with the respondents' daily functioning.	Rating scale Genetics specific

Adapted from Payne *et al.* (2008) Outcome Measurement in Clinical Genetics Services: A systematic review of validated measures.

2.3.1 Outcome Measures: General Properties

Table 2.1 presents all 82 validated outcome measures identified in this literature review. Half (n=41; 50.0%) of the measures can be described as being 'genetics-specific', i.e. they contain items which specifically refer to genetics or a genetic condition. Similarly, over half (n=46; 56.1%) were used in studies that evaluated CGS with respect to inherited cancers, primarily breast cancer. The style of questionnaire varies, but in general they are composed of a series of statements that require a rating on a scale. For example, in Benkendorf's measure 'Knowledge about genetic testing for inherited cancer' (Benkendorf *et al.,* 1997) one statement says: "A person should be able to get a genetic test even if their doctor recommends against it." Respondents are then asked to (i) Strongly Agree; (ii) Agree; (iii) Disagree; or (iv) Strongly Disagree. These types of rating scales are known as 'Likert Scales'. Four instruments provided respondents with multiple choice options, for example the measure of Decision-making process developed by Michie *et al.* (1997). This scale contained three multiple-choice questions, designed to assess the time spent thinking about whether or not to have a test, the number of people this was discussed with, and how many reasons (for or against) were considered by the respondent. Three measures offered True/False options.

2.3.2 Outcome Measures: Outcome Domains

A variety of different outcome domains are captured by these instruments, for example satisfaction with genetic counselling (Shiloh *et al.,* 1990), knowledge about genetic testing for inherited cancer (Lerman *et al.,* 1996), and psychological adaptation to genetic information (Read *et al.,* 2005). Psychological or emotional domains were particularly common, with over 20 measures being specifically designed to capture concepts such as depression, anxiety or worry. Similarly, 11 measures examine patient knowledge with regard to the condition, risk figures, or testing interventions, and 12 measures examine patient satisfaction. Two instruments study outcomes from the perspective of the physician: the modified Tolerance for Ambiguity Scale (Geller *et al.,* 1993) and the Pharmacogenetics in Psychiatry Follow-up Questionnaire (PIP-FQ) (Walden *et al.,* 2015). Of the 82 instruments identified, only three encompass a wide range of potential patient benefits from CGS: The Audit Tool for Genetics Services (Skirton *et al.,* 2005), the Perceived Personal Control (PPC) questionnaire (Berkenstadt

et al., 1999), and the Genetic Counselling Outcome Scale (GCOS-24) (McAllister *et al.,* 2011b).

2.3.3 Outcome Measures: Validation

Table 2.2 summarises the extent of psychometric validation for the 82 outcome measures identified in this review. Approximately one quarter (n=21; 25.6%) were assessed for internal consistency alone. Internal consistency is a reliability statistic, denoting the degree of correlation between items in a scale. It has become the primary method of estimating the reliability of multi-item scales, and is indexed using Cronbach's coefficient alpha (Frost *et al.,* 2007). The internal consistency value is commonly interpreted as indicating whether items which propose to measure a certain dimension do in fact measure the same dimension as each other. The remaining measures underwent more extensive psychometric assessment, for example content validity (n=25) and construct validity (n=29), but there was limited assessment of sensitivity to change (n=6) or interpretability (n=2) – key requirements for any questionnaire intended for use as a PROM (Mokkink *et al.,* 2010; Terwee *et al.,* 2012). Definitions for these terms are provided in Table 2.3.

Table 2.2: Validation of Outcome Measures Identified in the Literature Review

Outcome measure	Primary Source(s)	Validation
Anticipated impact of results	Hailey *et al.* (2000) Lerman *et al.* (1995)	**Internal Consistency**
Appropriateness of Genetic Testing Delivery	Andrea *et al.* (2018)	**Face Validity (part)**
Assessment of benefits and risk of breast cancer testing	Hailey *et al.* (2000) Lerman *et al.* (1995)	**Internal Consistency**
Audit Tool for Genetic Services	Skirton *et al.* (2005)	**Face Validity** **Content Validity**
Beck Depression Inventory (BDI)	Su *et al.* (2009)	**Internal Consistency** **Content Validity** **Construct Validity** **Criterion Validity**
Beliefs About Breast Cancer Genetic Testing	Bowen *et al.* (2002)	**Internal Consistency**
Body Image/Sexuality Scale (BISS)	Van Oostrum *et al.* (2003) Lodder *et al.* (2002)	**Internal Consistency** **Retest Reliability**
(Breast) Cancer Attitude Inventory (CAI)	Berrenberg (1991)	**Internal Consistency**

and Anxiety sub-scale (BCANX)	Hailey *et al.* (2000)	**Retest Reliability** **Construct Validity**
Breast cancer (hereditary) concern	Stalmeier *et al.* (1999)	**Internal Consistency**
Breast Cancer Genetic Counselling Knowledge Questionnaire (BCGKQ-27)	Erblich *et al.* (2005)	**Internal Consistency** **Content Validity** **Criterion Validity**
Breast Cancer Worry	Lerman *et al.* (1991); Van Oostrum *et al.* (2003)	**Internal Consistency** **Retest Reliability**
Brief Symptom Inventory	Derogatis & Melisaratos (1983)	**Internal Consistency** **Retest Reliability**
Cancer Anxiety and Helplessness Scale	Kash *et al.* (1992)	**Internal Consistency**
Center for Epidemiologic Studies Depression-Scale (CES-D)	Radloff (1977) Ross & Mirowsky (1984)	**Internal Consistency** **Retest Reliability** **Construct Validity**
Clinical Genetics Satisfaction (CGS) Indicator	Zellerino *et al.* (2009)	**Internal Consistency**
Decision Evaluation Scale	Stalmeier *et al.* (2005)	**Internal Consistency** **Content Validity** **Construct Validity**
Decisional Conflict Scale (DCS)	O'Connor (1995)	**Internal Consistency** **Retest Reliability** **Construct Validity** **Criterion Validity**
Decision making process	Brain *et al.* (2005); Michie *et al.* (1997)	**Internal Consistency**
Desire to participate in the program	Stalmeier *et al.* (1999)	**Internal Consistency**
Emotional reaction to the program	Stalmeier *et al.* (1999)	**Internal Consistency**
Evaluation of practical issues and responsibilities	Otten *et al.* (2016)	**Content Validity**
Expectations of online counselling	Otten *et al.* (2016)	**Content Validity**
Family Environment Scale (FES)	Moos & Moos (1994) Halvorsen (1991)	**Internal Consistency** **Retest Reliability** **Face Validity** **Content Validity** **Construct Validity**
Functional Assessment of Cancer Therapy-General (FACT)	Cella *et al.* (1993) Brady *et al.* (1997)	**Internal Consistency** **Retest Reliability** **Construct Validity** **Sensitivity**
General Health Questionnaire (GHQ)	Goldberg & Williams (1988). Goldberg & Hillier (1979)	**Internal Consistency** **Retest Reliability**

		Content Validity **Construct Validity** **Criterion Validity** **Interpretability** **Sensitivity**
Genetics Appointment Patient Satisfaction Score (GAPPS)	Westwood *et al.* (2012)	**Content Validity**
The Genetic Counseling Outcome Scale (GCOS-24)	McAllister *et al.* (2011b)	**Internal Consistency** **Face Validity** **Content Validity** **Construct Validity** **Retest Reliability** **Sensitivity**
Genetic Counseling Satisfaction Scale (GCSS)	Tercyak *et al.* (2001)	**Internal Consistency** **Face Validity**
Genetic Knowledge Index (GKI)	Furr & Kelly (1999)	**Internal Consistency** **Construct Validity**
Global Severity Index (GSI) of the Symptom Check List-90 (SCL90)	Derogatis (1983)	**Internal Consistency** **Retest Reliability** **Construct Validity** **Criterion Validity** **Sensitivity**
Health Beliefs Model	Kash *et al.* (1992)	**Face Validity**
Health Orientation Scale	Woolridge & Murray (1989)	**Internal Consistency** **Retest Reliability** **Construct Validity**
Hopkins Symptom Checklist (HSCL)	Derogatis *et al.* (1974)	**Internal Consistency** **Retest Reliability** **Construct Validity** **Criterion Validity**
Hospital Anxiety and Depression Scale (HADS)	Zigmond & Snaith (1983)	**Internal Consistency** **Content Validity**
Illness Perception Questionnaire (IPQ)	Cho *et al.* (2012)	**Content Validity**
Impact of Event Scale (IES)	Horowitz *et al.* (1979)	**Internal Consistency** **Retest Reliability** **Sensitivity**
Intention to act upon program	Stalmeier *et al.* (1999)	**Internal Consistency**
Knowledge About Breast Cancer	Donovan & Tucker (2000) Stager (1993) Vaeth (1993)	**Internal Consistency** **Content Validity**
Knowledge about genetic testing for inherited cancer	Lerman *et al.* (1996)	**Internal Consistency**

Knowledge about genetic risk for breast cancer	Donovan & Tucker (2000)	**Internal Consistency**
Knowledge Scale about Breast (and Ovarian) Cancer and Hereditary	Ondrusek *et al.* (1999)	**Retest Reliability** **Content Validity**
Life Orientation Test (LOT)	Scheier *et al.* (1994) Carver *et al.* (1994)	**Internal Consistency** **Retest Reliability** **Construct Validity**
Measure of Counselees' Knowledge of Down Syndrome	Braitman & Antley. (1978)	**Face Validity** **Content Validity** **Internal Consistency**
Medical Communication Behaviour System (MCBS)	Wolraich *et al.* (1986)	**Content Validity** **Construct Validity** **Criterion Validity**
Medical Interview Satisfaction Scale - modified (MISS)	Wolf *et al.* (1978)	**Internal Consistency** **Content Validity**
Medical Outcomes Short-Form Survey (SF-36 and SF-12)	Ware (1993) Jenkinson *et al.* (1996)	**Internal Consistency** **Content Validity** **Criterion Validity** **Construct Validity**
Medical Outcomes Study Social Support Scale (MOSS)	Sherbourne *et al.* (1991)	**Internal Consistency** **Retest Reliability** **Construct Validity**
Miller Behavioural Style Scale	Miller (1987)	**Internal Consistency** **Retest Reliability**
Minnesota Multiphasic Personality Inventory (MMPI)	Graham (1987)	**Internal Consistency** **Retest Reliability** **Content Validity** **Construct Validity** **Criterion Validity**
Modified Maternal Serum Screening Knowledge Questionnaire (MSSKQ)	Goel *et al.* (1996)	**Internal Consistency**
modified Tolerance for Ambiguity Scale (TFA)	Geller *et al.* (1993)	**Internal Consistency** **Content Validity**
Multidimensional Impact of Cancer Risk Assessment (MICRA)	Cella *et al*. (2002)	**Internal Consistency** **Construct Validity** **Criterion Validity**
Openness to Discuss Cancer in the Family Scale (ODCFS)	Mesters *et al.* (1997)	**Internal Consistency** **Content Validity** **Criterion Validity**
Patient Health Questionnaire (PHQ-9)	Meiser *et al.* (2013)	**Content Validity**
Patient Satisfaction with Genetic	Brain *et al.* (2000); Shiloh *et al.*	**Internal Consistency**

Counselling	(1990)	
Penn State Cancer Genetics Program Survey	Kausmeyer *et al.* (2006)	**Content Validity**
Perceived Devaluation Discrimination Scale (PDDS)	Meiser *et al.* (2013)	**Internal Consistency**
Perceived personal control (PPC)	Berkenstadt *et al.* (1999) Otten *et al.* (2016)	**Internal Consistency** **Construct Validity** **Content Validity** **Sensitivity**
Perceived Risk of Breast Cancer	Brain *et al.* (1999)	**Internal Consistency**
Perceptions of the benefits, limitations and risks of genetic testing	Donovan & Tucker (2000); Hughes *et al.* (1997) ; Audrain *et al.* (1995)	**Internal Consistency**
Pharmacogenetics in Psychiatry follow-up questionnaire (PIP-FQ)	Walden *et al.* (2015)	**Internal Consistency**
Profile of Mood State (POMS)	McNair *et al.* (1981)	**Internal Consistency** **Retest Reliability** **Face Validity**
Prostate cancer genetic screening survey	Doukas (2004)	**Internal Consistency**
Psychological Adaptation to Genetic Information Scale (PAGIS)	Read *et al*. (2005)	**Internal Consistency** **Content Validity**
Psychological Consequences Questionnaire (PCQ)	Cockburn *et al.* (1992)	**Internal Consistency** **Content Validity** **Construct Validity**
Quality of Care Through the Patients' Eyes (QUOTE)-gene^{CA}	Pieterse *et al.* (2005)	**Internal Consistency** **Content Validity** **Construct Validity**
Risk comprehension and subjective knowledge	Stalmeier *et al.* (1999)	**Internal Consistency**
Rosenberg Self-Esteem Scale	Rosenberg (1965) Curbow & Somerfield (1991)	**Internal Consistency** **Retest Reliability**
Satisfaction with Decision Scale	Brain *et al.* (2005) Holmes-Rovner *et al.* (1996)	**Internal Consistency**
Satisfaction with Genetic Counselling Questionnaire	Hilgart *et al.* (2012)	**Content Validity**
Satisfaction with shared decision making program	Stalmeier *et al.* (1999)	**Internal Consistency** **Construct Validity**
Self-rating Depression Scale (SDS)	Zung (1965)	**Internal Consistency** **Content Validity** **Face Validity** **Construct Validity**

Shared decision making program rationale acceptability	Stalmeier *et al.* (1999)	**Internal Consistency**
Short-form Health Survey (SF-12)	Hubalek *et al.* (2016)	**Internal Consistency** **Retest Reliability** **Face Validity** **Content Validity** **Construct Validity**
Short-form Health Survey (SF-36)	Bowen & Powers (2010)	**Internal Consistency** **Retest Reliability** **Face Validity** **Content Validity** **Construct Validity**
Spielberger State Trait Anxiety Inventory (STAI) and state scale (STAI-State)	Spielberger et al. (1970) Marteau & Bekker (1992)	**Internal Consistency** **Retest Reliability** **Face Validity** **Content Validity** **Construct Validity** **Criterion Validity** **Sensitivity** **Interpretability**
Spiritual Well-Being Scale (SWBS)	Ellison & Smith (1991) Gioiella *et al.* (1998)	**Internal Consistency** **Retest Reliability** **Criterion Validity**
Subjective Quality of Life Profile (SQLP)	Dazord (1995)	**Internal Consistency** **Retest Reliability** **Criterion Validity** **Construct Validity**
Telemedicine Satisfaction Questionnaire (TSQ)	Otten *et al.* (2015)	**Internal Consistency**
Tennessee Self-Concept Scale	Fitts (1965)	**Internal Consistency** **Retest Reliability** **Construct Validity** **Criterion Validity**
Utrecht Coping List (UCL)	Westbrook (1979)	**Internal Consistency**
Worry Interference Scale (WIS)	Trask (2001)	**Internal Consistency** **Retest Reliability** **Content Validity**

Table 2.3 Definitions of scale psychometric properties.

Content Validity	A non-statistical assessment of whether the measure covers the totality of the underlying theoretical construct.
Concurrent Validity	The extent to which the results of a test correspond to those of a previously established test for the same construct.
Construct Validity	The extent to which a measure captures the underlying theoretical construct.
Criterion Validity	The extent to which a measure is related to an outcome, i.e. the correlation between a test and an outcome.
Face Validity	The degree to which a scale appears effective with respect to its aim.
Internal Consistency	The degree of correlation between items in a scale. Indexed using Cronbach's alpha.
Interpretability	Assigns a numerical value to represent the degree to which a meaning is derived from a term, item or measure. Usually assessed using minimal important change (MIC) or minimal important difference (MID).
MIC / MID	The smallest change in a PRO that patients perceive as important. See interpretability
Preference-based	Reflecting the value or priority which is placed on each item by the target population. This allows changes in health state to be interpreted.
Responsiveness / Sensitivity to Change	Also called 'responsiveness'. The ability of an instrument to accurately assess change in the measured construct.
Test-Retest Reliability	The degree to which the test produces consistent results over two time periods.

2.3.4 Results Summary

In summary, this literature review identified 82 validated outcome measures used in the evaluation of CGS. A variety of different domains are captured by these measures, but many only pertain to a specific outcome and so represent a limited perspective of what CGS can offer patients. The Audit Tool for Genetics Services (Skirton *et al.,* 2005), the Perceived Personal Control (PPC) questionnaire (Berkenstadt *et al.,* 1999), and the Genetic Counselling Outcome Scale (GCOS-24) (McAllister *et al.,* 2011b) are the only instruments which capture a range of potential CGS patient outcomes. Additionally, the extent of psychometric validation was often low, with approximately one quarter being assessed for internal consistency alone. The results will now be discussed.

2.4 Discussion

This literature review has identified 82 validated outcome measures, either developed or used in the evaluation of CGS. Generic measures of physical health status were not commonly used, which is not surprising given that interventions offered by CGS are generally not able to provide physical health benefits. A small number of studies, however, utilised the generic Short-form Health Survey (SF-36) and the reduced version SF-12 to measure health status in the context of cancer genetics. Hubalek *et al.* (2016), for example, included SF-12 in a bundle of seven PROMs sent out to patients in order to investigate the long-term psychosocial consequences of genetic counselling and testing for hereditary breast and ovarian cancer. Bowen & Powers (2010) included SF-36 as part of a before-and-after study, in which six separate measures were applied to gather data on cancer worry, estimated risk for breast cancer, quality of life, knowledge of breast cancer, and awareness and perception of genetic testing. Items common to both SF-36 and SF-12 include: 'In general, would you say your health is...' (Excellent – Poor) and 'Does your health now limit you in climbing several flights of stairs?' (Yes, a lot – No, not limited at all). All studies in this review which utilised a generic health measure did so in conjunction with other measures, emphasising the fact that generic health measures are not sufficient to capture CGS outcomes. Indeed the majority of outcomes measures used to evaluate CGS capture Health-Related Quality of Life (HRQoL) outcomes, including the physical, emotional, psychological, and social domains of health.

Almost half of the instruments refer to genetics or a genetic condition. An example of a genetics-specific instrument is Erblich *et al.*'s Breast Cancer Genetic Counselling Knowledge Questionnaire (BGKQ) (Erblich *et al.,* 2005), a 27-item instrument developed with the aim of assessing women's knowledge of information presented during breast cancer genetic counselling. Some items are scored using a True / False / I don't know system, e.g. '50% of inherited genetic information (about breast cancer risk) is passed down from a person's mother' and 'One in 10 women has a breast cancer gene mutation', and some items offer multiple choice, e.g. 'What is the approximate risk that the average woman in the United States will develop breast cancer in her lifetime? (a. 12%; b. 24%; c. 58%; d. 72%; e. I don't know)'. One of the benefits of genetics-specific measures is that they have often been designed to include specialised items, capturing distinct outcomes relevant to the intended context. If the

specialised items are condition-specific, however, as with the BGKQ, the wider application of the instrument in CGS is limited.

Aside from the generic health measures, the majority of non-genetics-specific measures were used to capture a singular outcome domain, known to be relevant in the context of CGS. The revised Life Orientation Test (Scheier *et al.,* 1994), for example, is a ten-item measure of optimism versus pessimism. Respondents are asked to designate their level of agreement ('I agree a lot' to 'I disagree a lot') with items such as 'It's easy for me to relax' and 'I'm always optimistic about my future'. The 20-item Medical Outcomes Study (MOS) Social Support Survey (Sherbourne & Stewart, 1991) was designed to comprehensively assess various dimensions of social support. A five-point Likert scale ranging from 'None of the Time' to 'All of the Time' is presented, with items including 'How often would someone be able to help you if you were confined to bed?' and 'How often does someone show you love and affection?'. Any measure intended for use in CGS evaluations should capture a range of potential patient outcomes provided by the service.

In summary, generic measures of health will likely not be appropriate in the context of CGS. Both genetics-specific and non-genetics-specific instruments were identified which measure relevant HRQoL outcomes, but if an instrument is to be used as a universal PROM in CGS it must be applicable to all potential CGS patients and must capture a range of potential patient outcomes.

2.4.1 The Narrow Scope of Existing Measures

The majority of measures identified in this study are designed to capture a specific outcome or a restricted number of outcomes. Common outcome domains include patient knowledge regarding the condition, patient satisfaction with the genetic counselling process, anxiety and depression. Whilst such measures may be valid and robust, and highly relevant in specific contexts, they fail to take into account the range of potential benefits that CGS can offer. The Psychological Consequences Questionnaire (PCQ), for example, was developed by Cockburn *et al.* (1992) to assess the consequences of breast mammography on well-being. It contains 12 items, each rated on a four-point scale with options ranging from 'not at all' to 'quite a lot of the time', and respondents are instructed to indicate how often they had experienced

social, physical, and emotional reactions in the previous week as a result of concerns about breast cancer (e.g. 'have you experienced a change in appetite'; 'have you been scared or panicky'; and 'have you felt worried about your future'). The instrument has good construct validity, concurrent validity and internal consistency, and has since been used in subsequent studies examining emotional well-being in women receiving counselling for breast cancer risk (Kent *et al.,* 2010; Rijnsburger *et al.,* 2006). As a universal PROM for CGS evaluations, the PCQ is too specific to be suitable.

A number of measures were specifically designed to capture depression, for example the Self-rating Depression Scale (SDS) (Zung, 1965), the Beck Depression Inventory (Su *et al.,* 2009), and the Hospital Anxiety and Depression Scale (Zigmond & Snaith, 1983). The 20-item SDS was constructed for the purpose of assessing the physiological and psychological symptoms of depression, and contains items such as: 'I feel downhearted and blue'; 'I have trouble sleeping at night'; and 'I am more irritable than usual' (Zung, 1965). Respondents are asked to select one of four options from 'A Little of the Time' to 'Most of the Time'. Depression is certainly relevant in the context of CGS, with several studies indicating that a substantial proportion of individuals seeking genetic counselling for hereditary cancer have high levels of anxiety and depression (Geirdal *et al.,* 2005; Reichelt *et al.,* 2004; Schlich-Bakker *et al.,* 2006). Genetic counselling has also been shown to reduce depression levels in individuals at risk for hereditary cancer (Bjorvatn *et al.,* 2008). None of the measures of depression identified in this review, however, are sufficient to evaluate the complex range of potential patient benefits from CGS.

Patient knowledge is another important element of genetic counselling and a valuable outcome in the eyes of the NHS. Indeed in Resta *et al.*'s (2006) definition for the speciality it states that genetic counselling integrates 'Education about inheritance, testing, management, prevention, resources and research' (p77). In 1989, information giving was listed by the NHS during their proposed reforms, stating that hospitals should offer patients 'clear and sensitive explanations of what is happening, on practical matters such as where to go and who to see, and on clinical matters such as the nature of an illness and its proposed treatment' (DoH, 1989, paragraph 1.13).

The measures of knowledge used to date have mainly been specific to a certain condition. The 'Measure of Counselees' Knowledge of Down Syndrome' constructed by Braitman & Antley (1978), for example, is a 26-item test with items such as: 'What are

the chances that the brother or sister of a person with Down syndrome will have a baby with Down syndrome?' and 'Children with Down syndrome always have an extra chromosome or an extra piece of a chromosome (True / False)'. Similarly, the 'Risk Comprehension and subjective knowledge' test used by Stalmeier *et al.* (1999) is specific to breast cancer. An example item reads, 'What percentage of women (average women in the general population) get breast cancer before the age of 70?' The wider application of these condition-specific measures is limited. Additionally, using measures of knowledge or information recall to evaluate CGS can be problematic. The value placed on certain pieces of information will vary from person to person, as will the interpretation of information, particularly risk figures (Clarke *et al.,* 1996). Several findings also indicate that educational or informational elements of genetic counselling provide fewer benefits and are relatively less important to CGS users than supportive or emotional elements (Bowen *et al.,* 2004; Lerman *et al.,* 1997; Edwards *et al.,* 2008).

Twelve PROMs were designed to capture patient satisfaction. This may be satisfaction with the genetic counselling process (Otten *et al.,* 2016), satisfaction with a medical decision (Holmes-Rovner *et al.,* 1996), or satisfaction with respect to quality of life (Ellison & Smith, 1991). The wider literature suggests that CGS patients are generally highly satisfied with the service, finding genetic counselling to be informative and helpful (Bleiker *et al.,* 1997; DeMarco *et al.*, 2004; Nordin *et al.,* 2002; Sagi *et al.,* 1998; Schneider *et al.,* 1999; Shiloh *et al.,* 1990; Stadler & Mulvihill*,* 1998; Veach *et al.,* 1999). Patient satisfaction, however, may be dependent on a number of factors, and it is often not clear what aspects of the service are driving satisfaction levels. Bernhardt *et al.* (2000) found that one of the things the majority of clients liked most about their genetic counselling experience was their genetic counsellor, and clients spent a considerable amount of time during the follow-up interviews talking about how well they 'connected' with their counsellor. In contrast, the information provided to patients regarding a condition may cause significant distress. Whilst it is important to measure CGS outcomes from the patients' perspective, global patient satisfaction levels are not widely seen as a suitable metric for success in CGS (Clarke *et al.*, 1996, Payne *et al.,* 2008). Attention must instead be focused upon specific elements of the service, for example in Stalmeier *et al.*'s (1999) Satisfaction with the Shared Decision Making Program (SDMP) scale. Items include, 'Did the SDMP give you more/less insight in the treatment choice?' and 'Did the SDMP enable you to discuss your problem better/worse with others?'

Of the 82 instruments identified in this review, only three incorporate a range of outcome domains relevant to CGS: The Audit Tool for Genetics Services (Skirton *et al.,* 2005), GCOS-24 (McAllister *et al.,* 2011b), and the PPC questionnaire (Berkenstadt *et al.,* 1999). The 18-item Audit Tool was the result of a study aiming to develop a practical research and audit tool to measure outcomes of CGS (Skirton *et al.,* 2005). The questionnaire addresses six outcome domains (with example items in parenthesis): (i) Enhanced understanding ('I have more understanding of what causes the condition'); (ii) Positive psychological change ('I feel more positive'); (iii) Respect for autonomy ('My main questions were answered'); (iv) Adaptation ('I feel I can adapt better to changes'); (v) Disequilibrium ('I did not feel comfortable'); (vi) Value of contact ('I felt treated as an individual'). Responses are assessed on a seven-point Likert scale ranging from 'totally agree' to 'totally disagree'. However, for a health measurement scale to be suitable for use in service evaluation it must be sensitive to change. Due to item wording, The Audit Tool can only be used post-counselling, and is therefore unable to measure pre/post change.

The concept of PPC, established by Averill (1973) to reflect the extent to which a person believes that they are in control of a situation and that they are able to bring about positive changes to the situation, was operationalised as a measure for genetic counselling by Berkenstadt *et al.* (1999). The instrument captures a range of outcomes in genetic counselling, asking counselees their subjective perceptions of how much control they believe they have with regard to their genetic problem. More specifically, the PPC scale contains nine items representing three dimensions of control: Cognitive Control (e.g. 'I think I understand what problem brought me to genetic counselling'); Behavioural Control (e.g. 'I feel I know what to do to ease the situation'); and Decisional Control (e.g. 'I feel I have the tools to make decisions that will influence my future'). The PPC scale is valid, reliable, and responsive to change pre/post genetic counselling, and has been shown to be highly relevant as a patient reported outcome, valued by both patients and genetics clinicians (Payne *et al.,* 2007; McAllister *et al.,* 2012). Great Ormond Street Hospital, in their most recent biennial CGS questionnaire, used an adapted version of the PPC measure to evaluate CGS, with results suggesting that CGS appointments improve patients' understanding of what the genetic condition means for them and their families, as well as patients' sense of confidence in having the information to make choices.

GCOS-24 (McAllister *et al.,* 2011b) is a 24-item questionnaire which captures empowerment (Table 2.4) (McAllister *et al.,* 2011a). Empowerment includes all three PPC dimensions, as well as two further dimensions, Hope and Emotional Regulation, which represent elements such as anxiety, guilt, and hope for the future. It was developed through extensive qualitative research with genetics clinicians and those affected by having a genetic condition in the family. In an initial study, seven focus groups and 19 interviews were conducted with patients, patient group representatives, and health professionals (McAllister *et al.,* 2008). Following on from this, empowerment was validated and refined through further qualitative research with 12 patients, 15 representatives from patient support groups, 10 genetics clinicians and 4 service commissioners (McAllister *et al.,* 2011a). GCOS-24 has been shown to have a high degree of clinical utility, being used for service evaluation (Inglis *et al.,* 2014; McAllister *et al.,* 2016) and quality improvement (Costal-Tirado *et al.,* 2017) in CGS. It has also received international attention, being translated into Danish (Diness *et al.,* 2017) and Spanish (Munoz-Caballo *et al.,* 2017).

Table 2.4: Empowerment. (McAllister *et al.* 2011a).

Empowerment Dimension	**Definition (The belief that one**...)
Cognitive Control	...has sufficient information about the condition, including risks to oneself and one's relatives, and any treatment, prevention and support available.
Decisional Control	...can make important life decisions in an informed way.
Behavioural Control	...can make effective use of the health and social care systems
Emotional Regulation	...can manage their feelings about having a genetic condition in the family
Hope	...can look to the future having hope for a fulfilling family life, for oneself, one's family, and/or one's future descendents

In summary, the majority of measures identified in this review encompass only a narrow scope of potential patient outcomes which CGS can provide. Frequently observed outcome domains included patient knowledge, patient satisfaction, and depression, but each only represents a certain element within the complex array of CGS outcomes. Additionally, objective measures of information recall and of satisfaction can be problematic when used as indicators of service quality or patient

benefit. The 9-item PPC and the 24-item GCOS-24 both capture multi-dimensional constructs, incorporating outcomes relating to 'Cognitive Control', 'Decisional Control', and 'Behavioural Control'. GCOS-24 goes even further, including 'Emotional Regulation' and 'Hope' (Table 2.4). Extensive qualitative research suggests that these outcomes are relevant and valued by CGS users, and both instruments have a high degree of clinical utility.

2.4.2 The Heterogeneity of Existing Measures

The results from this review demonstrate a noticeable lack of consensus over the best way to evaluate patient outcomes from CGS, a sentiment echoed by other authors (Clarke *et al.,* 1996; McAllister *et al.,* 2008; McAllister & Dearing, 2015; Munoz-Cabello *et al.,* 2018; Payne *et al.,* 2008; Wang *et al.,* 2004). Indeed the high degree of heterogeneity in outcome measures has been identified as being a principal methodological limitation in reviews of economic evaluations in CGS, causing difficulties when making comparisons and drawing conclusions (Carlson *et al.,* 2005; Djalalov *et al.,* 2011). In a recent review of evaluations of predictive genetic testing programs (D'Andrea *et al.*, 2015), the variety of results produced by the various outcome measures was such that results could not be pooled and statistical methods could not be applied; a descriptive approach was taken instead.

Over half of the measures identified in this review were used in the evaluation of CGS for inherited cancers, primarily breast cancer. Many instruments were developed for use in a specific study, and would not be applicable in any other context. In Kausmeyer *et al.* (2006), for example, the aim of the study was to explore patient expectations, experiences and satisfaction with the Penn State Cancer Genetics Program, and a bespoke survey was developed accordingly. 'The Penn State Cancer Genetics Program Survey' contains 80 multiple choice items, including: 'How did you hear about the Penn State Cancer Genetics Program?' and 'Did the Cancer Genetics Packet and appointment letter mailed prior to your visit provide useful information regarding the cancer risk assessment process?' Similarly, Stalmeier *et al.* (1999) composed a number of novel bespoke measures to evaluate a Shared Decision Making Program (SDMP) for women suspected to have a genetic predisposition to breast cancer. Outcome domains included desire to participate in the SDMP, satisfaction with the SDMP, and the intention to act upon the SDMP, with items such as 'Did the SDMP give you more/less

insight in the treatment choice?' and 'Imagine that a close friend would have a high risk for breast cancer. Would you recommend the SDMP?' Instruments such as these are bound by their wording to be relevant only in a specific context, and comparisons with other instruments are difficult if not unworkable.

Two instruments studied outcomes from the perspective of the physician, the modified Tolerance for Ambiguity Scale (Geller *et al.,* 1993) and the Pharmacogenetics in Psychiatry Follow-up Questionnaire (PIP-FQ) (Walden *et al.,* 2015). The PIP-FQ was designed to assess physicians' perceptions of pharmacogenetic testing and their experience using the test results. Items include 'Has the information been easy to understand?' and 'Based on your experience, would you refer additional patients into our study?' Evaluating CGS from the perspective of the provider is not considered to be best practice (Clarke *et al.,* 1996). In a study by Wertz *et al.* (1988), patient outcomes as judged by the provider appeared to be associated with the education level of the patient rather than whether the needs of the patient had been met. Bernhardt *et al.* (2000) describe the idea of counsellor expectations influencing their perception of patient outcomes, saying that some counsellors expect their counsellees to show some level of engagement, and are often dissatisfied if there is a reduced level of response. From the patient's perspective, a reduced response may simply mean that they are listening and taking in the information they have been given.

In the absence of a universal instrument, a number of studies evaluating CGS chose to adapt an existing measure rather than develop a novel one. Van Oostrum *et al.* (2003) adapted the Openness to Discuss Cancer in the Family Scale originally constructed by Mesters *et al.* (1997) to assess the impact of genetic testing for cancer susceptibility on family relationships; Bowen *et al.* (2002) modified certain questions from the Tolerance for Ambiguity Scale (Geller *et al.,* 1993), as well as certain questions from a scale measuring fear of stigma associated with cystic fibrosis, for use in the context of breast cancer. The extent of scale adaptation varied from study to study, but as a whole this practice emphasises the lack of harmony regarding measurement scales in the context of CGS.

Having a suitable PROM accepted as the standard in CGS will enable patient outcomes to be compared and contrasted between separate interventions. It will help to identify which services are effective and of value, to encourage common practice, and to provide robust evidence for audit and service development. This was emphasised by

the National Institute for Clinical Excellence (NICE) in 2004, who called for the establishment of generic outcome measures to allow separate interventions to be compared directly (NICE, 2004). Ultimately, a standard measure will help to ensure that CGS patients are receiving optimal medical care.

Of the three measures designed to capture a range of outcome domains from CGS (The Audit Tool (Skirton *et al.,* 2005); PPC (Berkenstadt *et al.,* 1999); GCOS-24 (McAllister *et al.,* 2011b)) only the PPC and GCOS-24 were identified in more than one study. GCOS-24 is of particular note, since it was created with the intention of filling the gap generated by the lack of a universal PROM within CGS (McAllister *et al.,* 2011b). Despite being developed relatively recently, it has gone on to be used in multiple studies, both within the UK and internationally (Costal-Tirado *et al.,* 2017; Diness *et al.,* 2017; Munoz-Caballo *et al.,* 2017).

2.4.3 The Limited Validation of Existing Measures

Internal consistency, test-retest reliability and validity (Table 2.3) are essential properties for any measurement scale (Aaronson *et al.*, 2002; Mokkink *et al.,* 2010). The 2010 Consensus-based Standards for the selection of health measurement instruments (COSMIN) checklist, which provides guidelines for assessing the methodological quality of measurement scales, also describes how content validity, construct validity, responsiveness and interpretability are also relevant criteria to be considered when assessing a measurement scale (Mokkink *et al.,* 2010). Many of the measures identified in the review had undergone limited psychometric evaluation, with over half being assessed for internal consistency alone. For the purposes of this review any form of validation was sufficient for inclusion, but it could be argued that internal consistency alone is not sufficient evidence to confirm a measure as validated. Since the calculation is based upon item correlations, random error averages out as one adds more items, so in practice scales over 20 items generally have acceptable values of α (>.7) (Streiner, 2003). Shorter scales will have fewer correlations from which to draw upon and in turn may present with lower values. Cronbach's α, the index for internal consistency, would be higher for a 20-item measure with a mean

inter-item correlation of 0.1, than for a 5-item measure with a mean inter-item correlation of 0.3.

Traditional psychometric tests such as internal consistency and test-retest reliability fall into a category of tests known as Classical Test Theory (CTT). CTT approaches have guided the construction, refinement and validation of measurement scales for decades, and continue to remain the dominant paradigm (Petrillo *et al.,* 2015). There are, however, some issues with CTT that concern the calibration of item difficulty, sample dependence of coefficient measures, and estimates of measurement error (Magno, 2009). In short, CTT is a theory about test scores that introduces three concepts: (i) test score, often called the observed score (TO); true score (T), and error score (E), where the true and error scores are independent. These variables within CTT are best illustrated in the formula: TO = T+E. Because for each examinee there are two unknowns to the equation, some simplifying assumptions are made. The assumptions in the CTT model are that: (a) true scores and error scores are uncorrelated; (b) the average error score in the population of examinees is zero; (c) error scores on parallel tests are uncorrelated. In other words, the theory starts from the assumption that systematic effects between test responses are due only to variation in the ability of interest; all other potential sources of variation existing in the testing materials such as external conditions are assumed either to be constant or to have an effect that is random by nature (Linden & Hambleton, 2004). In other formulations of this model (e.g. Lord & Novick, 1968), true score is defined as the expected test score over parallel forms, and then the resulting properties of the error are derived.

Advantages of many CTT models are that they are based on relatively weak assumptions (i.e. they are easy to meet in real test data), and they are well known and have a long track record. On the other hand, both person parameters and item parameters are dependent on the test and the examinee sample, respectively, and these dependencies can limit the utility of the person and item statistics in practical test development.

Item Response Theory (IRT) is a relatively recent approach to psychometric design, developed to overcome the problems with CTT approaches (Wiberg, 2004). In IRT, it is assumed that an examinee has some latent unobservable trait (also called ability), which cannot be studied directly. The purpose of IRT is to propose models that permit to link this latent trait to some observable characteristics of the examinee. According

to Sohn (2009), one of the distinguishing characteristics of item indices under CTT and IRT frameworks is whether they are sample dependent or invariant. Whereas in CTT one uses a common estimate of the measurement precision that is assumed to be equal for all individuals irrespective of their ability level, in IRT the measurement precision depends on the ability (latent trait) value. As a result, IRT models will theoretically produce item statistics which are independent of examinee samples, and person statistics independent of the particular set of items administered. This invariance property of item and person statistics of IRT has been illustrated by Hambleton & Swaminathan (1985); Hambleton, Swaminathan and Rogers (1991).

The calculations involved in IRT models also make them preferable to CTT when analysing ordinal responses; e.g. Likert rating scales. For although the response categories in Likert scales have a rank order, it is not necessarily correct to presume that the intervals between values are equal. By way of example, would the 'difference' between Disagree and Strongly Disagree be the same as that of Agree and Strongly Agree? Treating ordinal scales as interval scales has long been controversial, and the subjective and ordinal nature of Likert scale data has proven problematic for formal statistical analysis (Jamieson, 2004). IRT methods were specifically developed to address the issue of subjective ordinal responses and the need to create robust measures.

Thus, IRT has been considered to hold a number of advantages over CTT, and from a practical perspective IRT methods can greatly assist in the construction and refinement of PROMs (Hays *et al.,* 2000; Nguyen, 2014). Indeed they are already being applied to some of the major PROMs, such as the EQ-5D, HUI2, HUI3, and SF-6D (Fryback *et al.,* 2009; Gibbons *et al.,* 2014; Johnsen *et al.,* 2013; van Hout *et al.,* 2012). No measure included in this review was developed using IRT, and no study utilised IRT.

Only six measures identified in this review have been assessed for sensitivity. Both PPC and GCOS-24 are well-validated in this respect, as well as for internal consistency, test-retest reliability and content validity (McAllister *et al.,* 2011b; McAllister *et al.,* 2012; Berkenstadt *et al.,* 1999). Neither instrument, however, has been studied for interpretability, which is not unusual since only two of the 82 have (STAI (Spielberger *et al.,* 1970); GHQ (Goldberg & Hillier, 1979)).

2.4.4 Valuing Health States: Preference weights

Over recent years, national decision-making bodies in the UK involved in the appraisal of cost-effectiveness of healthcare interventions have called for outcome measures used in service evaluation to be 'preference based' (NICE, 2004). A preference-based measure is a measure of HRQoL that has a set of 'preference weights' which reflect the value that individuals attach to each item and response option. This allows more desirable outcomes to receive greater weight in the analysis, and enables changes in score to be interpreted.

Nowadays, preference-based measures are being widely used in health economic evaluations and health technology assessments (HTA) within the UK system. Indeed there is in fact a dedicated HTA programme, funded by the NHS, which utilises preference-based measures to examine the clinical effectiveness, the cost effectiveness, and the broader impact of healthcare treatments. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The EQ-5D (Brooks, 1996), for example, is a preference-based measure of health, widely used in cost-effectiveness analysis. The five items relate to domains of mobility, self-care, pain/discomfort, usual activities, and anxiety/depression, and there are three levels of severity: 'no problems', 'some problems', and 'severe problems'. Each response pattern has a preference weight attached, and such is the popularity of the EQ-5D that many separate countries have assigned their own preference weights (Badia *et al.,* 2001 (Spain); Goudarzi *et al.,* 2016 (Iran); Lamers *et al.,* 2006 (Netherlands); Lee *et al.,* 2013 (Taiwan); Wu *et al.,* 2016 (China)).

None of the measures identified in this literature review are preference-based. If a CGS intervention were to be appraised by NICE, no instrument would meet their suggested requirements. With generic measures of health being of limited applicability in CGS, the lack of a relevant preference-based measure is seriously impeding rigorous audit of the service and comparison of different models of service delivery.

2.5 Limitations

One limitation of this review was the exclusion of non-validated outcome measures. This was a practical decision since validated measures are required for robust

evaluations, but may have resulted in potentially relevant instruments being excluded. Additionally, it could be argued in some studies that non-statistical properties such as face validity or content validity were implied. Thompson *et al.* (2015), for example, developed a five-item survey to examine psychiatrist attitudes towards pharmacogenetic testing and integrating genetic counselling into psychiatric patient care. Over 100 surveys were completed by practicing psychiatrists, with results strongly indicating that genetic data would be useful in making pharmaceutical decisions. Due to time constraints the measure was not piloted, and no validation was reported, but the process of construction and subsequent relevance implies face and content validity.

Determining scale validity was not always a straightforward process, particularly with adapted scales. A purist approach would require any changes to a scale to be separately validated, but for the purposes of this review a more flexible, inclusive approach was taken. Therefore not all reported scales were uniquely validated in their own right. The inclusion of a second reviewer during the screening process would have been beneficial. A further limitation is that only studies reported in English were included.

2.6 Refined Research Problem and Study Aims

GCOS-24 (Figure 2.1) emerged from the literature review as being the outstanding candidate for routine use in audit and clinical evaluations of CGS. GCOS-24 items are grounded in extensive qualitative research with CGS patients and providers, and the measure has been demonstrated to be valid, reliable and responsive, with no floor or ceiling effects observed (McAllister *et al.,* 2011b). GCOS-24 has previously been used for service evaluation (Inglis *et al.,* 2015; McAllister *et al.,* 2016) and quality improvement (Costal-Tirado *et al.,* 2017) in genetic counselling services, and it has also received international attention, having been translated into Danish (Diness *et al.,* 2017) and Spanish (Munoz-Caballo *et al.,* 2018). Perhaps most importantly, GCOS-24 captures a range of patient outcome domains from CGS.

However, if GCOS-24 is to meet NICE requirements for use in cost-effectiveness and HTA evaluations, it must have preference-weights attached. At a present length of 24 items each with 7 response options, GCOS-24 produces a substantial number of

possible response permutations ($1.92x10^{20}$). Since preference weights are assigned to each response pattern, it is impossible to design a study to elicit preference weights with such a vast number. The aim of this study was therefore refined, to develop a valid and reliable short form of the GCOS-24, amenable to future development by the addition of preference weights. Standardised and widely-validated PROMs such as the EQ-5D (Brooks, 1996) or SF-6D (Brazier et al., 2002), used in the preference-based evaluation of other branches of healthcare, suggest a five- or six-item measure would be of appropriate length.

Additionally, the wording of GCOS-24 means it is currently unsuitable for use outside of CGS. The first item, for example, reads: 'I am clear in my own mind why I am attending the clinical genetics service', with responses scored on a seven-point Likert scale ranging from 'Strongly Disagree' to 'Strongly Agree'. Genetic testing is increasingly being performed outside the existing models of service provision within CGS and is now moving into other specialities. This process is referred to as 'mainstreaming genetics' and is occurring in the context of cancer predisposition genes (Rahman, 2014), paediatrics (Valente *et al.,* 2008) and neurogenetic testing (Lo *et al.,* 2014). It is therefore becoming ever more important to have a valid and reliable PROM which can be used to evaluate genetic and genomic counselling and testing both within and outside of CGS.

The Genetic Counselling Outcome Scale (GCOS-24)

Using the scale below, circle a number next to each statement to indicate how much you agree with the statement.
Please answer all the questions. For questions that are not applicable to you,
please choose option 4 (neither agree nor disagree).

1 = strongly disagree
2 = disagree
3 = slightly disagree
4 = neither disagree nor agree
5 = slightly agree
6 = agree
7 = strongly agree

		strongly disagree	disagree	slightly disagree	neither agree nor disagree	slightly agree	agree	strongly agree
1	I am clear in my own mind why I am attending the clinical genetics service.	1	2	3	4	5	6	7
2	I can explain what the condition means to people in my family who may need to know.	1	2	3	4	5	6	7
3	I understand the impact of the condition on my child(ren)/any child I may have.	1	2	3	4	5	6	7
4	When I think about the condition in my family, I get upset.	1	2	3	4	5	6	7
5	I don't know where to go to get the medical help I / my family need(s).	1	2	3	4	5	6	7
6	I can see that good things have come from having this condition in my family.	1	2	3	4	5	6	7
7	I can control how this condition affects my family.	1	2	3	4	5	6	7
8	I feel positive about the future.	1	2	3	4	5	6	7
9	I am able to cope with having this condition in my family.	1	2	3	4	5	6	7
10	I don't know what could be gained from each of the options available to me.	1	2	3	4	5	6	7
11	Having this condition in my family makes me feel anxious.	1	2	3	4	5	6	7
12	I don't know if this condition could affect my other relatives (brothers, sisters, aunts, uncles, cousins).	1	2	3	4	5	6	7
13	In relation to the condition in my family, nothing I decide will change the future for my children / any children I might have.	1	2	3	4	5	6	7
14	I understand the reasons why my doctor referred me to the clinical genetics service.	1	2	3	4	5	6	7
15	I know how to get the non-medical help I / my family needs (e.g. educational, financial, social support).	1	2	3	4	5	6	7
16	I can explain what the condition means to people outside my family who may need to know (e.g. teachers, social workers).	1	2	3	4	5	6	7
17	I don't know what I can do to change how this condition affects me / my children.	1	2	3	4	5	6	7
18	I don't know who else in my family might be at risk for this condition.	1	2	3	4	5	6	7
19	I am hopeful that my children can look forward to a rewarding family life.	1	2	3	4	5	6	7
20	I am able to make plans for the future.	1	2	3	4	5	6	7
21	I feel guilty because I (might have) passed this condition on to my children.	1	2	3	4	5	6	7
22	I am powerless to do anything about this condition in my family.	1	2	3	4	5	6	7
23	I understand what concerns brought me to the clinical genetics service.	1	2	3	4	5	6	7
24	I can make decisions about the condition that may change my child(ren)'s future / the future of any child(ren) I may have.	1	2	3	4	5	6	7

Figure 2.1: The Genetic Counselling Outcome Scale

2.7 Conclusion

This literature review has used existing sources to examine validated outcome measures used in the evaluation of CGS. The majority of existing measures were tailored to capture a specific outcome such as patient knowledge or satisfaction, or to be relevant to a specific condition such as breast cancer. Outcome-specific instruments only represent a limited scope of what CGS can offer patients, and condition-specific instruments are limited in their wider application. The extent of psychometric validation was largely very limited, with over half of the identified measures being assessed for internal consistency alone. None of the measures were preference-based - a requirement of NICE for any instrument used in the appraisal of efficacy and cost effectiveness of healthcare interventions.

One of the aims of the literature review was to identify any candidates which may suitable for use as a standard measure in CGS evaluations. Three validated measures emerged which take into account a range of CGS patient outcomes: The Audit Tool for Genetics Services (Skirton *et al.,* 2005), GCOS-24 (McAllister *et al.,* 2011b), and the PPC questionnaire (Berkenstadt *et al.,* 1999). Due to item wording The Audit Tool is unsuitable for pre/post intervention analysis, but the GCOS-24 and PPC are both well validated and have a high degree of clinical utility. GCOS-24 stands out as the stronger candidate since it captures empowerment, a concept which encompasses all three dimensions of PPC (Cognitive Control; Decisional Control; Behavioural Control), as well as two further dimensions (Emotional Regulation and Hope). GCOS-24 is grounded in extensive qualitative data and, despite being developed relatively recently, has gained international recognition and has been translated into multiple languages.

If GCOS-24 is to meet NICE requirements for use in cost-effectiveness evaluations of CGS, it must have preference-weights attached, reflecting the value that individuals attach to each GCOS-24 item and response option. However, at its present length of 24 items each with 7 response options, it is impossible to design a study to elicit preference weights. The aim of this study was therefore to develop a valid and reliable short form of the GCOS-24, five or six items in length. The short-form should be applicable both within and outside the context of CGS.

3. Methods

The purpose of this chapter is to describe the methods used to answer the research objectives, including why the specific methods were chosen and how they were used. In some situations there were multiple potential approaches from which to choose, justifications as to why the chosen methods were most appropriate will be clarified.

Research aim: to develop a valid and reliable short form of the GCOS-24, amenable to future development by the addition of preference weights.

3.1 Study Design Overview

There were four phases to this study. Phase I: Cognitive interviews (Ericsson & Simon, 1980) were used to explore interpretability of GCOS-24 items and which GCOS-24 items were most valued by the target population. Phase II: Quantitative analysis of an existing data set of GCOS-24 responses (n = 395), using Classical Test Theory (CTT) methods to identify underlying traits, and Item Response Theory (IRT) methods to examine item discrimination. Phase III: Item Selection. The results from Phases I & II were used to inform the selection of a set of five or six GCOS-24 items. The Rasch rating scale model (Andrich, 1978) was also used to explore functional problems with the seven-point Likert Scale. In Phase IV the reliability and discriminative ability of the new instrument was tested through a test-retest study. The overall study design is presented as a flow chart in Figure 3.1.

3.2 Phase I: Qualitative Research Methods

The overall aim of this research study was to develop a valid and reliable short-form of GCOS-24. One of the specific aims was to capture outcomes which are relevant to, and valued by, those affected by a genetic condition within the family. Items which represent highly valued outcomes, for example, could be considered for selection over those which are less valued. A second aim was to explore the meaning and wording of GCOS-24 items, again using the perspective, attitudes and opinions of the target population. Items showing as hard to interpret may benefit from rewording.

In order to obtain this information, qualitative research methods were most appropriate. Qualitative methods produce rich, detailed datasets, providing effective ways to analyse the intricacies and variability of human emotion and beliefs (Fink, 2016). They can be used to provide information directly from the individual's perspective, making it possible to examine the relevant issues in a manner which quantitative analysis cannot offer (Beeson, 1997). In other words, if the purpose of the research is to understand the perceptions of participants, their experiences and interpretations, without destroying the complexity and context of the data, qualitative methods are most appropriate (Atieno, 2009).

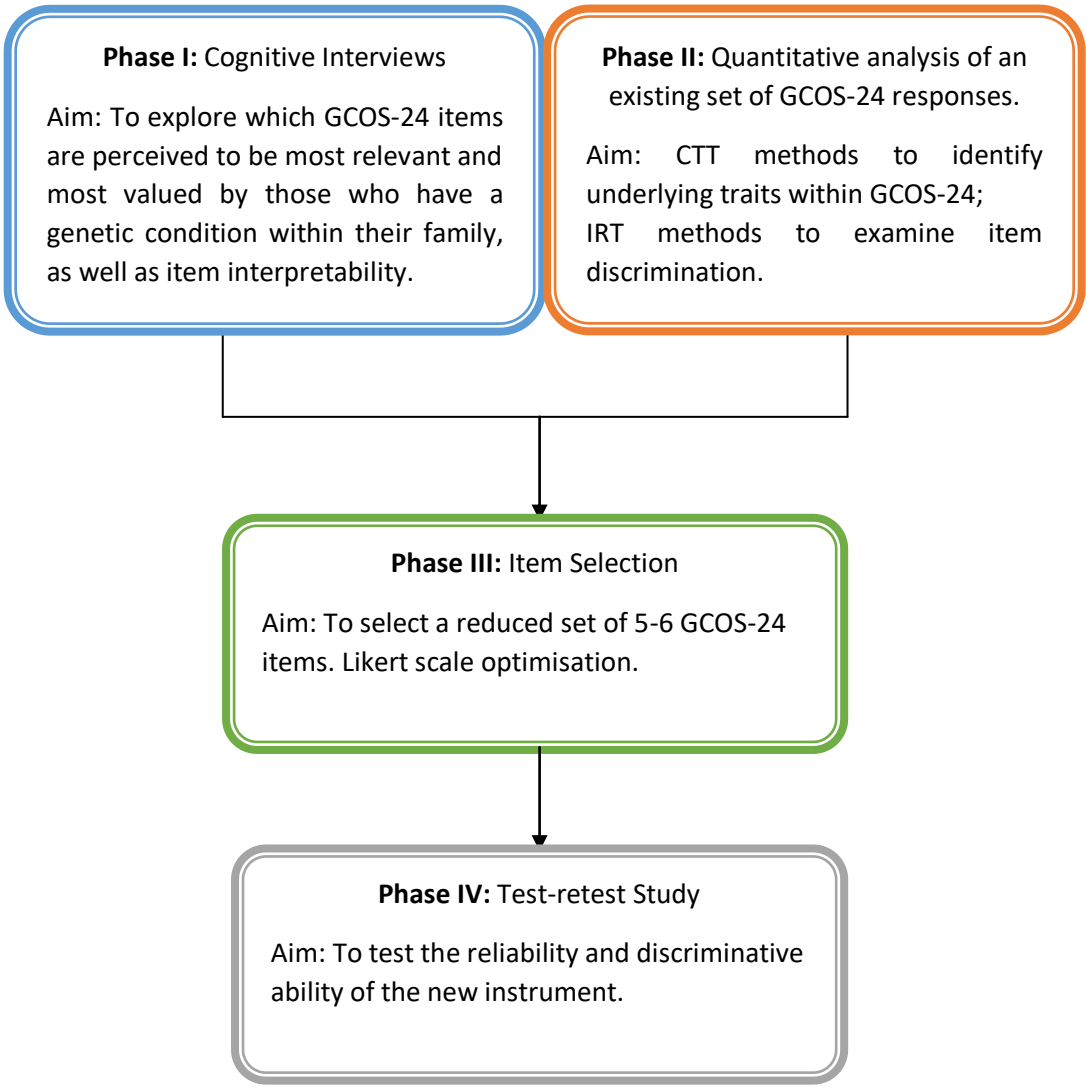


Figure 3.1: Flowchart of Overall Study Design

3.2.1 Cognitive Interviews

The cognitive interview (also called 'think-aloud' interview) derives from the psychological procedures described by Ericsson and Simon (1980), and involves subjects being explicitly instructed to 'think aloud' as they answer the questions. The interviewer interjects infrequently where possible, and encourages interviewees to explain their thoughts and to expand on their answers. The great advantage of cognitive interviewing over other qualitative methods for the purpose of this study is conferred by the think-aloud premise. Olson *et al.* (1984) stated that using the think-aloud technique is one of the most effective ways to assess higher-level cognitive processes (i.e. those which involve thought or memory), and that it was a valuable method for studying individual perspectives. Ericsson and Simon (1980) conclude that the data produced from think-aloud methods are 'thoroughly reliable' as a source of information about thought processes (p. 247). More recently, cognitive interviewing has emerged as one of the more prominent methods for analysing survey questions, with numerous academic, government and commercial research centres incorporating cognitive interviews into their usual procedures for questionnaire development (Beatty & Willis, 2007).

If carried out with a single interviewee, cognitive interviews give each individual an opportunity to speak in detail and in turn allow for more data to be collected from each participant than focus groups (Gill *et al.,* 2008). With respect to this study, some of the topics could be perceived to be sensitive or personal, and individual interviews allow these to be explored in private without the pressure of a group. It was also expected that participants may have different perspectives depending on the specific genetic condition in their family, and whether they are affected by, at risk for, or unaffected by said condition. Individual interviews give each participant a chance to speak freely on each question and provide their honest opinion. In short, cognitive interviews carried out on an individual basis were chosen as the most appropriate qualitative method for this study.

Before collecting data using cognitive interviews, it is important to decide on the interview structure, for example the appropriate degree of prompting (Charters, 2003). A non-directive, semi structured method was chosen because it gives interviewees the opportunity to speak freely and expand on their answers whilst still ensuring that the researcher has some control of the interview content. 'Non-directive' is a term

denoting a technique in which the interviewer refrains from asking leading questions, or from directing the interviewee in their responses (Rogers, 1945). 'Semi-structured' is an interview style which allows a degree of openness and flexibility in the line of questioning (Longhurst, 2003). While a structured interview involves a predetermined set of questions from which one is not allowed to divert, a semi structured interview is more conversational, allowing participants to raise and explore new ideas. The questions and overall structure will likely be predetermined, but modifications can be made by the interviewer depending upon what seems most appropriate. This can be beneficial for data collection as it allows for a comprehensive commentary from the perspective of the participant, and novel and unexpected points may arise (Barriball & While, 1994).

Face-to-face interviews, characterised by synchronous communication in time and place (Opdenakker, 2006), were chosen as preferable, but not a necessity. On the one hand, they allow for social cues such as body language to enrich the data. The interviewer and interviewee can directly react to what the other says or does, and this can help to create a good ambience and cultivate a good relationship between both parties (Opdenakker, 2006). With that said, telephone or video interviews also have certain benefits. They extend access to participants who would otherwise be hard to reach, for example mothers at home with small children, or people with disabilities who cannot travel. They can also be easier and cheaper to arrange and perform, with neither party having to travel. Whilst the ability of the interviewer to pick up on social cues may be reduced, telephone interviews can allow people to relax and feel able to disclose sensitive information (van Teijlingen, 2014).

Regarding sample size, cognitive interview guidelines (Malterud *et al.,* 2015) suggest that 10-20 participants should be sufficient to achieve data saturation (the stage at which the researcher can see no new themes emerging from the data). When looking to studies of a similar nature, a recent study by Diness *et al.* (2017) carried out 18 cognitive interviews with genetic counselling patients as part of a study to translate and adapt GCOS-24 for use in Denmark. This lies in accordance with Guest *et al.* (2006), who reported that 'data saturation often occurs following about 12 interviews with members of homogeneous groups' (p.74). Failure to reach data saturation will likely have a negative impact on the quality of research as well as content validity (Bowen, 2008; Kerr *et al.,* 2010), so the aim of this study was to recruit a minimum of 10 participants.

All interviews were audio-recorded with permission of the interviewee. Recording allows for an accurate and detailed method of data collection (Opdenakker, 2006). Coupled with this, notes were taken during interview, for example to record non-verbal cues and to keep track of the topics covered.

In summary, Phase I of this study aimed to explore which GCOS-24 items are perceived to be most relevant and most valued by those who have a genetic condition in their family. Cognitive interviews present the most appropriate means of satisfying this objective, and the aim was to recruit 10 – 15 participants. Non-directive, semi-structured cognitive interviews were conducted on an individual basis to provide information on GCOS-24 item valuation and interpretability from the perspective of the target population. The interview guide (Appendix D) was adapted from the guide used by Irwin *et al.* (2009), intending to last around 45 minutes. All interviews would be audio-recorded with permission of the interviewee, and no reward or financial compensation was provided. Letters of thanks were emailed to all interview participants (Appendix E).

3.2.2 Cognitive Interview Recruitment

A study sample for cognitive interviews was identified using a sampling frame provided by Genetic Alliance UK (GAUK), a national charity comprising over 180 support groups for genetic conditions, aiming to provide information and support to families and individuals with genetic conditions, as well as influencing the services needed by these people. The sample was an adaptation of the GAUK 'Rare Disease Patient Network' (a collection of patients, families, health care professionals and researchers in the South-Wales region who are interested in genetic diseases) with only patients and families included.

To ensure recipient anonymity, as required by the Data Protection Act 1998, recruitment materials including information about the project (Appendix A) were dispersed by Steven Blunt, the Public Engagement and Policy Officer for GAUK. An email recruitment method was used in an attempt to maximise responses, and to save costs seeing as the project was unfunded. If an expression of interest was received, contact was then made by Peter Grant to arrange an interview. Informed consent was

confirmed immediately prior to interview through a consent form (Appendix B). For telephone interviews, this process was done in advance by post.

Cognitive interview inclusion criteria were that participants:

- are at risk of, or affected by, a genetic condition within the family;
- are over 18 years old;
- have expressed an interest in participating in research.

Participants were excluded if they failed to meet these criteria, and also if they were unable to speak or read English. Ethics approval for the recruitment of human participants for cognitive interviews was granted by Cardiff University School of Medicine, 12th May 2017 (Appendix H).

3.2.3 Qualitative Data Analysis

With cognitive interviews in place in the study design, an appropriate method of data analysis had to be selected. Table 3.1 lists the common methodological approaches to qualitative analysis with brief descriptions.

Table 3.1 Common methodologies for qualitative analysis. Definitions adapted from Dey (1993)

Method	Description
Discourse Analysis	The study of meanings or ideas around a topic, and how these are established, used, and changed. Detailed analysis of discourses.
Ethnography	Observational study of people in their natural environment.
Framework Analysis	Mostly deductive. A theoretical framework provides structure to data analysis. Patterns are identified, reported and analysed.
Grounded Theory	Entirely inductive, no preconceived idea. Theory developed from data.
Interpretive Phenomenological Analysis	How individuals make meaning of their life and experiences.
Thematic Analysis	A method for identifying, analysing and reporting patterns within data. A descriptive approach; can be either inductive or deductive.

Three methods from Table 3.1 could be used to analyse the cognitive interview data produced in this study: Grounded theory, Thematic analysis, and Framework analysis. Grounded theory (Glaser & Strauss, 1967) is a popular approach for exploring new areas, as it focuses on developing a theory purely from the data collected. The researcher should not be influenced by any preconceived ideas, and does not specify a theory beforehand. Thematic analysis (Braun & Clarke, 2006) follows a somewhat similar methodology, albeit less interpretative, involving the identification of themes within the qualitative data. Both approaches were considered, but were judged to be rather too inductive considering that the interview data was expected to be structured by GCOS-24 items and the underlying construct of empowerment; novel themes were not expected to arise.

Framework analysis (Ritchie & Spencer, 1994) is a superior alternative to grounded theory and thematic analysis if the research has specific questions or issues, and if the research is primarily based on the observation and accounts of the participants (Srivastava & Thomson; 2009). It is a method for analysing and reporting patterns (themes) within qualitative data, and is becoming an increasingly popular approach in medical and health research (Gale *et al.,* 2013). Its defining feature is that the researcher analyses data with a theoretical structure already in place to provide guidance. In-depth analyses of key themes can still take place, but the data provided by each research participant remains connected to the theoretical framework so that the context is not lost. It is most commonly applied for the analysis of semi-structured interview transcripts, allowing for easy comparisons and contrasts to be made across different participants (Gale *et al.,* 2013).

Framework analysis was selected as the most appropriate method. Empowerment was chosen as the framework since GCOS-24 was specifically designed to capture empowerment, and also to help ensure that the shortened questionnaire captures a range of CGS outcome domains.

There are five steps to Framework Analysis.

1. First, the researcher must become familiarised with the qualitative data. Transcription from an audio recording will usually satisfy this step (Srivastava & Thomson, 2009).
2. Secondly, a theoretical framework must be identified and applied (empowerment). Although data will likely reflect the *a priori* issues, an open mind must be maintained and data should not be forced to fit into preconceived notions.
3. Third, data is 'indexed', which means identifying themes within the data. Ritchie & Spencer (1994) recommend that a numerical system (coding) be used.
4. Charting, the forth step, involves a more detailed examination of indexed data. Sub-themes are labelled, and data may be placed in charts or tables headed by the thematic framework.
5. The final stage is termed mapping and interpretation. This involves the holistic analysis of the themes and subthemes. The researcher is cognisant of the objectives of Framework Analysis: "to define concepts, to map the nature of phenomena, to create typologies, to find associations, to provide explanations, and to develop strategies" (Ritchie & Spencer, 1994, p186).

In summary, Framework Analysis was selected as being the most appropriate method of qualitative analysis. The defining feature of this method is that a theoretical framework is used during analysis, providing structure and enabling comparisons between participants. In this study, empowerment was the natural choice of framework, since GCOS-24 was specifically developed to capture it.

3.3 Phase II: Quantitative Research Methods

3.3.1 Parallel Analysis, Maximum-Likelihood & Rotation

One of the aims of this project was to produce a measure which captures the breadth of the underlying construct, empowerment (Table 2.4) (McAllister *et al.*, 2011a). Rather than using subjective judgement to assess this aim, Factor analysis (FA) was chosen as an appropriate quantitative technique. FA, first introduced by Thurstone (1931), is a generic term given to a class of statistical methods which aim to identify correlations between variables. Observed correlations are then used to group variables, with the concept being that correlations may be explained by latent traits. In other words, FA determines whether the data produced by the variables is a result of just a few underlying factors (Beukelman & Brunner, 2016).

One of the main applications of FA is in the process of scale reduction (Costello & Osborne, 2005). Variables (questionnaire items) will 'load' onto the underlying factors differentially depending upon the observed correlations, representing the relationship of each variable to the underlying factor. Retaining variables with higher loading values will ensure that the underlying traits are being captured as best as possible.

The alternative quantitative method of identifying underlying traits is called principal components analysis (PCA). PCA has long been a popular alternative to FA, due to it being quicker and less computationally intensive, and because it was the default option for early software programs (Gorsuch, 1990). Nowadays, however, with modern computing power, these benefits are insignificant, and many researchers argue in favour of FA (Bentler & Kano, 1990; Floyd & Widaman, 1995; Ford *et al.,* 1986; Gorsuch, 1990; Loehlin, 1990; MacCallum & Tucker, 1991; Mulaik, 1990; Snook & Gorsuch, 1989; Widaman, 1990; 1993). A major flaw of PCA is that it is does not discriminate between shared variance (present amongst all variables) and unique variance (particular to each variable) (Ford *et al.,* 1986). It therefore has a tendency to produce inflated values of variance for each item (Ford *et al.,* 1986; Gorsuch, 1997; McArdle, 1990).

When applied to this context, FA methods represent potentially valuable tools to assist in the scale reduction of GCOS-24. Indeed previous results have suggested that the

items within GCOS-24 can be divided into factors, although FA results have not been consistent (McAllister *et al.,* 2011b; Costal-Tirado *et al.,* 2017).

The first step was to select an FA method for identifying the optimal number of factors present within the data. Traditionally, default choices have been the eigenvalues-greater-than-one rule (Kaiser, 1960), or the scree plot (Cattell, 1966). These, however, present problems. The eigenvalues-greater-than-one rule typically over overestimates, and sometimes underestimates, the number of factors (Cliff, 1988; Zwick & Velicer, 1986), and there is a broad consensus in the current literature that this is one of the least accurate methods for selecting the number of factors to retain (Velicer & Jackson, 1990). The scree test involves an eye-ball search of a plot, and as such is liable to poor accuracy and reliability (Crawford & Koopman, 1979; Streiner, 1998). Parallel Analysis (Horn, 1965) has emerged as a superior method of finding the optimal number of factors (Dinno, 2009; Lance *et al.,* 2006; O'Connor *et al.,* 2000; Velicer *et al.,* 2000; Wood *et al.,* 1996; Zwick & Velicer, 1982, 1986). Although once computationally intensive (Costello & Osborne, 2005), Parallel Analysis can now be carried out quickly using modern computers on common statistical software such as SPSS and SAS (O'Connor, 2000). The concept of Parallel Analysis is to identify the number of factors which account for more variance than can be explained by random chance.

Although Parallel Analysis can be used to identify the number of underlying factors within a set of variables, it cannot be used to assign variables to the factors and produce factors loadings. The next decision, therefore, was to select a method for this purpose. Available methods include alpha factoring, generalised least squares, image factoring, maximum likelihood, and principal axis factoring. Articles by Fabrigar *et al.* (1999), Costello & Osborne (2005), Field (2013) and Sullivan *et al.,* (2005) argue that if data are normally distributed, maximum likelihood is the best choice. For one, it is the only method which does not treat the sample as the entire population, instead assuming that participants are randomly selected. This allows for inferences to be made about the larger population from the sample (Felsenstein, 1981). Additionally, maximum likelihood shows lower variation and better reliability than other methods as the calculations are least affected by error (Felsenstein, 1981; Sullivan *et al.,* 2005).

The final decision with respect to FA methods was the rotation method. 'Rotation' in this context is a process which helps to align the observed correlations with the actual data points, making the factors more clearly defined and interpretable. For variables

which are theoretically expected to correlate (as with GCOS-24), oblique rotation is most appropriate. There is no widely preferred specific method of oblique rotation, all tend to produce similar results (Costello & Osborne, 2005; Fabrigar *et al.,* 1999).

In summary, one of the aims of this study was to maintain the ability of the reduced scale to capture the breadth of the GCOS-24 underlying construct, empowerment. FA methods were chosen as being appropriate for this purpose, specifically Parallel Analysis, Maximum Likelihood, and oblique rotation. These methods examine item correlations to identify any underlying traits within the instrument, and show which items correspond to those underlying traits. Items with stronger correlations ('higher loadings') better represent the underlying trait, and this information can be used to prioritise items for selection.

3.3.2 Item Response Theory

A further aim of this study was to examine the discriminative ability of GCOS-24 items. In other words, if an item states 'I feel positive about the future', as with item 8 in GCOS-24, it should cause people who do not feel positive about the future to answer differently compared with those who do feel positive about the future. Items which are unable to discriminate between individuals of different trait levels would make poor candidates for selection in the reduced scale. Although the development, validation, and refinement of outcome measures have traditionally been guided by a set of quantitative approaches known as CTT (Gulliksen, 1950; Hambleton, 2000; Nguyen, 2014; Wiberg, 2004), the issues of CTT and the advantages of IRT as outlined on page 38 led to IRT methods emerging as the preferred choice.

There are a number of models within the IRT family, all designed to fit a certain purpose. Table 3.2 lists the available methods for measurement scales with polytomous item response formats (more than two options, as with GCOS-24), along with a summary of their appropriate use.

Table 3.2 Polytomous Item Response Theory (IRT) Models. Adapted from Nguyen *et al.* (2014)

IRT Model	Model Characteristics
Bock's Nominal Model	Used for unordered responses. Discrimination allowed to vary across items.
Generalised PCM	Used for ordered responses. Discrimination varies across items.
Graded Response Model	Used for ordered responses. Discrimination varies across items.
Partial Credit Model (PCM)	Equal discrimination across all items. Separate category location parameters estimated for each item.
Rasch Rating Scale Model	Equal discrimination across all items. A single set of categorical location parameters estimated for all items.

Bock's Nominal Model (Bock, 1972) operates on unordered response options which are in the form of nominal categories. The 7-point Likert scale of GCOS-24, ranging from strongly disagree to strongly agree, is an ordered rating scale response format, therefore Bock's cannot be applied. The Partial Credit Model (PCM) (Masters, 1982) and Rasch Rating Scale Model (Andrich, 1978) assume equal discrimination across all items, so cannot be used to test item discrimination. The two models which allow for separate discrimination parameters are the Graded response model (GRM) (Samejima, 1969) and the Generalised PCM (Muraki, 1992). Both methods are very similar and will generally agree very closely (Nguyen, 2014). The slight difference is that the Generalised PCM uses 'local estimation' during calculation (i.e. not all data are incorporated when estimating boundary parameters), which means there is no guarantee that the response categories will be ordered in the output (Muraki, 1992). The GRM, on the other hand, forces the response categories to be ordered (Samejima, 1969), which is more appropriate for analysing the strictly ordered seven-point Likert scale in GCOS-24. Therefore, the GRM was selected for assessing item discrimination.

3.3.3 Application of Quantitative Methods

Phase II used an existing dataset, comprising a set of responses to GCOS-24 (n = 395), collected in 2010 for the original psychometric validation (McAllister, 2011b). Specific details (e.g. gender, ethnicity, condition type, reason for referral) can be found in McAllister *et al.* (2011b). FA methods were conducted using IBM SPSS Statistics 23.0 (IBM Corp., 2015); the GRM used R statistics 3.5.0 and the package ltm (Rizopoulos, 2006); and Rasch Analysis used the Winsteps Rasch Measurement software version

4.3.2 (Linacre, 2017). Ethics approval for the secondary use of GCOS-24 responses was granted by the National Research Ethics Service (NRES) Committee North West.

In summary, one of the aims of Phase II was to assess the discriminative ability of GCOS-24 items. Items with better discriminative properties would then be prioritised for inclusion in the reduced scale. IRT methods provide a means of accomplishing this aim, as they are able to examine instruments at the item-level. The GRM was selected as most appropriate IRT method.

3.4 Phase III: Item Selection

Three principles guided the approach to item selection. (i) Items with an unjustifiably low discrimination parameter (>1.34) were not selected; (ii) Items with factor loadings <0.55 were not selected; (iii) To avoid redundancy, items capturing a similar outcome were not selected together; FA, GRM and cognitive interview findings were used to establish superior items.

The Likert scale within GCOS-24 was also examined with a view to reduction. The GRM naturally provides this information as part of the output, but only in the form of a graph. An eyeball assessment must then be made. The Rasch Rating Scale Model (Andrich, 1978) offers a useful supplementary method, providing numerical information on rating scale statistics from which purely objective conclusions can be drawn.

3.5 Phase IV: Validity and Reliability Testing

Internal consistency, test-retest reliability and validity are essential properties for any measurement scale (Aaronson *et al.*, 2002; Mokkink *et al.,* 2010). Content validity, a subjective assessment of whether the instrument measures the appropriate content and represents the variety of attributes that make up the measured construct (Frost *et al.,* 2007), was assured by the qualitative research underpinning GCOS-24 (McAllister *et al.,* 2008; 2011a).

Internal consistency, described on page 21, has become the primary method of estimating the reliability of multi-item scales. Indeed Streiner (2003) stated that, "Internal consistency is necessary in scales that measure various aspects of personality" (p.103).

Test-retest reliability is a different form of reliability, in which the test is administered at two time points. The scores from each time point are then correlated, estimating the extent to which scores are stable over time. Test-retest reliability is a valuable tool in scale development, as a scale should theoretically produce the same results if administered to the same group of people (McCrae *et al.,* 2011; Schmidt *et al.,* 2003). Choosing an appropriate time interval for a test-retest study is important. It should not be so soon that responses at the second assessment are influenced by memories of the first assessment, yet not so long that a change in the measured construct has occurred amongst respondents during the time interval. A time interval of two weeks is often considered appropriate for the evaluation of PRO instruments (Streiner & Norman, 2015). The Consensus-based Standards for the selection of health measurement instruments (COSMIN) guidelines suggest a minimum sample size of 50 for reliability studies (Terwee *et al.,* 2007).

In order to calculate internal consistency and test-retest reliability, an online test-retest study was designed. An online method was used in an attempt to maximise responses, and to save costs seeing as the project was unfunded. Firstly, a version of the new scale was created using SurveyMonkey. The survey was then advertised by GAUK to their membership in their weekly online newsletter. The advertisement contained a brief description of the research study, as well as links to the survey and further project information (Appendix F). When a survey was completed, the respondent was emailed after a period of 14 – 21 days requesting them to complete the survey a second time (Appendix G); a final reminder email was sent if no response was received within a week (Appendix G). Responses were used to calculate internal consistency and test-retest reliability. The GRM was also used to examine item discrimination within the new scale. Ethics approval for the recruitment of participants for the test-retest reliability study was granted by Cardiff University School of Medicine, 12th May 2017 (Appendix H).

In summary, this project aimed to produce a reliable short-form of GCOS-24. A test-retest study was designed with results used to calculate internal consistency, test-retest reliability and item discrimination. The time interval for the test-retest study was 14 – 21 days, aiming for a sample size of 50 amongst those affected by a genetic condition in the family.

3.6 Summary

This chapter has described the methods used in this study to meet the study objective of developing a valid and reliable reduced version of GCOS-24. Justifications as to why the chosen methods were most appropriate were clarified, and their implementation was explained. The final design consisted of four phases: Phase I used cognitive interviews to explore the interpretability of GCOS-24 items and which GCOS-24 items were most valued by the target population; Phase II utilised CTT methods to examine underlying traits within GCOS-24, and IRT methods to examine item discrimination; in Phase III the results from Phases I & II were used to inform the selection of set of 5-6 GCOS-24 items; and in Phase IV the reliability and discriminative ability of the new instrument was tested through a test-retest study. Chapter 4 will present the results of the study.

4. Results

This chapter will present the results obtained over the course of the project. For clarity the structure will follow the four-phase structure as described in the methods chapter. Phase I will present the results from the cognitive interviews, including participant demographics and framework analysis. In Phase II the results from the quantitative analyses will be described, and then in Phase III how both the qualitative and quantitative data were used to inform item selection for the short-form GCOS-24. Finally, in Phase IV, results from the test-retest reliability study will be presented, including an assessment of item discrimination using the Graded Response Model.

4.1 Phase I: Cognitive Interviews

Recruitment was carried out across June and July 2017. Thirty-five individuals were invited to participate in the study, ten of whom replied expressing their interest in participating (response rate 28.6%). Of these, all ten were successfully interviewed.

Think-aloud cognitive interviews were conducted on an individual basis across June, July and August 2017. Five face-to-face interviews took place at the Institute of Medical Genetics at Cardiff University, three face-to-face interviews took place at the participant's place of work or residence, one was conducted by telephone and one was conducted by Skype. Participant characteristics are summarised in Table 4.1. For anonymity, participants are identified with the letter P followed by a number. Proof of diagnosis of genetic condition was not requested, but all participants believed that their condition was genetic.

Qualitative data was analysed using Ritchie & Spencer's framework analysis method. This is described in detail in section 3.2.3, but in short interviews were transcribed and empowerment was applied as the theoretical framework. Data was then indexed to identify themes within the data, and findings were 'charted' which involved a more detailed process of labelling and sorting, bearing in mind the empowerment framework.. The qualitative framework analysis findings are presented below. Item numbering will be referred to in GCOS-24 (Figure 2.1), and a summary of the most highly-valued items is provided in Table 4.2 on page 65.

Table 4.1: Cognitive interview participant characteristics

Participant	Sex	Condition	Affected, At risk, Unaffected	Has a child?	Received Genetic Counselling?
P1	Male	Nystagmus	Affected	No	No
P2	Male	Ataxia	Affected	No	Yes
P3	Female	Tubular Sclerosis	Unaffected	Yes	Yes
P4	Male	Glaucoma	Affected	No	Yes
P5	Female	Thalassemia Intermedia	Affected	Yes	Yes
P6	Female	Episodic Ataxia	Unaffected	Yes	No
P7	Female	Ehlers Danlos Syndrome	Affected	Yes	No
P8	Female	Dystonia & Ataxia	Affected	Yes	No
P9	Female	Huntington's Disease	At risk	Yes	Yes
P10	Male	Leber's Hereditary Optic Neuropathy	Affected	No	No

4.1.1 Cognitive Control

Part of feeling empowered in relation to a genetic condition in the family is having a belief that you have sufficient knowledge and understanding about the condition (Cognitive Control) (McAllister *et al.*, 2008). This could be knowledge about how the condition is inherited, what causes it, what the signs and symptoms are, and what the implications are for the rest of the family, both at present and in the future. All ten participants spoke of their desire to learn more about their condition, both at the time of diagnosis and as an ongoing pursuit, and of the benefits that this knowledge could have on their lives. On an item level, six GCOS-24 items had originally been designed to capture cognitive control: items 1, 3, 12, 14, 18, and 23. Of these, items 18 ('I don't know who else in my family might be at risk for this condition') and 12 ('I don't know if this condition could affect my other relatives (brothers, sisters, aunts, uncles, cousins)') appeared to be valued most highly by participants. Knowing how the condition might affect one's relatives was judged to be very useful information. This participant spoke of item 18:

> "That's a really good question because if you, if this was day one, so you ask someone before their first session, they're probably going to answer that quite

high. If you ask them after 5 sessions, then actually the answer could be completely the opposite, so I think that's a valuable question to ask, because you can show the progress they've made and what they've learnt from the session. I think that's a really good question. And before I had the genetic counselling I would have answered I don't know, and now I can answer I do know, because I had the service and got the information that I needed." (P4)

Another participant, when considering item 12, stated:

> "I think it is an important piece of knowledge to have. If I didn't know that information I would be worried, and I could see how people would get worried about that type of thing" (P5).

Items 12 and 18 emerged as strong candidates for retention (Table 4.2).

4.1.2 Decisional Control

Decisional Control within the empowerment framework is not restricted only to decisions made about healthcare. It can include any major or minor decision which is influenced by having a genetic condition within the family (McAllister *et al.,* 2008). This might involve decisions on marriage, whether or when to have children, or on seemingly unrelated decisions such as buying a car or whether to take on a mortgage. Outcomes relating to Decisional Control were discussed by some participants in this study, for example this interviewee who had a daughter with episodic ataxia:

> ... "To me, reading that [item 24], it's just what I do anyway, I make decisions for her. If I feel she can't do something in the normal way, then I find other routes which enable her to do everything anybody else is doing. To me that is making a decision. So you're always decision making, always, you can never stop decision making for the child." (P6)

The corresponding GCOS-24 items (10, 13 and 24), however, were rarely chosen by participants as being of high value. More specifically, items 13 ('In relation to the condition in my family, nothing I decide will change the future for my children / any children I might have') and 24 ('I can make decisions about the condition that may

change my child(ren)'s future / the future of any child(ren) I may have') suffered because they were not seen as relevant by the 40% of participants who did not have children. Item 10 ('I don't know what could be gained from each of the options available to me') was of unclear meaning to many:

> ... [Interviewer: "Are any items difficult to understand?"] "I suppose number 10: I don't know what could be gained from each of the options available to me. That's a little bit, what options are we talking about?" (P1)

> ... [Interviewer: "What does item 10 mean to you?"] "Umm. Well the first thing that comes to mind after reading that question is, I don't know what options it means. Umm. As far as I'm concerned I had genetic counselling, and now I've just got to see my consultant, take my medication... and that's it. I don't know any options that are available to me at all. So, it doesn't mean a lot to me." (P3)

No items within Decisional Control emerged as strong candidates for selection.

4.1.3 Behavioural Control

Behavioural Control is perhaps the most diverse dimension of empowerment, representing the perception of an individual that they are able to take action to improve their situation. This includes making effective use of the health and social care systems which are available, managing the condition day to day, or communicating about genetic risks with relatives (McAllister *et al.,* 2008). All participants spoke at length about the importance of outcomes corresponding to Behavioural Control. Topics included their experiences with the NHS, the vital importance of both medical and non-medical services following diagnosis with a genetic condition, and how important it is to be able to communicate with others about the condition, whether that be with family, work colleagues, or with a school on behalf of their child. This participant, for example, spoke of her experience with local support groups:

> ... "The [support] groups are a massive help. I forced my sister to join. I do think that having a network of people going through the same thing, it doesn't matter what your situation is, whether it's, you know, cancer, depression, or

anything that's happened, if you've got a group of people going through the same thing you are, it's ultimately just support and it will always help." (P5)

On an item level, almost all items designed to capture Behavioural Control were valued by participants. Especially popular were items 2 ('I can explain what the condition means to people in my family who may need to know') and 16 ('I can explain what the condition means to people outside my family who may need to know'), with all ten expressing the benefits of being able to talk about the condition. This participant contextualised this outcome within social situations:

... "Most people I think are naturally inquisitive. If they can see or know that an individual has a condition, disability, call it what you want, and if you're able to talk comfortably about it, and other people around you can talk comfortably about it, and answer what may sometimes seem ignorant or silly questions, and you're happy to take those questions, then that's in the best interest of everybody. Whether it is family, friends, or work colleagues, whatever, it makes life easier for everybody." (P1)

These participants spoke of item 16:

... "For example my son is starting comp in September and I've had to put a thing on his medical notes saying that he's got thalassemia intermedia. His school then rang me, asked what that entailed and would he suffer in any aspects, so I told them about it." (P5)

... "I feel very passionate about doing that [being able to explain the condition] and sort of being out there and making sure everyone knows about the condition." (P2)

A comparable problem was observed with respect to items 7 ('I can control how this condition affects my family') and 22 ('I am powerless to do anything about this condition in my family'), due to the contrasting interpretations over the meanings of 'control' and 'powerless':

... "I've got control over how people react to it, over how much people need to know, or how they act around my child. Obviously I can't control how ill he'll

get. [...] I think that's the bit of control I've got, he will be very confident. He's going to be brought up very confident, and very, you know, not embarrassed about anything." (P3)

... "See I don't like, I would take out control and I would put manage. I can't, we have no control. We can only do things to lessen the impact, or try to lessen the impact. [...] So I don't think you can control, ever control it. I think you can try and manage the condition. But control, no." (P7)

... "You've got no control over it [the condition]. Get over yourself." (P2)

... "I don't know about control. [...] It's more empowerment and advocacy of ownership, those are the things. Those are the terms I would be more likely to use over control. I don't use control, or very rarely." (P1)

... "I mean, you can't change genetics can you. I don't know ... are you powerless? I think we're all dealt with a hand of cards and how you deal with the hand you've got is the bit that determines whether you're powerless. You can't suddenly change your genetics, you know if I wanted to have a different colour skin I can't change that genetically can I. Uh powerless, such a strange term. I guess in pure genetic terms I can't change my own genetics so in that instance yeah, I guess I would be powerless." (P4)

... "I am powerless about this condition in my family. That can mean numerous things though can't it. Because like I can't control genetics, but I think I can affect change in my life now I know about it." (P6)

Aside from item 7 and 22, all items in Behavioural Control emerged as candidates for selection (Table 4.2).

4.1.4 Emotional Regulation

Emotional Regulation in the empowerment framework refers to the ability to manage the emotional aspects of a genetic condition, both individually and within the family (McAllister *et al.,* 2011a). The diagnosis of a genetic condition can raise significant emotional challenges, and it is important that these emotions are addressed by any provider of genetic counselling or testing services. In this study, outcomes such as guilt,

anxiety, blame, helplessness and sadness were brought up by participants as well as feeling “lost” (P6), “damaged”(P4), or “broken” (P4).

Items 4, 11, and 21 were designed to capture Emotional Regulation. Of these, item 4 (‘When I think about the condition in my family I get upset’) emerged as the strongest candidate. This interviewee spoke of the item:

> ... “That [item 4] is very, very important. Because it’s just like you’ve walked into a brick wall. All of a sudden you’re going forward and somebody will put this brick wall in front of you, and to me that brick wall is ataxia, and oh you just don’t know how to get through that brick wall. So you need people there to say well actually it’s not such a brick wall there is a doorway just over by here, let’s get around it and go through. So yeah you do need people in place for that especially when they’re first diagnosed.” (P6)

Item 21 (‘I feel guilty because I (might have) passed this condition on to my children’) was irrelevant to those without children. Item 11 (‘Having this condition in my family makes me feel anxious’) was highly valued by some, but overall the findings suggest that anxiety levels do not necessarily reduce in the long term, and instead fluctuate depending on the situation. This would not be a desirable item to have in an instrument which measures patient benefits from CGS.

> ... “Having the condition in my family makes me feel anxious... I don’t think that will ever fully go away. I’m anxious for him [the son] when he starts a family. What if he gets someone pregnant and he hasn’t stopped the hydroxycarbamide? It does happen. I’m anxious for my future grandchildren. [...] Knowing more about it [the condition] has made me less anxious in one respect, but more anxious in other respects. I am very anxious because, even though I know a lot, I don’t feel I know everything.” (P5)

> ... [Interviewer: “Does the anxiety improve over time?”] “The anxiety? ... no. Because the minute she has another episode you sort of take a deep breath and you hold it until she comes out of this episode.” (P6)

Item 4 emerged as the strongest candidate for retention (Table 4.2).

4.1.5 Hope

Positivity, or a positive mind-set, was the chief manifestation of hope in this study. Item 19 ('I am hopeful that my children can look forward to a rewarding family life') is only applicable to those with children, and even within that demographic appeared to be of questionable relevance:

> ... "I don't see why you're asking that as part of genetic counselling [laughs]. I just think that everybody, who would say no to that? [Continues laughing] 'I'm going to have kids and I hope that they have a c**p life!' So I don't, if you're trying to evaluate the results of these questions, I can't see how that would help the service at all; because everybody always wants the best for their children. I can't imagine anybody not answering positively to that." (P4)

Item 6 ('I can see that good things have come from having this condition in my family') was criticised for being irrelevant:

> ... "I can see that good things have come from having this condition? No, no I don't see that. Because we're a close family anyway and whether this condition was there or not there we would still be the same close family, so that to me, that's an irrelevant question. I can see good things that have come? No." (P6)

> ... "Ah... well we've got Cal [the son], ummm, I guess it makes you feel thankful in different ways doesn't it, makes you appreciate little things. But... no I wouldn't think that was very relevant to this sort of thing, personally." (P3)

Item 8 ('I feel positive about the future') and item 20 ('I am able to make plans for the future') emerged as the strongest candidates. Both received some criticism for being vague, but nevertheless were highly valued.

> ... "I mean number 8 is good, I feel positive about the future. Possibly a little bit vague. It doesn't actually specify; somebody might think that their horse is going to come in tomorrow at some race and they're going to win a whole lot of money. They might be positive because of that." (P1)

> ... "I like number 8, it's a nice all encompassing statement. But what if somebody has some other issue or some other hope in their mind and they think, 'oh yeah I'm positive', it doesn't have anything to do with this." (P1)

... “I think it’s a good question to have in there [item 8], but it’s just a case of, I feel positive about the future, it’s like what aspect of the future? I feel positive about the future of my health, or the future of my mental health, or just feeling positive about the future in general. I think it’s just a bit too open ended.” (P5)

... [Item 20] “That’s why we went really. You know, we were thinking about having a family. We needed the information before we put the plan in place. So yeah, that’s quite a valid question.” (P4)

... [Item 20] “How far in the future do they mean? Do they mean a few weeks, or a few years into the future? Our lives are constant planning, everything has to be risk assessed and planned in advance, there is very little that we can do spontaneously. Umm. I don’t know. It’s a difficult one because how far in advance are they asking you to look? Am I looking to plan 5 years, or am I looking to when my kids are adults? It’s difficult, I can’t really answer it.” (P7)

Items 8 and 20 emerged as the strongest candidates for retention.

Table 4.2: Cognitive interview item valuation. Items have been grouped according to which sub-dimension of McAllister’s five-dimensional empowerment framework each was designed to capture.

Dimension of Empowerment	Corresponding GCOS-24 Items	Highly Valued Items*
Cognitive Control	1, 3, 12, 14, 18, 23	12 or 18
Decisional Control	10, 13, 24	None
Behavioural Control	2, 5, 7, 9, 15, 16, 17, 22	2 or 16; 5 or 15; 17; 9
Emotional Regulation	4, 11, 21	4
Hope	6, 8, 19, 20	8 or 20
*Highly valued items with a similar meaning are separated by ‘or’.		

In summary, Table 4.2 presents the items which were most valued by cognitive interview participants. Empowerment was sufficient to integrate all themes which arose, however no item capturing Decisional Control was highly valued. Considering the diversity of outcomes within Behavioural Control, and the high value given to them by participants, it was observed that multiple items may merit inclusion if they capture different aspects of the dimension.

One noticeable trend throughout the interview process was the confusion experienced by participants when answering items beginning with 'I don't know'. Of the ten participants, seven selected a response option contrary to what they meant when asked about such an item. One individual recognised this when asked about item 18 ('I don't know who else in my family might be at risk for this condition'):

> ... "Erm... so I would disagree to that [item 18]. Because it's like a negative isn't it. Umm if I've read it correct. So it says I don't know, but actually I'm saying I do know, so I would have to disagree with that statement. So that might be slightly confusing to someone. You may get a couple of false positives, if someone misunderstands the question. It is common for these, I've done it myself when I've had to write these kind of evaluations, as soon as you put in the word 'I do not' or 'I don't', you know those kind of things, you sometimes get people who misunderstood the question. So I would I would say strongly disagree or disagree. Because after the service I did know [who else in my family might be at risk for this condition]." (P4)

The result of this finding was an agreement within the research team that any items containing 'I don't know' would be reworded to 'I know' if selected for inclusion in the reduced scale.

4.2 Phase II: Quantitative Analysis

4.2.1 Parallel Analysis & Maximum Likelihood

Table 4.3 shows the results of Parallel Analysis. The first five raw data eigenvalues all exceed the eigenvalues produced for random data at the 95th, the statistically significant, percentile. The sixth raw data eigenvalue however (1.138) does not exceed that produced for random data. This shows that the variance in the raw data is greater than can be explained by random variation up until the 5th, but not the 6th eigenvalue, suggesting an optimal five-factor structure.

Maximum Likelihood was used to determine how the five-factor structure recommended by Parallel Analysis fits to GCOS-24. The pattern matrix is presented in

Table 4.4, results ordered and loadings <.3 excluded. For clarity, Likert scale responses to negatively worded questions were reversed and are labelled with the suffix 'P'. In order to capture empowerment in the new scale, it may be beneficial to select items with high loading values.

Table 4.3 (left): Parallel analysis. The Root column lists the number of factors, with a maximum of 24, one for each variable. The Raw Data column lists the eigenvalues produced by the raw data set. The Percentile column lists the eigenvalues produced by the parallel analysis method for random data at the 95th percentile (statistically significant).

Table 4.4 (right): Maximum Likelihood. Factor loadings have been ordered and loadings < .3 have been excluded. The letter P indicates that the Likert scale responses have been reversed for these items.

Root	Raw Data	Percentile
1	4.706	1.550
2	3.090	1.452
3	1.970	1.384
4	1.451	1.332
5	1.292	1.284
6	1.138	1.241
7	.964	1.200
8	.929	1.163
9	.865	1.128
10	.781	1.091
11	.717	1.061
12	.693	1.025
13	.633	.992
14	.605	.961
15	.561	.933
16	.525	.901
17	.493	.871
18	.478	.842
19	.442	.812
20	.411	.784
21	.371	.753
22	.324	.721
23	.289	.687
24	.271	.648

	Factor				
	1	2	3	4	5
GCOS 8	.753				
GCOS 9	.691				
GCOS 4P	.617				
GCOS 11P	.606				
GCOS 20	.567				
GCOS 19	.400				
GCOS 21P	.385			.375	
GCOS 2		.799			
GCOS 3		.764			
GCOS 16		.364			
GCOS 14			.830		
GCOS 23			.778		
GCOS 1			.517		
GCOS 17P				.641	
GCOS 18P				.628	
GCOS 22P				.603	
GCOS 12P				.515	
GCOS 10P				.509	
GCOS 13P				.494	
GCOS 5P				.411	
GCOS 7					.659
GCOS 6					.490
GCOS 15					.372
GCOS 24					.357

4.2.2 The Graded Response Model

Tables 4.5 - 4.9 present the numeric GRM results for GCOS-24 items, grouped by empowerment dimension. The extremity parameters (Extrmt n) show the latent trait score at which people have a 50/50 chance of selecting certain responses. The discrimination parameter (Dscrmn) represents the slope of the curve at the point where the probability of endorsing an item is 50% (also referred to as item difficulty), and describes how well an item can differentiate between individuals of varying ability.

Table 4.5 – GRM results for GCOS-24 items within Cognitive Control

	Extrmt 1	Extrmt 2	Extrmt 3	Extrmt 4	Extrmt 5	Extrmt 6	Dscrmn
GCOS1	-2.87	-2.32	-1.91	-1.54	-1.16	0.15	2.48
GCOS3	-4.03	-2.79	-2.19	-1.00	-0.32	1.53	0.94
GCOS12	-45.64	-11.55	-0.72	14.79	20.70	43.39	0.04
GCOS14	-2.46	-2.28	-2.21	-1.69	-1.00	0.29	3.51
GCOS18	4904.26	17.62	-1307.90	-3508.13	-4319.39	-9424.74	0.00
GCOS23	-3.54	-2.63	-2.39	-1.66	-1.03	0.50	2.04

Table 4.6 – GRM results for GCOS-24 items within Behavioural Control

	Extrmt 1	Extrmt 2	Extrmt 3	Extrmt 4	Extrmt 5	Extrmt 6	Dscrmn
GCOS2	-3.09	-2.29	-1.89	-1.20	-0.59	1.01	1.39
GCOS5	-4.03	-2.54	-1.75	-0.55	0.07	1.94	0.73
GCOS7	-2.16	-0.33	0.18	2.67	3.74	6.84	0.59
GCOS9	-3.95	-2.30	-1.55	-0.25	0.56	2.76	1.05
GCOS15	-2.33	-0.99	-0.65	0.58	0.95	2.31	1.30
GCOS16	-2.65	-1.85	-1.44	-0.29	-.18	1.41	1.79
GCOS17	-3.07	-0.81	0.06	1.80	2.36	4.33	0.84
GCOS22	-2.47	-0.28	0.66	3.15	4.26	6.98	0.47

Table 4.7 – GRM results for GCS-24 items within Decisional Control

	Extrmt 1	Extrmt 2	Extrmt 3	Extrmt 4	Extrmt 5	Extrmt 6	Dscrmn
GCOS10	5.01	2.93	1.96	-2.83	-3.33	-5.54	0.63
GCOS13	4.67	1.43	0.78	-2.49	-3.69	-7.35	0.46
GCOS24	3.75	2.43	1.92	-0.49	-1.09	-3.20	0.87

Table 4.8 - GRM results for GCS-24 items within Emotional Regulation

	Extrmt 1	Extrmt 2	Extrmt 3	Extrmt 4	Extrmt 5	Extrmt 6	Dscrmn
GCOS4	-0.98	-0.24	0.28	0.94	1.09	1.94	3.65
GCOS11	-1.23	-0.22	0.47	1.18	1.39	2.02	2.47
GCOS21	-2.27	-1.10	-0.38	1.57	1.73	2.67	0.91

Table 4.9 – GRM results for GCOS-24 items within Hope

	Extrmt 1	Extrmt 2	Extrmt 3	Extrmt 4	Extrmt 5	Extrmt 6	Dscrmn
GCOS6	-2.17	2.49	3.30	10.16	12.20	17.60	0.20
GCOS8	-2.22	-1.48	-0.97	-0.29	0.37	1.67	1.69
GCOS19	-4.97	-4.47	-3.76	-1.51	-1.08	0.52	0.95
GCOS20	-1.96	-1.52	-1.30	-0.75	-0.33	0.65	4.32

Using the guidelines provided by Baker (2001) for interpreting item discrimination parameter values, verbal labels can be applied. Table 4.10 states the thresholds for each verbal label, and table 4.11 lists all GCOS-24 items in rank order by discrimination parameter and the associated verbal label.

Table 4.10 – Verbal labels for item discrimination parameters.

Verbal Label	None	Very Low	Low	Moderate	High	Very High	Perfect
Range	0	.01 - .34	.35 - .64	.65 - 1.34	1.35 - 1.69	>1.70	Infinity

Table 4.11 – GCOS-24 items ranked by discrimination parameter and verbal label.

Item	Dscrmn	Label	Item	Dscrmn	Label
20	4.32	Very High	3	0.94	Moderate
4	3.65	Very High	21	0.91	Moderate
14	3.51	Very High	24	0.87	Moderate
1	2.48	Very High	17	0.84	Moderate
11	2.47	Very High	5	0.73	Moderate
23	2.04	Very High	10	0.63	Low
16	1.79	Very High	7	0.59	Low
8	1.69	High	22	0.47	Low
2	1.39	High	13	0.46	Low
15	1.30	Moderate	6	0.20	Very Low
9	1.05	Moderate	12	0.04	Very Low
19	0.95	Moderate	18	0.00	Very Low

The GRM item characteristic curves (Appendix I) provide an illustration of the numeric results. Figure 4.1 presents the GRM output for item 15 as an example of an item with moderate to high discriminative ability (1.30). Clear peaks can be seen ordered from 'Strongly Disagree' at low levels of the latent trait to 'Strongly Agree' at high levels of the latent trait. Each curve, however, is not especially distinct, largely overlapping with its neighbour. Figure 4.2 presents the GRM output for item 22, an item of low discrimination (0.47).

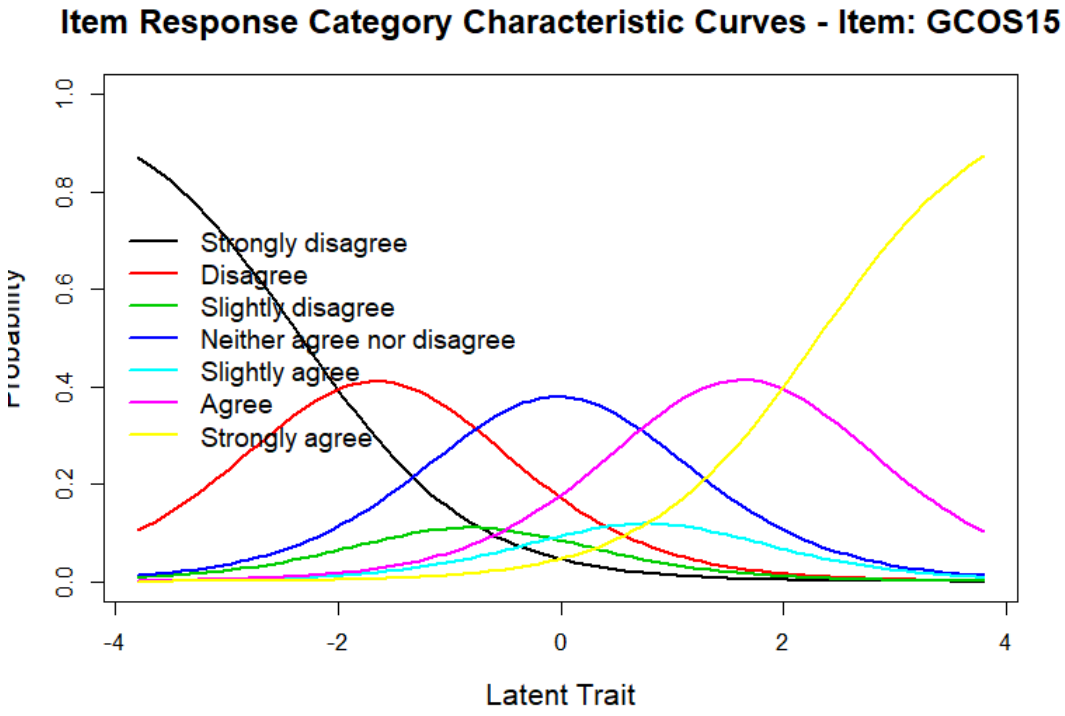


Figure 4.1: Example GRM item characteristic curve. Item 15 – Dscrmn = 1.30.

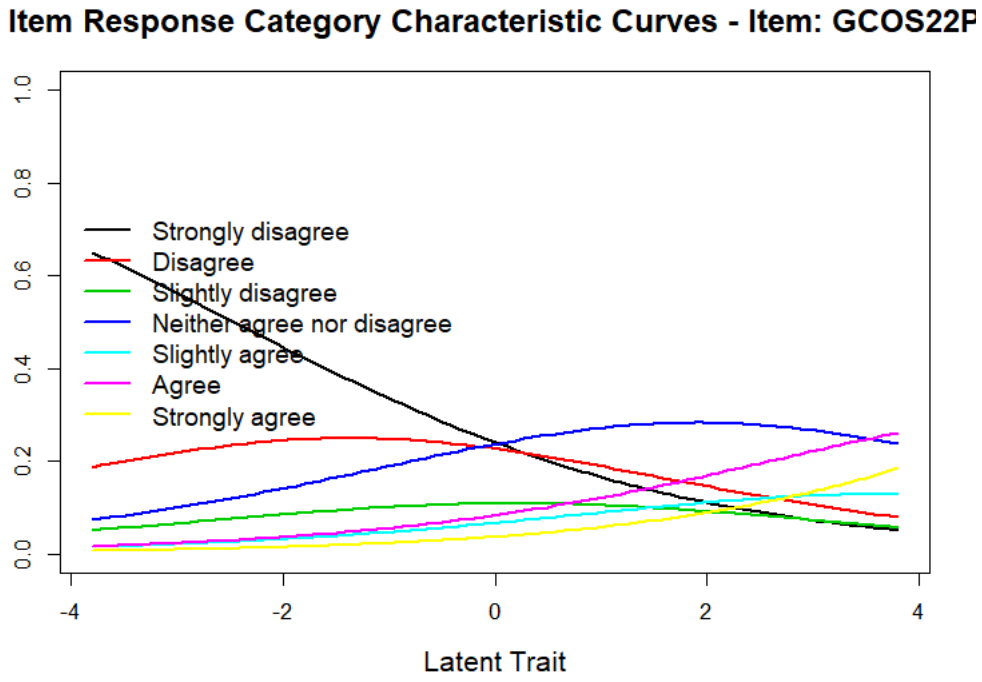


Figure 4.2: Example GRM item characteristic curve. Item 22 – Dscrmn = 0.47

The GRM item information curves (Appendix I) show how well and precisely each item measures the latent trait across various levels of said trait. Certain items may provide information at low levels of the trait, while others may provide more information at higher levels. Comparing item information curves allows a comparison to be made between items on how well the latent trait is represented by the item. An example item information curve is provided in Figure 4.3. The plot includes items designed to capture 'Emotional Regulation', and shows that item 4 would be the best candidate. The application of the item information curves to item selection is described in Phase III.

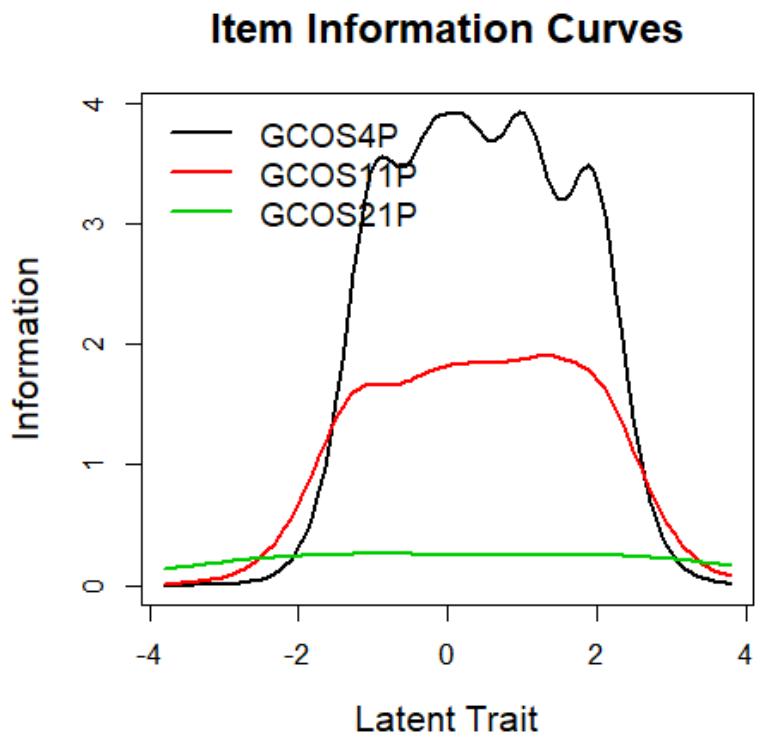


Figure 4.3: Example GRM item information curve (Emotional Regulation)

4.3 Phase III: Item Selection

One aim of this study was to develop a measure which could be used outside the context of CGS. Items 1, 14 and 23 were therefore not considered for selection because they specifically refer to 'clinical genetics services'. The three principles of item selection will now be addressed.

(i) Items with an unjustifiably low discrimination parameter (>1.34) were not selected:

Item discrimination parameters and associated curves were assessed, and those items with a 'High' or 'Very High' discrimination parameter were retained. For the other items, a flexible approach was used for data interpretation, since it was recognised that a number of factors may contribute to an item's quantitative properties. For example, because the cognitive interview findings indicated that a significant proportion of respondents experience confusion when answering items beginning with 'I don't know', such items were not immediately rejected for displaying inferior discrimination. Additionally, items asking specifically about children were expected to show a prominent peak for Option 4 ('Neither Agree nor Disagree / Not Applicable'). Following

consideration, item 3 (Dscrmn = 0.94); item 5 (0.73); item 6 (0.20); item 7 (0.59); item 10 (0.63); item 13 (0.46); item 19 (0.95); item 21 (0.91); item 22 (0.47) and item 24 (0.87) showed an unjustifiable inability to discriminate and were therefore removed from further consideration.

(ii) Items with factor loadings <0.55 were not selected.

Comrey and Lee (1992) suggest the following threshold values for factors analysis: 0.32 (Poor), 0.55 (Good), 0.63 (Very good), 0.71 (Excellent). Of the 11 items remaining in consideration, item 15 and item 16 presented with factor loadings <.55, in factors 5 and 2 respectively (Table 4.4). These items are poor representatives of empowerment and were therefore removed from further consideration. At this stage of item selection, nine items remained in consideration (Table 4.12).

Table 4.12: Items in consideration following the second principle of item selection.

Dimension of Empowerment	GCOS-24 Item
Cognitive C	(12) I don't know if this condition could affect my other relatives (brothers, sisters, aunts, uncles, cousins).
Cognitive C	(18) I don't know who else in my family might be affected by this condition.
Behavioural C	(2) I can explain what the condition means to people in my family who may need to know.
Behavioural C	(9) I am able to cope with having this condition in my family.
Behavioural C	(17) I don't know what I can do to change how this condition affects me/my children.
Emotional R	(4) When I think about the condition in my family, I get upset.
Emotional R	(11) Having this condition in my family makes me feel anxious.
Hope	(8) I feel positive about the future.
Hope	(20) I am able to make plans for the future.

(iii) To avoid redundancy, items capturing a similar outcome were not selected together; FA, GRM and cognitive interview findings were used to establish superior items.

Cognitive Control: Items 12 ('I don't know if this condition could affect my other relatives (brothers, sisters, aunts, uncles, cousins)') and 18 ('I don't know who else in my family might be at risk for this condition'). Both items were highly valued by cognitive interview participants (Table 4.2), and quantitative results were very similar. The descriptive information included in the parenthesis appeared to improve interpretability for item 12 so item 18 was removed from further consideration.

Behavioural Control: All remaining items were considered sufficiently distinct.

Emotional Regulation: Item 4 ('When I think about the condition in my family, I get upset') was selected over item 11 ('Having this condition in my family makes me feel anxious') because it was valued more highly by interviewees (Table 4.2), because it has a superior item discrimination parameter (3.65 to 2.47) (Table 4.11), and because the qualitative data suggested that anxiety levels may not reduce over time in people affected by a genetic condition, but instead fluctuate depending on the situation. This is not a desirable property in a scale designed to measure outcomes.

Hope: Item 8 ('I feel positive about the future') was selected over item 20 ('I am able to make plans for the future') for two reasons. Firstly, although both display high / very high discrimination, item 20 has significant ceiling effects (Figure 4.4). Secondly, item 8 has a factor loading of .753, compared to .567 of item 20 (Table 4.4).

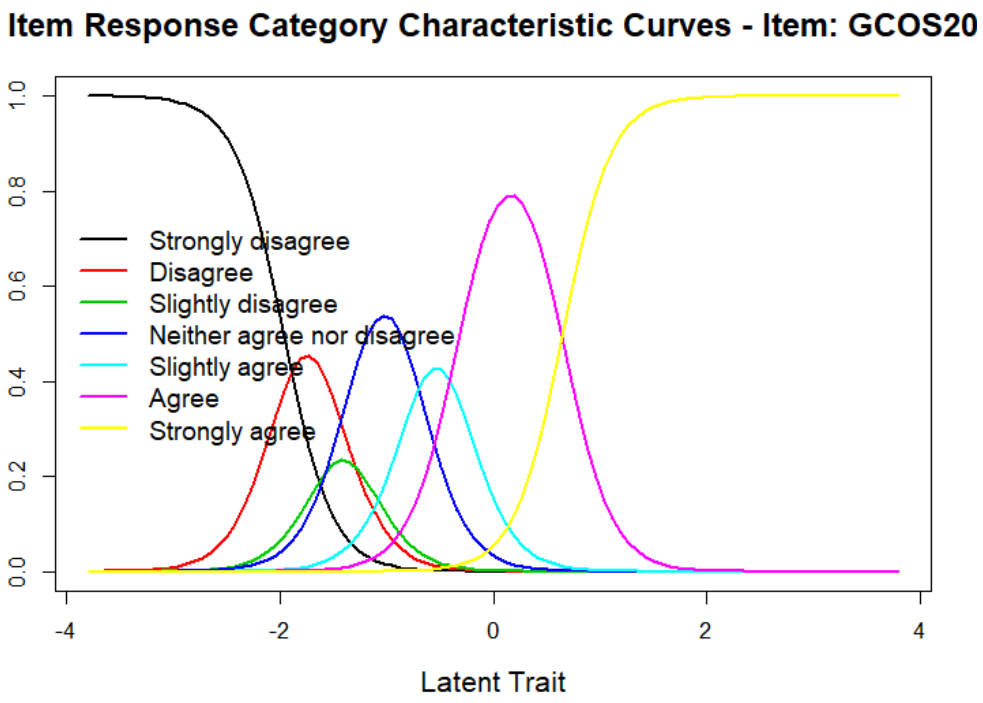


Figure 4.4: Item 20 item characteristic curve showing ceiling effects.

The reduced scale was constructed using the six items remaining with no justifiable reason for exclusion (2, 4, 8, 9, 12, 17) (Table 4.13).

Table 4.13: The final six items.

GCOS-24 item	Retained Items	Empowerment Dimension
2	I can explain what the condition means to people in my family who may need to know.	Behavioural Control
4	When I think about the condition in my family, I get upset.	Emotional Regulation
8	I feel positive about the future.	Hope
9	I am able to cope with having this condition in my family.	Behavioural Control
12	I don't know if this condition could affect my other relatives (children, brothers, sisters, aunts, uncles, cousins).	Cognitive Control
17	I don't know what I can do to change how this condition affects me/my children.	Behavioural Control

. 4.3.1 Likert Scale Optimisation

Figure 4.5 presents the results of the Rasch Rating Scale analysis. At low levels of empowerment, option 1 (Strongly Disagree) has the highest probability of response. Likewise option 7 (Strongly Agree), has the highest probability of response at positive levels of empowerment. Option 4 (Neither Agree nor Disagree) is the most likely to be chosen at the zero point. Options 3 and 5 (Slightly Disagree and Slightly Agree) have low probabilities of being chosen and do not show distinct peaks, suggesting that they could be removed without compromising scale quality; GRM results support this suggestion. It was decided that a five-point scale would be adopted, with a view to possible further shortening following results from the test-retest study.

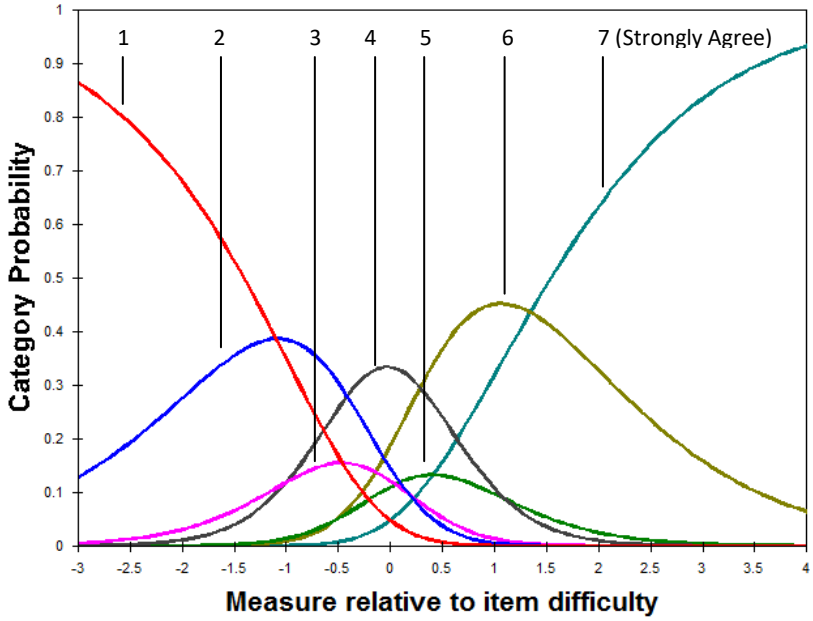


Figure 4.5: Rasch Rating Scale results. Each curve corresponds to a GCOS-24 response option, (1) Strongly Disagree to (7) Strongly Agree moving left to right.

Figure 4.6 shows the proposed scale, termed GCOS-6. Based on cognitive interview results, items were reworded to change 'I don't know' to 'I know' to eliminate the possibility for confusion over double-negatives.

		Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
1	I can explain what the condition means to people in my family who may need to know.	1	2	3	4	5
2	I know the chance that this condition could affect my other relatives (children, brothers, sisters, aunts, uncles, cousins).	1	2	3	4	5
3	When I think about the condition in my family, I get upset.	1	2	3	4	5
4	I am able to cope with having this condition in my family.	1	2	3	4	5
5	I know what I can do to change how this condition affects me/my children.	1	2	3	4	5
6	I feel positive about the future.	1	2	3	4	5

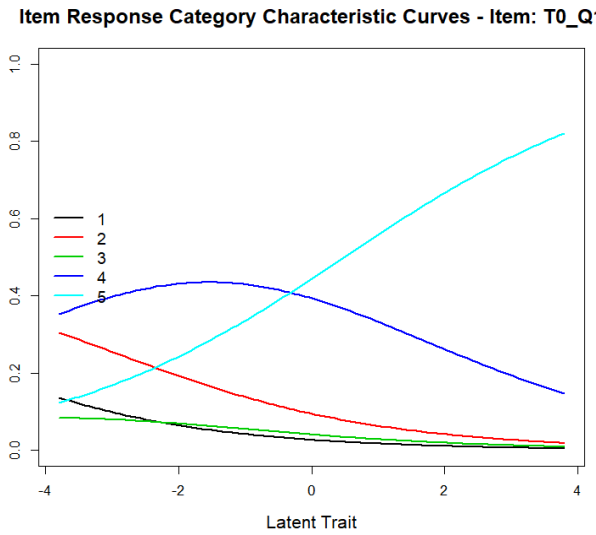
Figure 4.6: GCOS-6

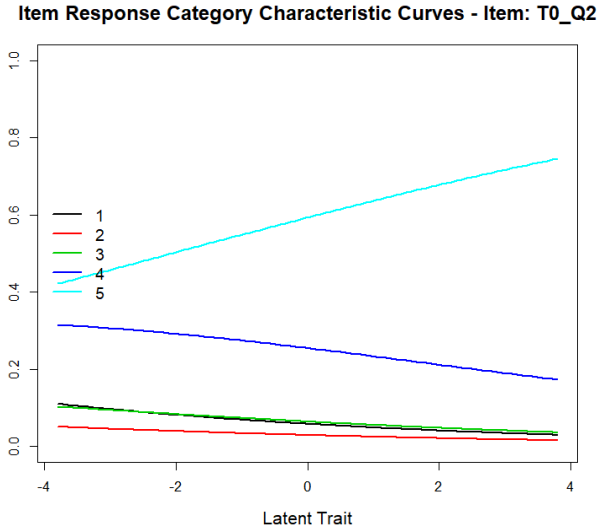
4.4 Phase IV: Validity and Reliability Testing

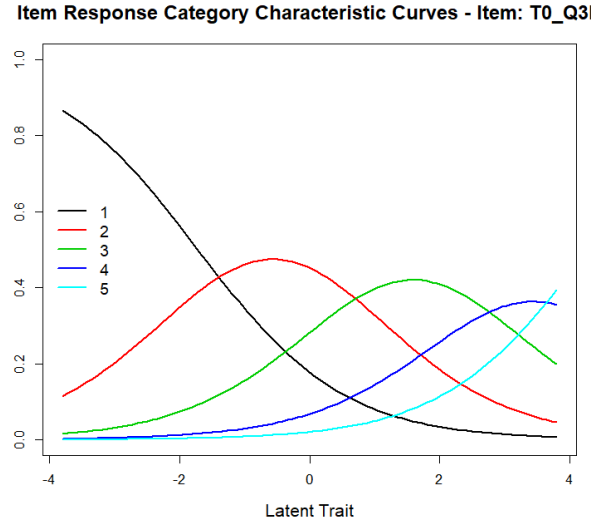
Face validity and content validity of GCOS-6 was assessed within the research team: Prof Angus Clarke (Clinical Professor) and Dr Marion McAllister (Senior Lecturer and Programme Director for the Genetic and Genomic Counselling MSc) at Cardiff University; Prof Katherine Payne (Professor of Health Economics) and Dr Maria Pampaka (Senior lecturer and psychometrician) at the University of Manchester. GCOS-6 content validity was also supported by the existing GCOS-24 content validity (McAllister *et al.,* 2011b).

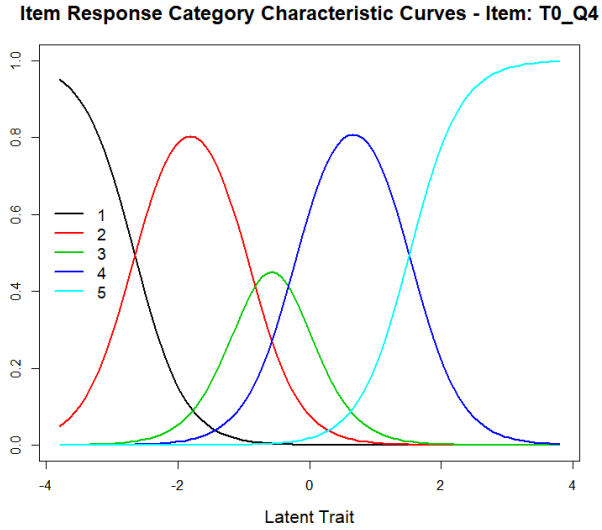
In the test-retest reliability study, 170 GAUK members affected by a genetic condition in their family responded to the advertisement and completed the online measure at T0. Of these, 96 (56.5%) completed the measure again at T1. Reliability as measured by the intraclass correlation coefficient was 0.788. Internal consistency (Cronbach's α) was α = .570. GRM item characteristic curves are presented in Figures 4.7 – 4.12. Item 1 ('I can explain what the condition means to people in my family who may need to know') (Fig 4.7) and item 2 ('I know the chance that this condition could affect my other relatives (children, brothers, sisters, aunts, uncles, cousins)') (Fig 4.8) show significant ceiling effects. Indeed item 2 suggests that respondents select 'Strongly Agree' across all empowerment levels. Considering that respondents were all active GAUK members with an interest in research, these findings are not surprising. Item 3 ('When I think about the condition in my family, I get upset') displays clear peaks but some positive skew. Items 4, 5 & 6 (Fig 4.10 – 4.12) display clear peaks and no skew.

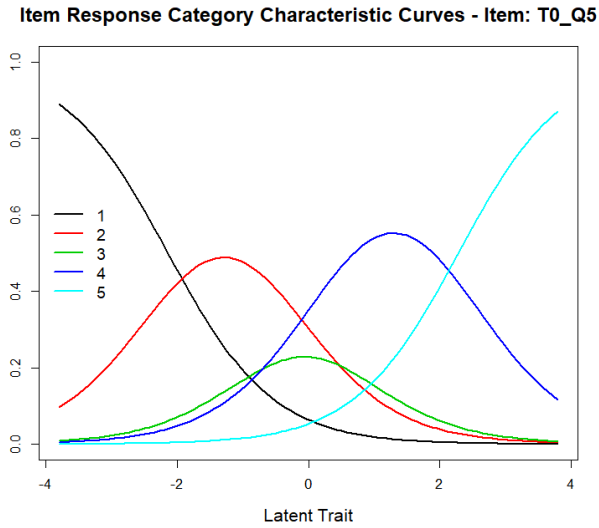
In summary, this chapter presents the results obtained over the course of the project. Qualitative and quantitative data were used to create a six-item, five-level version of GCOS-24: GCOS-6 (Figure 4.6). Three principles were used to guide item selection: (I) (i) Items with an unjustifiably low discrimination parameter (>1.34) were not selected; (ii) Items with factor loadings <0.55 were not selected; (iii) To avoid redundancy, items capturing a similar outcome were not selected together; FA, GRM and cognitive interview findings were used to establish superior items. GCOS-6 displays good test-retest reliability (0.788) and moderate internal consistency (α = .570). Item discrimination was generally good, with some understandable ceiling effects given the study sample of active GAUK members.











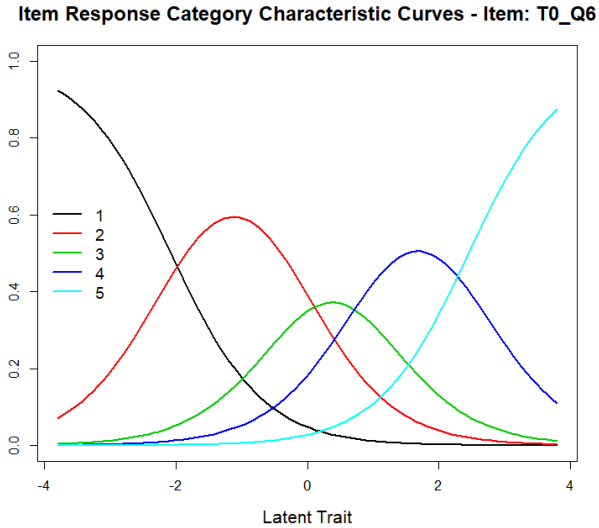


Figure 4.7 (top left) – 4.12 (bottom right): GCOS-6 GRM results. T0 = Time point zero. Q = Question e.g. Q1 (I can explain what the condition means to people in my family who may need to know).

5. Discussion

This chapter will discuss the reported results in the context of published research regarding patient outcomes in genetic counselling and testing services (CGS). The potential impact of the study within CGS will be considered, as well as the range of implications for future research and clinical practice. The discussion will include an assessment of the strengths and limitations of the study, and will conclude by commenting on the whether the aims and objectives of the study have been achieved.

5.1 Results in Context of Published Research

This research has developed a new short-form (6-item) version of the Genetic Counselling Outcome Scale (GCOS-24), potentially suitable for use in research, clinical audit, and clinical evaluations of CGS. The new scale, GCOS-6, shows good test retest reliability (0.788), whilst providing a less burdensome measurement scale for respondents and producing a significantly reduced number of response permutations ($1.56x10^{4}$) compared to GCOS-24 ($1.92x10^{20}$). Additionally, with genetic testing increasingly being performed in contexts outside the traditional models of service provision (Lo *et al.,* 2014, Rahman*,* 2014; Valente *et al.,* 2008), GCOS-24 items specifically referring to clinical genetics services were omitted from GCOS-6, making the new instrument appropriate for use both within and outside the context of clinical genetics. Ultimately this study represents a step towards the development of a preference-based patient-reported outcome measure (PROM) which could be used for the economic evaluation of CGS.

This study reports the first use of Item Response Theory (IRT) analysis on GCOS-24, contributing to the growing body of evidence that IRT methods confer many benefits over the traditional approaches of classical test theory (CTT), and supporting the call for wider use of IRT methods in PROM development (Embretson, 1996; Hambleton *et al.,* 1991; Hambleton & Jones, 1993; Nguyen, 2014; Reeve, 2002). Nevertheless, IRT findings should be interpreted with caution when analysing subjective topics. Item performance may be influenced by a variety of factors, and may be representing a minor issue in wording rather than the importance or quality of the underlying outcome. GCOS-24 items asking about children, for example, showed a prominent peak for the 'Not Applicable' response option in their item characteristic curves regardless of

the outcome domain being captured, reflecting the reality that not all CGS users have children; items beginning with 'I don't know' generally performed very poorly, reflecting the qualitative results which suggest that many people misinterpret the double-negative. Rather than taking IRT results at face value, reasonable judgement should be applied.

The same rule is true for classical methods, and the factor analysis carried out in Phase II provides a good example. Factor analysis is a tried and tested method, using correlations between response patterns to determine which items capture similar underlying traits. Correlations, however, may be due to unexpected causes. In this study, all five GCOS-24 items beginning with 'I don't know' were grouped into the same factor. Three dimensions of empowerment are represented, including themes of decision-making, knowledge of the condition, and powerlessness. Taking into account the qualitative findings which indicate participant confusion over items beginning with 'I don't know' it is likely that the correlations in this factor were due to this rather than any underlying trait.

Even internal consistency, a key requirement for any questionnaire intended for use as a PROM (Mokkink *et al.,* 2010; Terwee *et al.,* 2012), is open to interpretation. Since the calculation is based upon item correlations, random error averages out as one adds more items, so in practice Cronbach's α is affected by the length of the scale (Streiner, 2003). Scales over 20 items will generally have acceptable values of α (>.7), whereas scales with fewer items will have fewer correlations from which to draw upon and in turn may present with lower internal consistencies (Streiner, 2003). It is therefore not entirely unexpected that the internal consistency of GCOS-6 (α=.570) is significantly lower than that of GCOS-24 (α=.870). Whilst it is understandable that internal consistency is highly recommended for a new measure, holding all scales to the same threshold is problematic. In short, quantitative methods offer powerful tools for PROM development, but results must be interpreted with reason.

Evaluations of CGS have traditionally examined such outcome variables as information recall, reproductive intentions and decisions made, and patient satisfaction. Measures of process such as waiting times and numbers of patients seen have also been used, as well as the performance characteristics of genetics tests (e.g. sensitivity, specificity and predictive values) (Clarke *et al.,* 1996; Payne *et al.,* 2008; Wang *et al.,* 2004). It is widely argued by clinical genetics professionals that traditional approaches to CGS evaluation

are neither relevant nor appropriate, nor are they highly valued as outcomes by patients and their families (Clarke *et al.,* 1996; McAllister *et al.,* 2008; Payne *et al.,* 2008). GCOS-6 captures outcomes which are relevant to and valued by the population of individuals who use CGS, demonstrated through the extensive qualitative research collected in GCOS-24 development and the cognitive interviews in this study.

Moreover, existing outcome measures used in evaluations of CGS have generally been designed to capture a specific outcome or a restricted number of outcomes, often with respect to a single genetic condition (Payne *et al.,* 2008). Indeed over half of the measures identified in the literature review were used in the evaluation of CGS for inherited cancer, commonly breast cancer, and many were bespoke measures developed for use in a specific study (Section 2.4.2: p33). GCOS-6 has been designed to capture a range of potential patient outcomes relevant to any potential CGS user with any condition, and to provide information to clinicians on patient benefits which may be useful for service development and audit of process. With that said, the omission of an item capturing the Decisional Control outcome domain is concerning. Outcomes relating to Decisional Control, such as informed decision making, have been proven to be valued by CGS users (Clarke *et al.,* 1996; Legare *et al.,* 2016; McAllister *et al.,* 2008; McAllister *et al.,* 2011a*;* Metcalfe, 2018; Miller *et al.,* 2005), and were discussed by participants at interview. Clinical use of GCOS-6 may be held back by this omission, and a potential area of future research could be the development of a valued and relevant item which can capture Decisional Control for all potential CGS patients.

Finally, this study supports the use of both quantitative and qualitative methods when approaching the task of scale development. In its ideal form, a mixed methods approach can represent a happy marriage between theory and empirical confirmation, providing unique strengths and offsetting the weaknesses of either approach alone. For instance, qualitative research is strong when used to understand the context or setting in which people behave, including their perspectives, attitudes and opinions. Quantitative methods can offer little information in this respect. On the other hand, a weakness of qualitative research is the potential for researcher-bias to affect data collection and interpretation; quantitative methods do not have these weaknesses. Taken together, incorporating a mixed methods approach into the study design can help to provide a complete and comprehensive understanding of the research problem.

5.2 Implications for Practice

Patient engagement is increasingly acknowledged as a key component in the process of service improvement in healthcare, with recent evidence affirming that patients who are engaged in their care perceive improved outcomes (Remmers *et al.,* 2009). Historically, with the exception of collecting feedback on satisfaction or experience with care, patients have been an untapped resource when evaluating the quality of healthcare and of long-term support services (National Quality Forum, 2013). The introduction of routine use of PROMs to the NHS in 2009 was a landmark development, reflecting the growing recognition throughout the world that the patient's perspective is highly relevant to efforts to improve the quality and effectiveness of healthcare. To this end, valid and reliable instruments such as GCOS-6 provide essential tools.

The routine use of PROMs by the NHS has generated considerable interest from other countries, including Canada, Germany, New Zealand and Sweden (Devlin & Appleby, 2010). With that said, although PROMs offer enormous potential, there are at present only four procedures that are covered by the National PROMs programme, accounting for only around 3.3% of all elective activity in the UK (Devlin & Appleby, 2010). Extending the coverage of PROMs, especially into areas of NHS activity which have traditionally lacked universal measures of quality and effectiveness, is a challenge which can only be met through the establishment of valid outcome measures. Designing and implementing PROMs in certain branches of healthcare, such as CGS and mental health services, may prove more challenging since care pathways and patient outcomes may be significantly more complex in comparison to those branches which involve discrete treatment events, the success of which can be measured in objective terms e.g. surgical interventions.

As well as widening the scope of PROMs coverage, developing valid and reliable PROMs such as GCOS-6 could offer other potential benefits, such as the encouragement of a more coordinated system for comparisons of healthcare quality, and the provision of additional sources of information for NICE evaluations of healthcare interventions. Indeed a significant problem faced by NICE is the dearth of appropriate or robust evidence from healthcare practice. PROMs are also appreciated by healthcare professionals as a tool to complement their own clinical judgement and encourage their professional development (Boyce & Browne, 2014; Costal-Tirado, 2017). The introduction of routine PROMs collection has potentially important

implications for enlarging the base of real-world evidence on cost-effectiveness that NICE can draw upon to inform its guidance to the NHS.

The National Quality Forum (NQF) is a neutral standards-setting organisation in the US which endorses outcome measures used to assess the quality of healthcare based on well-vetted, widely accepted criteria. Along with the baseline requirements of validity and reliability, these criteria include being 'Person Centred' and 'Meaningful'. The concepts measured by the PROM should be relevant and important from the perspective of patients and their families, ideally capturing health-related quality of life (HRQoL) impacts. GCOS-6 satisfies these criteria, and indeed GCOS-24 is currently in the process of being endorsed by the NQF.

In an economic context, the NHS faces the sizeable challenge of bringing about £22 billion worth of productivity improvements by the year 2020/21 (ref) and as such there is currently a considerable focus on efficiency and cost-effectiveness within the healthcare system. As financial and workload pressures increase, it is important that both human and monetary resources are targeted where they are most effective, and that the provision of services is modified according to need. In this time of budget constraints, and rising costs, in healthcare, the rapid advances in CGS are a source of both hope and concern. On the one hand, these services have the potential to benefit the population in many ways, for example by enabling the early detection of hereditary predispositions to specific diseases, and by offering support and guidance to those affected by a genetic condition within the family. From a financial perspective however, genetic interventions can be extremely costly (Ref?). Any increases in funding would likely come at the expense of another service, and so such actions must be thoroughly justified in the eyes of decision makers looking to maximise benefits per unit cost.

In order to help determine the allocation of resources amongst competing healthcare interventions, the NHS uses economic evaluations. One of the most common methods is cost-effectiveness analysis, which compares interventions in terms of their cost per quality adjusted life years (QALYs) gained. A QALY value is calculated through the use of two variables: (i) the change in a patient's health status, quality of life (QoL), or health related quality of life (HRQoL); (ii) the change in a patient's length of life (in terms of 'years'). With this in mind, if a service is going to be amenable for economic evaluation, it must have a clear set of patient outcomes laid out in terms of their health, QoL or HRQoL, as well as a means of measuring the outcomes. Additionally,

NICE have called for outcome measures used in service evaluation to be preference-based, to reflect the value that individuals attach to each item and response option. GCOS-6 serves as a promising first step in the development of a preference-based PROM which can be accepted as the standard for use in economic evaluations of CGS.

This research is especially timely since the rate at which economic evaluations of CGS are being carried out and published has increased greatly over recent years (Carlson *et al.,* 2005; Djalalov *et al.,* 2011; Andrea *et al.,* 2015). Rapid advances in genetics technology, coupled with the current financial pressures, have led to a demand for economic evaluations to help identify which interventions confer greater health gains per unit cost. This was exemplified by Andrea *et al.* (2015), who carried out a systematic review of primary economic evaluations of predictive genetic and pharmacogenetic testing programs from inception until 2012. Of the 128 articles identified, almost 40% were published in the three years from 2010-2012.

5.3 Implications for Future Research

The development of GCOS-6 opens up a number of avenues for future research. One option would be to construct a relevant and valued item to capture the Decisional Control dimension. Decisional Control is a vital dimension of patient empowerment (McAllister *et al.,* 2011a), and one of the central tenets of genetic counselling is that the counsellor should adopt a 'non-directive' approach, trying to help the client arrive at the best decisions from their own perspective, rather than guiding them towards any particular decision (Elwyn *et al.,* 2000). The process of genetic counselling may involve the facilitation of a decision making process in relation to prenatal diagnosis or the termination of a pregnancy; it may relate to a decision about predictive genetic testing. Some decisions may be particularly complex, involving a balance between the risk of a procedure and the benefit of obtaining diagnostic information.

Recent studies have found that, although patients are better informed today than in the past, there is dissatisfaction and frustration due to inadequate personal input into their decisions about treatment (Jun *et al.,* 2016; Nicholls *et al.,* 2013; Pae *et al.,* 2014). In the US, empowering patients and families to actively engage in decision-making has been emphasised in the Patient Protection and Affordable Care Act (2010), and by national agencies such as the Agency for Healthcare Research and Quality (2015). It is

therefore likely that the construction and implementation of an item representing Decisional Control to GCOS-6 will be of benefit to the scale.

Sensitivity to change (responsiveness) is one of three quality domains, along with reliability and validity, recommended for all new measures by COSMIN guidelines. It relates to the ability of an instrument to detect change over time. The minimally important difference (MID) is another important concept in measurement scale development, providing a measure of the smallest change in the outcome that patients perceive as important. Neither the MID nor the responsiveness of GCOS-6 were tested in this study. Assessment of these measurement properties will help to ascertain whether GCOS-6 is a robust instrument and to identify possible areas for improvement.

COSMIN guidelines also recommend the use of IRT methods in the development and evaluation of measurement properties (Terwee *et al.,* 2012). Whilst the graded response model (GRM) was used in this study to assess item discrimination, there are a number of other methods which fall within the scope of IRT, each offering unique characteristics. One particular avenue for future research could be to use the Rasch model (Rasch, 1966) to explore the extent to which GCOS-24 and GCOS-6 measure the same construct, thereby indicating how well GCOS-6 captures empowerment. An appropriate sample size for IRT analysis is around 100 individuals (Terwee *et al.,* 2012). The Rasch model is a goodness-of-fit test, applying constant item discriminating powers and calculating a result based on item scores and overall estimates of item difficulties. Going further, were an item to be implemented which represents Decisional Control, the Rasch model could indicate the degree to which this new instrument agrees with GCOS-24 in comparison to GCOS-6.

IRT methods could also be used to examine the rating scale statistics of the five-point Likert scale within GCOS-6, with a view to further reduction. Two potential benefits may arise from this. Firstly, reduction of the Likert scale would result in an even lower number of potential health state values, streamlining the process of attaching preference weights (Brazier *et al.,* 2002). At its current length, of six items each with five levels, GCOS-6 has 1.56×10^{4} possible response permutations, greatly reduced from GCOS-24 (1.92×10^{20}) and similar to the widely-used preference-based EQ-5D instrument (3.13×10^{3}). Secondly, rating scale statistics may help to optimise the GCOS-6 Likert scale by identifying redundant response options.

5.4 Strengths and Limitations

One of the great strengths of GCOS-6 is that it was developed from GCOS-24: an internationally recognised PROM with demonstrated validity, reliability and responsiveness which specifically measures patient outcomes valued by CGS patients and service providers (McAllister *et al.,* 2011b). Indeed GCOS-24 content validity has previously been demonstrated for CGS (McAllister *et al.,* 2011b), and clinical utility has been demonstrated both in the UK and internationally (Diness *et al.,* 2017; Inglis *et al.,* 2014; McAllister & Dearing, 2015; Munoz-Cabello *et al.,* 2017). The substantial qualitative research underpinning GCOS-24, coupled with the further qualitative research in the current study, supports the potential implementation of GCOS-6 for service evaluation in CGS and in future research.

A further strength of this study was the large sample size achieved for the test-retest reliability analysis (n=96), a figure meeting the COSMIN quality criteria for assessment of the methodological quality of studies on measurement properties of health instruments (Terwee *et al.,* 2012). A weakness, however, was sample homogeneity. During the interview stage all ten participants were white-British and resident in South-Wales. Cognitive interview results are therefore limited to the perspective of families who live in a specific region and speak English. Additionally, only one of the ten interviewees was classified as 'at-risk' for a genetic condition. Seven were affected and two were unaffected themselves but had an affected child. At-risk individuals are a key target demographic for CGS, and higher representation would have been beneficial.

For the test-retest reliability study, sample homogeneity could be a possible explanation for the ceiling effects observed with certain items. Individuals who join GAUK, take an active interest in research projects, and volunteer themselves for such projects, are likely to have a good understanding of their condition. Such individuals may also have higher levels of empowerment than the majority of patients referred to CGS, a suggestion supported by McAllister *et al.* (2011b) who found that active patient support group members were significantly more empowered than those who did not attend support groups. In order to develop an outcome measure, and indeed a service, which caters to all needs it will be necessary to collect data from a wide range of CGS users, particularly those which could be classified as 'hard to reach'.

5.5 Conclusion

In conclusion, GCOS-24 has been shortened to a six-item measure with a five-point Likert scale (GCOS-6). GCOS-6 offers a genetics-specific measure which is applicable both within and outside the context of clinical genetics, capturing a range of potential patient outcomes for individuals affected by any genetic condition. The new instrument will be less burdensome to patients than GCOS-24 and psychometric testing indicates that GCOS-6 has good test-retest reliability. Further testing, however, for example to examine interpretability and responsiveness, will be necessary before GCOS-6 can be recommended unreservedly for routine evaluations of genetic counselling and testing services. Obtaining data from hard to reach demographics will be of particular benefit in constructing an instrument relevant to the entire population of service users. Future developments to GCOS-6 could involve the addition of an item representing Decisional control, and the attachment of preference weights reflecting the value placed on items by CGS users. Overall, this study represents the first step in developing a preference-based measure which could be used in the evaluation of genetic counselling and associated testing services.

6. Reflective Discussion

In this chapter I shall reflect more informally on the reality of the research process compared to what was planned. I will outline the practical pitfalls which I have experienced and hypothetically what I would do differently. Lessons learnt during the process will be scattered throughout, before concluding with some advice I might pass on to any future students undertaking an MPhil.

6.1 Starting Out

Prior to starting this project I had no experience with clinical genetics research, and little idea of what the genetic counselling process involved. Undergraduate studies had included modules on cell biology, molecular biology, and genetics, but clinical genetics seemed to occupy an individual niche, separate from other areas of Biology. Looking back now, I was far too slow to build up a foundation of knowledge in this new field. Rather than taking the time early on to read extensively and gain a thorough understanding of the basic principles, I rushed ahead and began learning how to perform the quantitative methods.

My lack of understanding was evidenced through some sub-par work in the initial stages of the project for which I can only apologise to my supervisor for subjecting her to. One essay in particular comes to mind, where I neither took the time nor had the mental awareness to realise that the empowerment construct I was meant to be studying was distinct from empowerment in other areas of science. Over time I caught up, but I am sure that I could have saved myself a great deal of time and a great deal of misplaced effort had I taken the necessary steps to properly inform myself in the early weeks.

It was also during these initial stages that I was advised to scrutinise the proposed study design; to examine every planned method and provide a justification for its use over other available methods. This was excellent advice, which shall stay with me moving forward. At the time, however, as a lowly student with no background in clinical genetics, the idea of seriously questioning the study design did not occur to me. Instead, I took the study design as gospel and worked backwards, supplying justifications based on a known end-point. In short, I did not take ownership of the

project fast enough. I followed directions and did as instructed, but I failed to step up and apply independent thought in the initial stages.

More excellent advice came my way early on when I was told to construct a Gantt chart showing the project timeline, and to circulate it around the research team. Having never used a Gantt chart before I was unsure how much value one would be, but again this proved to be a lesson which I will take with me and apply to future work. Of particular benefit was the ability to visualise the timescale for each stage of the project and to adjust deadlines accordingly.

Linked on to this, one major difference I found with this MPhil, as opposed to undergrad, was the necessity to take other parties into account when aiming for a deadline. An undergrad semester is only 11 weeks in total, and for most assignments the student is entirely responsible for their own work. If an essay is to be written, for example, no-one else can influence the time which one's essay is handed in. At postgrad, however, other people must be considered. The person you need to contact may be on holiday; the interview participant may ask to delay until next week. Depending on the circumstances it may be polite to circulate a piece of work around the research team prior to submission. If this is the case, they may ask for a couple of weeks to provide feedback, and then their feedback will have to be applied. Suddenly a submission deadline which is four weeks in the future becomes a top priority. I began to build in buffer periods to the study design in case of unforeseen delays which are so common in research. The importance of prior planning was a valuable lesson from this project.

6.2 Recruitment

The online recruitment method proved to be an excellent choice. Response rates were pleasing and the process was much faster than if materials had been sent out by post. It was also much cheaper, which is not an unimportant consideration for an unfunded project. With that said I perhaps should have explored other recruitment sources to try and overcome sample homogeneity. A wider catchment area, for example, could have been used. Local events, clinics or focus groups could have been contacted. Again this goes back to my sluggishness in taking ownership of the project. The application for

ethics approval was submitted in March, and by the time I realised that more recruitment options could be beneficial, the window of opportunity had long closed.

The only serious complication to occur during the project came in the recruitment phase. A representative from GAUK had very kindly agreed to help recruit participants for cognitive interviews by contacting GAUK members on my behalf. Months later, however, following a disappointing campaign in which only ten participants were recruited, it was discovered that the mailing list was not exclusively for individuals affected by genetic conditions. Once academics, GAUK staff and the like had been removed, only 35 of the initial 130 email addresses remained. With time marching on it was not possible to explore other avenues or apply for further ethical approval. Although a specific example, I have taken away a wider warning to leave no stone unturned during the study design and to make sure that back-up options are in place in case the initial plan unexpectedly fails.

The recruitment process, coupled with consistently excellent advice from my supervisor, helped to develop my confidence in email communication. Indeed at the start of the project I was often reluctant to send emails for fear of troubling the recipient. Likewise If I didn't receive a response to an email, I would be very hesitant to chase it up. Over the course of this year I feel that I have become more confident in drafting correspondence, and in finding the balance between being polite, grateful, and concise.

6.3 Methodology

The most enjoyable moments of the project came when carrying out the methods. The cognitive interviews were immensely interesting as they provided a real-world perspective of living with a genetic condition, and working with statistics is a particular pleasure of mine. Nevertheless, I believe in hindsight that some elements of the methodology could have been changed.

For one, I believe that Factor Analysis (FA) could have been discarded from the study design. FA was selected to provide an objective approach to identifying underlying traits within GCOS-24, a decision which in theory was sound. FA is a popular method in scale development, and had previously been used in the construction of GCOS-24 (McAllister *et al.,* 2011b) and the Audit Tool for clinical genetics (Skirton *et al.,* 2005).

The problem with using FA in this context was that GCOS-24 had been specifically designed to capture empowerment, a five dimensional-construct developed through extensive qualitative research. Whilst I recognise the benefit of FA methods for researchers wishing to avoid subjective or arbitrary criteria for factor retention, I believe that the qualitative evidence supporting the dimensions of empowerment was sufficient to supersede FA for the purposes of this study. Additionally, FA is not a foolproof method of determining underlying traits, since the correlations used to produce factor loadings may arise as a result of unexpected influences. This was evidenced in this study by the fact that all items beginning with 'I don't know' fell into the same factor, regardless of the apparent outcome domain. All in all, I believe in hindsight that FA was superfluous given the existing presence of empowerment.

It is unclear whether a more liberal approach to altering GCOS-24 would have resulted in a superior final scale. Operating under the reasoning that any changes made to the meaning of an item would require separate validation, an initial decision was made that only minor alterations to improve item interpretability would be permitted. A purist, however, might argue that any change whatsoever could affect the meaning and therefore would require separate validation, and in turn that the decision to replace 'I don't know' with 'I know' was mistaken. The possibilities of a more liberal approach are interesting to consider. Similar items could have been combined to encompass a broader range of outcome e.g. Item 2 ('I can explain what the condition means to people in my family who may need to know') and item 16 ('I can explain what the condition means to people outside my family who may need to know') could perhaps have been combined to read 'I can explain what the condition means to other people who may need to know'). Alternatively, item 24 ('I can make decisions about the condition that may change my child(ren)'s future / the future of any child(ren) I may have') could have been reworded to 'I can make decisions about the condition that may change my future' and included as a representative of Decisional Control. With that said, providing solid justifications for such changes could prove challenging.

With the rules on GCOS-24 alterations as they were, certain GCOS-24 items could have been immediately discounted from analysis. Items specifically referring to the 'clinical genetics service', for example, were never going to be selected for the short-form as the aim was to create a measure appropriate for use both within and outside the context of clinical genetics. Similarly it could have been recognised that items pertaining to children would not be relevant to a significant proportion of CGS users.

Removing unsuitable items *ab* initio would have reduced participant burden during interview and would have simplified the quantitative analysis.

6.4 Advice to Peers

In the interest of brevity I have narrowed my advice down to two points.

1. Pursue other interests outside of the project.

It is very easy for a prolonged PhD or MPhil project to take over one's life. There are always deadlines looming, there are always pages to be written, and there is always work to be done. An uncompromising work ethic may pay off in the short term, and may have proven fruitful during the short semesters of undergrad study, but in my opinion is not conducive to a happy and productive life when faced with a project lasting up to four years.

It may sound trivial, but the recent push by universities to consider mental health is no accident. If left unchecked, the cloud of a PhD project can hang overhead at all hours of the day. It can surround you and consume you and can be hard to escape from, especially if the project is oriented on independent research. Working alone on a project, staring at a laptop all day can make for a lonely time. This is particularly relevant in a period where there is so much entertainment available through the screen of a phone or computer.

So to any future student I would pass on the advice to join a society, join a sport, and make a point of attending on a regular basis. It could be badminton or ballroom dancing, wine tasting or poetry reading, find something which interests you, something which maybe doesn't involve a computer screen, and get involved. University offers a wealth of opportunity in this respect and I certainly believe that extra-curricular activities are an important dimension of university life. Furthermore, whether you're new to the area or not, get outside of the Cathays bubble once in a while, leave work behind and explore Wales. It will be refreshing and invigorating.

With regards to the work, many people have found great success in treating PhD research like a nine-til-five job, confining the project to working hours and keeping evenings and weekends free. Personally, sport has always played a large part of my life and served as my escape; I also had a part-time job in a cocktail bar. For me, regular

extra-curricular commitments helped to provide a structure to my time and to cultivate the much mentioned ‘work-life balance’ which despite its clichéd usage represents an important philosophy.

2. Plan ahead with specific, time-bounded goals.

Self-discipline was crucial to my research project. Aside from a weekly meeting with my supervisor, which later became a monthly meeting, my time was my own. I could get out of bed at any time, I could work as much or as little as desired, or not work at all. Whilst this had certain benefits, for example being able to plan my time as I wished and fit work around my other commitments, it was often difficult to find the motivation to work. Therefore I would advise any student carrying out independent research to generate a mentality of self-discipline rather than self-motivation. Motivation is fleeting, unpredictable, and too often absent altogether. Motivation can be distracted by YouTube or delayed by a hangover. Discipline, I feel, is a much better alternative. That said, discipline is not always easy, so I would advise someone with no structure or pressure to their time to generate a structure and a pressure. Write a to-do list and set discrete objectives which can be achieved in the short term. Plan deadlines ahead of time and note the relevant dates. For example, if your supervisor gives you a month to write a Literature Review chapter, take the time before getting started to break down the deadline into smaller chunks e.g. “From 2pm to 5pm every day I will read at least 10 papers, making notes in MS word and logging the details in a spreadsheet. After one week I will design a preliminary structure to the review and plan the to-do list for the next week.” Vague, amorphous goals will not prove fruitful.

As well as helping to maintain my work ethic, artificial deadlines provided a structure to the project timeline. I would encourage any research student to expend some effort early on in planning realistic and detailed deadlines, to take the deadlines seriously, and to update the deadlines if circumstances change. As mentioned earlier, Gantt charts and buffer periods have proved useful in my experience to visualise the timescales and to allow some flexibility. I would advise a new student to pay particular attention to the time required for ethics applications and to get those applications in as early as possible. On a smaller scale, I would advise them to not develop a habit of unnecessarily delaying the minor tasks e.g. replying to emails. If indulged, this can easily lead to more significant delays in the project as a whole.

Finally, if I may, I'd like to end with a quote from Alexandre Dumas' Count of Monte Cristo. I first read the book not long before starting this project, and the following lines have stayed with me as I have tried with varying levels of success to get to grips with the field of genetic counselling. The protagonist, Dantes, who has wrongfully been thrown into solitary confinement, has managed to make contact with an old man in the neighbouring cell. The old man possesses considerable knowledge and wisdom, proficient in mathematics, physics, history and languages, and has just offered to teach Dante everything he knows over the next two years:

> "Two years!" exclaimed Dantes; "do you really believe I can acquire all these things in so short a time?"
>
> "Not their application, certainly, but their principles you may; to learn is not to know; there are the learners and the learned. Memory makes the one, philosophy the other."

I thank you for your consideration of this thesis.

Appendix

Contents

Appendix A – Cognitive Interview Recruitment Email

Dear *|FNAME|* *|LNAME|*

Researchers from the Cardiff University School of Medicine would like to invite you to take part in a research project. You have been chosen because either you or members of your family are affected by or at risk of a genetic condition. Because of this, we believe that you have the personal experience we are looking for to help with our research. The project will involve one short interview, in which you will have the opportunity to speak about your experience of having an inheritable genetic condition in the family, and give your opinions on how we can improve our research. In the long run, we hope to use your comments to improve our research, which in turn can be used to improve how clinical genetics services responds to patient needs.

I hope you will consider taking part in this important study,

Yours Sincerely,

Steven Blunt.
Public Engagement and Policy Officer - Wales
Genetic Alliance UK, Wales Gene Park, Cardiff University, Cardiff, CF14 4XN
E: blunts@cardiff.ac.uk T - +44 (0)2920 748 154

Yes, I would like to take part

If you are interested in being part of our research, please click the box above. If you would like to find out more about the project, please continue reading.

Appendix A – Cognitive Interview Recruitment Email

Project Title: Developing a Questionnaire for the Evaluation of Clinical Genetics Services

You are being invited to take part in a research study which aims to help improve the quality of care provided by clinical genetics services. Before you make a decision we would like you to understand why the research is being done and how you would be involved. Please take the time to read the following information carefully, and you're welcome to talk to other people about it before you decide. If you have any questions, or if you would like more information, please feel free to get in touch using the contact details below.

Information

Why are we doing this study?
We want to develop a short questionnaire which can be used to evaluate genetic counselling and testing services. More specifically, we would like to hear feedback from patients, family members, and others so that we can build the best questionnaire possible. We want to help genetic counselling and testing services focus on what patients really want from their healthcare. The study will begin on May 1st 2017 and finish on January 1st 2018. It is an 'unfunded' study, forming part of the research of Cardiff University MPhil student Peter Grant.

Information

Why have I been chosen?
You have been chosen because you are, or have a connection to, an individual with a genetic condition. We believe that you have the relevant knowledge and experience to help us in our research.

Do I have to take part?
Taking part is entirely voluntary, and there will be no consequences if you decide not to take part.

Can I withdraw?
You are free to leave the study at any time, with no consequences.

What will I have to do?
There will be one meeting lasting 30-50 minutes. You can come to us or we can come and meet you. Alternatively interviews can be carried out by phone or Skype. You will be asked to read through a questionnaire and answer five or six questions, each to do with clinical genetics, and discuss your thoughts about each one. We will then ask you some further questions about the questionnaire, and we encourage you to speak freely and honestly. With your consent, the meeting will be audio-recorded. Unfortunately, as this is an unfunded student research project, we cannot offer reimbursement for any travel costs incurred as part of the research study.

How much time will the study take?
The study will take 30-50 minutes in a single meeting.

Who is carrying out the study?
The study is being carried out by MPhil student Mr Peter Grant, and is supervised by Dr Marion McAllister (Senior Lecturer in Genetic Counselling, Cardiff University); Prof Angus Clarke (Clinical Professor in Medical Genetics, Cardiff University); Prof Katherine Payne (Professor in Health Economics, University of Manchester).

Will the study benefit or disadvantage me?
We do not expect that the study will harm or disadvantage you in any way. Also we do not guarantee that there will be any benefits to you from taking part in the study. You will not have to answer any questions which make you uncomfortable, and you can choose to leave the study at any time without giving a reason. We cannot guarantee compensation in the event of something going wrong, and unfortunately, as this is an unfunded student research project, we cannot offer reimbursement for any travel costs incurred as part of the research study.

What will happen to the results of the study?
The results will be used by the researchers to help develop the questionnaire, and for no other reason. Data will not be shared with anyone outside of the research study team. Your answers and comments will be analysed and compared to see if other people have said similar things. What you say may be used to re-word some of the questions to help make them more clear. All aspects of the study, including results, will be strictly confidential and only the researchers will have access to information on participants. If the results form part of a report, it will not be possible to identify individual participants.

Will my personal information be confidential?
All data will be strictly confidential and only the researchers will have access to your personal information. Cardiff University has strict rules and standards on confidentiality, and these rules will be followed. The results from completed questionnaires will be made anonymous and all personal information removed. Likewise, your name will not be used and you will not be identifiable on any interview transcripts or interview excerpts in any publications arising from the research.
All personal information will be kept in a locked cabinet in the office of the Principal Invesitgator Dr Marion McAllister, located in the Institute of Medical Genetics, Cardiff University Heath Park Campus.

I have ethical concerns about the project
For any concerns about how the research is being / was conducted, please contact the Public Engagement and Policy Officer for Genetic Alliance UK - Steven Blunt. Email: blunts@cardiff.ac.uk

Can I tell other people about the study?
Yes you are welcome to tell other people about the study.

Yes, I wouldlike to take part

Contact Us
If you would like more information or have questions about this study, please contact: Steven Blunt, Email: blunts@cardiff.ac.uk, Telephone: 029 2074 8154

Division of Cancer & Genetics
School of Medicine
Cardiff University
Heath Park Campus

Clinical Genetics Research Study

Participant Consent Form

I would like to be a participant

Thank you! If you are happy to help with the study, please complete the consent form below.

If you would like to learn more about the study, or if you have any questions, please contact: Peter Grant, Email: grantp2@cardiff.ac.uk Telephone: 029 2074 4055

Title of Project: Developing a Questionnaire for the Evaluation of Clinical Genetics Services

Name of Researcher: Peter Grant **Please initial box**

1. I confirm that I have read and understand the information sheet for the above study and have had the opportunity to ask questions. ☐
2. I understand that taking part is voluntary and that I do not have to answer any questions do not wish to. I understand that I can leave the study at any time without explanation. ☐
3. I consent to my details and data being used by the research group, as laid out in the information sheet. I understand that data will not be shared beyond the research group. ☐
4. I agree to the interview being audio-recorded ☐
5. I agree to take part in the above study. ☐

_____	_____	_____
Name of Participant	Date	Signature
_____	_____	_____
Name of Person taking consent (if different from researcher)	Date	Signature
_____	_____	_____
Researcher	Date	Signature

Genetic Alliance UK
Wales Gene Park
Cardiff
CF14 4XN

Title: Developing a Questionnaire for the Evaluation of Clinical Genetics Services

Invitation

You are being invited to participate in a research study which aims to help improve the quality of care provided by clinical genetics services. Before you make a decision we would like you to understand why the research is being done and what your participation will involve. Please take the time to read the following information carefully and discuss with others if you wish. If you have any questions, or if you would like more information, please feel free to ask.

Why are we doing this study?

We want to develop a short questionnaire which can be used to evaluate genetic counselling and testing services. More specifically, we would like to hear feedback from patients, family members, and others so that we can build the best questionnaire possible. We want to help genetic counselling and testing services focus on what patients really want from their healthcare. The study will begin on May 1st 2017 and finish on January 1st 2018. It is an 'unfunded' study, forming part of the research of Cardiff University MPhil student Peter Grant.

Why have I been chosen?

You have been chosen because you are, or have a connection to, an individual with a genetic condition. We believe that you have the relevant knowledge and experience to help us in our research.

What will I have to do?

There will be one meeting lasting 30-50 minutes. You can come to us or we can come and meet you. Alternatively interviews can be carried out by phone or Skype. You will be asked to read through and answer five or six questions, each to do with clinical genetics, and discuss your thoughts about each one. We will then ask you some further questions about the questionnaire, and we encourage you to speak freely and honestly. With your consent, the meeting will be audio-recorded. Unfortunately, as this is an unfunded student research project, we cannot offer reimbursement for any travel costs incurred as part of the research study.

How much time will the study take?

The study will take 30-50 minutes in a single meeting.

Do I have to take part?

Taking part is entirely voluntary, and there will be no consequences if you decide not to take part.

Can I withdraw?

You are free to leave the study at any time, with no consequences.

Appendix C – Cognitive Interview PIS

Who is carrying out the study?

The study is being carried out by MPhil student Mr Peter Grant, and is supervised by Dr Marion McAllister (Senior Lecturer in Genetic Counselling, Cardiff University); Prof Angus Clarke (Clinical Professor in Medical Genetics, Cardiff University); Prof Katherine Payne (Professor in Health Economics, University of Manchester).

Will the study benefit or disadvantage me?

We do not expect that the study will harm or disadvantage you in any way. Also we do not guarantee that there will be any benefits to you from taking part in the study. You will not have to answer any questions which make you uncomfortable, and you can choose to leave the study at any time without giving a reason. We cannot guarantee compensation in the event of something going wrong, and unfortunately, as this is an unfunded student research project, we cannot offer reimbursement for any travel costs incurred as part of the research study.

Can I tell other people about the study?

Yes you are welcome to tell other people about the study.

What will happen to the results of the study?

The results will be used by the researchers to help develop the questionnaire, and for no other reason. Data will not be shared with anyone outside of the research study team. Your answers and comments will be analysed and compared to see if other people have said similar things. What you say may be used to re-word some of the questions to help make them more clear. All aspects of the study, including results, will be strictly confidential and only the researchers will have access to information on participants. If the results form part of a report, it will not be possible to identify individual participants in any way.

Will my personal information be confidential?

All data will be strictly confidential and only the researchers will have access to your personal information. Cardiff University has strict rules and standards on confidentiality, and these rules will be followed. The results from completed questionnaires will be made anonymous and all personal information removed. Likewise, your name will not be used and you will not be identifiable on any interview transcripts or interview excerpts in any publications arising from the research.

All personal information will be kept in a locked cabinet in the office of the Principal Investigator Dr Marion McAllister, located in the Institute of Medical Genetics, Cardiff University Heath Park Campus.

I have concerns about the project

For any concerns about how the research is being / was conducted, please contact the public engagement and policy officer for Genetic Alliance UK - Steven Blunt. Email: blunts@cardiff.ac.uk

Contact Us

If you would like more information or have questions about this study, please contact: Steven Blunt, Email: blunts@cardiff.ac.uk. Telephone: T: +44 (0)2920 748 154

Appendix D – Interview Guide

Interview Guide

Interviews will follow a semi-structured format. This guide provides an outline of discussion topics and will act as a reference to prompt conversation where necessary.

A non-directive approach shall be taken. The participant will be encouraged to 'think out loud', verbalising his/her thoughts as s/he answers survey questions. The interviewer will be primarily passive, providing prompts and encouragement where necessary and asking open-ended questions.

Introduction

- Welcome
- Brief explanation of research project
- Give participant GCOS-24, explain and ensure understanding
- Description of the interview format
- Hand participant PIS, review selected points
- Participant questions
- Signing of consent form
    - Permission for audio recording included

Recorded Interview

- Introductory questions (~5mins) e.g.:
    - Ask the participant to talk about themselves
    - Connection to a genetic disorder or clinical genetics.
    - Encourage to expand upon answers.
    - Give participant time to become comfortable and speak freely

- Ask participant to read a GCOS-24 item. Question interpretability and meaning e.g.:
    - Could you re-phrase the question in your own words?
    - Was this question hard to understand, if so, why?
    - How would you make the question more clear/easy to understand?
    - What does [XXX] mean to you?
- Prompt participant with open-ended questions e.g.:
    - Could you tell me more about [X]?

- General GCOS-24 Questions e.g.
    - What are your overall thoughts of the questionnaire?
    - Are there any questions which you feel don't fit in or seem different the others?
    - Is the layout / format of the questionnaire easy to understand?
    - Anything you would change to the questionnaire as a whole?
    - Did the questionnaire evoke any emotions?

- Debrief
    - Thank them for taking part
    - Any further questions

GCOS-24 questions will be addressed in a random order. Interviews should last around 45 minutes. A brief email of thanks will be sent to all participants.

Appendix E – Cognitive Interview Email of Thanks

Text copied from this letter of thanks will be used in email format.

Date: XXXX 2017

Dear XXXX,

Many thanks for your recent participation in our research study. We are immensely grateful to have members of the public such as yourself who are willing to give up their time to support local research. Your contributions and comments were most helpful and will assist us in improving our questionnaire.

Again I would like to pass on our thanks for taking part in our study, for your time and effort.

If you have any questions, or if I can help in any way, please feel free to get in touch.

Sincerely,

Peter Grant
Email: grantp2@cardiff.ac.uk T: +44 (0)2920 744 055

Appendix F – Test-retest Recruitment Email

Genetic Alliance UK
Supporting. Campaigning. Uniting.

Dear *|FNAME|* *|LNAME|*

Researchers from the Cardiff University School of Medicine would like to invite you to take part in a research project. You have been asked because either you or members of your family are affected by or at risk of a genetic condition. Because of this, we believe that you have the personal experience we are looking for to help with our research. The project will involve a short (less than 10 minutes) questionnaire, in which you will be asked about your experience of having a genetic condition in the family. In the long run, we hope to use your comments to improve the questionnaire, which in turn can be used to improve how clinical genetics services responds to patient needs.

I hope you will consider taking part in this important study.

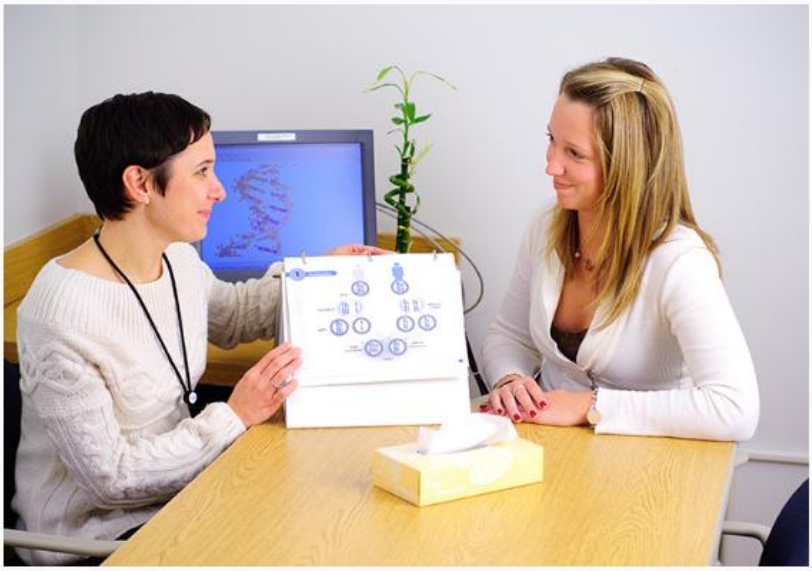
Yours Sincerely,

Steven Blunt
Public Engagement and Policy Officer - Wales
Genetic Alliance UK, Wales Gene Park, Cardiff University, Cardiff, CF14 4XN
E: blunts@cardiff.ac.uk T- +44 (0)29 2074 8154

Yes, I would like to take part

If you are interested in being part of our research, please click the box above. If you would like to find out more about the project, please continue reading.

Appendix F – Test-retest Recruitment Email



Project Title: Developing a Questionnaire for the Evaluation of Clinical Genetics Services

You are being invited to take part in a research study which aims to help improve the quality of care provided by clinical genetics services. Before you make a decision, we would like you to understand why the research is being done and how you would be involved. Please take the time to read the following information carefully, and you're welcome to talk to other people about it before you decide. If you have any questions, or if you would like more information, please feel free to get in touch using the contact details below.

Information

Why are we doing this study?
We want to develop a short questionnaire which can be used to evaluate genetic counselling and testing services. More specifically, we would like to hear from patients, family members, and others so that we can build the best questionnaire possible. We want to help genetic counselling and testing services focus on what patients really want from their healthcare. The study will begin on May 1st 2017 and finish on January 1st 2018. It is an 'unfunded' study, forming part of the research of Cardiff University MPhil student Peter Grant.

Information

Why have I been chosen?
You have been chosen because you are, or have a connection to, an individual with a genetic condition. We believe that you have the relevant knowledge and experience to help us in our research.

Do I have to take part?
Taking part is entirely voluntary, and there will be no consequences if you decide not to take part.

Can I withdraw?
You are free to leave at any time during the study, with no consequences.

Appendix F – Test-retest Recruitment Email

What will I have to do?
If you choose to click the link within this email, you will be taken to our survey which will first ask you for consent. After consent, the questionnaire has six questions to answer, each to do with having a genetic condition within the family. After two weeks, you will be sent the same questionnaire again, so we can compare your answers from two different times. This will give us an idea of how 'reliable' the questionnaire is. Please try to answer as honestly as you can.

How much time will the study take?
The questionnaire should take less than 10 minutes to answer. Around two weeks later, you will be sent the same questionnaire again.

Who is carrying out the study?
The study is being carried out by MPhil student Mr Peter Grant, and is supervised by Dr Marion McAllister (Senior Lecturer in Genetic Counselling, Cardiff University); Prof Angus Clarke (Clinical Professor in Medical Genetics, Cardiff University); Prof Katherine Payne (Professor in Health Economics, University of Manchester).

Will the study benefit or disadvantage me?
We do not expect that the study will harm or disadvantage you in any way. Also we do not guarantee that there will be any benefits to you from taking part in the study. You will not have to answer any questions which make you uncomfortable, and you can choose to leave the study at any time without giving a reason. We cannot guarantee compensation in the event of something going wrong.

What will happen to the results of the study?
The results will be used by the researchers to help develop the questionnaire, and for no other reason. Data will not be shared with anyone outside of the research study team. After you have completed the questionnaire for a second time, your answers will be analysed to see if they are the same or different. All aspects of the study, including results, will be strictly confidential and only the researchers will have access to information on participants. If the results form part of a report, it will not be possible to identify individual participants in any way.

Will my personal information be confidential?
All data will be strictly confidential and only the researchers will have access to your personal information. Cardiff University has strict rules and standards on confidentiality, and these rules will be followed. The results from completed questionnaires will be made anonymous and all personal information removed. Likewise, your name will not be used and you will not be identifiable on any interview transcripts or interview excerpts in any publications arising from the research.
All personal information will be kept in a locked cabinet in the office of the Principal Investigator Dr Marion McAllister, located in the Institute of Medical Genetics, Cardiff University Heath Park Campus.

Can I tell other people about the study?
Yes you are welcome to tell other people about the study.

I have ethical concerns about the project
For any concerns about how the research is being / was conducted, please contact the public engagement and policy officer for Genetic Alliance UK - Steven Blunt. Email - blunts@cardiff.ac.uk

Yes, I would like to take part

If you are happy to take part in this study, please click the button above to be taken to the consent form and survey.
Contact Us
If you would like more information or have questions about this study, pleae contact: Steven Blunt. Email: blunts@cardiff.ac.uk Tel: 02920 748 154

Appendix G – Test-retest Email of Thanks

Date: XXXX 2017

Dear XXXX,

Many thanks for your decision to take part in our research study. We are immensely grateful to have members of the public such as yourself who are willing to give up their time to support local research.

You are halfway there!

In two weeks you will be sent the same questionnaire again. This is the most important part of the study, because we are trying to judge whether people will give the same answers if they do it again. When the email comes through, we would greatly appreciate it if you took the time to fill out the questionnaire again.

Again I would like to pass on our thanks for taking part in our study, for your time and effort.

If you have any questions, or if I can help in any way, please feel free to get in touch.

Sincerely,

Peter Grant
Email: grantp2@cardiff.ac.uk T: +44 (0)2920 744 055

Appendix H – SMREC Ethics Approval

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

School of Medicine
Ysgol Meddygaeth

Cardiff University
Main Building
Heath Park
Cardiff CF14 4XN
Wales UK

Friday 28th April 2017

Peter Grant, Prof Angus Clarke,
Dr Marion McAllister, and Prof Katherine Payne
Institute of Medical Genetics,
School of Medicine,
Heath Park.

Prifysgol Caerdydd
Prif Adeilad
Parc y Mynydd Bychan
Caerdydd CF14 4XN
Cymru, Y Deyrnas Unedig

Dear Peter,

Re: Assessing the psychometric properties of a short-form Genetic Counselling Outcome

SMREC Reference Number: 17/37

This application was reviewed by the Committee on Friday 21st April 2017.

Ethical Opinion

On review, the Committee have asked for the following issues to be addressed:

1. Replace Dr Jonathan Hewitt's contact details on the Participant Information Sheet with an independent, impartial contact.
2. Remove any intention of sharing the data with anyone outside of the research study team. If the sharing of the data is essential to the study then please provide the Committee with justification as to why it has been included.
3. Provide clarification as to whether you will provide reimbursement for travel for any participants coming to the Heath Park campus.
4. For any participants taking part via Skype or the phone, the Committee would ask that you arrange for the participant to return a signed consent form. This can be undertaken by sending out Participant Information Sheets and Consent Forms with a stamped addressed envelope to potential recruits.
5. Provide further clarification as to how the two groups will be recruited to the study.

Once the issues above have been addressed, please resubmit your application via email to the Committee Secretary, Miss Claire Batten.

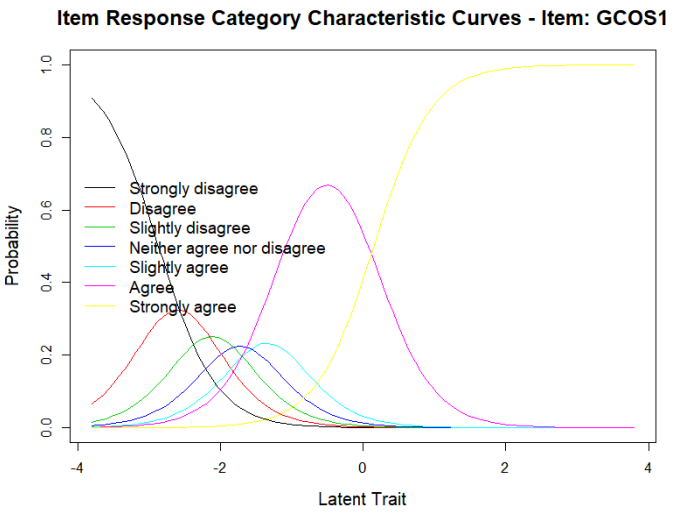
Documents Considered

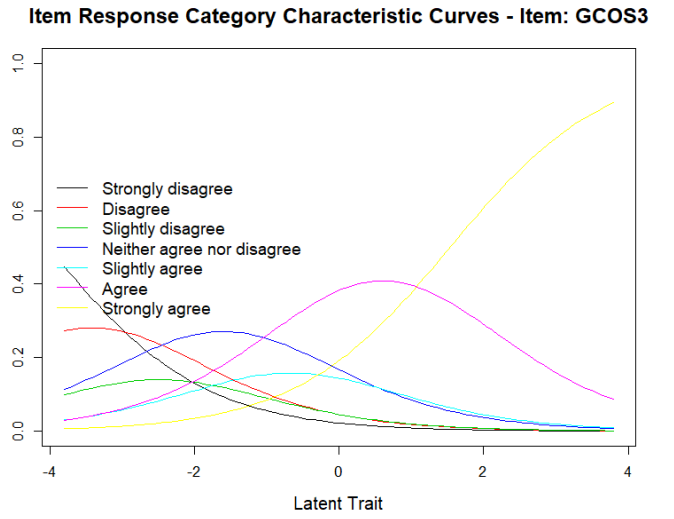
Document Type:	Version:	Date Considered:
Application	Signed 24/03/2017	21/04/2017
Research Proposal	V1.1 24/03/2017	21/04/2017
Appendix 1: Email Invite and Participant Information	V1.1 24/03/2017	21/04/2017
Appendix 2: Reply Email	V1.1 24/03/2017	21/04/2017
Appendix 3: Participant Consent Form	V1.1 24/03/2017	21/04/2017
Appendix 4: Invitation Letter	V1.1 24/03/2017	21/04/2017
Appendix 5: Participant Information Sheet	V1.1 24/03/2017	21/04/2017
Appendix 6: Letter of thanks	V1.1 24/03/2017	21/04/2017
Appendix 7: Outcome Scale	No date or version	21/04/2017
Appendix 8: Interview Guide	V1.0 24/03/2017	21/04/2017
Appendix 9: Email Invite and Participant Information	V1.1 24/03/2017	21/04/2017
Appendix 10: Halfway Letter	V1.1 24/03/2017	21/04/2017
Appendix 11: Halfway Email	V1.1 24/03/2017	21/04/2017
Appendix 12: Letter of thanks	V1.1 24/03/2017	21/04/2017

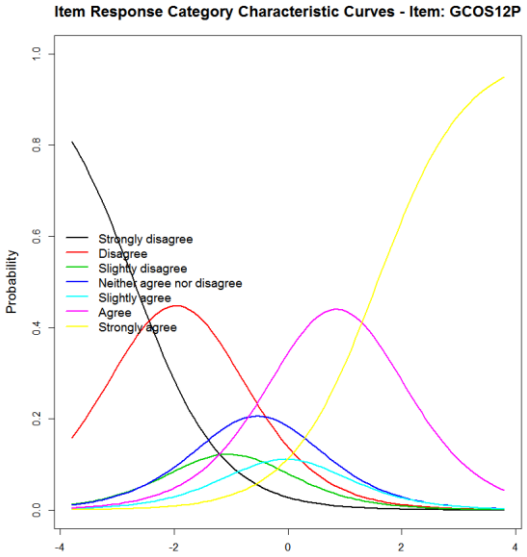
GCOS-24 Items: Item Characteristic Curves and Item Information Curves

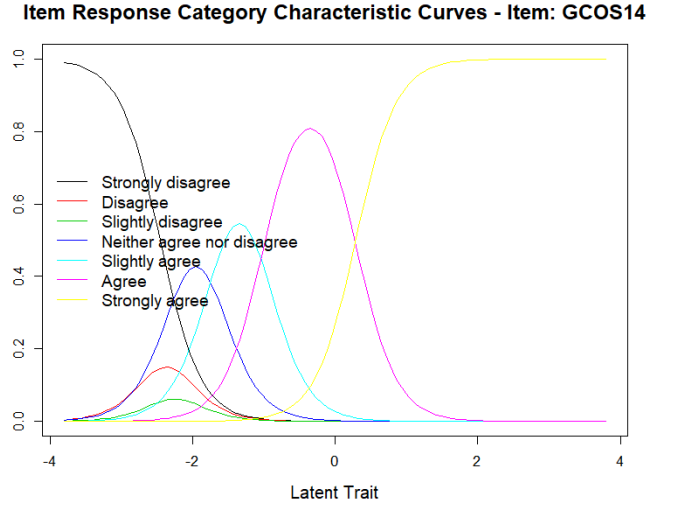
Item Characteristic Curves

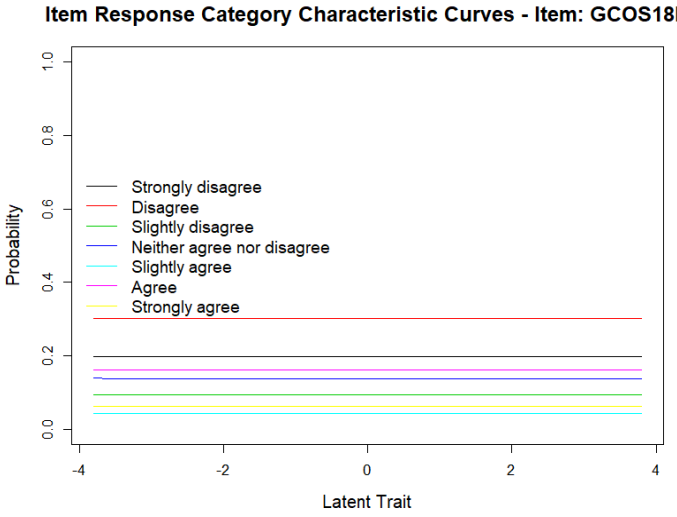
Cognitive Control: Items 1; 3; 12; 14; 18; 23

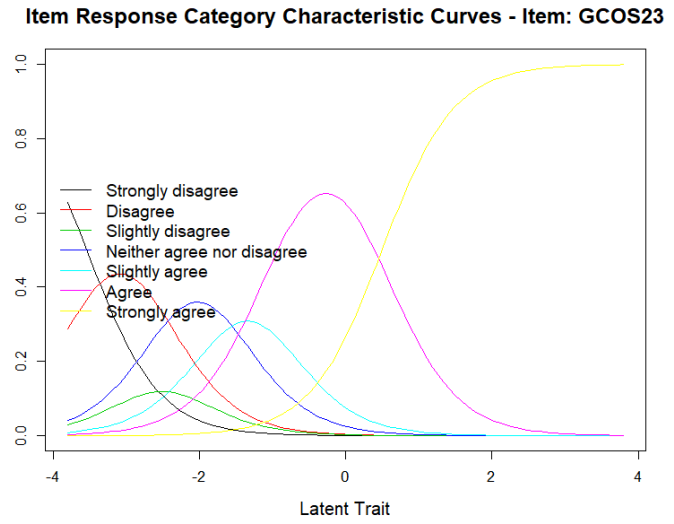




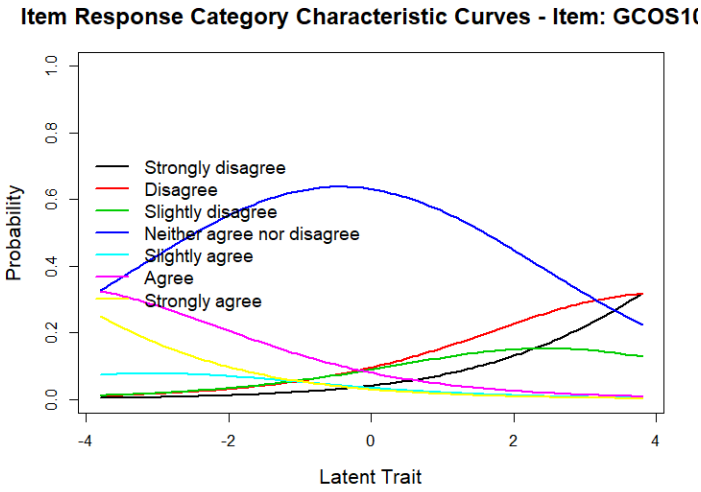


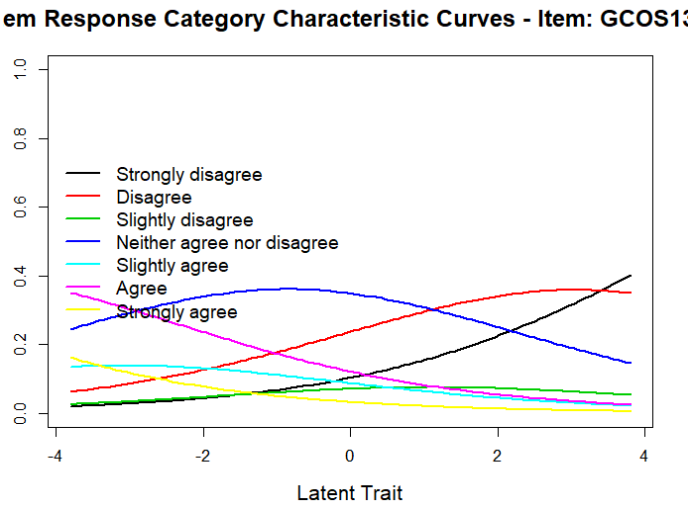


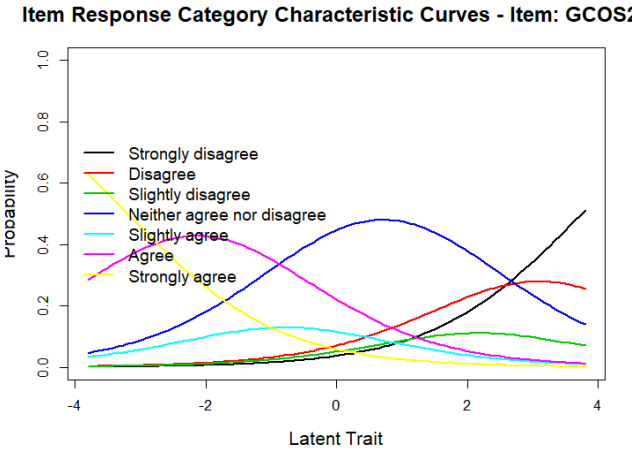




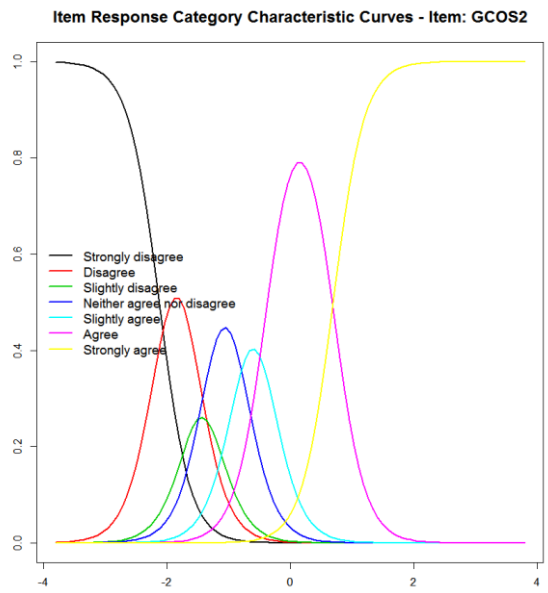
Decisional Control: Items 10; 13; 24

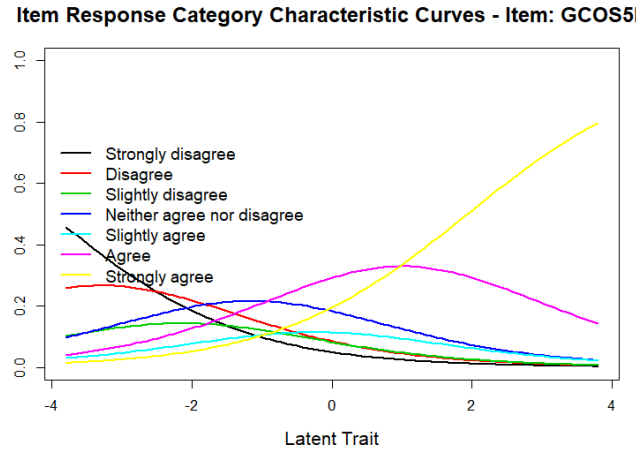




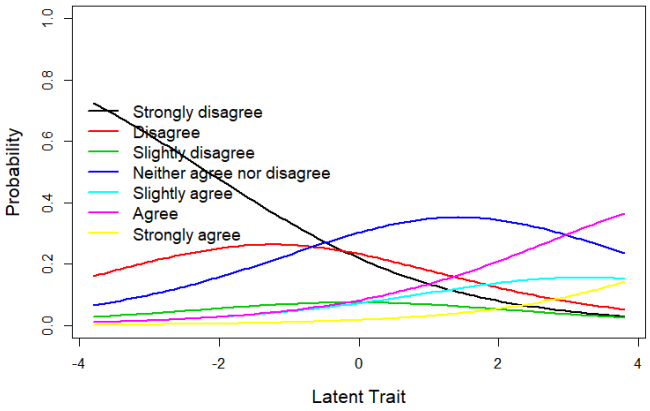


Behavioural Control: Items 2; 5; 7; 9; 15; 16; 17; 22

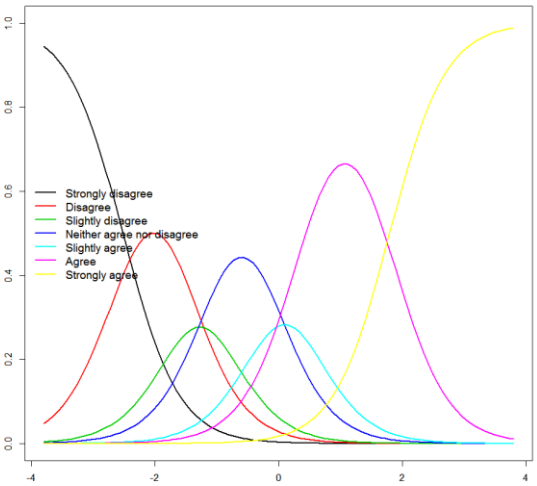




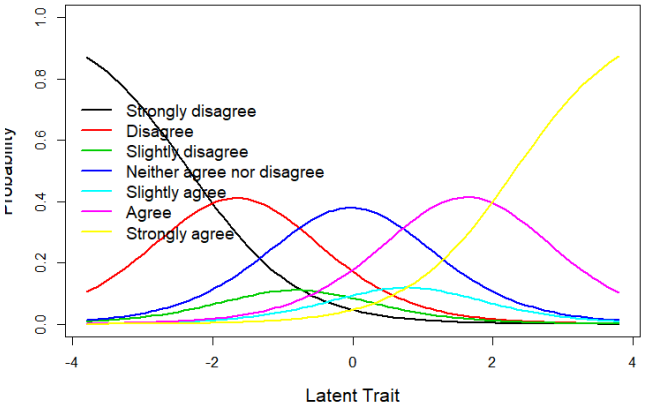
Item Response Category Characteristic Curves - Item: GCOS7



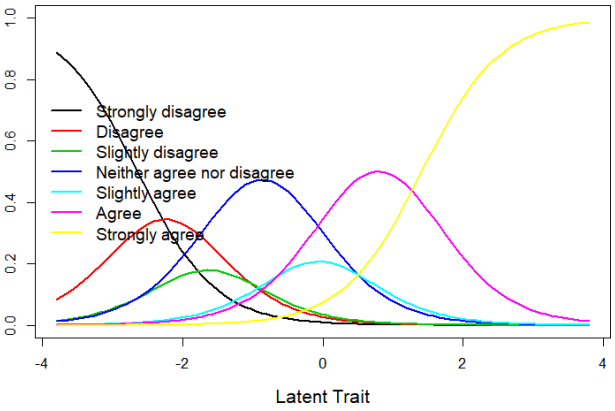
Item Response Category Characteristic Curves - Item: GCOS9



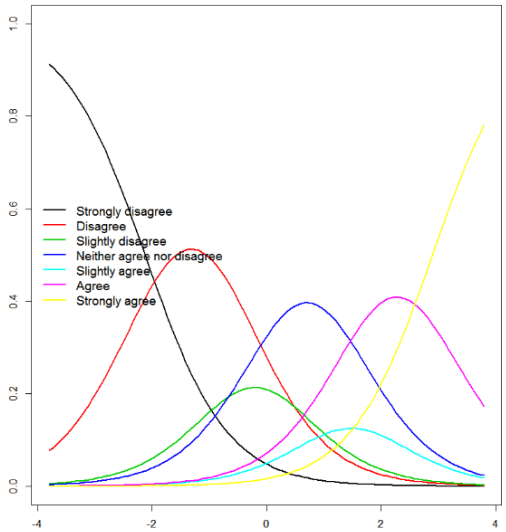
Item Response Category Characteristic Curves - Item: GCOS15



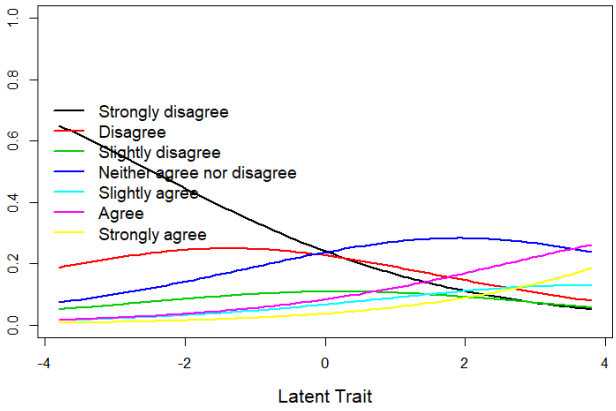
Item Response Category Characteristic Curves - Item: GCOS16



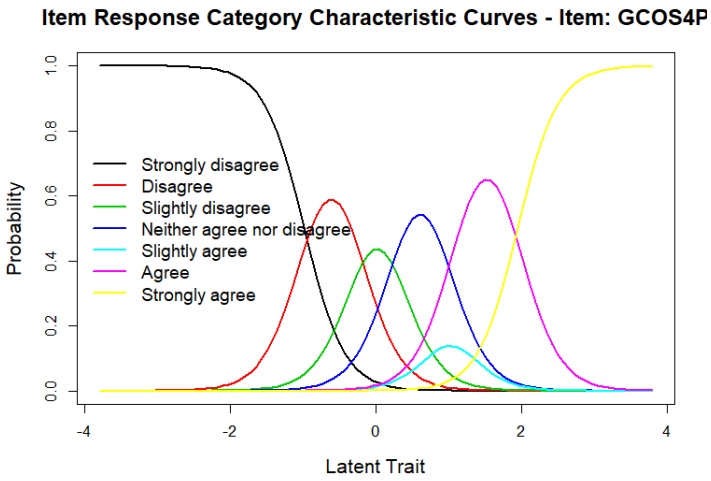
Item Response Category Characteristic Curves - Item: GCOS17P

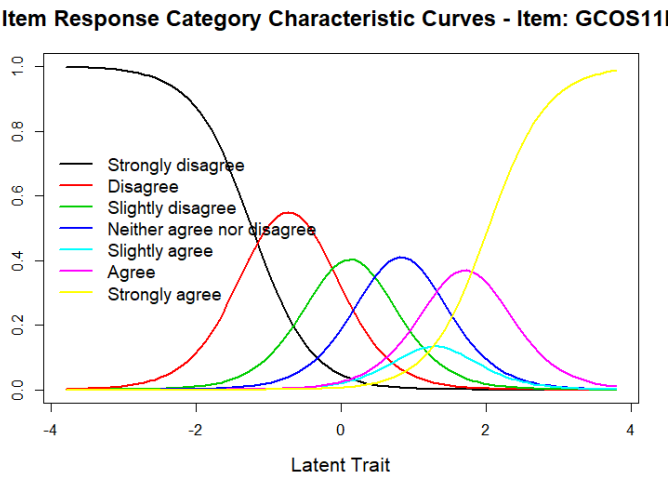


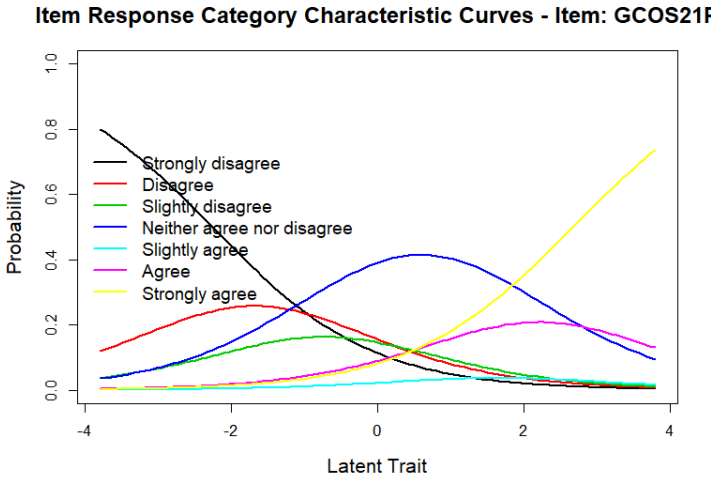
Item Response Category Characteristic Curves - Item: GCOS22P



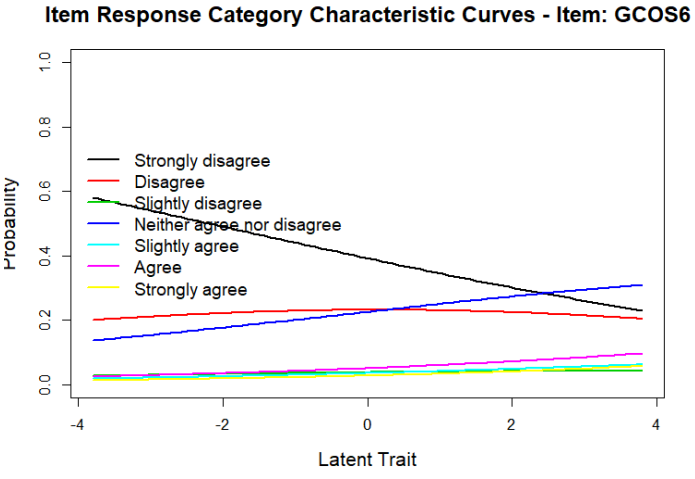
Emotional Regulation: Items 4; 11; 21

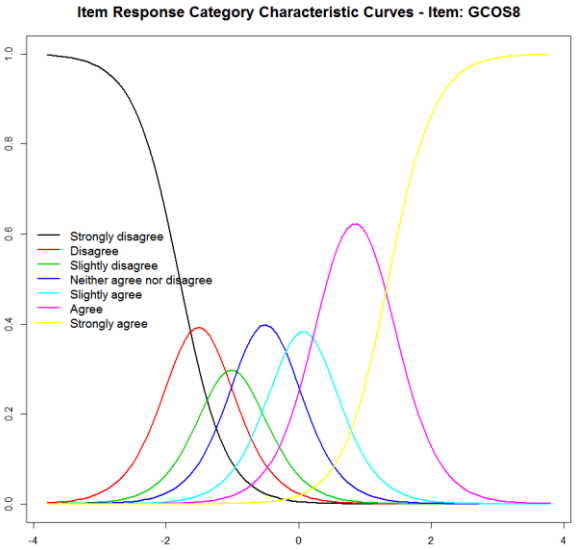


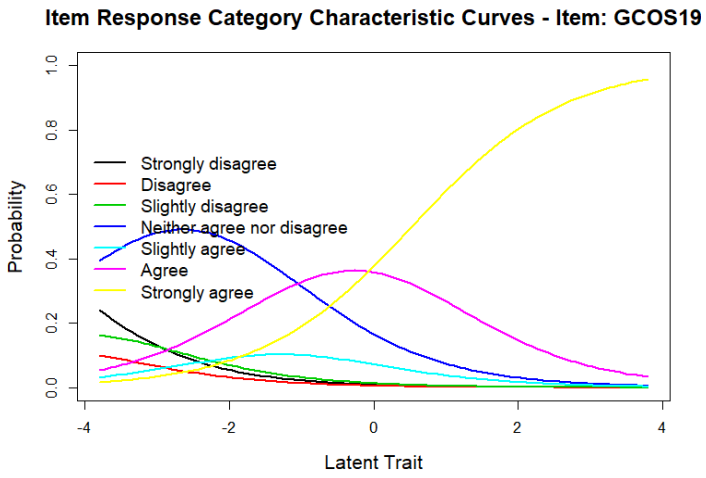


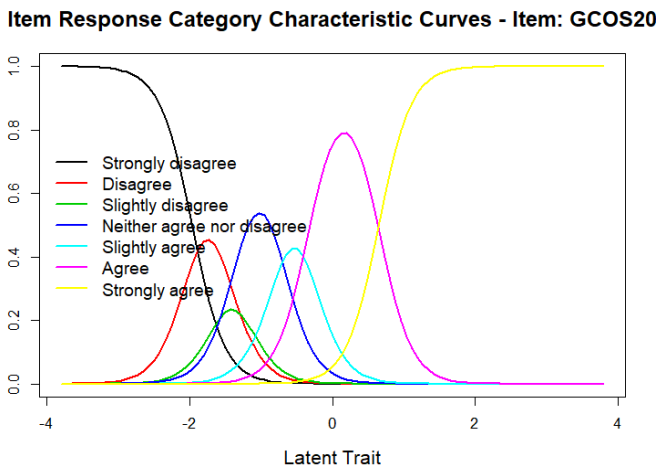


Hope: Items 6; 8; 19; 20



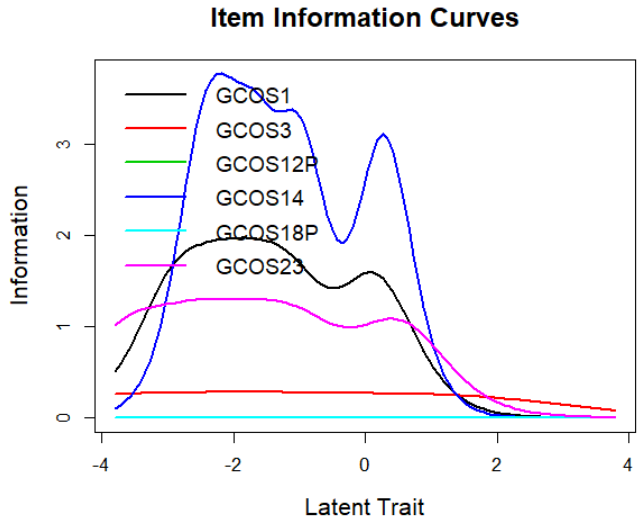




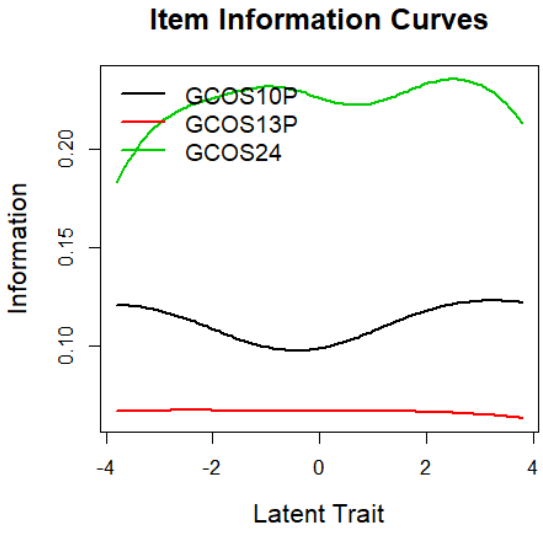


Item Information Curves

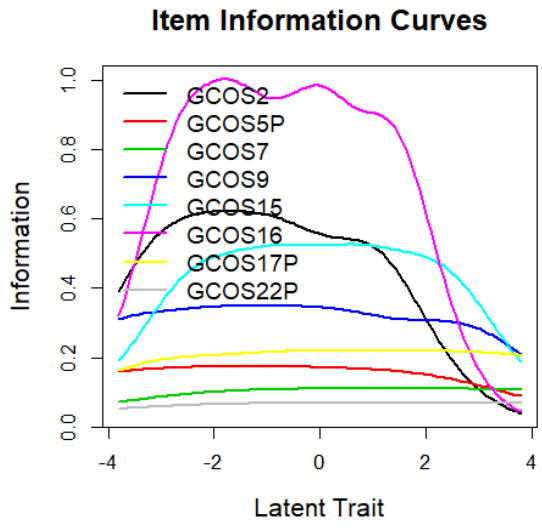
Cognitive Control



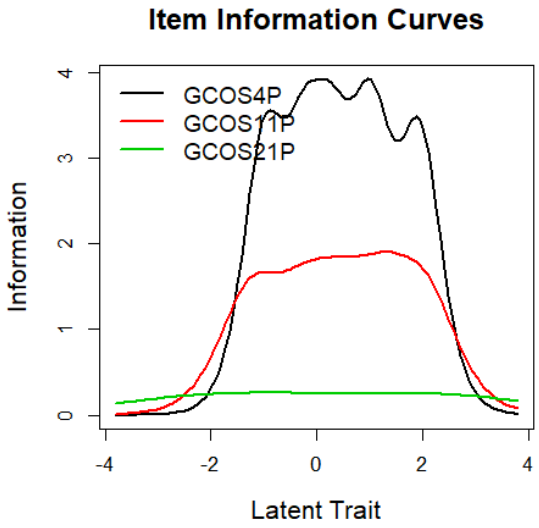
Decisional Control



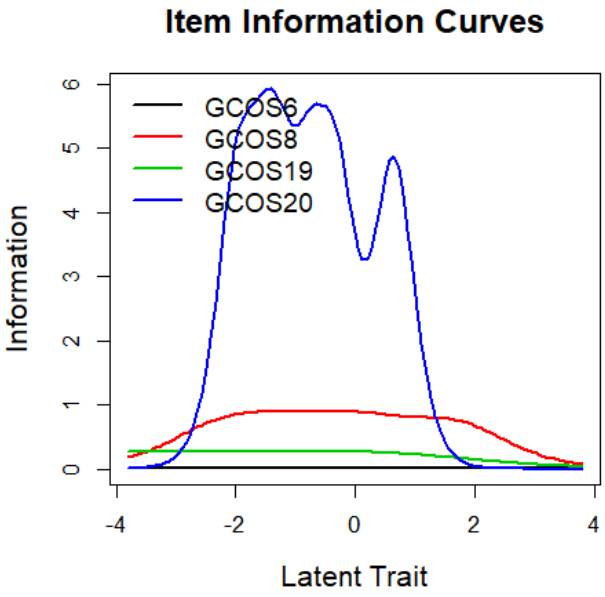
Behavioural Control



Emotional Regulation



Hope


Item Information Curves
GCOS6
GCOS8
GCOS19
GCOS20
Information
0
1
2
3
4
5
6
-4
-2
0
2
4
Latent Trait

Aaronson, N., Alonso, J., Burnam, A., Lohr, K., Patrick, D., Perrin, E., Stein, R. (2002) Assessing health status and quality if life instruments: attributes and review criteria. *Qual Life Res* **11(3)**: 193-205.

Andrea, E., Lagerberg, T., De Vito, C., Pitini, E., Marzuillo, C., *et al.* (2016) Patient Experience and utility of genetic information: a cross-sectional study among patients tested for cancer susceptibility. *Eur J Hum Gen* **26**: 518-526.

Andrich, D. (1978) A rating formulation for ordered response categories. *Psychometrika*, **43**, 561-73

Atieno, O. (2009) An Analysis of the Strengths and Limitation of Qualitative and Quantitative Research Paradigms. *Problems of Education in the 21st Century* **13**, 13-18.

Audrain J, Lerman C, Rimer B, Cella D, Steffens R, Gomez-Caminero A. (1995) Awareness of heightened breast cancer risk among first-degree relatives of recently diagnosed breast cancer patients. *Cancer Epidemiology, Biomarkers & Prevention* **4**:561-565.

Averill, J. (1973) Perceived control over aversive stimuli and its relationship to stress. *Psychological Bulletin* **80**:286-303.

Badia, X., Roset, M., Herdman, M., Kind, P. (2001) A comparison of UK and Spanish general population time trade-off values for EQ-5D health states. *Med Decis Making* **21(1)**: 7-16.

Barriball, K., While, A. (1994) Collecting data using a semi-structured interview: a discussion paper. *Journal of Advanced Nursing* **19**, 328-335.

Beatty, P., Willis, G. (2007) Research Synthesis: The practice of Cognitive Interviewing. *Public Opinion Quarterly* **71(2)**, 287-311.

Beeson, D. (1997) Nuance, complexity and context: qualitative methods in genetic counseling research. *Journal of genetic counselling* **6(1),** 21-43.

Benkendorf, J., Reutenauer, J., Hughes, C., Eads, N., Williason, J., Powers, M., Lerman, C. (1997) Paitents' Attitudes about Autonomy and Confidentiality in Genetic Testing for Breast-Ovarian Cancer Susceptibility. *AJMG* **73**:296-303.

Bennett, R., Hampdel, H., Mandell, J., Marks, J. (2003) Genetic counsellors: translating genomic science into clinical practice. *J Clin Invest* **112**: 1274-1279.

Bentler, M., Kano, Y. (1990) On the equivalence of factors and components. *Multivariate Behavioural Research* **25(1)**, 67-74.

Berkenstadt, M., Shiloh, S., Barkai, G., Katznelson, M., Goldman, B. (1999) Perceived personal control (PPC): a new concept in measuring outcome of genetic counseling. *Am J Med Genet* **82(1)**: 53-59.

Bernhardt, B., Biesecker, B., Mastromarino, C. (2000) Goals, benefits and outcomes of genetic counselling: Client and genetic counsellor assessment. *American Journal of Med Gen* **94**, 189-197.

Berrenberg, J. (1991) The Cancer Attitude Inventory: development and validation. *Journal of Psychosocial Oncology* **9(2)**:35-44.

Beukelman, T., Brunner, H. (2016) Trial Design, Measurement, and Analysis of Clinical Investigations: Textbook of Pediatric Rheumatology, 7th ed. *Elsevier*, 54-77.

Beulen, L., van den Berg, M., Faas, B., Feenstra, I., Hageman, M., van Vugt, J., Bekker, M. (2016) The effect of a decision aid on informed decision-making in the era of non-invaisve prenatal testing: a randomised controlled trial. *EJHG* **24**: 1409-1416.

Bjorvatn, C., Eide, G., Hanestad, B., Havik, O. (2008) Anxiety and depression among subjects attending genetic counselling for hereditary cancer. *Patient Educ Couns* **71(2)**: 234-243.

Bleiker, E., Aaronson, N., Menko, F., Hahn, D., van Asperen, C., Rutgers, E., ten Kate, L., Leschot, N. (1997) Genetic counseling for hereditary cancer: A pilot study on experiences of patients and family members. *Patient Educ Couns* **32**: 107-116.

Bock, R. (1972) Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika* **37**: 29-51.

Bowen, D., Burke, W., Yasui, Y., McTiernan, A., McLeran, D. (2002) Effects of risk counselling on interest in breast cancer genetic testing for lower risk women. *Genetics in Medicine* **4**:359-65.

Bowen, D., Burke, W., McTiernan, A., Yasui, Y., Andersen, M. (2004) Breast cancer risk counselling improves women's functioning. *Patient Educ Counsel* **53**: 79-86.

Bowen, G. (2008) Naturalistic enquiry and the saturation concept: A research note. *Qualitative research.* **8(1)** 137-152.

Bowen, D., Powers, D. (2010) Effects of a mail and telephone intervention on breast health behaviours. *Health Education & Behaviour* **37(4)**:479-489.

Brady M, Cella DF, Bonomi AE, Tulsky DS, Lloyd SR, Deasy S, *et al*. (1997) Reliability and validity of the functional assessment of cancer therapy-breast quality of life instrument. *Journal of Clinical Oncology* **15**:974-986.

Brain, K., Norman, P., Gray, J., Mansel, R. (1999) Anxiety and adherence to breast self-examination in women with a family history of breast cancer. Psychosomatic medicine. **61(2)**: 181-187.

Brain, K., Gray, J., Norman, P., France, E., Anglim, C., *et al.* (2000) Randomised trial of a specialist genetic assessment service for familial breast cancer. *Journal of the National Cancer Institute* **92(16)**: 1345-51.

Brain, K., Sivell, S., Bennert, K., Howell, L., France, L., Jordan, S., *et al.* (2005) An exploratory comparison of genetic counselling protocols for HNPCC *Clin Genet* **68**:255-61.

Braitman A, Antley R. (1978) Development of instruments to measure counselees' knowledge of Down syndrome. *Clinical Genetics* **13**:25-36.

Braun, V., Clarke, V. (2006) Using thematic analysis in psychology. *Qualitative Research in Psychology* **3(2)**: 77-101.

Brazier, J., Roberts, J., Deverill, M. (2002) The estimation of a prefence-based measure of health from the SF-36. *Journal of Health Economics* **21(2)**: 271-292.

Brooks, R. (1996) EuroQol: the current state of play. *Health Policy* **37(1)**: 53-72.

Carlson, J., Henrikson, N., Veenstra, D., Ramsey, S. (2005) Economic analyses of human genetics services: a systematic review. *Genet Med* **7(8)**:519-23.

Carver C, Pozo-Kaderman C, Harris S, Noriega V, Scheier M, Robinson D, et al. (1994) Optimism versus pessimism predicts the quality of women's adjustment to early stage breast cancer. *Cancer* **73**:1213-1120.

Cattell, R. (1966) The Scree test for the number of factors. *Multivariate Behavioural Research* **1(2)**, 245-276.

Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, *et al*. (1993) The functional assessment of cancer therapy scale: development and validation of the general measure. *Journal of Clinical Oncology* **11**:570-579.

Cella D, Hughes C, Peterman A, Chang C, Peshkin B, Schwartz M, *et al.* (2002) Brief assessment of concerns with genetic testing for cancer: The Multidimensional Impact of Cancer Risk Assessment (MICRA) Questionnaire. *Health Psychology* **21(6):**564-572.

Charters, E. (2003) The use of think-aloud methods in qualitative research: An introduction to think-aloud methods. *Brock Education Journal* **12(2):** 68-82.

Cho, A., Killeya-Jones, L., O'Daniel, J., Kawamoto, K., Gallagher, P., Haga, S., *et al.* (2012) Effect of genetic testing for risk of type 2 diabetes mellitus on health behaviours and outcomes. *BMC Health Services Research* **12**:16.

Ciarleglio, L., Bennett, R., Williamson, J., Mandell, J., Marks, J. (2003) Genetic counseling throughout the life cycle. *J Clin Invest* **112**: 1280-1286.

Clarke, A., Parsons, E., Williams, A. (1996) Outcomes and process in genetic counselling. *Clinical Genetics* **50,** 462-469.

Cliff, N. (1988) The eigenvalues-greater-than-one rule and the reliability of components. *Psychological Bulletin* **103(2),** 276-279.

Cockburn J, De Luise T, Hurley S, Clover K. (1992) Development and validation of the PCQ: a questionnaire to measure the psychological consequences of screening mammography. *Social Science and Medicine* **34(10):** 1129-1134.

Costal-Tirado, A., McDermott, A., Thomas, C., Ferrick, D., Harris, J., Edwards, A., McAllister, M. (2017) Using patient-reported outcome measures for quality improvement in clinical genetics: an exploratory study. *Journal of Genetic Counseling*, 1-12.

Costello, A., Osborne, J. (2005) Best Practice in Exploratory Factor Analysis. *Practical Assessment, Research & Evaluation* **10(7)**.

Crawford, C., Koopman, P. (1979) Inter-rater reliability of scree test and mean square ration test of number of factors. *Perceptual and Motor Skills* **49**, 223-226.

Curbow B, Somerfield M. (1991) Use of the Rosenberg self-esteem scale with adult cancer patients. Journal of Psychological Oncology **9(2):**113-131.

D'Andrea, E., Marzuillo, C., Pelone, F., De Vito, C., Villari, P. (2015) Genetic testing and economic evaluations: a systematic review of the literature. *Epidem Prev* **39(4)**: 45-50.

Dazord A. (1995) Quality of life assessment in medicine: presentation of a new instrument (Subjective Quality of Life Profile (SQLP)). *European Respiratory Review* **5(25):** 66-71.

DeMarco, T., Peshkin, B., Mars, B., Tercyak, K. (2004) Patient satisfaction with cancer genetic counselling: a psychometric analysis of the genetic counseling satisfaction scale. *J Genet Couns* **13(4)**: 293-304.

Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. (1974) The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behavioural Science* **19**:1-5.

Derogatis, L., Melisaratos, N. (1983) The Brief Symptom Inventory: an introductory report. *Psychological Medicine* **13(3):** 595-605.

DeVellis, R. (2011) Scale Development: Theory and Applications. SAGE.

Devlin, N., Appleby, J. (2010) Getting the most out of PROMS. The King's Fund. London.

Dey, I. (1993) Qualitative Data Analysis: A user friendly guide for social scientists. Routledge: London.

Diness, B., Overbeck, G., Hjortshoj, T., Hammer, T., Timshel, S., Sorensen, E., McAllister, M. (2017) Translation and Adaptation of the Genetic Counselling Outcome Scale (GCOS-24) for use in Denmark. *J Genet Couns* **26(5)**: 1080-1089.

Dinno, A. (2009) Exploring the sensitivity of Horn's Parallel Analysis to the distributional form of random data. *Multivariate Behavioural Research* **44(3)**, 362-388.

Djalalov, S., Musa, Z., Mendelson, M., Siminovitch, K., Hoch, J. (2011) A review of economic evaluations of genetic testing services and interventions. *Genet Med* **13(2)**: 89-94.

DoH Department of Health (1989) *Working for Patients*. Department of Health, London.

DoH Department of Health (2008) *High quality care for all. NHS next stage review final report.* Department of Health, London.

Donovan, K., Tucker, D. (2000) Knowledge about genetic risk for breast cancer and perceptions of genetic testing in a sociodemographically diverse sample. *J Behav Med* **23(1)**: 15-36.

Doukas D, Li Y. (2004) Men's values-based factors on prostate cancer risk genetic testing: a telephone survey. *BMC Medical Genetics* **5**:28-35.

Duric, V., Butow, P., Sharpe, L., Lobb, E., Meiser, B., Barratt, A., Tucker, K. (2003) Reducing psychological distress in a genetic counseling consultation for breast cancer. *J Genet Couns* **12(3)**: 243-264.

Edwards, A., Gray, J., Clarke, A., Dundon, J., Elwyn, G., Gaff, C., Hood, K., Iredale, R., Sivell, S., Shaw, C., Thornton, H. (2008) Interventions to improve risk communication in clinical genetics: systematic review. *Patient Educ Couns* **71(1)**: 4-25.

Ellison C, Smith J. (1991) Towards an integrative measure of health and well-being. *Journal of Psychology and Theology* **19(1):**35-48.

Embretson, S. (1996) The new rules of measurement. *Psychol Assess* **8(4)**, 341-349.

Erblich, J., Brown, K., Kim, Y., Valdimarsdottir, H., Livingston, B., Bovbjerg, D. (2005) Development and validation of a breast cancer genetic counseling knowledge questionnaire. *Patient Education and Counselling* **56**: 182-191.

Ericsson, K., Simon, H. (1980) Verbal reports as data. *Psychological Review* **87,** 215-250.

Fabrigar, R., Wegener, T., MacCallum, C., Strahan, J. (1999) Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods* **4(3)**, 272-299.

Fathzadeh, M., Babaie, B., Bazrgar, M., Yavarian, M., Tabatabaee, H., Akrami, S. (2008) Genetic counseling in southern Iran: consanguinity and reason for referral. *J Genet Couns* **17(5)**: 472-479.

FDA Food and Drug Administration. Guidance for industry patient-related outcome measures: use in medical product development to support labelling claims. Washington: US. DoHaH FDA.

Felselstein, J. (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* **17(6)**, 368-376.

Field, A. (2013) Discovering Statistics Using SPSS (4th Ed)

Fink, A. (2016) *How to Conduct Surveys: A step by step guide.* 6 arg. Los Angeles: SAGE, 197.

Fitts W. (1965) Tennessee Self-Concept Scale manual. Counsellor recordings and tests. Western Psychological Services

Floyd, J., Widaman, F. (1995) Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment* **7(3)**, 286-299.

Ford, K., MacCallum, C., Tait, M. (1986) The application of exploratory factor-analysis in applied psychology – a critical review and analysis. *Personnel Psychology* **39(2)**, 291-314.

Frost, M., Reeve, B., Liepa, A., Stauffer, J., Hays, R. (2007) What is sufficient evidence for the reliability and validity of patient reported outcome measures? *Value in Health* **10(2)**, 94-105.

Frost, M., Reeve, B., Liepa, A., Stauffer, J., Hays, R. (2007) What is sufficient evidence for the reliability and validity of patient reported outcome measures? *Value in Health* **10(2)**, 94-105.

Fryback, D., Palta, M., Cherepanov, D., Bolt, D., Kim, J. (2010) Comparison of 5 Health-Related Quality of Life Indexes Using Item Response Theory Analysis. *Medical Decision Making* **30(1)**: 5-15.

Furr LA, Kelly SE, Furr LA, Kelly SE. (1999) The Genetic Knowledge Index: developing a standard measure of genetic knowledge. *Genetic Testing* **3**:193-9.

Gale, N., Heath, G., Cameron, E., Rashid, S., Redwood, S. (2013) Using the framework method for the analysis of qualitative data in multi-disciplinary health research. *BMC Medical Research* **13:** 117.

Geirdal, A., Reichelt, J., Dahl, A., Heimdal, K., *et al.* (2005) Psychological distress in women at risk of hereditary breast/ovarian or HNPCC cancers in the absence of demonstrated mutations. *Familial Cancer* **4(2)**: 121-126.

Geller G, Tambor E, Chase G, Holtzman N. (1993) Measuring physicians' tolerance for ambiguity and its relationship to their reported practices regarding genetic testing. *Medical Care* **31(11)**:989-1001.

Gibbons, R., Perraillon, M., Kim, J. (2014) Item Response Theory Approaches to Harmonisation and Research Synthesis. *HSORM* **14(4)**: 213-231.

Gill, P., Stewart, K., Treasure, E., Chadwick, B. (2008) Methods of data collection in qualitative research: interviews and focus groups. *BDJ* **204**, 291-295.

Gioiella M, Berkman B, Robinson M. (1998) Spirituality and quality of life in gynecologic oncology patients. *Cancer Practice* **6(6):**333-38.

Glaser, B., Strauss, A. (1967) The Discovery of Grounded Theory. Aldine Publishing. Hawthorne NY

Goel V, Glazier R, Holzapfel S, Pugh P, Summers A. (1996) Evaluating patient's knowledge of maternal serum screening. Prenatal Diagnosis **16**: 425-430.

Goldberg D, Hillier V. (1979) A scaled version of the general health questionnaire. *Psychological Medicine* **9**:139-145.

Goldberg D, Williams P. (1988) A users guide to the GHQ. Berkshire: NFER Nelson Publishing.

Gorsuch, R. (1990) Common Factor Analysis versus Component Analysis. *Multivariate Behavioural Research* **25(1)**, 33-39.

Gorsuch, R. (1997) Exploratory factor analysis: Its roel in item analysis. *Journal of Personality Assessment*, **68(3)**, 532-560.

Goudarzi, R., Zeraati, H., Sari, A., Rashidian, A., Mohammad, K. (2016) Population-based preference weights for the EQ-5D health states using the VAS in Iran. *IRCMJ* **18(2)**

Graham, J. (1987) The MMPI. A practical guide. Oxford: Oxford University Press.

Guest, G., Bunce, A., Johnson, L. (2006) How Many Interviews are Enough? *Field Methods* **18(1)**: 59-82.

Gulliksen, H. (1950) Theory of mental tests. Wiley: NY.

Hailey, B., Carter, C., Burnett, D. (2000) Breast cancer attitudes, knowledge, and screening behaviour in women with and without a family history of breast cancer. *Health Care Women International* **21**:701-15.

Halvorsen JG. (1991) Self-report family assessment instruments: an evaluative review. *The Family Practise Research Journal* **11(1):**21-55.

Hambleton, R., Swaminathan, H., Rogers, W. (1991) Fundamentals of item response theory. Newbury Park: Sage Publications.

Hambleton, R., Jones, R. (1993) Comparison of classical test theory and item response theory and their applications to test development. *Instructional Topics in Educational Measurement*, 38-47.

Hambleton, R. (2000) Emergence of item response modelling in instrument development and data analysis. *Med Care* **38(9)** II60-5.

Hamilton, J., Lobel, M., Moyer, A. (2009) Emotional distress following genetic testing for hereditary breast and ovarian cancer: a meta-analytic review. *Health Psychol* **28(4)**: 510-518.

Harding, M. (2016) Genetic Counselling – a guide for GPs. *Patient*

Hays, R., Morales, L., Reise, S. (2000) Item Response Theory and health outcomes measurement in the 21st century *Med Care* **38(9)**: 28-42.

Hilgart, J., Hayward, J., Coles, B., Iredale, R. (2012) Telegenetics: a systematic review of telemedicine in genetics services. *Genet Med* **14(9)**:765-776.

Holmes-Rovner, M., Kroll, J., Schmitt, N., Rovner, D., Breer, L., Rothert, M., Faan, R. *et al.* (1996) Patient Satisfaction with Health Care Decisions: The satisfaction with decision scale. *Med Decis Making* **16**: 58-64

Horn, J. (1965) A rationale and test for the number of factors in factor analysis. *Psychometrika*, **30**, 179-185.

Horowitz FW, Wilner N, Alvarez W. (1979) Impact of event scale: a measure of subjective stress. *Psychosomatic Medicine* **41(3):**209-18.

Hough, M. (2002) Person-centred counselling. A practical approach to counselling. Pearson Education: Harlow: 51-71.

Hubalek, M., Sztankay, M., Meraner, V., Martini, C., Sperner-Unterweger, B., *et al.* (2016) Long-term psychosocial consequences and counsellees' satisfaction after genetic counselling for hereditary breats and ovarian cancer. *Cancer Research* **76(4)**: 9-22

Hughes C, Gomez-Caminero G, Benkendorf J, Kerner J, Isaacs C, Barter J, Lerman C. (1997) Ethnic differences in knowledge and attitudes about BRCA1 testing in women at increased risk. *Patient Education & Counselling* **32**:51-62.

IBM Corp. (2015) IBM SPSS Staitistics for Windows, Version 23.0. Armonk, NY: IBM.

Irwin, D., Varni, J., Yeatts, K., DeWalt, D. (2009) Cognitive interviewing methodology in the development of a pediatric item bank. *Health Qual Life Outcomes* **7(3)**.

Jenkinson C, Layte R, Wright L, Coulter A. (1996) The UK SF-36: an analysis and interpretation manual. Health Services Research Unit, University of Oxford.

Johnsen, L., Hellum, C., Nygaard, O., Storheim, K., *et al.* (2013) Comparison of the SF6D, the EQ5D, and the oswestry disability index in patients with chronic low back pain and degenerative disc disease. *BMC Musculoskeletal Disorders* **14**: 148

Judge, A., Arden, N., Kiran, A., Price, A., Javaid, M. *et al.* (2012) Interpretation of patient-reported outcomes for hip and knee replacement surgery: identification of thresholds associated with satisfaction with surgery. *J Bone Joint Surg* **94**: 412-418.

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

Kash KM, Holland JC, Halper MS, Miller DG. (1992) Psychological distress and surveillance behaviours of women with a family history of breast cancer. *Journal of the National Cancer Institute* **84**:24-30.

Kausmeyer, D., Lengerich, E., Kluhsman, B., Morrone, D., Harper, G., Baker, M. (2006) A survey of patients' experiences with the cancer genetic counselling process: recommendations for cancer genetics programs. *Journal of Genetic Counselling* **15(6)**: 409-431.

Kent, G., Howie, H., Fletcher, M., Newbury-Ecob, R., Hosie, K. (2010) The relationship between perceived risk, thought intrusiveness and emotional well being in women receiving counselling for breast cancer risk. *BJHP* **5(1)**

Kerr, C., (2010) Assessing and demonstrating data saturation in qualitative inquire supporting patient-reported outcomes research. *Expert review of Pharmacoeconomics & Outcomes Research.* **(10) 3**, 269-281.

Kipling, W. (2008)

Lamers, L., McDonnell, J., Stalmeier, P., Krabbe, P., Busschbach, J. (2006) The dutch tariff: results and arguments for an effective design for national EQ-5D valuation studies. *Health Econ* **15(10)**: 1121-32.

Lance, C., Butts, M., Michels, L. (2006) The sources of four commonly reported cutoff criteria. *Organisational Research Methods* **9**, 202-220.

Lee, H., Hung, M., Hu, F., Chang, Y., Hsieh, C., Wang, J. (2013) Estimating quality weights for EQ-5D health states with the time trade off method in Taiwan. *JFMA* **112(11)**: 699-706.

Légaré, F., Robitaille, H., Gane, C., Hébert, J., Labrecque, M., Rousseau, F. (2016) Improving Decision making about genetic testing in the clinic: an overview of effective knowledge translation intervention. *PLoS One* **11(3)**

Lerman C, Trock B, Rimer BK, Boyce A, Jepson C, Engstrom PF. (1991) Psychological and behavioural implications of abnormal mammograms. *Annals of Internal Medicine* **114**:657-661.

Lerman C, Seay J, Balshem A, Audrain J. (1995) Interest in genetic testing among first-degree relatives of breast cancer patients. *American Journal of Medical Genetics* **57**:385-392.

Lerman C, Narod S, Schulman K, Hugghes C, Gomez-Caminero A, Bonney, *et al.* (1996) BRCA1 testing in families with hereditary breast-ovarian cancer: A prospective study of patient decision making and outcomes. *JAMA* **275(24)**: 1885-1892.

Lerman, C., Biesecker, B., Benkendorf, J., Kerner, J., Gomez-Caminero, A., Hughes, C., Reed, M. (1997) Controlled trial of pretest education approaches to enhance informed decision-making for BRCA1 gene testing. *JNCI* **89**: 148-157.

Linacre, J. (2018) Winsteps® Rasch measurement computer program. Beaverton, Oregon.

Lo, C., Martindale, J., Hadjivassiliou, M., Martin, P., Dalton, A., Bandmann, O. (2014) The Documentation of Consent and Disclosure of Neurogenetic Testing Outside Clinical Genetics. *Neurogenetics* **15(1)**: 19-21.

Lodder, L., Frets, P., Trijsburg, R., Tibben, A., Meijers-Heijboer, E., Duivenvoorden, H., et *al.* (2002) Men at risk of being a mutation carrier for hereditary breast/ovarian cancer: an exploration of attitudes and psychological functioning during genetic testing. *European Journal of Human Genetics* **9**:492-500.

Loehlin, C. (1990) Component analysis versus common factor analysis – a case of disputed authorship. *Multivariate Behavioural Research* **25(1)**, 29-31.

Longhurst, R. (2003) Semi-structured interviews and Focus Groups. *Key Methods in Geography.* CA: SAGE, 117-132.

MacCallum, C., Tucker, R. (1991) Representing sources of error in the common factor model – implications for theory and practice. *Psychological Bulletin* **109(3)**, 502-511.

Macleod, R., Craufurd, D., Booth, K. (2002) Patients' Perceptions of what makes Genetic Counselling effective: An interpretative phenomenological analysis. *J Health Psy* **7(2)**: 145-156.

Malterud, K., Siersma, V., Guassora, A. (2015) Sample size in qualitative interview studies. *Qualitative Health Research* **26(13)**, 1753-1760.

Marteau T, Bekker H. (1996) The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology* **31**:301-306.

Masters, G. (1982) A Rasch model for partial credit scoring. *Psychometrika*, **47**, 149-74.

Matloff, E. (1994) Practice variability in prenatal genetic counseling. *J Genet Couns* **3(3)**: 215-231.

McAllister, M., Dearing, A. (2015) Patient reported outcomes and patient empowerment in clinical genetic services. *Clin Genet* **88**, 114-121.

McAllister, M., Payne, K., Macleod, R., Nicholls, S., Donnai, D., Davies, L. (2008) Patient empowerment in clinical genetics services. *J Health Psychology* **13(7)**, 895-905.

McAllister, M., Dunn, G., Todd, C. (2011a) Empowerment: qualitative underpinning of a new clinical genetics-specific patient-reported outcome. *Eur J Hum Gen* **19(2)**, 125-130.

McAllister, M., Wood, A., Dunn, G., Shiloh, S., Todd, C. (2011b) The Genetic Counseling Outcome Scale: a new patient-reported outcome measure for clinical genetics services. *Clin Genet*, **79**, 413-424.

McAllister, M., Wood, A., Dunn, G., Shiloh, S., Todd, C. (2012) The perceived personal control (PPC) questionnaire: Reliability and validity in a sample from the United Kingdom. *Am J Med Genet* **158**: 367-372.

McArdle, J. (1990) Principles versus principals of structural factor analyses. *Multivariate Behavioural Research* **25(1)**, 81-87.

McCarthy-Veach, P., LeRoy, B., Bartels, D. (2003) Facilitating the Genetic counseling process: A practice manual. Springer NY.

McNair D, Lorr M, Droppleman L. (1981) Manual for the Profile of Mood States. Educational and Industrial testing Service (EdITS).

McRae, R., Kurtz, J., Yamagata, S., Terracciano, A. (2011) Internal consistency, retest reliability, and their implications for personality scale validity. *Pers Soc Psychol Rev* **15(1)**, 28-50.

Meiser, B., Schofield, P., Trevena, L., Wilde, A., Barlow-Stewart, K. *et al.* (2013) Cluster randomised trial of a psycho-educational intervention for people with a family history of depression for use in general practice. *BMC Psychiatry* **13**: 325.

Merenda, P. (1997) A guide to the proper use of factor analysis in the conduct and reporting of research: pitfalls to avoid. *MECD* **30(3)**, 156-164.

Mesters I, van den Borne H, McCormick L, Pruyn J, de Boer M, Imbos T. (1997) Openness to discuss cancer in the nuclear family: scale, development and validation. *Pychosomatic Medicine* **59**:269-279.

Metcalfe, S. (2018) Genetic counselling, patient education, and informed decision-making in the genomic era. *Semin Fetal Neonatal Med* **23(2)**: 142-149.

Michie, S., Axworthy, D., Weinman, J., Marteau, T. (1996) Genetic counselling: predicting patient outcomes. *Psychology and Health*, **11(6)**, 797-809.

Michie S, Smith D, McClennan A, Marteau TM. (1997) Patient decision making: an evaluation of two different methods of presenting information about a screening test. *British Journal of Health Psychology* **2**:317-326.

Middleton, A., Hall, G., Patch, C. (2015) Genetic counsellors and genomic counseling in the UK. *Mol Genet Gen Med* **3(2)**:79-83.

Miller S. (1987) Monitoring and blunting: validation of a questionnaire to assess styles of information seeking under threat. *Journal of Personality and Social Psychology* **52(2):**345-353.

Mohanty, D., Das, K. (2011) Genetic counselling in tribals in India. *IJMR* **134(4)**:561-571.

Mokkink, L., Terwee, C., Patrick, D., Alonso, J., Stratford, P. *et al.* (2010) The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* **19(4)**: 539-549.

Moos RH, Moos BS. (1994) Family Environment Scale Manual. Consulting Psychologists Press.

Mulaik, S., (1990) Blurring the Distinctions between component analysis and common factor analysis. *Multivariate Behavioural Research* **25(1)**, 53-59.

Munoz-Cabello, P., Garcia-Minaur, S., Espinel-Vallejo, M., *et al.* (2018) Translation and cross-cultural adaptation with preliminary validation of GCOS-24 for use in Spain. *J Genet Couns* **27(3)**: 732-743.

Muraki, E. (1992) A generalised partial credit model: application of an EM algorithm. *Appl Psychol Meas*, **17**: 159-176.

Nguyen, T., Han, H., Kim, M., Chan, K. (2014) An introduction to Item Response Theory for Patient-reported Outcome Measurement. *Patient*, **7(1)**, 23-35.

NICE. (2004) A guide to technology appraisals. London

Nordin, K., Liden, A., Hansson, M., Rosenquist, R., Berglund, G. (2002) Coping style, psychological distress, risk perception, and satisfaction in subjects attending genetic counseling for hereditary cancer. *J Med Genet* **39**: 689-694.

Nordin, K., Roshanai, A., Bjorvatn, C., Wollf, K., Mikkelsen, E., *et al.* (2011) Is genetic counseling a stressful event? *Acta Oncol* **50(7)**: 1089-97.

Nunnally, J., Bernstein, I. (1994) Psychometric Theory. New York: McGraw-Hill.

Nuttall, D., Parkin, D., Devlin, N. (2013) Inter-provider comparison of patient-reported outcomes: Developing an adjustment to account for differences in patient case mix. *Health Economics* **24**:41-54.

O'Connor AM. (1995) Validation of a decisional conflict scale. *Medical Decision Making* **15**:25-30.

O'Connor, B. (2000) SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Met, Ins & Comp,* **32(3)** 396-402.

Olson, G., Duffy, S., Mack, R. (1984) Thinking aloud as a method for studying real-time comprehension processes. *New methods in reading comprehension research,* 253-286.

Ondrusek N, Warner E, Goel V. (1999) Development of a knowledge scale about breast cancer and heredity (BCHK). *Breast Cancer Research and Treatment* **53**:69-75.

Opdenakker, R. (2006) Advantages and disadvantages of four interview techniques in qualitative research. *In Forum Qualitative Sozialforschung*: Qualitative Social Research **7(4)**.

Ormond, K. (2013) From genetic counseling to genomic counseling. *Mol Genet Gen Med* **1**: 189-193.

Otten, E., Birnie, E., Ranchor, A., van Langen, I. (2016) Telegenetics use in presymptomatic genetic counselling: patient evaluations on satisfaction and quality of care. *EJHG* **24(4)**: 513-520.

Pampols, T., Ramos, F., Lapunzina, P., Gozalo-Salellas, I., Perez-Jurado, L., Pujol, A. (2016) A view on clinical genetics and genomics in Spain: of challenges and opportunities. *Mol Genet Gen Med* **4(4)**: 376-391.

Payne, K., Nicholls, S. G., McAllister, M., *et al*. (2007) Outcome measures for clinical genetics services: a comparison of genetics healthcare professionals and patients' views. *Health Policy* **84(1),** 112-122.

Payne, K., Nicholls, S., McAllister, M., MacLeod, R., Donnai, D., Davies, L. (2008) Outcome Measurement in Clinical Genetics Services: A systematic review of validated measures. *Value in Health* **11(3)**: 497-508.

Petrillo, J., Cano, S., McLeod, L., Coon, C. (2015) Using classical test theory, item response theory and rasch measurement theory to evaluate PROMs. *Value in Health***18(1)**: 25-34.

Piertse, A., Dulmen, S., Ausems, M., Schoemaker, A., Beemer, F., Bensing, J. (2004) QUOTE-gene: development of a counselee-centered instrument to measure needs and preferences in genetic counselling for hereditary cancer. *Psycho-Oncology* **14(5)**: 361-375.

Radloff LS. (1977) The CES-D scale: a self-report depression scale for research in the general population. *Applied Psychological Measures* **1(3)**:385-401.

Rahman, N. (2014) Realizing the promise of cancer predisposition genes. *Nature* **505(7483)**: 302-308.

Read, C., Perry, D., Duffy, M. (2005) Design and psychometric evaluation of the Psychological Adaptation to Genetic Information Scale. *J Nurs Scholarsh* **37(3)**: 203-208.

Reed, S. (1955) *Counseling in Medical Genetics.* Philadelphia, PA: Saunders.

Reeve, B. (2002) An introduction to modern measurement theory. *National Cancer Institute*

Reichelt, J., Heimdal, K., Moller, P., Dahl, A. (2004) BRCA1 testing with definitive results: a prospective study of psychological distress in a large clinic-based sample. *Familial Cancer* **3(1)**: 21-28.

Remmers, C., Hibbard, J., Mosen, D., Wagenfield, M., Hoye, R., Jones, C. (2009) Is patient activation associated with future health outcomes and healthcare utilisation? *Journal of Ambulatory Care Management* **32(4)**: 320-327.

Resta, R., Biesecker, B., Bennet, R., Blum, S., Hahn, S., Strecker, M., Williams, J. (2006) A new definition of genetic counselling: National society of genetic counselors' task for report. *J Genet Couns* **15(2)**: 77-83.

Ritchie, J., Spencer, L. (1994) Qualitative data analysis for applied policy research. *Analysing qualitative data,* 173-194.

Rizopoulos, D. (2006) ltm: An R package for Latent Variable Modeling and Item Response Theory Analyses. *Journal of Statistical Software*, **17(5)**.

Rjinsburger, A., Essink-Bot, M., van As, E., Cockburn, J., de Koning, H. (2006) Measuring psychological consequences of screening: adaptation of the psychological consequences questionnaire into Dutch. *Qual Life Res* **15(5)**: 933-940.

Rogers, C. (1945) The Nondirective Method as a Technique for Social Research. *American Journal of Sociology* **50(4)**, 279-283.

Rosenberg M. (1965) Society & the Adolescent Self-Image. Princeton University Press.

Ross CE, Mirowsky J. (1984) Components of depressed mood in married men and women: The Center for Epdemiologic Studies Depression Scale. *American Journal of Epidemiology* **119**:997-1001.

Sagi, M., Kaduri, L., Zlotogora, J., Peretz, T. (1998) The effect of genetic counseling on knowledge and perceptions regarding risks for breast cancer. *J Genet Couns* **7**: 417-434.

Samejima, F. (1969) Estimation of latent ability using a response pattern of graded scores. *Psychom Monogr* **34**, 386-415.

Scheier M, Carver C, Bridges M. (1994) Distinguishing optimism from neuroticism (and Trait anxiety, self-mastery and self-esteem): a re-evaluation of the Life Orientation Test. *Journal of Personality and Social Psychology* **67(6):**1063-1078.

Schlich-Bakker, K., Warlam-Rodenhuis, C., van Echtelt, J., van den Bout, J., Ausems, M., Kroode, H. (2006) Short term psychological distress in patients actively approached for genetic counselling after diagnosis of breast cancer. *EJC***42(16)**: 2722-2728.

Schmidt, F., Le, H., Ilies, R. (2003) Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates. *Psychological Methods* **8(2),** 206-224.

Schneider, K., DiGianni, L., Kieffer, S., Chittenden, A., Shannon, K., Patenaude, A., Garber, J. (1999) Assessing patient satisfaction in a BRCA1/2 testing program. *Am J Hum Genet* **65**

Scientific Advisory Committee of the Medical Outcomes Trust (2002) – See Aaronson *et al.* (2002).

Sherbourne C, Stewart A. (1991) The MOS social support survey. *Social Science and Medicine* **32(6):**705-714.

Shiloh S, Avdor O, Goodman R. (1990) Satisfaction with genetic counselling: dimensions and measurement. *AJMG* **37**:522-9.

Skirton, H. (2001) The Client's Perspective of Genetic Counselling – A Grounded Theory Study. *J Genet Couns* **10(4)**: 311-329.

Skirton, H., Parsons, E., Ewings, P. (2005) Development of an audit tool for genetic services. *American Journal of Medical Genetics* **136A**:122-7.

Slomp, C., Morris, E., Inglis, A., Lehman, A., Austin, J. (2017) Patient outcomes of genetic counselling: Assessing the impact of different approaches to family history collection. *Clin Genet* **93(4)**: 830-836.

Snook, S., Gorsuch, R. (1989) Component analysis versus common factor analysis – a Monte Carlo study. *Psychological Bulletin***106(1)**, 148-154.

Spielberger C, *et al.* (1970) Manual for the STAI. Consulting Psychologists Tests.

Srivastava, A., Thomson, S. (2009) Framework Analysis: A qualitative methodology for applied policy research. *Journal of Administration and Governance* **4(2)**, 72-79.

Stadler, M., Mulvihill, J. (1998) Cancer risk assessment and genetic counseling in an academic medical center: Consultants' satisfaction, knowledge, and behaviour in the first year. *J Genet Couns* **7**: 279-297.

Stager JL. (1993) The comprehensive Breast Cancer Knowledge Test: validity and reliability. *Journal of Advanced Nursing* **18**:1133-1140.

Stalmeier PF, Unic IJ, Verhoef LC, van Daal WA. (1999) Evaluation of a shared decision making program for women suspected to have a genetic predisposition to breast cancer: preliminary results. *Medical Decision Making* **19**:230-41.

Stalmeier, P., Roosmalen, M., Verhoef, L., Hoekstra-Weebers, J., *et al.* (2005) The Decision evaluation scales. *Pat Educ Couns* **57(3)**: 286-293.

Stevenson, D., Carey, J. (2009) Health-Related Quality of Life measures in genetics disorders: an outcome variable for consideration in clinical trials. *AJMG* **151(3)**: 255-260.

Streiner, D. (1998) Factors affecting reliability of interpretations of scree plots. *Psychological Reports* **83**, 687-694.

Streiner, D. (2003) Starting at the beginning: An inteoduction to coefficient alpha and internal consistency. *Journal of personality assessment* **80(1)**, 199-103.

Streiner, D., Norman, G. (2008) Health measurement scales – A practical guide to their development and use. Oxford University Press. Oxford.

Streiner, D., Norman, G. (2015) Health Measurement Scales (5th Edition). Oxford University Press, Oxford.

Su, S., Miller, A., Sneider, H., Bremner, J., Ritchie, J., Masiano, C., *et al.* (2009) Common genetic contributions to depressive symptoms and inflammatory markers in middle-aged men: the twins heart study. *Psychosom Med* **71(2):** 152-158.

Sullivan, J., Abdo, Z., Joyce, P., Swofford, D. (2005) Evaluating the performance of a successive-approximations approach to parameter optimisation in a maximum-likelihood phylogeny estimation *Mol Biol Evol* **22(6)**, 1386-1392.

Temtamy, S., Hussen, D. (2017) Genetics and Genomic Medicine in Egypt: steady pace. *Mol Genet Gen Med* **5(1)**: 8-14.

Tercyak KP, Johnson SB, Roberts SF, Cruz AC. (2001) Psychological response to prenatal genetic counselling and amniocentesis. *Patient Education and Counselling* **43**:73-84.

Terwee, C., Bot, S., de Boer, M., van der Windt, D., Knol, D., Dekker, J. (2007) Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidem* **60(1)**, 34-42.

Terwee, C., Mokkink, L., Knol, D., Ostelo, R., Bouter, L., de Vet, H. (2012) Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* **21(4)**: 651-657.

Thomson, C., Hamilton, S., Hippman, C. (2015) Psychiatrist attitudes towards pharmacogenetic testing, direct to consumer genetic testing, and integrating genetic counseling into psychiatric patient care. *Psy Res* **226**: 68-72

Thurstone, L. (1931) Multiple factor analysis. *Psychological Review* **38(5)**, 406-427.

Tinsley, H., Tinsley, D. (1987) Uses of factor analysis in counseling psychology research. *Journal of counseling psychology* **34(4)**, 414-424.

Tluczek, A., Zaleski, C., Stachiw-Hietpad, D., Modaff, P., Adamski, C., Nelson, M., Reiser, C., *et al.* (2011) A tailored approach to family centred genetic counseling for cystic fibrosis newborn screening. *J Genet Couns* **20(2)**: 115-128.

Trask P, Paterson AG, Wang C, Hayasaka H, Milliron K, Blumberg L, *et al*. (2001) Cancer-specific worry interference in women attending a breast and ovarian cancer risk evaluation program: Impact on emotional distress and health functioning. *Psychooncology* **10**:349-60.

Vaeth PA. (1993) Women's knowledge about breast cancer. Dimensions of knowledge and scale development. *American Journal of Clinical Oncology* **16(5)**:446-454.

Valente, E., Ferraris, A., Dallapiccola, B. (2008) Genetic Testing for Paediatric Neurological Disorders. *Lancet Neurol* **7(12)**: 1113-26.

Van Hout, B., Janssen, M., Feng, Y., Kohlmann, T., Busschbach, Golicki, D. (2012) Interim Scoring for the EQ-5D-5L. *Value in Health* **15(5)**: 708-715.

Van Oostrum, I., Meijers-Heijboer, H., Lodder, L., Duivenvoorden, H., van Gool, A., Seynaeve, C. (2003) Long-term psychological impact of carrying a BRCA1/2 mutation and prophylactic surgery: a 5-year follow-up study. *Journal of Clinical Oncology* **21**:3867-74.

van Teijlingen, E. (2014) Semi-structured Interviews. *Bournemouth University*

Veach, P., Truesdell, S., LeRoy, B., Bartels, D. (1999) Client perceptions of the impact of genetic counseling: An exploratory study. *J Genet Couns* **8**: 191-216.

Velicer, W., Jackson, D. (1990) Component analysis versus common factor analysis. *Multivariate Behavioural Research* **25(1)**, 1-28.

Velicer, W., Eaton, C., Fava, J. (2000) Construct explication through factor or component analysis: A review and evaluation of alternative procedures. *Problems and Solutions in Human Assess.*, 41-71.

Walden, L., Brandi, E., Changasi, A., Sturgess, J., Soibel, A., *et al.* (2015) Physicians' opinions following pharmacogenetic testing for psychotropic medication. *Psychiatry Research* **229**: 913-918.

Wang, C., Gonzalez, R., Merajver, S. (2004) Assessment of genetic testing and related counselling services: current research and future directions. *Soc Sci Med* **58**: 1427-42.

Ware J. (2003) SF-36 Health survey: manual and interpretation guide. Health Institute, New England Medical Centre.

Wertz, D., Sorenson, J., Heeren, T. (1988) "Can't Get No (Dis) Satisfaction": Professional Satisfaction with Professional-Client Encounters. *Work and Occupations* **15(1)**

Westbrook M. (1979) A classification of coping behaviour based on multidimensional scaling of similarity ratings. *Journal of Clinical Psychology* **35(2):**407-409.

Westwood, G., Pickering, R., Latter, S., Little, P., Gerard, K., Lucassen, A., Temple, K. (2012) A primary care specialist genetics service: a cluster randomised factorial trial. *J Gen Pract* **62(596)**: 191-197.

Wiberg, M. (2004) Classical test theory vs item response theory: an evaluation of the theory test in the Swedish driving licence test. *DiVA* **50**:1-26.

Widaman, K. (1990) Bias in pattern loadings represented by common factor analysis and component analysis. *Multivariate Behavioural Research* **25(1)**, 89-95.

Widaman, K. (1993) Common factor analysis versus principal component analysis – differential bias in representing model parameters. *Multivariate Behavioural Research* **28(3)**, 263-311.

Wolf M, Putnam S, James S, Stiles W. (1978) Medical Interview Satisfaction Scale: development of a scale to measure patient perceptions of physician behaviour. *Journal of Behavioural Medicine* **1(4):**391-401.

Wolraich, M., Albanese, M., Stone, G., Nesbitt, D., Thomson, E., Shymansky, J. *et al.* (1986) Medical Communication Behaviour System: An interactional analysis system for medical interactions. *Medical Care* **24(10)**: 891-903.

Wood, J., Tataryn, D., Gorsuch, R. (1996) Effects of under and over extraction on principal axis factor analysis with varimax rotation. *Psychological Methods* **1**, 354-365.

Woolridge EQ, Murray RF. (1989) The health orientation scale: a measure of feelings about sickle cell trait. *Social Biology* **35**:123-136

Wu, C., Gong, Y., Wu, J., Zhang, S., Yin, X., Dong, X., Li, W., Cao, S., Mkandawire, N., Lu, Z. (2016) Chinese version of the EQ-5D preference weights. *PLoS One* **11(10)**

Zellerino, B., Milligan, S., Brooks, R., Freedenberg, D., Collingridge, D., Williams, M. (2009) Development, testing and validation of a patient satisfaction questionnaire for use in the clinical genetics setting. *AJMG* **151C:** 191-199.

Zigmond AS, Snaith AP. (1983) The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandanavica* **67**:361-70.

Zung, W. (1964) A self-rating Depression Scale. *General Psychiatry* **12**: 63-70.

Zwick, W., Velicer, W. (1982) Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioural Research* **17**, 253-269.

Zwick, W., Velicer, W. (1986) Comparison of five rules for determining the number of components to retain. *Psychological Bulletin* **99(3)**, 432-442.